Annals of Mathematics Studies
165

Spherical CR Geometry and Dehn Surgery

Richard Evan Schwartz

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press, 3 Market Place, Woodstock, Oxfordshire OX20 1SY

Library of Congress Cataloging-in-Publication Data

(this section to be filled in later)
(later)
p. cm. — (later)
Includes bibliographical references and index.
ISBN: (later)
(later)

(later)
536′.7—dc22 2004066029

British Library Cataloging-in-Publication Data is available

This book has been composed in FONT? in LaTeX

The publisher would like to acknowledge the authors of this volume for providing the camera-ready copy from which this book was printed.

Printed on acid-free paper. ∞

pup.princeton.edu

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

I dedicate this monograph to my daughters, Lucina and Lilith.

Contents

Preface

A general theme in geometry is the search for connections between the topological properties of a space and the geometrical properties of finer structures on that space. Thurston's hyperbolic Dehn surgery theorem (see [**T0**]) is a great theorem along these lines: All but finitely many Dehn fillings performed on a cusp of a hyperbolic 3-manifold result in new hyperbolic 3-manifolds. See Section 1.1. The purpose of this monograph is to prove an analogue of Thurston's result in the setting of spherical CR geometry and then to derive some consequences from it. We call our result the *Horotube Surgery Theorem*, or HST for short. See Theorem 1.2.

Spherical CR geometry is the $PU(2,1)$-invariant geometry of S^3, the 3-sphere. Here $PU(2,1)$ is the group of complex projective automorphisms of the unit ball in $\boldsymbol{C}^2$. The unit ball in $\boldsymbol{C}^2$ has a $PU(2,1)$-invariant Kähler metric, known as the *complex hyperbolic metric*. See Section 2.3. Spherical CR geometry is the "limit at infinity" of 4-dimensional complex hyperbolic geometry much in the same way that the Möbius-invariant geometry of S^2 is the "limit at infinity" of 3-dimensional (real) hyperbolic geometry.

While there are close connections between the Möbius geometry on S^2 and 2-dimensional hyperbolic geometry, there has not seemed to be much connection between spherical CR geometry and 3-dimensional hyperbolic geometry. However, in [**S0**] and [**S2**], respectively, we constructed a cusped hyperbolic 3-manifold and a closed hyperbolic 3-manifold that admit complete spherical CR structures. A *complete spherical CR structure* on a 3-manifold M is a homeomorphism between M and a quotient of the form Ω/Γ, where $\Gamma \subset PU(2,1)$ is a discrete subgroup and $\Omega \subset S^3$ is its *domain of discontinuity*. See Section 3.4. So far these are the only examples known; the closed example in [**S2**] required a computer-aided proof.

Using the HST we will construct many more closed manifolds that admit both hyperbolic and complete spherical CR structures. Unlike the proof given in [**S2**], the proofs we give here are traditional–aside from a few routine calculations in *Mathematica* [**W**]. One highlight of our results is that a positive density set of Dehn fillings of the Whitehead link complement give rise to closed 3-manifolds that admit both hyperbolic and spherical CR structures. Theorems 1.5 and 1.7 give precise statements along these lines.

As in [**S0**] and [**S2**], our examples are derived from representations of triangle groups into $PU(2,1)$. The work of Goldman and Parker [**GP**] is a seminal paper on these representations. Also see [**S3**]. Aside from arithmetic lattices and sporadic nonarithmetic lattices (see [**ACT**], [**DM**], [**T1**]), the

triangle groups provide some of the most nontrivial examples of complex hyperbolic discrete groups.

The subject of nonlattice complex hyperbolic discrete groups promises to be very interesting, though it has still not been explored as deeply as, say, real hyperbolic Kleinian groups. The HST incidentally makes an advance in this subject. We will use the HST to prove the (p, q, r) *Goldman-Parker conjecture* for $\min(p, q, r)$ large and generally to understand the complex hyperbolic (p, q, r)-triangle groups when $\min(p, q, r)$ is large. Theorems 1.10 and 1.12 give precise statements.

We don't expect a complete dictionary between 3-dimensional hyperbolic geometry and 3-dimensional spherical CR geometry. Our point of view is more conservative: Both geometries arise from big, interesting group actions on 3-dimensional spaces, so it's plausible that they should overlap. Second, the HST is not as ambitious as Thurston's theorem, which has two points: The cusp of a hyperbolic 3-manifold admits certain deformations, and these deformations result in Dehn fillings. The HST does not deal with the existence of these deformations, only the consequences of having them.

ACKNOWLEDGMENTS

I would like to thank Bill Goldman, friend and colleague, whose work has been an inspiration; Bill Thurston, my thesis advisor from 1987 to 1991, whose work has likewise been an inspiration; and John Millson for his encouragement. I would like to thank Martin Bridgeman, Nathan Dunfield, Galia Dafni, Peter Doyle, Elisha Falbel, Nikolay Gusevskii, Henry King, Sean Lawton, Pierre Pansu, John Parker, Blake Pelzer, Anna Pratoussevitch, and Justin Wyss-Gallifent for helpful and interesting discussions related to this work.

I started this work at the University of Maryland and completed it while on sabbatical at the Institute for Advanced Study in Princeton. This work was supported by the National Science Foundation grant DMS-0305047 and also by a fellowship from the J. S. Guggenheim Memorial Foundation. I would like to thank these institutions for their generous support.

Spherical CR Geometry and Dehn Surgery

PART 1
Basic Material

Chapter One

Introduction

1.1 DEHN FILLING AND THURSTON'S THEOREM

Dehn filling is a basic surgery one can perform on a 3-manifold. Let M be a 3-manifold that is the interior of a compact manifold with boundary $\overline{M}$. We say that a *torus end* of M is a torus boundary component of $\overline{M}$. Let E be such a torus end, and let $\alpha \in H_1(E)$ be a primitive homology element; i.e., α is not a multiple of another $\beta \in H_1(E)$. Let Σ be a solid torus with boundary $\partial\Sigma$. Let $f : E \to \partial\Sigma$ be a homeomorphism such that $f_*(\alpha) = 0$ in $H_1(\Sigma)$. Then the identification space

$$M_\alpha = (M \cup \Sigma)/f \tag{1.1}$$

is called the *α-Dehn filling* of E. The homeomorphism type of M_α only depends on α. When there is some implicit identification of $H_1(E)$ with $\boldsymbol{Z}^2$ carrying α to (p, q), we call M_α the *(p, q)-Dehn filling* of E. Such an identification is called a *marking* of E.

A *hyperbolic 3-manifold* M is a 3-manifold equipped with a Riemannian metric locally isometric to hyperbolic 3-space $\boldsymbol{H}^3$. See Section 2.1. We assume that M is oriented, has finite volume, and is metrically complete. When M is not closed, we call M *cusped*. In this case M is the interior of $\overline{M}$, as above. M is the union of a compact set and finitely many ends, each one being the quotient of a horoball by a $\boldsymbol{Z}^2$ subgroup. We call these ends *horocusps*. Each horocusp is homeomorphic to a torus cross a ray and is bounded by a component of $\partial\overline{M}$. Here is Thurston's celebrated hyperbolic Dehn surgery theorem.

Theorem 1.1 (See [**T0**]; cf. [**NZ**], [**PP**], [**R**]): *Suppose M is a cusped hyperbolic 3-manifold and E is a horocusp of M. All but finitely many Dehn fillings of E result in another hyperbolic 3-manifold.*

Let $\rho : \pi_1(M) \to \mathrm{Isom}(\boldsymbol{H}^3)$ be the representation whose image is the universal covering group of M. A main step in the proof of Theorem 1.1 is the analysis of representations $\widehat{\rho} : \pi_1(M) \to \mathrm{Isom}(\boldsymbol{H}^3)$, which are suitably nearby to ρ. We will describe the HST in such perturbative terms.

1.2 DEFINITION OF A HOROTUBE GROUP

$PU(2, 1)$ is the holomorphic isometry group of $\boldsymbol{CH}^2$, the complex hyperbolic plane. A *parabolic element* $P \in PU(2, 1)$ is one that acts so as to fix a point

on the ideal boundary of $\boldsymbol{CH}^2$ but no points in $\boldsymbol{CH}^2$ itself. For instance, one kind of parabolic has the form $P(z,t) = (uz, t+1)$ (where $|u| = 1$) when we normalize so that the ideal boundary of $\boldsymbol{CH}^2$ is identified with $(\boldsymbol{C} \times \boldsymbol{R}) \cup \infty$. See Section 2.3 for details.

The group in the HST, which plays the role analogous to the hyperbolic isometry group $\rho(\pi_1(M))$, is what we call a *horotube group*. Suppose that P is a parabolic element. We say that a *horotube* is a P-invariant open subset $T \subset S^3 - \{p\}$ such that $T/\langle P \rangle$ has a compact complement in $(S^3 - \{p\})/\langle P \rangle$. In the special case mentioned above, the set $\{|z| > 1\} \times \boldsymbol{R}$ is a good example of a horotube.

In general, we call the quotient $T/\langle P \rangle$ a *horocusp*. It is a consequence of Lemma 2.7 that any horocusp E is the union of a compact set and a horocusp E', which is homeomorphic to a torus cross a ray. In other words, after some pruning, a CR horocusp is topologically the same as a hyperbolic horocusp.

Let G be an abstract group, and let $\rho : G \to PU(2,1)$ be a discrete and injective representation. Let $\Gamma = \rho(G)$. Let Λ be the limit set of Γ, and let $\Omega = S^3 - \Lambda$ be the regular set. We say that Γ has *isolated type* if the elliptic elements of Γ have isolated fixed points. We say that Λ is *porous* if there is some $\epsilon_0 > 0$ such that $g(\Omega)$ contains a ball of spherical diameter ϵ_0 for any $g \in PU(2,1)$. See Section 3.5 for more details about these definitions.

Definition: ρ is a *horotube representation* and Γ is a *horotube group* if the following hold: Γ has isolated type, Λ is porous, and Ω/Γ is the union of a compact set together with a finite pairwise disjoint collection of horocusps.

It follows from Lemma 3.4 that Ω/Γ is a manifold when Γ is a horotube group. The following is another basic structural result about horotube groups:

Lemma 10.1 (Structure): *Let E_1,...,E_n be the horocusps of Ω/Γ. There are horotubes $\widetilde{E}_1$,...,$\widetilde{E}_n$ and elements γ_1,...,$\gamma_n \in \Gamma$ such that $E_j = \widetilde{E}_j/\langle \gamma_j \rangle$. Furthermore, every parabolic element of Γ is conjugate to a power of some γ_j. Thus, any maximal $\boldsymbol{Z}$-parabolic subgroup of Γ is conjugate in Γ to some $\langle \gamma_j \rangle$.*

1.3 THE HOROTUBE SURGERY THEOREM

1.3.1 The Result

Let ρ be a horotube representation and let $\Gamma = \rho(G)$ be the associated horotube group. In this case, we have the collection of maximal $\boldsymbol{Z}$ subgroups $\{H_\alpha\}$ of G such that $\rho(H_\alpha)$ is a parabolic group. We call these groups the *peripheral subgroups*. By Lemma 10.1, the conjugacy classes of peripheral subgroups are in natural bijection with the horocusps of Ω/Γ.

Let $\mathrm{Rep}(G)$ be the space of homomorphisms of G into $PU(2,1)$. We say that a sequence $\{\rho_n\}$ *converges nicely* to ρ iff $\rho_n(g)$ converges to $\rho(g)$ for each individual $g \in G$ and $\rho_n(H)$ converges geometrically to $\rho\,(H)$ for each peripheral subgroup of H. That is, the set $\rho_n(H)$ converges to the set $\rho\,(H)$ in the Hausdorff topology. See Section 3.1 for a definition of the Hausdorff topology. We make one additional technical requirement: In the case where $\rho_n(H)$ is a finite group, we require that each element of $\rho_n(H)$ acts freely on S^3, which is to say that $\rho_n(H)$ acts with an isolated fixed point in $\boldsymbol{CH}^2$.

In the elliptic case of the nice convergence, it turns out–at least for n large–that there is a preferred generator h of H such that $\rho_n(h)$ is $PU(2,1)$-conjugate to

$$\begin{bmatrix} \exp(2\pi i/m_n) & 0 \\ 0 & \exp(2\pi i k_n/m_n) \end{bmatrix} \tag{1.2}$$

with k_n, m_n relatively prime and $|k_n| < m_n/2$. We say that $\rho_n(H)$ has *type* (m_n, k_n) in this case.

Theorem 1.2 (Horotube Surgery) *Suppose $\rho \in \mathrm{Rep}(G)$ is a horotube representation and $\Gamma = \rho\,(G)$ has no exceptional cusps. If $\widehat{\rho} \in \mathrm{Rep}(G)$ is sufficiently far along in a sequence of representations converging nicely to ρ, then $\widehat{\Gamma} = \widehat{\rho}\,(G)$ is discrete and $\widehat{\Omega}/\widehat{\Gamma}$ is obtained from Ω/Γ by performing a Dehn filling on each horocusp E_H of Ω/Γ corresponding to a peripheral subgroup H such that $\widehat{\rho}(H)$ is not parabolic. Relative to a canonical marking, the filling has type $(0,1)$ when $\widehat{\rho}(H)$ is loxodromic and type (m,k) when $\widehat{\rho}(H)$ is elliptic of type (m,k). If at least one cusp is not filled, then $\widehat{\rho}$ is a horotube representation of $\widehat{G} = G/\ker(\widehat{\rho})$.*

We say that a horocusp of Ω/Γ is *exceptional* if the generator of the corresponding group is conjugate to the map $(z,t) \to (-z, t+1)$. When some cusp of Γ is exceptional, there is an ambiguity of sign, and (m,k) needs to be replaced by $(\pm m, k)$. The choice of sign depends on which of two equally canonical markings we choose. See Section 8 for details.

1.3.2 Some Remarks on the HST

Relative to the canonical marking, all the possible filling slopes–i.e., the quantities m/k in the elliptic case of HST–lie in $(-1/2, 1/2)$. Thus, for a given group with flexible cusps, the HST at best gives an open cone's worth of fillings in Dehn filling space. The HST doesn't quite have the same range as Thurston's theorem, which produces essentially all possible filling slopes. One structural reason for the difference is that the elliptic case of the HST involves a transition from $\boldsymbol{Z}/n$ to $\boldsymbol{Z}$, whereas Thurston's theorem involves a transition from $\boldsymbol{Z}$ to $\boldsymbol{Z}^2$.

Our proof of the HST only uses fairly general properties of $\boldsymbol{CH}^2$. The important ingredient is a kind of thick-thin decomposition of the manifold at infinity, where the thick part does not change much with the representation

and the thin part undergoes Dehn surgery. This ought to occur in a fairly general setting. Here are some settings related to the HST.

First, some version of the HST should work for complex hyperbolic discrete groups with rank-2 cusps. Actually, in the rank-2 case, one ought to be able to get the kind of transitions that more closely parallel the one in Thurston's theorem,–i.e., a $\boldsymbol{Z}$-loxodromic subgroup limiting to a $\boldsymbol{Z}^2$-parabolic subgroup. Not many groups of this kind have been studied. See [**F**] for some recent examples.

Second, a version of the HST should work in real hyperbolic 4-space $\boldsymbol{H}^4$. Indeed, one might compare the HST with results in [**GLT**]. From a practical standpoint, one difference between $\boldsymbol{H}^4$ and $\boldsymbol{CH}^2$ is that we can easily obtain the kind of convergence we need for the HST just by controlling the traces of certain elements in $PU(2,1)$. This makes it easy to show that particular examples satisfy the hypotheses of the HST. The situation seems more complicated in $\boldsymbol{H}^4$.

Third, there is some connection between the HST and contact topology. A complete spherical CR manifold Ω/Γ has a *symplectically semifillable* contact structure (See [**El**]) because a finite cover Ω/Γ bounds the symplectic orbifold $\boldsymbol{CH}^2/\Gamma$. Thus, the HST gives a geometric way to produce symplectically semifillable (and, hence, tight) contact structures on some hyperbolic 3-manifolds. Compare Eliashberg's Legendrian surgery theorem (see [**El**], [**Go**]).

We don't pursue these various alternate settings because (at least when we wrote this monograph) we had in mind the specific applications listed below. However, the interested reader might want to keep some of the above connections in mind while reading our proof of the HST.

1.4 REFLECTION TRIANGLE GROUPS

The *complex hyperbolic reflection triangle groups* provide our main examples of horotube groups. We will discuss them in more detail in Chapter 4.

Let $\boldsymbol{H}^2$ denote the real hyperbolic plane. Let $\zeta = (\zeta_0, \zeta_1, \zeta_2) \in (\boldsymbol{N} \cup \infty)^3$ satisfy $\sum \zeta_i^{-1} < 1$. Let G'_ζ denote the usual ζ reflection triangle group generated by reflections in a geodesic triangle $T_\zeta \subset \boldsymbol{H}^2$ with internal angles π/ζ_0, π/ζ_1, π/ζ_2. (When $\zeta_i = 0$ the ith vertex of the triangle is an ideal vertex.) Let G_ζ denote the index-2 *even subgroup* consisting of the even-length words in the three generators $\iota_0, \iota_1, \iota_2$ of G'_ζ.

A *complex reflection* is an element of $PU(2,1)$ which is conjugate to the map $(z,w) \to (z,-w)$. See Section 2.2.3. A ζ-*complex reflection triangle group* is a representation of G'_ζ that maps $\iota_0, \iota_1, \iota_2$ to *complex reflections* I_0, I_1, I_2 so that $\iota_i\iota_j$ and I_iI_j have the same order. We call a (∞,∞,∞)-complex reflection triangle group an *ideal complex reflection triangle group*.

Let $\mathrm{Rep}(\zeta)$ denote the set of ζ-complex reflection triangle groups, modulo conjugacy in $\mathrm{Isom}(\boldsymbol{CH}^2)$. It turns out that $\mathrm{Rep}(\zeta)$ is a 1-dimensional half-

open interval. The endpoint of $\mathrm{Rep}(\zeta)$ is the element ρ_0, which stabilizes a totally geodesic slice, like a Fuchsian group. In [**GP**] Goldman and Parker initiated the study of the space $\mathrm{Rep}(\infty,\infty,\infty)$.

In [**S1**] and [**S5**], we proved the Goldman-Parker conjecture (see below) about $\mathrm{Rep}(\infty,\infty,\infty)$. Here is a mild strengthening of our main result from [**S1**].

Theorem 1.3 *An element $\rho \in \mathrm{Rep}\,(\infty,\infty,\infty)$ is a horotube representation, provided that $I_0I_1I_2$ is loxodromic. Any parabolic element of $\Gamma = \rho\,(G)$ is conjugate to $(I_iI_j)^a$ for some $a \in \boldsymbol{Z}$.*

There is one element $\rho \in \mathrm{Rep}(\infty,\infty,\infty)$ for which $I_0I_1I_2$ is parabolic. ρ is the "endpoint" of the representations considered in Theorem 1.3. We call the corresponding group $\boldsymbol{\Gamma'}$ the *golden triangle group* because of its special beauty. Let $\boldsymbol{\Gamma}$ be the even subgroup. Let $\boldsymbol{\Gamma_3}$ be the group obtained by adjoining the "vertex-cycling" order-3 symmetry to $\boldsymbol{\Gamma}$. Let G_3 denote corresponding real hyperbolic group and $\rho : G_3 \to \boldsymbol{\Gamma_3}$ the corresponding representation. Here is a mild strengthening of our main result in [**S0**].

Theorem 1.4 *ρ is a horotube representation. Any parabolic element $g \in \boldsymbol{\Gamma_3}$ is conjugate to either $(I_0I_2)^a$ for some $a \in \boldsymbol{Z}$ or $(I_0I_1I_2)^b$ for some $b \in 2\boldsymbol{Z}$. The quotient $\Omega/\boldsymbol{\Gamma_3}$ is homeomorphic to the Whitehead link complement.*

1.5 SPHERICAL CR STRUCTURES

A *spherical CR structure* on a 3-manifold is a system of coordinate charts into S^3, such that the overlap functions are restrictions of elements of $PU(2,1)$. The complete spherical CR structures mentioned in the Preface are special cases. One can ask the general question, Which 3-manifolds admit spherical CR structures? Here are some answers.

- The unit tangent bundle of a closed hyperbolic surface admits a "tautological" complete spherical CR structure. We will discuss the construction in Section 4.3, in the related context of triangle groups. Many other circle bundles over surfaces admit complete spherical CR structures (see [**GKL**] [**AGG**]) as do many Seifert fiber spaces (see [**KT**]).

- In very recent work Anan'in and Gusevskii [**AG**] have announced that there is a complete spherical CR structure on the trivial circle bundle over a closed hyperbolic surface. This has been a long-standing open problem.

- In recent work, Falbel [**F**] has shown that there is no complete spherical CR structure on the figure-eight knot complement whose cusps have purely parabolic holonomy. In the language above, this means that there is no horotube group Γ such that Ω/Γ is the figure-8 knot complement.

- Here is an example of the kind of result that can be proved using contact-/symplectic-based manifold invariants: Let P be the Poincaré homology 3-sphere. Then P is the quotient of S^3 by a finite subgroup of $PU(2,1)$, and hence, it admits a complete spherical CR structure. In [**L**] it is proved that the oppositely oriented version of P is not *symplectically semifillable*. Hence, the oppositely oriented version of P does not admit a complete spherical CR structure compatible with that orientation.

- In [**S2**], we gave a computer-aided construction of a closed hyperbolic 3-manifold with a complete spherical CR structure. The associated group is the $(4,4,4)$-complex reflection triangle group in which $I_0I_1I_0I_2$ is elliptic of order $n = 7$. Experimental evidence suggests that the same result holds when $n > 7$, and this evidence led us to the HST. (We originally wanted to use the HST to deal with the $(4,4,4)$ triangle groups, but we were unable to prove the analogue of Theorem 1.4 in the $(4,4,4)$ case.)

If M is a 3-manifold with n torus ends, then we can identify the set of closed fillings of M (all cusps filled) with the subset $\mathcal{D}(M) \subset \boldsymbol{Z}^{(2n)}$ consisting of the $2n$-tuples $(p_1, q_1, ..., p_n, q_n)$, where p_j and q_j are relatively prime for all j. We say that a cusped manifold is *CR positive* if there is an open cone in $\boldsymbol{R}^{2n}$ such that all but finitely many points in $\mathcal{D}(M) \cap C$ result in fillings that have complete spherical CR structures. If M is both hyperbolic and CR positive, then a positive density subset of fillings of M admit both hyperbolic and spherical CR structures. Combining the HST with an analysis of the flexibility of the group $\boldsymbol{\Gamma_3}$ from Theorem 1.4, we prove the following theorem.

Theorem 1.5 *This Whitehead link complement is CR positive.*

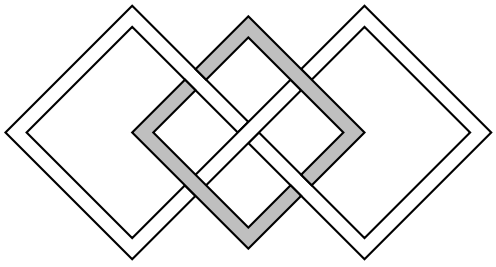

Figure 1.1: The Whitehead link

In [**A**], it is shown that any infinite list of closed hyperbolic 3-manifolds, with uniformly bounded volume, contains an infinite number of commensurability classes. Thus, Theorem 1.5 and Theorem 1.1 combine to prove the following corollary.

Corollary 1.6 *There is an infinite list of pairwise incommensurable closed hyperbolic* 3*-manifolds which admit complete spherical CR structures.*

Generalizing Theorem 1.5, we will prove the following.

Theorem 1.7 *Let T be a finite tree, with an odd number N of vertices, such that every vertex has valence at most 3. Suppose also that at least one vertex of valence 1 is incident to a vertex of valence 2. Then there is an $(N+3)$ cusped finite cover of the Whitehead link complement, canonically associated to T, which is CR positive.*

The side condition about the valence-2 vertex seems not to be necessary; it is an artifact of our proof.

One of the great consequences of Theorem 1.1 is that the set of volumes of hyperbolic 3-manifolds is well ordered and has the ordinal structure of ω^ω. See [**T0**]. Using Theorem 1.7, we easily get a related statement.

Corollary 1.8 *Let S be the set of volumes of closed hyperbolic 3-manifolds, which admit complete spherical CR structures. Let $S^{(0)} = S$, and let $S^{(n+1)}$ be the accumulation set of $S^{(n)}$. Then $S^{(n)} \neq \emptyset$ for all n.*

The basic idea in proving Corollary 1.8 is to note that $\mathrm{vol}(M_n) \to \mathrm{vol}(M)$ if $\{M_n\}$ is a nonrepeating sequence of Dehn fillings of M. If we have a CR positive manifold with many cusps, then we can produce a lot of limit points to the set S simply by controlling the rates at which we do the fillings on different cusps. This is the essentially the same trick that Thurston performs in [**T0**].

1.6 THE GOLDMAN-PARKER CONJECTURE

We continue the notation from Section 1.4. Say that a word in a reflection triangle group G has *genuine length* k if the word has length k in the generators and is not conjugate to a shorter word. Referring to the representation space $\mathrm{Rep}(\zeta)$, the endpoint ρ_0 maps the words of genuine length 3 and 4 to loxodromic elements. The following is a central conjecture about the complex hyperbolic reflection triangle groups.

Conjecture 1.9 *Suppose that $\rho \in \mathrm{Rep}(\zeta)$ maps all words of genuine length 3 and 4 to loxodromic elements. Then ρ is discrete.*

The (∞, ∞, ∞) case of Conjecture 1.9, the first case studied, was introduced by Goldman and Parker in [**GP**]. Goldman and Parker made substantial progress on this case, and we gave a computer-aided proof of the conjecture in [**S1**]. Then we gave a better and entirely traditional proof in [**S5**]. Computer experiments done by Wyss-Gallifent [**W-G**] and me suggested the general version of the conjecture.

In Part 3 we will combine the HST with Theorems 1.3 and 1.4 to prove the following result.

Theorem 1.10 *Suppose that $\rho \in \mathrm{Rep}(\zeta)$ maps all words of genuine length 3 to loxodromic elements. If $|\zeta|$ is sufficiently large, then ρ is a horotube representation and, hence, discrete.*

Here $|\zeta| = \min(\zeta_0, \zeta_1, \zeta_2)$. Unfortunately, we don't have an effective bound on $|\zeta|$. For $|\zeta| > 14$, it turns out that the words of genuine length 4 are automatically loxodromic if the words of genuine length 3 are loxodromic. Compare [**P**] or [**S3**].

To help understand the triangle groups, we will prove the following *addendum* to the HST.

Theorem 1.11 *Let ρ, $\widehat{\rho}$, and G be as in the HST. Additionally suppose that $\widehat{\rho}$ is injective and that $\widehat{\rho}(g)$ is parabolic iff $\rho(g)$ is parabolic for all $g \in G$. If $\widehat{\rho}$ is sufficiently far along in a sequence that converges nicely to ρ, then there is a homeomorphism from Ω to $\widehat{\Omega}$ that conjugates Γ to $\widehat{\Gamma}$.*

With a bit more work, we deduce the following corollary.

Corollary 1.12 *Suppose that $|\zeta|$ is sufficiently large and Γ_1 and Γ_2 are even subgroups of two ζ-complex reflection triangle groups, whose words of genuine length 3 are loxodromic. Then Γ_1 and Γ_2 have topologically conjugate actions on S^3. In particular, the limit sets of these groups are topological circles.*

1.7 ORGANIZATIONAL NOTES

The monograph has four parts. Parts 2, 3, and 4 all depend on Part 1 but are essentially independent from each other. Parts 2, 3, and 4 can be read in any order after part I has been finished.

Part 1 (Chapters 1–5) is the introductory part. Chapter 2 presents some background material on complex hyperbolic geometry. Chapter 3 presents some background material on discrete groups and topology. Chapter 4 introduces the reflection triangle groups, relating them to the HST. In Chapter 5 we give a heuristic explanation of the HST, comparing it with Thurston's theorem. Chapter 5 is the conceptual heart of the HST. Most of our proof of the HST amounts to making the discussion in Chapter 5 rigorous.

In Part 2 (Chapters 6–13), we prove the HST and Theorem 1.11. We give a more extensive overview of Part 2 just before starting Chapter 6.

In Part 3 (Chapters 14–18), we derive all our applications, using Theorems 1.3 and 1.4 as black boxes.

In Part 4 (Chapters 19–23), we prove Theorems 1.3 and 1.4. Our proof of Theorem 1.4, which is fairly similar to what we did in [**S0**], is almost self-contained. We omit a few minor and tedious calculations, and in those places we refer the reader to [**S0**] for details. Our proof of Theorem 1.3 is sketchier but tries to hit the main ideas in [**S5**]. Most of Part 4 appears in our other published work, but here we take the opportunity to improve on the exposition given in our earlier accounts and also to correct a few glitches. (See Sections 21.1.2 and 22.4.)

We have written an extensive Java applet that illustrates the constructions in Part 4 graphically and in great detail. We encourage the reader to use this applet as a guide to the mathematics in Part 4. As of this writing, in 2006, the applet is called *Applet 45*. It currently resides on my Web site, at http://www.math.brown.edu/∼res/applets.html.

Chapter Two

Rank-One Geometry

2.1 REAL HYPERBOLIC GEOMETRY

The reader may wish to consult [**B**] and [**R**], which are good references for real hyperbolic geometry. Let $\boldsymbol{H}^n$ denote n-dimensional hyperbolic space. There are several natural models for $\boldsymbol{H}^n$, and we will discuss three of these.

Klein Model: In the *Klein model*, $\boldsymbol{H}^n$ is the open unit ball in $\boldsymbol{R}^n$, which in turn is considered an affine patch of real projective space $\boldsymbol{RP}^n$. In this model the isometries of $\boldsymbol{H}^n$ are given by projective (i.e., line–preserving) transformations, which stabilize the open unit ball. In the invariant Riemannian metric, the geodesics are Euclidean line segments.

Poincaré Model: In the *Poincaré model*, $\boldsymbol{H}^n$ is the open unit ball and is considered a subset of S^n. The isometries of $\boldsymbol{H}^n$ are given by conformal (i.e., angle –preserving) transformations of S^n, which stabilize the open unit ball. In the invariant Riemannian metric, the geodesics are circular arcs that meet the boundary at right angles. The angle between two curves in the Poincaré model coincides with their Euclidean angle.

Upper Half-Space Model: In the *upper half-space model*, one chooses a stereographic projection from the unit ball to the open upper half-space in $\boldsymbol{R}^n$. In this way, one transfers the Poincaré metric on the open ball to a metric on the upper half-space. In this model, the isometries of $\boldsymbol{H}^n$ are conformal automorphisms of $\boldsymbol{R}^n \cup \infty$, which stabilize $\boldsymbol{R}^{n-1} \cup \infty$. The isometries of $\boldsymbol{H}^n$ that stabilize ∞ are restrictions of similarities of $\boldsymbol{R}^n$.

In the Klein and Poincaré models, S^{n-1} is the *ideal boundary* of $\boldsymbol{H}^n$. In the upper half-space model, the ideal boundary is $\boldsymbol{R}^{n-1} \cup \infty$. The isometries of $\boldsymbol{H}^n$ can be classified into three types: *elliptic*, *parabolic* and *hyperbolic*. Elliptic elements have fixed points in $\boldsymbol{H}^n$. Parabolic elements have one fixed point in the ideal boundary and no fixed points in $\boldsymbol{H}^n$. Hyperbolic elements have two fixed points in the ideal boundary and no fixed points in $\boldsymbol{H}^n$.

A *horoball* in $\boldsymbol{H}^n$ is the geometric limit of unboundedly large metric balls, provided that this limit is a nonempty proper subset of $\boldsymbol{H}^n$. A *horosphere* is the boundary, in $\boldsymbol{H}^n$, of a horoball. A horoball or horosphere accumulates, on the ideal boundary, at a single point. This point is called the *basepoint* of the horoball or horosphere.

If $x \in \boldsymbol{H}^n$ is a point and p is an ideal point, then there is a unique

function $\beta_{x,p} : \boldsymbol{H}^n \to \boldsymbol{R}$, normalized so that $\beta_{x,p}(x) = 0$, such that we have the following.

- The level sets of $\beta_{x,p}$ are horospheres whose basepoints are p.
- The map $\beta_{x,p}$ is an orientation-preserving isometry (between lines) when restricted to any geodesic that limits at p and is oriented towards p.

The function $\beta_{x,p}$ is called a *Busemann function*. The simplest example of a Busemann function is given by

$$\beta(x_1, \ldots, x_{n-1}, y) = \log y \tag{2.1}$$

in the upper half-space model, where $\boldsymbol{H}^n$ is identified with $\boldsymbol{R}^{n-1} \times \boldsymbol{R}^+$. Here $\beta = \beta_{x,p}$, with $x = (0, \ldots, 0, 1)$ and $p = \infty$.

2.2 COMPLEX HYPERBOLIC GEOMETRY

[**E**] and [**G**] are good references for complex hyperbolic geometry.

2.2.1 The Ball Model

$\boldsymbol{C}^{n,1}$ denotes the standard $n+1$ complex dimensional vector space, equipped with the Hermitian form

$$\langle u, v \rangle = u_1 \overline{v}_1 + \cdots + u_n \overline{v}_n - u_{n+1} \overline{v}_{n+1}. \tag{2.2}$$

$\boldsymbol{CH}^n$ and its ideal boundary are the projective images, in the complex projective plane $\boldsymbol{CP}^n$, of

$$N_- = \{v \in \boldsymbol{C}^{n,1} |\ \langle v, v \rangle < 0\}, \qquad N_0 = \{v \in \boldsymbol{C}^{n,1} |\ \langle v, v \rangle = 0\}, \tag{2.3}$$

respectively. (The set N_+ has a similar definition.) The projectivization map

$$(v_1, \ldots, v_n, v_{n+1}) \to \left(\frac{v_1}{v_{n+1}}, \ldots, \frac{v_n}{v_{n+1}} \right) \tag{2.4}$$

takes N_- and N_0 to the open unit ball and unit sphere in $\boldsymbol{C}^n$, respectively. Henceforth, we identify $\boldsymbol{CH}^n$ with the open unit ball.

Up to scale, there is a unique Kähler metric on $\boldsymbol{CH}^n$, which is invariant under complex projective automorphisms, known as the *complex hyperbolic metric*. With respect to the Riemannian part of this metric, $\boldsymbol{CH}^n$ has 1/4-pinched negative sectional curvature. We do not use this metric in the monograph.

Henceforth, we take $n = 2$. In this case, $\boldsymbol{CH}^2$ is called the *complex hyperbolic plane*.

2.2.2 Visual Diameter

Here we will talk about the isometry group of $\boldsymbol{CH}^2$ in detail. One basic feature we mention right away is that the stabilizer of the origin in $\boldsymbol{CH}^2$ consists of Euclidean isometries, and the full isometry group acts transitively on the space; one can move any point to any other point by an isometry. Here is one metric concept we will use frequently. If $x \in \boldsymbol{CH}^2$ and $S \subset \partial \boldsymbol{CH}^2$, then we choose an isometry g, which carries $\boldsymbol{CH}^2$ to the unit ball in $\boldsymbol{C}^2$, such that $g(x) = (0, 0)$. We then define $\mathrm{DIAM}_x(S)$ to be the diameter of $g(S)$, as measured in the round metric on S^3. This quantity is the *visual diameter of* S, as seen from x. It only depends on x and S.

2.2.3 Geodesic Slices

There are two kinds of totally geodesic 2-planes in $\boldsymbol{CH}^2$.

- The $\boldsymbol{R}$*-slices* are 2-planes, $PU(2,1)$-equivalent to $\boldsymbol{RH}^2 = \boldsymbol{R}^2 \cap \boldsymbol{CH}^2$. The $\boldsymbol{R}$-slices naturally carry the Klein model of $\boldsymbol{H}^2$.
- The $\boldsymbol{C}$*-slices* are 2-planes, $PU(2,1)$-equivalent to $\boldsymbol{CH}^1 = \boldsymbol{CH}^2 \cap \boldsymbol{C}^1$. The $\boldsymbol{C}$-slices naturally carry the Poincaré model of $\boldsymbol{H}^2$.

The accumulation set on S^3, of an $\boldsymbol{F}$-slice, is called an $\boldsymbol{F}$*-circle*.

An element of $PU(2,1)$ is determined by its action on an $\boldsymbol{R}$-slice, whereas there is an S^1-family of elements, which act the same on a $\boldsymbol{C}$-slice. For instance, the maps $(z, w) \to (z, uw)$, for $|u| = 1$, all fix $\boldsymbol{CH}^1$. Another difference between these slices is that there is a $PU(2,1)$-invariant contact distribution on S^3, given by the complex lines tangent to S^3. The $\boldsymbol{C}$-circles are transverse to this distribution, and the $\boldsymbol{R}$-circles are tangent to it.

2.2.4 Isometries

$SU(2,1)$ is the $\langle , \rangle$-preserving subgroup of $SL_3(\boldsymbol{C})$, the special complex linear group. $PU(2,1)$ is the projectivization of $SU(2,1)$. Elements of $PU(2,1)$ act isometrically on $\boldsymbol{CH}^2$ and are classified according to the same scheme as given in Section 2.1. Elements of $SU(2,1)$ are classified by their images in $PU(2,1)$. There are various names for parabolic elements, and we recall them all here.

- A parabolic element is called $\boldsymbol{C}$*-parabolic* (or ellipto-parabolic) if it stabilizes a $\boldsymbol{C}$-slice and rotates the normal bundle to this slice by a nontrivial amount. This is equivalent to the condition that the element stabilizes a *unique* $\boldsymbol{C}$-slice.
- A parabolic element is called *unipotent* if it is not $\boldsymbol{C}$-parabolic. Thus, every parabolic element is either $\boldsymbol{C}$-parabolic or unipotent.
- We call a parabolic element $\boldsymbol{R}$*-parabolic* if it stabilizes an $\boldsymbol{R}$-slice. An $\boldsymbol{R}$-parabolic is unipotent, but some unipotent parabolics are not $\boldsymbol{R}$-parabolic.

Here are some names for elliptic elements.

- An elliptic element is called *regular* if it is represented by a matrix in $SU(2,1)$ whose eigenvalues are all distinct.
- An elliptic element is called *irregular* if it is not regular.

We shall single out some additional kinds of elliptic elements below.

There is a computational method for classifying elements of $SU(2,1)$. Given $A \in SU(2,1)$, define

$$\delta(A) = f\big(\text{trace}(A)\big); \qquad f(z) = |z|^4 - 8\text{Re}(z^3) + 18|z|^2 - 27. \tag{2.5}$$

For a proof of the following beautiful result, see [**G**, p. 204].

Lemma 2.1 *Let $A \in SU(2,1)$. Then we have the following.*

- *A is loxodromic if and only if $\delta(A) > 0$.*
- *A is regular elliptic if and only if $\delta(A) < 0$.*
- *A is $\boldsymbol{C}$-parabolic if and only if A is not elliptic, $\delta(A) = 0$, and the trace of A is not 3 times a cube root of unity.*
- *A is unipotent if and only if the trace of A is 3 times a cube root of unity.*

If $\delta(A) = 0$, then we would like to conclude that A represents a parabolic element, but this need not be the case. It might happen that A represents an irregular elliptic element.

2.2.5 More about Elliptics

Here we include a very basic lemma about traces because of the important role the lemma plays in one of our applications in Part 3.

Lemma 2.2 *Suppose that $X \in SU(2,1)$ is an element of trace 0. Then X has order 3.*

Proof: This is a well-known result; we give the proof to keep our exposition more self-contained. We know that X is elliptic from Lemma 2.1. Hence, X is conjugate to a diagonal matrix with entries a, b, c such that $abc = 1$, $a+b+c = 0$, and $|a| = |b| = |c| = 1$. Multiplying the equation $a+b+c = 0$ by bc, we find that $bbc + ccb = -1$. Hence, $1 = |bc||b+c| = |b+c|$. Since $|b| = |c| = |b+c| = 1$, the points b and c must be two vertices of an equilateral triangle centered at 0. Cycling the vertices, we see that a, b, c are the three vertices of an equilateral triangle. Hence, there is some unit complex number u such that $ua = 1$, $ub = \omega$, and $uc = \omega^2$. Here ω is a cube root of unity. But then $1 = abc = u^3$. Hence, u itself is a cube root of unity. Hence,

a, b, c are all cube roots of unity. □

We distinguish certain special elliptic elements of $PU(2,1)$. Given a vector $C \subset N_+$, we define

$$I_C(U) = -U + \frac{2\langle U, C\rangle}{\langle C, C\rangle} C. \tag{2.6}$$

I_C is an involution fixing C and $I_C \in SU(2,1)$. See [**G**, p. 70]. The element of $PU(2,1)$ corresponding to I_C is called a *complex reflection*. Every complex reflection is conjugate to the map $(z, w) \to (z, -w)$, whose fixed-point set is $\boldsymbol{CH}^1$. Thus, every complex reflection fixes a unique $\boldsymbol{C}$-slice. We say that an elliptic element of $PU(2,1)$ is *lens-elliptic of type* (m, k) if it is conjugate to the map

$$\begin{bmatrix} \exp(2\pi i/m) & 0 \\ 0 & \exp(2\pi ik/m) \end{bmatrix} \tag{2.7}$$

with k relatively prime to m and $|k| < m/2$. Compare Equation 1.2.

We call an elliptic element $\boldsymbol{R}$*-elliptic* if it is conjugate to

$$\begin{bmatrix} \cos(2\pi/n) & \sin(2\pi/n) \\ -\sin(2\pi/n) & \cos(2\pi/n) \end{bmatrix}. \tag{2.8}$$

Here n need not be a natural number, but this is the main case of interest to us. Referring to Equation 2.7, an $\boldsymbol{R}$-elliptic of order $n \in \boldsymbol{N}$ is lens-elliptic of type $(n, -1)$.

2.3 THE SIEGEL DOMAIN AND HEISENBERG SPACE

Everything we say in this section can be found in [**G**] and also in [**S0**].

2.3.1 The Siegel Domain

In the ball model, $\boldsymbol{CH}^2$ is a ball sitting inside the complex projective space $\boldsymbol{CP}^2$. For this discussion we fix some $p \in S^3$, the ideal boundary of $\boldsymbol{CH}^2$. There exists a complex projective automorphism β of $\boldsymbol{CP}^2$ that maps p to a point in $\boldsymbol{CP}^2 - \boldsymbol{C}^2$ and that identifies $\boldsymbol{CH}^2$ with the *Siegel domain*:

$$Z = \{(z, w) |\ \mathrm{Re}(w) > |z|^2\} \subset \boldsymbol{C}^2 \subset \boldsymbol{CP}^2. \tag{2.9}$$

We write $\infty = \beta(p)$ in this case. The holomorphic isometries of $\boldsymbol{CH}^2$ that fix ∞ act as complex linear automorphisms of Z. Let ∂Z denote the intersection of the ideal boundary of Z with $\boldsymbol{C}^2$. Note that $\boldsymbol{C}^2$ is a neighborhood of $Z \cup \partial Z$ in $\boldsymbol{CH}^2$.

2.3.2 Heisenberg Space

We call $\mathcal{H} = \boldsymbol{C} \times \boldsymbol{R}$ the *Heisenberg space*. $\mathcal{H}$ is equipped with a group law:

$$(z_1, t_1) \cdot (z_2, t_2) = \big(z_1 + z_2, t_1 + t_2 + 2\mathrm{Im}(\overline{z}_1 z_2)\big). \tag{2.10}$$

A *Heisenberg stereographic projection from p* is a map $\boldsymbol{B} : S^3 - \{p\} \to \mathcal{H}$ of the form

$$\boldsymbol{B} = \pi \circ \beta, \qquad \pi(z,w) = \big(z, \mathrm{Im}(w)\big) \tag{2.11}$$

Here β is as above. We write $\infty = \boldsymbol{B}(p)$ in this case.

There is a canonical projection map from $\mathcal{H}$ to $\boldsymbol{C}$ given by $(z,t) \to z$. The fibers are the *vertical lines*. If V is a vertical line, then $V \cup \infty$ is the image of a $\boldsymbol{C}$-circle under Heisenberg stereographic projection.

In $\mathcal{H}$ the $\boldsymbol{C}$-circles that are not vertical lines are ellipses that project to circles in $\boldsymbol{C}$. The $\boldsymbol{R}$-circles that contain ∞ are straight lines. One of these $\boldsymbol{R}$-circles is $(\boldsymbol{R} \times \{0\}) \cup \infty$. The bounded $\boldsymbol{R}$-circles in $\mathcal{H}$ are such that their projections to $\boldsymbol{C}$ are lemniscates. The standard lemniscate in polar coordinates is given by the equation $r^2 = \cos(2\theta)$. It looks like the infinity symbol, ∞.

$\boldsymbol{B}$ carries the contact distribution on S^3 to a contact distribution on $\mathcal{H}$. We will give explicit formulas below. There is a natural path metric on $\mathcal{H}$ called the *Carnot-Caratheodory metric*, or Carnot metric for short. In brief, we put the unique Riemannian metric on each contact plane in $\mathcal{H}$ so that the projection to $\boldsymbol{C}$ is always an isometry. Then we define the distance between two points in $\mathcal{H}$ as the infimal length of a path that joins the points and is always tangent to the contact planes. We get the length of this path by integrating the Riemannian metric. We equip $\mathcal{H}$ with the Carnot metric.

2.3.3 Heisenberg Automorphisms

We now consider maps of the form $P = \boldsymbol{B} g \boldsymbol{B}^{-1}$, where $\boldsymbol{B}$ is as in Equation 2.11 and $g \in PU(2,1)$ is an element fixing $\boldsymbol{B}^{-1}(\infty)$. We call P a *Heisenberg automorphism*. When g is parabolic, we will sometimes call P a *parabolic Heisenberg automorphism*. Heisenberg automorphisms act as similarities on $\mathcal{H}$ (relative to the Carnot metric), and parabolic Heisenberg automorphisms act as isometries. All Heisenberg automorphisms commute with the projection $\mathcal{H} \to \boldsymbol{C}$ and induce similarities on $\boldsymbol{C}$. In the parabolic case, the induced maps on $\boldsymbol{C}$ are isometries.

Let P be a parabolic Heisenberg automorphism. If the action of P on $\boldsymbol{C}$ has a fixed point, then we can conjugate so that 0 is a fixed point and P has the form

$$(z,t) \to (uz, t+1), \qquad |u| = 1. \tag{2.12}$$

The "exceptional" case mentioned in connection with the HST in Chapter 1 corresponds to $u = -1$ in Equation 2.12. In general, we call u the *twist* of P. A parabolic element is $\boldsymbol{C}$-parabolic if and only if it is conjugate to the map in Equation 2.12 with $u \neq 1$.

If the action of P on $\boldsymbol{C}$ has no fixed point, then we can conjugate so that this action has the form $z \to z+1$. In this case, P has the form

$$(z,t) \to (1,s) \cdot (z,t) = \big(z+1, t + 2\mathrm{Im}(z) + s\big). \tag{2.13}$$

These parabolic elements are all unipotent. The map in Equation 2.13 stabilizes an $\boldsymbol{R}$-circle if and only if $s = 0$. When $s = 0$, the map stabilizes all the $\boldsymbol{R}$-circles that project to lines in $\boldsymbol{C}$ parallel to $\boldsymbol{R}$.

Lemma 2.3 *Any parabolic Heisenberg automorphism P includes as P_1 in a 1-parameter subgroup $\langle P \rangle = \{P_r | r \in \boldsymbol{R}\}$ of parabolic Heisenberg automorphisms such that $\langle P \rangle$ stabilizes at least one straight line and transitively permutes a family of parallel planes transverse to the line.*

Proof: It suffices to consider P as in Equation 2.12 or 2.13. For Equation 2.12 we make any choice of $\alpha = \log(u)$, then we define $P_r(z,t) = (\alpha^r z, t+r)$. This subgroup stabilizes $\{0\} \times \boldsymbol{R}$ and permutes the planes parallel to $\boldsymbol{C} \times \{0\}$. If $u \neq 1$ then the family of transverse parallel planes is unique. For Equation 2.13 we define P_r as left-multiplication by (r, sr). The line $\{(r, sr) |\ r \in \boldsymbol{R}\}$ is stabilized by the subgroup, and the planes parallel to $i\boldsymbol{R} \times \boldsymbol{R}$ are permuted. (There are other parallel families that are permuted, but this is a canonical choice.) In the case where $s = 0$, the map P_r stabilizes any $\boldsymbol{R}$-circle that projects to a straight line in $\boldsymbol{C}$, which is parallel to $\boldsymbol{R}$. □

Heisenberg automorphisms are generated by the parabolic Heisenberg automorphisms and maps of the form

$$(z,t) \to (\lambda z, |\lambda|^2 t), \qquad \lambda \in \boldsymbol{C} - \{0\}. \tag{2.14}$$

2.3.4 Action on the Siegel Domain

The maps on the Siegel domain corresponding to Equations 2.13 and 2.12 are

$$(z,w) \to (uz, w+i), \qquad |u| = 1 \tag{2.15}$$

$$(z,w) \to (z+1, w+2z+1+is). \tag{2.16}$$

For the purpose of visualizing things, it is sometimes useful to think of $\boldsymbol{C}^2$ as $\mathcal{H} \times \boldsymbol{R}$ and $Z = \mathcal{H} \times \boldsymbol{R}^+$. A nice isomorphism is given by

$$\big((z,t),x\big) \to (z, |z|^2 + it + x). \tag{2.17}$$

Here $(z,t) \in \mathcal{H}$ and $x \in \boldsymbol{R}$. The reader might recognize these coordinates as *horospherical coordinates*. Compare [**E**]. In these coordinates Equations 2.15 and 2.16 become

$$\big((z,t),x\big) \to \big((uz, t+1), x\big), \tag{2.18}$$

$$\big((z,t),x\big) \to \big((1,s) \cdot (z,t), x\big), \tag{2.19}$$

respectively. Note that the $\boldsymbol{R}$-factor just goes along for the ride.

2.4 THE HEISENBERG CONTACT FORM

Here we describe the contact form in $\mathcal{H}$ whose kernel is the contact distribution. We picture this distribution as a kind of pinwheel. The contact plane at the origin is just $\boldsymbol{C} \times \{0\}$. Let's work out the contact plane at the point $(1,0)$. This plane is spanned by the vectors

$$(1,0)\cdot(1,0)-(1,0)=(1,0), \qquad (1,0)\cdot(0,i)-(1,0)=(1,2i).$$

Hence, this contact plane lies in the kernel of the 1-tensor $\alpha = dt - 2dy$. Here we are using (x,y,t) coordinates in $\mathcal{H}$. If we want to extend α to be a 1-form on $\mathcal{H}$ that is invariant under the maps in Equation 2.12 and 2.14, then we must take

$$\alpha = dt - 2xdy + 2ydx. \tag{2.20}$$

By symmetry, our contact distribution is the kernel of this form.

Remark: Our formula differs by a sign from the one in [**G**, p. 55]. The cause of this is a similar sign difference in our group law. The specific calculations we make in $\mathcal{H}$ (in Part 4) are consistent with our conventions here.

If L is any loop in $\boldsymbol{C}$, then we can find a lift $\widehat{L} \subset \mathcal{H}$ such that $\pi(\widehat{L}) = L$. The curve $\widehat{L}$ is not necessarily a closed curve. Here is a quantitative version of this statement.

Lemma 2.4 (Lift Principle I) *The difference in the height between the two endpoints of $\widehat{L}$ is 4 times the signed area enclosed by L.*

Proof: Given our equation for α, the height difference between the endpoints of $\widehat{L}$ is

$$\int_L (2xdy - 2ydx) = 4\int_L \frac{1}{2}(xdy - ydx) = 4\text{Area}$$

by Green's theorem. □

Corollary 2.5 (Lift Principle II) *Let $C \subset \mathcal{H}$ be a finite $\boldsymbol{C}$-circle, with center of mass $c \in \mathcal{H}$. Let $x \in C$. The height difference $v(x) - v(c)$ is 4 times the signed area of the triangle with vertices 0, $\pi(c)$, and $\pi(x)$.*

Proof: Let L be the triangle with the three mentioned vertices. Then $\widehat{L}$ can be taken as the union of two arcs, one contained in the horizontal $\boldsymbol{R}$-circle through x and one contained in the horizontal $\boldsymbol{R}$-circle through c. □

2.5 SOME INVARIANT FUNCTIONS

2.5.1 The Box Product

Given $u = (u_1, u_2) \in \boldsymbol{C}^2$, let $U = (u_1, u_2, u_3)$ be a lift of u to $\boldsymbol{C}^{2,1}$. If u and v are two distinct points in $\boldsymbol{C}^2$, then we take lifts U and V and define

$$U \boxtimes V = (\overline{u_3 v_2 - u_2 v_3},\ \overline{u_1 v_3 - u_3 v_1},\ \overline{u_1 v_2 - u_2 v_1}). \tag{2.21}$$

This vector is such that $\langle U, U \boxtimes V \rangle = \langle V, U \boxtimes V \rangle = 0$. See [**G**, p. 45]. This product is sometimes called the *box product*. It is a Hermitian version of the ordinary cross product.

2.5.2 The Angular Invariant

We follow the treatment given in [**G**, p. 210–214]. Given 3 points $a, b, c \in S^3$, we take lifts $\widetilde{a}, \widetilde{b}, \widetilde{c} \in \boldsymbol{C}^{2,1}$ and consider the triple

$$\langle \widetilde{a}, \widetilde{b}, \widetilde{c} \rangle = \langle \widetilde{a}, \widetilde{b} \rangle \langle \widetilde{b}, \widetilde{c} \rangle \langle \widetilde{c}, \widetilde{a} \rangle \in \boldsymbol{C}. \tag{2.22}$$

It turns out that this quantity always has negative real part and changing the lifts multiplies the number by a positive real constant. Hence,

$$\boldsymbol{A}(a, b, c) = \arg(-\langle \widetilde{a}, \widetilde{b}, \widetilde{c} \rangle) \in \left[\frac{-\pi}{2}, \frac{\pi}{2} \right] \tag{2.23}$$

is independent of lifts and is $PU(2,1)$-invariant. This invariant is known as the *Cartan angular invariant*. It turns out that $\boldsymbol{A} = \pm \pi/2$ iff the points all lie in a $\boldsymbol{C}$-circle and $\boldsymbol{A} = 0$ iff the points all lie in an $\boldsymbol{R}$-circle. In general, two triples (a_1, b_1, c_1) and (a_2, b_2, c_2) are $PU(2,1)$-equivalent iff they have the same angular invariant.

2.5.3 The Cross Ratio

The cross ratio for points in S^3 is known as the *Koranyi-Reimann cross ratio*. See [**G**, p. 224] for details. Here is a real-valued variant. Given 4 distinct points $\alpha, b, c, d \in S^3$, we define

$$[a, b, c, d] = \left| \frac{\langle \widetilde{a}, \widetilde{c} \rangle \langle \widetilde{b}, \widetilde{d} \rangle}{\langle \widetilde{a}, \widetilde{b} \rangle \langle \widetilde{c}, \widetilde{d} \rangle} \right|. \tag{2.24}$$

Here $\widetilde{a}$ is a lift of a to $\boldsymbol{C}^{2,1}$, etc. The quantity doesn't depend on lifts. This quantity is also invariant under the action of $PU(2,1)$.

Here we use the cross ratio to prove a technical result that deals with nested families of sets in S^3. We say that two compact subsets $A_1, A_2 \subset S^3$ are *properly nested* if $A_2 \subset A_1$ and $\partial A_1 \cap \partial A_2 = \emptyset$. In this case we let

$$\delta(A_1, A_2) = \sup[a, b, c, d], \qquad b, c \in A_1, \qquad a, d \in A_2. \tag{2.25}$$

Here $[a, b, c, d]$ is the cross ratio above. It follows from compactness and the fact that $\partial A_1 \cap \partial A_2 = \emptyset$ and that $\delta(A_1, A_2)$ is positive and finite. The quantity $\delta(A_1, A_2)$ records a kind of $PU(2,1)$-invariant notion of the "amount" by which A_2 sits inside A_1.

Lemma 2.6 *Let $\{A_n\}$ denote an infinite nested family of compact sets. Suppose that there are infinitely many indices $\{k_n\}$ such that (A_{k_n}, A_{k_n+1}) are properly nested. Suppose also that there are only finitely inequivalent pairs (A_{k_n}, A_{k_n+1}) mod the action of $PU(2,1)$. Then the intersection $\bigcap A_n$ is a single point.*

Proof: Note the following property of the cross ratio: If $\{a_n\}$, $\{b_n\}$, $\{c_n\}$, and $\{d_n\}$ are sequences of points in S^3 such that $a_n, b_n \to x$, $c_n, d_n \to y$, and $x \neq y$, then $[a_n, b_n, c_n, d_n] \to \infty$. (The denominator converges to 0.)

Let $\delta_k = \delta(A_{k_n}, A_{k_n+1})$. The finiteness mod $PU(2,1)$ guarantees that $\{\delta_k\}$ is uniformly bounded. On the other hand, if the nested intersection of our sets contains at least 2 points $x \neq y$, then we can choose $b_k, c_k \in \partial A_{n_k}$ and $a_k, d_k \in \partial A_{n_k+1}$ such that $a_k, b_k \to x$ and $c_k, d_k \to y$. But then $\delta_k \to \infty$. This is a contradiction. □

2.6 SOME GEOMETRIC OBJECTS

2.6.1 Spinal Spheres

Spinal spheres are studied extensively in [**G**]. We will not really use these objects much, but they play a large motivational role in our construction of $\boldsymbol{R}$-spheres in Part 4.

A *bisector* is defined as a subset of $\boldsymbol{CH}^2$ equidistant between two distinct points. As it turns out, every two bisectors are isometric to each other. A *spinal sphere* is the ideal boundary of a bisector. Every two spinal spheres are equivalent under $PU(2,1)$. This is not necessarily obvious from the definition. Equivalently, a spinal sphere is any set of the form

$$\boldsymbol{B}^{-1}\big((\boldsymbol{C} \times \{0\}) \cup \infty\big). \tag{2.26}$$

Here $\boldsymbol{B}$ is a Heisenberg stereographic projection. Thus, $\Sigma_0 = (\boldsymbol{C} \times \{0\}) \cup \infty$ is a model in $\mathcal{H}$ for a spinal sphere.

Here are some objects associated to Σ_0.

- Σ_0 has a singular foliation by $\boldsymbol{C}$-circles. The leaves are given by $C_r \times \{0\}$, where C_r is a circle of radius r centered at the origin. The singular points are 0 and ∞. We call this the $\boldsymbol{C}$*-foliation*.

- Σ_0 has a singular foliation by $\boldsymbol{R}$-circles. The leaves are horizontal lines through the origin. The singular points are again 0 and ∞. We call this the $\boldsymbol{R}$*-foliation*. The singular points 0 and ∞ are called the *poles*.

- The *spine* of Σ_0 is defined as the $\boldsymbol{C}$-circle containing the poles. In our case, the spine is $(\{0\} \times \boldsymbol{R}) \cup \infty$. Note that the spine of Σ_0 only intersects Σ_0 at the singular points.

Any other spinal sphere has the same structure. The two foliations on a spinal sphere look topologically like lines of latitude and longitude on a globe. A spinal sphere is uniquely determined by its poles. However, the spine is not enough to determine the spinal sphere. Two spinal spheres are *cospinal* if they have the same spine.

2.6.2 Horotubes

Horotubes are peculiar to this monograph. Let $P \in PU(2,1)$ be a parabolic element, with fixed point p. Recall from Chapter 1 that a P-horotube is an open P-invariant subset $T \subset S^3 - p$ such that $T/\langle P \rangle$ has a compact complement in $(S^3 - \{p\})/\langle P \rangle$. We say that T is *based* at p. We can also consider horotubes as subsets of $\mathcal{H}$, when their basepoint is ∞.

Say that a horotube H is *nice* if ∂H is a smooth cylinder and H is stabilized by a 1-parameter parabolic subgroup $\{P_r\}$. We call the quotient H/P_1 a *nice horocusp*. Here P_1 is one of the members of $\{P_r\}$. A nice horocusp is homeomorphic to a torus cross a ray.

Lemma 2.7 *Let H be a P-horotube based at p. There is a nice P-horotube $H' \subset H$ such that $(H - H')/\langle P \rangle$ has compact closure in $(S^3 - p)/\langle P \rangle$.*

Proof: Let $\{P_r\}$ be a 1-parameter subgroup containing P. Let L be a line stabilized by $\{P_r\}$, and let Π be one of the planes in a parallel family permuted by $\{P_r\}$. By compactness, every point of ∂H is at most N units from L, for some N. Let $K \subset \Pi$ be a closed disk containing $L \cap \Pi$, chosen so that its boundary is more than N units from L. Then $H' = \bigcup P_r(\Pi - K)$ is a nice horotube contained in H. The intersection $(H - H') \cap \Pi$ is contained in K, and so $(H - H')/\langle P \rangle$ has compact closure in $(S^3 - p)/\langle P \rangle$. □

Let H be a nice horotube. There is a natural $\{P_r\}$-invariant retraction $H \to \partial H$. One first makes this retraction in $H \cap \Pi$ and then extends using $\{P_r\}$. Within $H \cap \Pi$ the retraction takes place along rays emanating from the center of mass of $\partial H \cap \Pi$. The fibers of the retraction are rays.

A *horotube function* for P is a smooth and P-invariant function $f : S^3 - \{p\} \to [0, \infty)$ whose superlevel sets

$$\langle f \rangle_s = f^{-1}(s, \infty) \tag{2.27}$$

are all horotubes. We will discuss these in detail in Part 2. For now, we just give a quick example: When P has the form given in Equation 2.12 the function $f(z,t) = |z|$ is a horotube function.

Chapter Three

Topological Generalities

3.1 THE HAUSDORFF TOPOLOGY

If X is a metric space, then we can equip the set of compact subsets of X with the *Hausdorff metric*. The distance between two compact $K_1, K_2 \subset X$ is defined as the infimal ϵ such that K_j is contained in the ϵ-tubular neighborhood of K_{3-j} for $j = 1, 2$. This metric induces the *Hausdorff topology* on closed subsets of X. A sequence $\{S_n\}$ of closed subsets converges to S if, for every compact $K \subset X$, the Hausdorff distance between $S_n \cap K$ and $S \cap K$ converges to 0 as $n \to \infty$.

One of the cases of interest to us is the case where $X = PU(2,1)$. We equip $PU(2,1)$ with a left-invariant Riemannian metric and put the Hausdorff topology on the space of its closed subsets. Any choice of such a metric induces the same topology. Let $\{P_n\}$ be a sequence of elements of $PU(2,1)$, and let P be a parabolic element of $PU(2,1)$. We say that $P_n \to P$ *algebraically* if P_n converges to P as an element of $PU(2,1)$. This is the usual kind of convergence.

Let $\langle P_n \rangle$ be the group generated by P_n and let $\langle P \rangle$ be the group generated by P. We say that P_n converges *geometrically* to P if the following occur.

- P_n converges to P algebraically.
- The set $\langle P_n \rangle$ converges in the Hausdorff topology to the set $\langle P \rangle$.
- $\langle P_n \rangle$ acts freely on S^3 when this group has finite order.

When P_n is elliptic of order m_n, we define the *true size* of an exponent e_n (as in $P_n^{e_n}$) to be the absolute value of the representative of e_n mod m_n, which lies between $-m_n/2$ and $m_n/2$. Thus, the true size of the exponent 31 for an element of order 41 is 10. In all other cases, the true size e_n is just its absolute value.

Lemma 3.1 *Suppose that $P_n \to P$ geometrically. Then there is no sequence of exponents $\{e_n\}$ whose true size is unbounded such that $\{P_n^{e_n}\}$ is a bounded subset of $PU(2,1)$.*

Proof: We will consider the case where $\langle P_n \rangle$ is an elliptic subgroup. The other cases are similar and easier.

We put some left-invariant Riemannian metric on $PU(2,1)$. There is some $\delta > 0$ such that every element of $\langle P \rangle$ is at least δ from the identity element.

Since the subgroup $\langle P_n \rangle$ converges to $\langle P \rangle$ in the Hausdorff topology, we can pass to a subsequence so that $P_n^{e_n} \to P^k$, for some fixed exponent k. But then $Q_n = P_n^{e_n - k}$ converges to the identity. On the other hand, Q_n is not the identity because $0 < |e_n - k| < m_n$ for n large. But then the sequence of elements $Q_n, Q_n^2, ...$ moves away from the identity element at a very gradual rate. Hence, we can find some exponent d_n so that the distance from $Q_n^{d_n}$ to the identity converges to $\delta/2$. But then $\langle P_n \rangle$ certainly does not Hausdorff-converge to $\langle P \rangle$. □

Remark: Our proof above did not use the condition in the elliptic case where $\langle P_n \rangle$ acts freely on S^3. This condition is present for another purpose entirely.

3.2 SINGULAR MODELS AND SPINES

In this section, we roughly follow [**Mat**] but work in the smooth category.

First, we describe a model for the simplest kind of singular space. Let $2 \leq k \leq n+1$. Let $\Pi \subset \boldsymbol{R}^{n+1}$ denote the hyperplane consisting of points whose first k coordinates sum to 0. Let $V_j \subset \boldsymbol{R}^{n+1}$ denote the set of points where the jth coordinate is largest. Let

$$M(k,n) = \bigcup_{i=1}^{k} (\Pi \cap \partial V_k). \tag{3.1}$$

$M(2,1)$ is a point, and for $k \geq 3$, the space $M(k, k-1)$ is a subset of a copy of $\boldsymbol{R}^{k-1}$ and it is the cone on the $(k-3)$-skeleton of a regular $(k-1)$-dimensional simplex. Moreover,

$$M(k,n) = \boldsymbol{R}^{n+1-k} \times M(k, k-1). \tag{3.2}$$

In the 3-dimensional case $n = 3$, the three examples of interest to us are

- $M(2,3) = \boldsymbol{R}^2$;
- $M(3,3) = \boldsymbol{R} \times Y$, where Y is a union of 3 rays meeting at a point;
- $M(4,3)$, the cone on the 1-skeleton of a tetrahedron.

Let M be a smooth noncompact 3-manifold. We say that a *good spine* on M is a complete Riemannian metric on M together with an embedded 2-complex $\Sigma \subset M$ such that we have the following.

- M deformation retracts onto Σ.
- Σ is partitioned into a union of points, smooth arcs, and smooth 2-cells.
- Every point $x \in \Sigma$ has a neighborhood U and a diffeomorphism $U \to \boldsymbol{R}^3$ which carries $\Sigma \cap U$ to one of $M(k,3)$ for $k = 2, 3, 4$.

- In the case where $k = 3, 4$, the map $U \to \boldsymbol{R}^3$ is an isometry on a smaller neighborhood of x.

These objects, considered in the PL category, are called *special spines* in [**Mat**].

Lemma 3.2 *Every noncompact* 3*-manifold* M *has a good spine.*

Proof: In [**Mat**, Theorem 1.1.13], it is shown that every noncompact triangulated manifold M has the polyhedral version of a smooth special spine. The construction involves a finite number of basic cut-and-paste operations. If M is smooth and we imitate the construction starting with a smooth triangulation of M, then we produce a spine for M with all the topological properties above, except that the cells are piecewise smooth rather than smooth. But then we can just take a smooth approximation to each cell.

To get the Riemannian metric, we note that the locus of singular points of Σ is just a graph. The vertices correspond to $M(4, 3)$, and the edges correspond to $M(3, 3)$. We can build a diffeomorphic copy of a neighborhood of this graph in Σ just by gluing together appropriate subsets of the Euclidean models *via* Euclidean isometries. We then choose a diffeomorphism from the abstract neighborhood to an actual neighborhood and push forward the Euclidean metric. We extend this Riemannian metric to all of Ω using a partition of unity. □

For the sake of exposition, we give a self-contained proof of Lemma 3.2 in a fairly broad special case that includes many of the examples of interest to us—in particular, the Whitehead link complement. Say that an *ideal tetrahedron* is a solid tetrahedron with its vertices deleted. Suppose that M is constructed by gluing together finitely many ideal tetrahedra. (In spite of our terminology, M need not be a hyperbolic manifold.) The regular ideal tetrahedron τ_0 has a canonical 2-complex σ_0 onto which it retracts. Namely, σ_0 is the set of points in τ_0 that are equidistant to at least two of the vertices. If M is partitioned into ideal tetrahedra τ_1,...,τ_n, then we define σ_j by picking an affine isomorphism from τ_j to τ_0 and pulling back σ_j. The union $\Sigma' = \bigcup \sigma_j$ has all the topological properties of a smooth special spine except that the cells are piecewise smooth rather than smooth. (The point is that σ_i and σ_j fit together continuously across a common face of τ_i and τ_j but not necessarily smoothly.) We then replace each cell by a smooth approximation. The Riemannian metric is constructed as above.

3.3 A TRANSVERSALITY RESULT

Let V be an open subset of $\boldsymbol{R}^n$ and let $F_1, ..., F_k$ be a collection of smooth functions on V. We say that $x \in V$ is *good* with respect to the collection if

the following implication holds:

$$F_1(x) = \cdots = F_k(x) \quad \Longrightarrow \quad \dim\Big(\text{Hull}\big(\nabla F_1(x), \ldots, \nabla F_k(x)\big)\Big) = k - 1. \tag{3.3}$$

Here Hull is the convex hull operation. We mean to take the convex hull of the *endpoints* of the gradients ∇F_j. We say that V is good with respect to the collection if every point is good. Note that our definition only has content if $k \geq 2$. Also Equation 3.3 is impossible unless $k \leq n+1$. These are the same constraints placed on our singular models $M(k,n)$, and we will see in Chapter 11 that good collections of functions are the building blocks for singular spaces that have the local structure of our models discussed above.

Here we prove a general transversality result that is used in Section 11.2 when we want to extend our spine for Ω/Γ into the 4-manifold $\boldsymbol{CH}^2/\Gamma$.

Lemma 3.3 *Let E_1,...,E_k be smooth functions on an open set $V \subset \boldsymbol{R}^n$. There exist arbitrarily small perturbations F_i of E_i such that V is good with respect to F_1,...,F_k.*

In Lemma 3.3, the term *perturbation* means that there is a single $\epsilon > 0$ so that all partial derivatives of F_j are within ϵ of the corresponding partial derivatives of E_j.

Lemma 3.3 has two cases, which we will establish in turn.

3.3.1 Case 1

Suppose that $k \geq n+2$. For any point $c = (c_1,...,c_k) \in \boldsymbol{R}^k$ and $y \in V$, define

$$\Theta(c, y) = \big(E_1(y) + c_1, \ldots, E_k(y) + c_k\big).$$

Here Θ is a map from $\boldsymbol{R}^k \times V$ to $\boldsymbol{R}^k$. Let L be the line spanned by the vector (1,...,1). The linear differential $d\Theta$ is clearly a surjection at each point. Hence, $\Theta^{-1}(L)$ is

$$(n + k) - k + 1 = n + 1$$

dimensional. Since $k \geq n + 2$, the projection of $\Theta^{-1}(L)$ into $\boldsymbol{R}^k$ must avoid points of $\boldsymbol{R}^k$ that are arbitrarily close to 0. In other words, there are points $c \in \boldsymbol{R}^k$ arbitrarily close to 0 such that

$$L \cap \Theta(\{c\} \times V) = \emptyset.$$

For arbitrarily small such values of c, we can take $F_j = E_j + c_j$. Then there is a pair of indices $i \neq j$ such that $F_i(x) \neq F_j(x)$ for all $x \in V$. This deals with the case $k \geq n + 2$.

3.3.2 Case 2

Now suppose that $k \in \{2, ..., n + 1\}$. Given $a = (a_1,...,a_k) \in \boldsymbol{R}^k$, $b \in (b_1,...,b_k) \in (\boldsymbol{R}^n)^k$, and $y \in V$, we define

$$\Theta(a, b, y) = (F_1, \ldots, F_k, G_1, \ldots, G_k),$$

$$F_j(y) = E_j(y) + a_j + b_j \cdot y, \qquad G_j(y) = \nabla E_j(y) + b_j. \tag{3.4}$$

Note that $\nabla F_j = G_j$. The map Θ is a smooth map from an open subset of $\boldsymbol{R}^A$ into $\boldsymbol{R}^B$, where

$$\boldsymbol{R}^A = \boldsymbol{R}^k \times \boldsymbol{R}^{nk} \times \boldsymbol{R}^n, \qquad \boldsymbol{R}^B = \boldsymbol{R}^k \times \boldsymbol{R}^{nk}. \tag{3.5}$$

Consider $d\Theta_z$ at some point $z \in V$.

- By varying just the a coordinates, we see that the image of $d\Theta_z$ contains $\boldsymbol{R}^k \times \{0\} \subset \boldsymbol{R}^B$.
- Let $\pi : \boldsymbol{R}^B \to \boldsymbol{R}^{nk}$ be the projection onto the last nk coordinates. By varying the b coordinates and keeping everything else fixed, we see that $\pi \circ d\Theta_z$ is onto $\boldsymbol{R}^{nk}$.

These two items imply that $d\Theta_z$ is surjective. Hence, Θ is a submersion.

We represent points in $\boldsymbol{R}^B$ by tuples of the form $(x_1,...,x_k, v_1,...,v_k)$, where $x_j \in \boldsymbol{R}$ and $v_j \in \boldsymbol{R}^n$. Let L denote those tuples for which $x_1 = \cdots = x_k$ and the vectors $\{v_1, ..., v_k\}$ are not in general position in $\boldsymbol{R}^n$. There is an $n(k-1)$ set of ways to choose the vectors $v_1,...,v_{k-1}$, and generically these vectors will span a $(k-2)$-plane. Hence, each generic choice of $v_1,...,v_{k-1}$ yields a $(k-2)$-dimensional set of choices for v_k that leads to $v_1,...,v_k$ not being in general position. Adding the single dimension for the condition $x_1 = \cdots = x_k$, we see that

$$\dim(L) = 1 + n(k-1) + (k-2). \tag{3.6}$$

Since Θ is a submersion, we get from Equations 3.5 and 3.6 that

$$\dim\left(\Theta^{-1}(L)\right) = A - B + \dim(L) = kn + k - 1 < \dim(\boldsymbol{R}^k \times \boldsymbol{R}^{nk}).$$

Hence, there are points of the form $(a, b) \in \boldsymbol{R}^k \times \boldsymbol{R}^{nk}$ that are arbitrarily close to 0, such that $(\{a\} \times \{b\} \times V) \cap \Theta^{-1}(L) = \emptyset$. This to say that

$$\Theta(\{a\} \times \{b\} \times V) \cap L = \emptyset.$$

We choose such an (a, b) and replace E_j with the map F_j. By construction, the functions $F_1,...,F_k$ have the desired properties. □

3.4 DISCRETE GROUPS

See [**CG**] for some foundational material on complex hyperbolic discrete groups. See [**B**] and [**M**] for the real hyperbolic case.

$PU(2,1)$ has the topology it inherits from its description as a smooth Lie group. A subgroup $\Gamma \subset PU(2,1)$ is *discrete* if the identity element of Γ is isolated from all other elements of Γ. A discrete group Γ acts *properly discontinuously* on $\boldsymbol{CH}^2$ in the usual sense: The set $\{g \in \Gamma |\ g(K) \cap K \neq \emptyset\}$ is finite for any compact $K \subset \boldsymbol{CH}^2$. The quotient $\boldsymbol{CH}^2/\Gamma$ is called a *complex hyperbolic orbifold*.

Let $\Gamma \subset PU(2,1)$ be a discrete group. The *limit set* Λ of Γ is defined as the accumulation set, on S^3, of an orbit Γx for $x \in \boldsymbol{CH}^2$. This definition is independent of the choice of x. The *domain of discontinuity* of Γ is defined as $\Omega = S^3 - \Lambda$. This set is also called the *regular set*. Γ acts properly discontinuously on Ω. The quotient Ω/Γ is called the *orbifold at infinity* in general. When Γ acts *freely* on Ω (that is, with no fixed points), then Ω/Γ is a manifold, and it is called the *manifold at infinity*. Passing to a finite index subgroup of Γ does not change Λ or Ω.

If G acts properly discontinuously on a space X, then we say that a *fundamental domain* for the action of G is a subset $F \subset X$ such that F is the closure of its interior, $X = \bigcup_{g \in G} g(F)$, and $F \cap g(F)$ is disjoint from the interior of F, for all nontrivial $g \in G$. In this case, X is *tiled* by the translates of F, and every compact subset of X intersects only finitely many translates of F.

Lemma 3.4 *If Γ is a discrete group of isolated type (as in the HST), then Ω/Γ is a manifold.*

Proof: It suffices to show that Γ acts freely on Ω. Suppose there is some $g \in \Gamma$ and some $p \in \Omega$ such that $g(p) = p$. Since Γ acts properly discontinuously on Ω, we must have that g is an elliptic element of finite order. But then g also fixes some $q \in \boldsymbol{CH}^2$. But then g fixes the complex slice containing p and q. This contradicts the isolated type condition. □

We end this section with a brief discussion of the porous limit set condition, which appears in the definition of a horotube group. For the case of real hyperbolic discrete groups, the condition is equivalent to the statement that the convex hull H of the limit set is within a bounded distance from its boundary ∂H. We think this is also true for hyperbolic discrete groups, but we haven't tried to prove it. We think that Γ has a porous limit set iff it is geometrically finite and has no full rank cusps, but we haven't tried to prove this either. See [**Bo**] for a definition of geometrical finiteness.

To explain the connection to full rank cusps, we note the following result: If Γ is a discrete group whose regular set is nonempty and whose limit set has more than one point, then Λ is not porous. To see this, note that we can normalize so that the rank-3 cusp fixes ∞ in $\mathcal{H}$ and has an ϵ-dense orbit as measured in the Carnot metric. But then the limit set is ϵ-dense.

The porosity condition also comes up in [**McM**, p. 20], in the context of Julia sets of complex maps.

3.5 GEOMETRIC STRUCTURES

[**T0**] and [**CEG**] give basic information about geometric structures. Let X be a homogeneous space, and let $G : X \to X$ be a Lie group acting transitively on X by analytic diffeomorphisms. A (G, X)-*structure* on M

is an open cover $\{W_\alpha\}$ of M and a system $\{(W_\alpha, f_\alpha)\}$ of coordinate charts $f_\alpha : W_\alpha \to X$ such that the overlap function $f_\alpha \circ f_\beta^{-1}$ coincides with an element of G when restricted to sets of the form $f_\beta(W)$, where W is a connected component of $W_\alpha \cap W_\beta$. Two (G, X)-manifolds are *isomorphic* if there is a diffeomorphism between them that is locally in G, when measured in the coordinate charts. Here are two examples.

- When $G = PU(2,1)$ and $X = \boldsymbol{CH}^2$, a (G, X)-manifold is the same as a manifold locally isometric to $\boldsymbol{CH}^2$.
- When $G = PU(2,1)$ and $X = S^3$, a (G, X)-manifold is known as a *spherical CR manifold*, as in Section 1.5. In particular, if Γ is a complex hyperbolic discrete group that acts freely on its nonempty domain of discontinuity Ω, then Ω/Γ canonically has the structure of a spherical CR manifold.

3.6 ORBIFOLD FUNDAMENTAL GROUPS

See [**T0**] for a very general existence theorem about orbifold fundamental groups. Here we will establish a special case, by hand, which suffices for our purposes.

Let $G \subset PU(2,1)$ be a finite group that acts freely on S^3 and has $0 \in \boldsymbol{CH}^2$ as its only fixed point. For any r, let B_r be the metric ball of radius r about 0, and let $Q(G, r) = B_r/G$. Then $Q(G, r)$ is the cone on a 3-manifold, and $Q(G, r) - \{0\}$ is a complex hyperbolic manifold.

A complete metric space M is a *complex hyperbolic orbifold with isolated singularties* if there are finitely many points p_1,...,$p_k \in M$ such that $M - \bigcup p_j$ is a complex hyperbolic manifold and sufficiently small metric balls around each p_j are isometric to $Q(G_j, r_j)$ for suitable choices of G_j and r_j.

Lemma 3.5 *Let M be a complex hyperbolic orbifold with isolated singularities. Then there exists a discrete group $\Gamma \subset PU(2,1)$ such that $M = \boldsymbol{CH}^2/\Gamma$.*

Proof: Let $Q_{j,r} \subset M$ be the metric ball of radius r about p_j, with r chosen so that all these metric balls are disjoint. Let M_r be the metric closure of $M - \bigcup Q_{j,r}$. Then M_r is a complex hyperbolic manifold with k disjoint boundary components. Let $\widetilde{M}_r$ be the universal cover of M_r. Let π be the covering map and let Γ be the covering group. The metric on M_r lifts to a complete metric on $\widetilde{M}_r$. In its interior, $\widetilde{M}_r$ is locally isometric to $\boldsymbol{CH}^2$, and every boundary component of $\widetilde{M}_r$ is locally isometric to $\partial B_r(0) \subset \boldsymbol{CH}^2$.

Let Σ be a boundary component of $\widetilde{M}_r$. Since Σ covers $\partial Q_{j,r}$ for some j, there is a finite group $G_j' \subset G_j$ acting freely on S^3 such that Σ is isometric to $\partial B_r(0)/G_j'$. Since $\widetilde{M}_r$ is simply connected, there is a local isometry $\phi : \widetilde{M}_r \to \boldsymbol{CH}^2$. The restriction $\phi|_\Sigma$ maps Σ to $\partial B_r(0)$ by a local isometry.

Hence, there is a continuous map from $\partial B_r(0)/G'_j$ to $\partial B_r(0)$ that is a local isometry. This is only possible if G'_j is trivial. Hence, Σ is globally isometric to $\partial B_r(0)$, and the restriction of π to each boundary component is equivalent to the universal covering map from $\partial B_r(0)$ onto $\partial Q_{j,r}$ for some j. Call this the *boundary covering property*.

We form a new space $\widetilde{M}$ by isometrically gluing $B_r(0)$ onto each boundary component of $\widetilde{M}_r$. The covering group Γ acts isometrically on $\widetilde{M}$, and by the boundary covering property, $\widetilde{M}/\Gamma$ is isometric to M. By construction $\widetilde{M}$ is a complete, simply connected metric space, locally isometric to $\boldsymbol{CH}^2$. Hence, $\widetilde{M}$ is isometric to $\boldsymbol{CH}^2$. □

3.7 ORBIFOLDS WITH BOUNDARY

A set $X = M \cup M^\infty$ is a *complex hyperbolic manifold with boundary* if M is a complex hyperbolic manifold and M^∞ is a spherical CR manifold. The two structures should be compatible in that they come from a single system of coordinate charts into $\boldsymbol{CH}^2 \cup S^3$ with transition functions in $PU(2,1)$.

When M is a complex hyperbolic orbifold with isolated singularities, we let $M_{\rm reg}$ denote M minus its singular points. We call $X = M \cup M^\infty$ a *capped orbifold with isolated singularities* if M is a complex hyperbolic orbifold with isolated singularities and $X_{\rm reg} = M_{\rm reg} \cup M^\infty$ is a complex hyperbolic manifold with boundary. We say that a *cap* in X is an open subset $C \subset X_{\rm reg}$ that is contained in a single coordinate chart.

Lemma 3.6 *Let $X = M \cup M^\infty$ be a capped orbifold with isolated singularities. Suppose that $\gamma \in M$ is a geodesic ray that exits every compact subset of M. Then γ accumulates at a single point of M^∞.*

Proof: We set $M = \boldsymbol{CH}^2/\Gamma$ and let $\pi : \boldsymbol{CH}^2 \to M$ be the quotient map. If γ has more than one accumulation point, then we can find caps C_1 and C_2, with closure $(C_1) \subset C_2$, such that γ enters C_1 infinitely often and exits C_2 infinitely often. Choosing C_1 and C_2 sufficiently small, we can find lifts $\widetilde{C}_1$ and $\widetilde{C}_2$ such that closure $(\widetilde{C}_1) \subset \widetilde{C}_2$ and the covering map $\pi : \widetilde{C}_j \to C_j$ is a homeomorphism. The properties of γ imply that infinitely many distinct lifts $\{\widetilde{\gamma}_k\}$ of γ intersect $\widetilde{C}_1$ and exit $\widetilde{C}_2$. Hence, there are infinitely many lifts of γ that have Euclidean diameter at least ϵ, a contradiction. □

Lemma 3.7 *Let $X = M \cup M^\infty$ be a capped orbifold with isolated singularities. Let Γ be the orbifold fundamental group for M. Then Γ has a nonempty domain of discontinuity Ω, and Ω/Γ is isomorphic to M^∞.*

Proof: To see that Ω is nonempty, choose a nontrivial cap $C \subset X$ that intersects M_∞ in an open set. Let $\widetilde{C}$ be a lift of C in $\boldsymbol{CH}^2$. Let $\widetilde{U}$ be

the accumulation set of $\widetilde{C}$ on S^3. Then $\widetilde{U}$ is an open subset of S^3. By construction $g(\widetilde{C}) \cap \widetilde{C} = \emptyset$ for all nontrivial $g \in \Gamma$. But then $g(\widetilde{U}) \cap \widetilde{U} = \emptyset$ as well. This shows that $\widetilde{U} \subset \Omega$. Hence, Ω is not empty.

The space $X' = (\boldsymbol{CH}^2 \cup \Omega)/\Gamma$ also has the structure of a capped orbifold. This follows from the fact that Γ acts freely and properly discontinuously on Ω. We write $X' = M' \cup M'_\infty$, with $M' = \boldsymbol{CH}^2/\Gamma$ and $M'_\infty = \Omega/\Gamma$. There is an isometric map $i : M \to M'$. We just need to see that i extends to a homeomorphism from M_∞ to M'_∞. Given a point $x \in M_\infty$, there is a geodesic ray γ that accumulates on x. Then $i(\gamma)$ accumulates at a unique point $x' \in M'_\infty$, by Lemma 3.6. We define $i(x) = x'$. Any two geodesic rays that accumulate to x are asymptotic in the sense that, for any $\epsilon > 0$, all but a compact part of one ray is contained in the ϵ neighborhood of the other. Asymptotic rays in M' accumulate at the same point, so the extension is well defined.

To show continuity, we choose a basepoint $y \in M$. If $x_1, x_2 \in M_\infty$ are close, then there are two geodesic rays γ_1, γ_2 emanating from y, which stay close together for a long distance before accumulating at x_1 and x_2 respectively. But then $i(\gamma_1)$ and $i(\gamma_2)$ have this same property in M'. Hence, $i(x_1)$ and $i(x_2)$ are close. In short, our extension is continuous. We can make all the same constructions, reversing the roles of M and M'. Hence, our extension is a homeomorphism from M_∞ to M'_∞. $\square$

Chapter Four

Reflection Triangle Groups

4.1 THE REAL HYPERBOLIC CASE

In this chapter we elaborate on the discussion in Section 1.4. Here is a well-known result from real hyperbolic geometry.

Lemma 4.1 *Let $\zeta = (\zeta_0, \zeta_1, \zeta_2)$ be a triple of positive numbers such that $\sum \zeta_i^{-1} < 1$. Then there exists a geodesic triangle $T_\zeta \subset \boldsymbol{H}^2$ whose angles are $\pi/\zeta_0, \pi/\zeta_1$, and π/ζ_2. This triangle is unique up to isometry.*

Proof: Let $\theta_i = \pi/\zeta_i$. The easiest way to see the existence is as follows: Start with two lines L_1 and L_2 through the origin in the disk model that make an angle of θ_0. Now take a line L_0 that makes an angle of θ_1 with L_2. If $L_0 \cap L_2$ is very close to the origin, then the angle sum of the triangle made by the three lines is close to π, and hence, L_0 makes an angle with L_1, which exceeds θ_2. On the other hand, we can adjust L_0 so that L_0 and L_1 are asymptotic and, hence, make an angle of 0 with each other. By the intermediate value theorem, we can find a location for L_0 that makes the angle at $L_0 \cap L_1$ exactly θ_2. The triangle formed by our three lines is the desired one. For the uniqueness, note that the sum of the angles of a geodesic hyperbolic quadrilateral is less than 2π. This property implies that the angle made by L_0 and L_1 is monotone in the position of L_2. □

As an extra case, we can allow the possibility that some or all of the ζ_i are infinite. In this case, the relevant lines will intersect on the ideal boundary, and the corresponding angle is measured as $0 = \pi/\infty$. The same existence and uniqueness results hold, but we omit the proof.

Let G'_ζ denote the group generated by reflections in the sides of the triangle T_ζ. If the ζ_i are all integers or infinite, then G'_ζ is a discrete group. The intuitive way to see this is that we can form a tiling of the hyperbolic plane by isometric copies of T_ζ, with 3 kinds of vertices. Around the ith kind of vertex there are $2\zeta_i$ triangles. The group G'_ζ then acts isometrically as a subgroup of the group of symmetries of the tiling. (Usually G'_ζ equals the group of symmetries, but sometimes the full group of symmetries is slightly larger.) A formal discretness proof usually involves the Poincaré polyhedron theorem. (See [**R**], for example.) The discreteness-proving machinery we develop in this monograph will also suffice, but this is overkill.

4.2 THE ACTION ON THE UNIT TANGENT BUNDLE

The *unit tangent bundle* to $\boldsymbol{H}^2$ is the set of pairs (p, v), where $p \in \boldsymbol{H}^2$ is a point and v is a unit vector based at p. We denote this space by $T_1\boldsymbol{H}^2$. Any isometry of $\boldsymbol{H}^2$ automatically extends to an action of $T_1\boldsymbol{H}^2$. In this section we will make some remarks about the action of G'_ζ on $T_1\boldsymbol{H}^2$. In particular, we will see how Dehn surgery comes up in a natural way.

To make our discussion easier, we will work with the subgroup G_ζ of G'_ζ consisting of even-length words. This group is nicer because the quotient

$$Q_\zeta = T_1\boldsymbol{H}^2/G_\zeta$$

is a manifold. Consider first the case of $G := G_{\infty,\infty,\infty}$. In this case $\boldsymbol{H}^2/G$ is the 3-punctured sphere S. The space $T_1\boldsymbol{H}^2/G$ is some circle bundle over S. However, we can certainly produce a unit vector field on S, trivializing the bundle. Hence, $Q := Q_{\infty,\infty,\infty} = S \times S^1$. Note that Q has 3 torus ends.

Suppose now we keep $\zeta_1 = \zeta_2 = \infty$, but we make ζ_0 a finite but large integer. We can adjust our triangle $\widehat{T} := T_{\zeta_0,\infty,\infty}$ so that its two ideal points coincide with the ideal points of $T := T_{\infty,\infty,\infty}$ and the third point is close (in the Euclidean sense) to the third ideal point of T. If we increase ζ_0 unboundedly, then the group $\widehat{G} := G_{\zeta_0,\infty,\infty}$ corresponding to $\widehat{T}$ will converge nicely to G, in the sense of the HST. In fact, this kind of convergence is the model for the kind that occurs in the HST.

To understand the quotient $\widehat{Q}$, we try to see how $\widehat{Q}$ fibers over $\boldsymbol{H}^2/\widehat{G}$. We use the projection $T_1\boldsymbol{H}^2 \to \boldsymbol{H}^2$, which descends to the quotients. If we cut off a small neighborhood $\overline{D}$ of the ζ_0-vertex of $\boldsymbol{H}^2/\widehat{G}$, then we again have topologically a 3-punctured sphere $\overline{S}$. The preimage $\widehat{S}$ in $\widehat{Q}$ must again be topologically $S \times S^1$. The difference

$$\widehat{\Sigma} = \widehat{Q} - \widehat{S}$$

is the set which projects to $\overline{D}$. Note that $\overline{D}$ is the quotient of a disk D in $\boldsymbol{H}^2$ by a rotation of order ζ_0. But then $\widehat{\Sigma}$ is the quotient of the solid torus $T_1D = D \times S^1$ by this same rotation. This rotation acts freely on T_1D and hence $\widehat{\Sigma}$ is a solid torus. In this way we see that $\widehat{Q}$ is obtained from Q by performing a Dehn filling. If we let all of the ζ_i be finite then we see Q_ζ as a Dehn filling on all 3 cusps of $Q_{\infty,\infty,\infty}$.

The Dehn filling we have just seen is a toy model for the kind of filling that occurs in the HST. In the next section we will place the example here in a complex hyperbolic context and see it as a special case of the HST.

4.3 FUCHSIAN TRIANGLE GROUPS

Here we describe the simplest kinds of complex reflection triangle groups, the *Fuchsian* ones. It is a basic principle that any real projective transformation of $\boldsymbol{RP}^2$ that preserves the unit disk $\boldsymbol{RH}^2 = \boldsymbol{CH}^2 \cap \boldsymbol{R}^2$ extends to an element of $PU(2,1)$ preserving $\boldsymbol{CH}^2$. This is really just an application of

the fact that a rational function defined over $\boldsymbol{R}$ automatically extends to a rational function defined over $\boldsymbol{C}$. In particular, the generators of G'_ζ extend to complex hyperbolic isometries. The generators in question are complex reflections. Their fixed points are the complex lines extending the real lines fixed by the original elements.

We call the above representation of G'_ζ into $PU(2,1)$ the *Fuchsian* ζ-*triangle group*, and we denote it by Γ'_ζ. We let Γ_ζ be the even subgroup. These are really the same groups as the ones considered in the previous section, assuming that our model of $\boldsymbol{H}^2$ is the Klein model. One justification for the name *Fuchsian* is that the limit set of the group is a round $\boldsymbol{R}$-circle $S^3 \cap \boldsymbol{R}^2$. A second justification is that it is possible to deform Γ_ζ within $PU(2,1)$ in a nontrivial way, in the same way that it is possible to deform a Fuchsian surface group in $PSL_2(\boldsymbol{R})$ to a quasi-Fuchsian one.

We would like to understand the quotient Ω/Γ_ζ, where Ω is the domain of discontinuity of Γ_ζ. Before we do this, we study more about the geometry of $\boldsymbol{CH}^2$. Any element of $PU(2,1)$ that stabilizes both $\boldsymbol{R}^2$ and the origin in $\boldsymbol{R}^2$ also stabilizes the orthogonal plane $i\boldsymbol{R}^2$. This follows from the fact that all such maps are complex linear isometries of $\boldsymbol{C}^2$. Indeed, any rotation of $\boldsymbol{R}^2$ about the origin also rotates $i\boldsymbol{R}^2$ "in the same way," much as a rotation of $\boldsymbol{H}^2$ rotates the tangent plane at a fixed point "in the same way" that it rotates the space.

If $p \in \boldsymbol{RH}^2$, then we define $\Pi_p = g(i\boldsymbol{RH}^2)$, where $g \in PU(2,1)$ stabilizes $\boldsymbol{R}^2$ and maps 0 to p. From what we have said about the stabilizers of 0, we see that our definintion of Π_p is independent of the choice of g. In this way we recognize $\boldsymbol{CH}^2$ as a disk bundle over $\boldsymbol{RH}^2$, with the fibers given by the planes $\{\Pi_p\}$.

Let $T\boldsymbol{H}^2$ denote the open unit disk bundle of $\boldsymbol{H}^2$. We can identify the fibers of $T\boldsymbol{H}^2$ with copies of the hyperbolic disk. There is a tautological map

$$I : T\boldsymbol{H}^2 \to \boldsymbol{CH}^2.$$

I is characterized, up to postcomposition by a rotation, by the following properties.

- I maps the 0-section of $T\boldsymbol{H}^2$ to $\boldsymbol{RH}^2$ isometrically, and $I(0) = 0$.
- I maps each fiber of $T\boldsymbol{H}^2$ isometrically to one of the planes Π_p.
- I intertwines the action of $\mathrm{Isom}^+(\boldsymbol{H}^2)$, the group isometries of $\boldsymbol{H}^2$ that preserve orientation, with the action of the $PU(2,1)$ stabilizer of $\boldsymbol{RH}^2$.

Indeed, these conditions tell you how to construct I.

The map I extends in a natural way to give a map

$$I : T_1\boldsymbol{H}^2 \to \Omega = S^3 - \boldsymbol{R}^2.$$

In this way we see that the action of Γ_ζ on Ω is topologically conjugate to the action of G_ζ on $T_1\boldsymbol{H}^2$. In particular, Ω/Γ_ζ is homeomorphic to the quotient

Q_ζ discussed in the previous sections. The two objects are really the same space, wearing a slightly different geometric structure.

Now we can view the example in the previous section as a special case of the HST. The group $\Gamma = \Gamma_{\infty,\infty,\infty}$ is a horotube group. The horotubes are sets of the form $I(T_1 H)$, where $H \subset \boldsymbol{H}^2$ is a horoball neighborhood of a cusp of G as it acts on $\boldsymbol{H}^2$. The perturbed group $\widehat{\Gamma} = \Gamma_\zeta$ is the group that is part of a sequence of groups converging nicely to Γ. The surgery follows the same pattern discussed in the previous section. We will revisit this kind of example in Chapter 5.

One main goal of this monograph is to see that the surgery just described is part of a somewhat more general kind of surgery, having the same flavor.

4.4 COMPLEX HYPERBOLIC TRIANGLES

Our discussion here parallels the discussion in Section 4.1.

Recall that a $\boldsymbol{C}$-slice is the intersection of a complex line with $\boldsymbol{CH}^2$, provided that this intersection is an open disk. For any pair (C_0, C_1) of distinct and intersecting $\boldsymbol{C}$-slices, we can find a unique $\theta \in (0, \pi/2]$ and some $g \in PU(2,1)$ such that

$$g(C_0) \cap g(C_1) = (0,0), \qquad (1,0) \in g(\overline{C}_0), \qquad \big(\cos(\theta), \sin(\theta)\big) \in g(\overline{C}_1). \tag{4.1}$$

Here $\overline{C}_j$ is the extension of C_j to S^3. The angle θ is the *complex angle* between C_0 and C_1. The product of the complex reflections determined by C_0 and C_1 is conjugate to a rotation through an angle of 2θ. As in the real hyperbolic setting, there is a limiting case (which cannot be normalized as above) in which $v = \overline{C}_0 \cap \overline{C}_1 \in S^3$. That is, the two $\boldsymbol{C}$-slices are asymptotic. In this case we set $\theta = 0$. The product of the two complex reflections here is an $\boldsymbol{R}$-parabolic element fixing v.

We say that a $(\zeta_0, \zeta_1, \zeta_2)$-*triangle* is a triple (C_0, C_1, C_2) of distinct $\boldsymbol{C}$-slices such that

$$\theta(C_{j-1}, C_{j+1}) = \frac{\pi}{\zeta_j}, \qquad j = 1,2,3. \tag{4.2}$$

(We allow the case $\zeta_i = \infty$.) The main goal in this section is to classify these triangles. The general classification result is due to Brehm [**Br**]. We will follow the essentially equivalent treatment given in [**P**].

It is convenient to set

$$r_j = \cos\left(\frac{\pi}{\zeta_j}\right); \qquad r_\infty = 1. \tag{4.3}$$

The number r_j has a Hermitian geometry interpretation. Let v_0, v_1, v_2 be the vertices of a complex hyperbolic triangle. Let $\widetilde{v}_j$ be any lift of v_j to $\boldsymbol{C}^{1,2}$. There is a vector $c_j \in N_+$ such that $\langle c_j, \widetilde{v}_{j\pm1}\rangle = 0$ and $\langle c_j, c_j\rangle = 1$. The vector c_j is unique up to multiplication by a unit complex number. We call it a *polar vector*. We have $r_j = |\langle c_{j-1}, c_{j+1}\rangle|$. Thus, the magnitude of $\langle c_{j-1}, c_{j+1}\rangle$ determines ζ_j.

Let ϑ_j be the argument of $\langle c_{j-1}, c_{j+1}\rangle$. The three arguments together have significance. The *polar angular invariant*[1]

$$\alpha = \arg\left(\prod_{j=0}^{2}\langle c_{j-1}, c_{j+1}\rangle\right) = (\vartheta_0 + \vartheta_1 + \vartheta_2) \pmod{2\pi} \tag{4.4}$$

is independent of the choices of c_0, c_1, c_2. The following classification result appears in [**P**].

Lemma 4.2 *A $(\zeta_0, \zeta_1, \zeta_2)$-triangle in $\boldsymbol{CH}^2$ is determined up to $PU(2,1)$-equivalence by its polar angular invariant. Moreover, for any $\alpha \in [0, 2\pi]$ there exists a $(\zeta_0, \zeta_1, \zeta_2)$-triangle with invariant α if and only if*

$$\cos\alpha < \frac{r_0^2 + r_1^2 + r_2^2 - 1}{2r_0 r_1 r_2}. \tag{4.5}$$

When $\alpha = \pi$ the vertices lie in an $\boldsymbol{R}$-slice.

Proof: Let $s_j = \sin(\pi/\zeta_j)$. Let C_j be the complex line containing the vertices v_{j-1} and v_{j+1}. We normalize C_0 and C_1 as in Equation 4.1. (Here $\theta = \pi/\zeta_2$.) This leads to vectors

$$c_0 = (0, 1, 0), \qquad c_1 = (-s_2, r_2, 0), \qquad c_2 = (z, r_1, z'), \tag{4.6}$$

with $z, z' \in \boldsymbol{C}$. We can choose any unit complex number u and apply the isometry $(z_0, z_1, z_2) \to (z_0, z_1, uz_2)$. This isometry has the effect of replacing z' by uz' and leaving c_0 and c_1 alone. Hence, we can assume that $z' \in [0, \infty)$.

Since C_2 does not contain the origin, we have $z' > 0$. The condition $\langle c_2, c_2\rangle = 1$ now forces $|z| > s_1$. The equation $|\langle c_1, c_2\rangle| = r_0$ and Equation 4.4 give

$$-s_2\overline{z} + r_1 r_2 = r_0 e^{i\alpha}. \tag{4.7}$$

Therefore,

$$|r_0 e^{i\alpha} - r_1 r_2|^2 = s_2^2|z|^2 > s_1^2 s_2^2 = (1 - r_1^2)(1 - r_2^2). \tag{4.8}$$

Expanding things out yields Equation 4.5. This shows that Equation 4.5 is necessary for the existence of a triangle with the above specifications.

We can always solve Equation 4.7 for z. If Equation 4.5 holds, then the solution will satisfy $|z| > s_1$. But then we can choose $z' > 0$ in such a way as to make all the relevant equations above true. Hence Equation 4.5 is sufficient.

For uniqueness, note that the triple ζ and the invariant α determine z and z', and these values in turn determine the geometry of the triangle.

[1] In the ideal case, the polar angular invariant gives the same information as the angular invariant of the three ideal vertices of the triangle, though the two quantities do not coincide. The term *polar angular invariant* is our own, used to distinguish it from the closely related quantity defined for points in S^3 in Section 2.4.

Finally, if $\alpha = \pi$, then we have $z \in \boldsymbol{R}$. Hence, all three vectors c_0, c_1, c_2 have real coordinates. This suffices to show that v_0, v_1, v_2 all have real coordinates as well. □

Fixing ζ, we note from Equation 4.7 that the map $(z, w) \to (\overline{z}, \overline{w})$ conjugates between the two groups with polar angular invariant α and $2\pi - \alpha$. For this reason, it suffices to take $\alpha \in [0, \pi]$.

4.5 THE REPRESENTATION SPACE

A complex hyperbolic triangle, as above, determines a $(\zeta_0, \zeta_1, \zeta_2)$-*complex reflection triangle group*, the group generated by the complex reflections fixing the given slices. Such a group comes from a *complex reflection triangle group representation*, which maps the standard generators of the corresponding real reflection triangle group to the complex reflections just mentioned. Let $\mathrm{Rep}(\zeta)$ denote the space of conjugacy classes of ζ-complex reflection triangle groups. The points of $\mathrm{Rep}(\zeta)$ are in bijection with the choices of $\alpha \in [0, \pi]$ for which Equation 4.5 holds.

For any choice of ζ, the point $\pi \in \mathrm{Rep}(\zeta)$ corresponds to the case when the representation stabilizes an $\boldsymbol{R}$-slice. We discussed this case in detail above. In this case, it is well known that the element $I_0 I_1 I_2$ is a loxodromic element. (In the cases of interest to us, we will give a proof.) Hence, by continuity, there is some nontrivial maximal interval $\diamond\mathrm{Rep}(\zeta)$ consisting of representations for which $I_0 I_1 I_2$ is loxodromic. We call the representations in $\diamond\mathrm{Rep}(\zeta)$ the *subcritical* representations.

One endpoint of $\diamond\mathrm{Rep}(\zeta)$ is π. If $\diamond\mathrm{Rep}(\zeta)$ is a proper subinterval of $[0, \pi]$, then we call the other endpoint of this interval the *critical representation*. The element $I_0 I_1 I_2$ is necessarily parabolic at the critical representation. The point here is that any continuous path from loxodromic elements to elliptic elements must pass through a parabolic element.

We define $|\zeta| = \min(\zeta_0, \zeta_1, \zeta_2)$. We are mainly interested in understanding $\diamond\mathrm{Rep}(\zeta)$ when $|\zeta|$ is large. For the rest of the chapter, we will deal with the ideal case when $|\zeta| = \infty$. Later in the monograph, we will treat the case where $|\zeta|$ is large as a perturbation.

4.6 THE IDEAL CASE

In terms of Lemma 4.2, the existence condition is $\cos(\alpha) < 1$. This gives us the interval $(0, \pi]$ from which to take α. As $\alpha \to 0$, the three vertices of our triangle come closer and closer to lying in a common $\boldsymbol{C}$-slice. Rather than use the parameter α, we will use a different parameter $s \in [0, \infty)$. The case $s = 0$ corresponds to $\alpha = \pi$, the Fuchsian case. The limit $s \to \infty$ corresponds to the limit $\alpha \to 0$. (We work out the exact correspondence between s and α below.)

4.6.1 A Formula for the Generators

Given $s \in [0, \infty)$ we define

$$\beta_s = \frac{s+i}{\sqrt{2+2s^2}}. \tag{4.9}$$

Sometimes we write β instead of β_s, when the dependence is clear.

Lemma 4.3 *Every ideal triangle in S^3 is conjugate to a triangle with vertices v_0, v_0, v_1 of the form*

$$v_0 = (\beta, \overline{\beta}), \qquad v_1 = (\beta, \beta), \qquad v_2 = (\overline{\beta}, \overline{\beta}). \tag{4.10}$$

Proof: The points in Equation 4.10 have lifts

$$\widetilde{v}_0 = (\beta, \overline{\beta}, 1), \qquad \widetilde{v}_1 = (\beta, \beta, 1), \qquad \widetilde{v}_2 = (\overline{\beta}, \overline{\beta}, 1). \tag{4.11}$$

Noting that we want $\langle c_j, \widetilde{v}_j \rangle = 0$ for $i \neq j$, we see that there are real constants x_0, x_1, x_2 such that the polar vectors are

$$c_0 = x_0(1, -1, 0), \qquad c_1 = x_1(0, 1, \beta), \qquad c_2 = x_2(1, 0, \overline{\beta}). \tag{4.12}$$

Looking at Equation 4.4 we see that $\alpha = \arg(\beta^2) = 2\arg(\beta) = 2\tan^{-1}(1/s)$, so that $s \in [0, \infty)$ corresponds to $\alpha \in (0, \pi]$. $\square$

Let I_j be the complex reflection that fixes v_{j-1} and v_{j+1}. We can compute I_j using Equations 2.6 and 4.12. The matrices for I_0, I_1, I_2 are

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \qquad \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & -4\overline{\beta} \\ 0 & 4\beta & -3 \end{bmatrix}, \qquad \begin{bmatrix} 3 & 0 & -4\beta \\ 0 & -1 & 0 \\ 4\overline{\beta} & 0 & -3 \end{bmatrix}, \tag{4.13}$$

respectively. Using the fact that $|\beta|^2 = 1/2$, you can verify the above equations directly by checking, for $j = 0, 1, 2$, that I_i is an involution whose rows and columns form an orthonormal basis relative to the Hermitian form, and $I_j(c_j) = c_j$.

4.6.2 Criticality

Letting δ be as in Equation 2.5, we compute that

$$\delta(I_1 I_2 I_0) = \frac{-1024(3s^2 - 125)}{(1+s^2)^3}. \tag{4.14}$$

From Lemma 2.1 we see that the subcritical interval is $[0, \overline{s})$, where $\overline{s} = \sqrt{125/3}$. This was first worked out in [**GP**]. We call the critical representation at $\overline{s}$ the *golden triangle group*. We denote this group by $\mathbf{\Gamma}'$, and we let $\mathbf{\Gamma}$ denote the even subgroup.

One can check by direct calculation that

$$M = \begin{bmatrix} -6\sqrt{5} + 2i\sqrt{3} & -6\sqrt{5} - 2i\sqrt{3} & 18 \\ 10\sqrt{5} + 2i\sqrt{3} & 10\sqrt{5} - 2i\sqrt{3} & -32 \\ 3\sqrt{5} - 3i\sqrt{3} & 3\sqrt{5} + 3i\sqrt{3} & 12 \end{bmatrix} \tag{4.15}$$

represents a Heisenberg stereographic projection such that

$$MgM^{-1} = \begin{bmatrix} \frac{i}{8}(7i+\sqrt{15}) & 0 & 0 \\ 0 & \frac{i}{4}(i+\sqrt{15}) & \frac{-i}{4}(-15i+\sqrt{15}) \\ 0 & 0 & \frac{i}{4}(i+\sqrt{15}) \end{bmatrix}.$$

From this equation we see that MgM^{-1} acts on the Siegel domain as

$$MgM^{-1}(z,w) = (uz, w+ki)$$

with $|u|=1$ and $k \in \boldsymbol{R}$. Hence, M conjugates g to a $\boldsymbol{C}$-parabolic automorphism of $\mathcal{H}$ that stabilizes $\{0\} \times \boldsymbol{R}$. The twist u—and any power of u—is an irrational multiple of π.

4.6.3 Another Construction of the Golden Group

Any triple of points in S^3 has an order-3 complex hyperbolic symmetry. This follows from the fact that the angular invariant from Chapter 2 is a complete invariant of such ideal triangles. Let A be the symmetry that permutes the points defining $\mathbf{\Gamma}'$. Let $\mathbf{\Gamma_3}$ denote the group obtained by adjoining A to $\mathbf{\Gamma}$. Conjugation by A permutes the complex reflections I_0, I_1, I_2. Here we will give another construction of $\mathbf{\Gamma_3}$ that highlights the role played by A. Actually, this is a bit of a trade-off: We can make A simpler at the cost of making the other generators more complicated.

First of all, we define

$$A = \begin{bmatrix} \omega & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.16}$$

We define

$$C(\tau) = \frac{1-\tau^2}{1+\tau^2}, \qquad S(\tau) = \frac{2\tau}{1+\tau^2}. \tag{4.17}$$

Here $(C(\tau), S(\tau))$ is the usual rational parametrization of the circle. Next, we define

$$x_1 = \big(C(\tau), S(\tau), 1\big), \qquad x_2 = A(x_1), \qquad x_0 = A(x_2). \tag{4.18}$$

These are null vectors in $\boldsymbol{C}^{2,1}$ that represent points on S^3. Let I_j be the complex reflection in the points x_{j-1} and x_{j+1}, computed using Equation 2.6. Then $\mathbf{\Gamma}'$ is conjugate in $PU(2,1)$ to the group generated by (the new versions of) I_0, I_1, I_2, once a is suitably chosen.

Our choice of τ leads to a conjugate of $\mathbf{\Gamma}'$, provided that

$$\frac{\mathrm{Im}^2(X)}{\mathrm{Re}^2(X)} = \frac{125}{3}, \qquad X = \langle x_1, x_2\rangle\langle x_2, x_3\rangle\langle x_3, x_1\rangle. \tag{4.19}$$

This leads to the equation

$$0 = Z(\tau) := 1 - 4566\tau^4 + 19023\tau^8 - 24820\tau^{12} + 19023\tau^{16} - 4566\tau^{20} + \tau^{24}. \tag{4.20}$$

The largest real solution is

$$\tau_0 = 8.218353195588304393042278767487165254931994226781.... \quad (4.21)$$

As a double check, we let δ be the function from Lemma 2.1, and we compute

$$D(\tau) := \delta(AB) = \frac{-27Z(\tau)}{(1+14\tau^4+\tau^8)^3}. \quad (4.22)$$

So $\delta(AB) = 0$ iff $Z(\tau) = 0$. For later use, we compute

$$D'(\tau) = \frac{-2^{11}3^5\tau^3(\tau^4-1)^3(\tau^4+1)(\tau^2-2\tau-1)(\tau^2+2\tau-1)(1+6\tau^2+\tau^4)}{(1+14\tau^4+\tau^8)^4} \quad (4.23)$$

The largest real root of this last equation is $1+\sqrt{2}$. Hence,

$$D'(\tau_0) \neq 0. \quad (4.24)$$

Setting $\tau = \tau_0$, we have that $\mathbf{\Gamma_3}$ is conjugate in $PU(2,1)$ to the group generated by (the new versions of) I_0I_1, I_1I_2, I_2I_0, and A. To see $\mathbf{\Gamma_3}$ a different way, we define $B = I_0AI_0^{-1}$. Then $\mathbf{\Gamma_3}$ is generated by the two order-3 elements A and B. For instance,

$$AB^{-1} = AI_0^{-1}A^{-1}I_0 = AI_0A^{-1}I_0 = I_1I_0,$$

$$(AB)^3 = (AI_0AI_0)^3 = (AI_0AI_0AI_0)^2 = (AI_0A^{-1}A^{-1}I_0AI_0)^2 = (I_1I_2I_0)^2. \quad (4.25)$$

In this way we see explicitly that $\mathbf{\Gamma_3}$ is the image of a certain representation of the free product $\boldsymbol{Z}_3 * \boldsymbol{Z}_3$ into $SU(2,1)$.

Remark: We don't know how to compute τ_0 exactly, but we can compute it with enough accuracy to tell that every entry in I_0 has absolute value at least $1/10$. In particular, every entry in the matrix I_0 is nonzero.

Chapter Five

Heuristic Discussion of Geometric Filling

5.1 A DICTIONARY

The proof of Theorem 1.1, at least as outlined in [**T0**], breaks down into the following three parts.

- Let M be a cusped hyperbolic 3-manifold. We have a discrete embedding $\rho : \pi_1(M) \to \mathrm{Isom}(\boldsymbol{H}^3)$ so that $M = \boldsymbol{H}^3/\Gamma$, where $\Gamma = \pi_1(M)$. If one perturbs the structure on M, then the resulting structure is determined by the *holonomy representation* $\widehat{\rho} : \pi_1(M) \to \mathrm{Isom}(\boldsymbol{H}^3)$.
- Let $\widehat{\Gamma} = \widehat{\rho}(\pi_1(M))$. Suppose $\widehat{\rho}$ maps the $\boldsymbol{Z}^2$ subgroups of $\pi_1(M)$ to discrete loxodromic $\boldsymbol{Z}$ subgroups of $\mathrm{Isom}(\boldsymbol{H}^3)$. Then $\boldsymbol{H}^3/\widehat{\Gamma}$ is a closed hyperbolic 3-manifold obtained by Dehn filling the cusp of M.
- There are enough perturbations of the hyperbolic structure on M to account for all but finitely many topological types of Dehn fillings.

The HST corresponds to the second item above, which is a comparison between Γ and $\widehat{\Gamma}$. In this chapter we will discuss the analogy in detail.

Here is a kind of dictionary that translates between certain spherical CR objects and certain real hyperbolic objects. The terms of the dictionary will be defined in the chapter.

Hyperbolic Object	**Spherical CR Object**
cusped hyperbolic 3-manifold group	horotube group
$\boldsymbol{H}^3$	domain of discontinuity $\Omega \subset S^3$
horoball	horotube
Busemann function	horotube function
Busemann complex	horotube complex
$\boldsymbol{Z} \oplus \boldsymbol{Z}$-parabolic subgroup	$\boldsymbol{Z}$-parabolic subgroup
torus cusp	torus cusp
perturbed holonomy representation	a representation converging nicely
$\boldsymbol{Z}$-loxodromic subgroup	$\boldsymbol{Z}/n$ lens-elliptic subgroup
$\boldsymbol{Z}$-invariant banana neighborhood	$\boldsymbol{Z}/n$-invariant solid torus
Dehn filling of cusp	Dehn filling of cusp

5.2 THE TREE EXAMPLE

5.2.1 Immersions of the Trivalent Tree

We will work in the Klein model of $\boldsymbol{H}^2$. See Section 2.1.

Let T denote the infinite trivalent tree, with some distinguished vertex v_0 and some distinguished edge e_0 incident to v_0. Given any length $L > 0$, there is a unique map $\psi_L : T \to \boldsymbol{H}^2$ such that we have the following.

- $\psi_L(v_0) = 0$, and $\psi_L(e_0)$ is contained in the nonnegative real axis.
- If e is any edge of T, then $\psi_L(e)$ is a geodesic segment of length L.
- If e_1 and e_2 are any two incident edges of T, then the angle between $\phi(e_1)$ and $\phi(e_2)$ is $2\pi/3$.

Let $\Psi_L = \psi_L(T)$. There is some value L_∞ such that ψ_L is an embedding iff $L \geq L_\infty$. Here L_∞ is twice the length of the compact side of the $(2, 3, \infty)$-triangle. Ψ_{L_∞} is invariant under the $(2, 3, \infty)$ reflection triangle group. Also, there is a countable sequence of lengths L_7, L_8, L_9,... such that Ψ_{L_n} is the 1-skeleton of the usual tiling by regular hyperbolic n-gons, meeting three per vertex. L_n is twice the length of the shortest side of the $(2, 3, n)$-triangle, and Ψ_{L_n} is invariant under the $(2, 3, n)$-reflection triangle group. For ease of notation, we set $\Psi_n = \Psi_{L_n}$ and $\Psi_\infty = \Psi_{L_\infty}$.

In a large-scale sense, the complexes Ψ_∞ and Ψ_n are very different from each other. The former object is simply connected, for instance, while the latter object has an infinitely generated fundamental group. However, there is a "midscale" sense in which the two objects are very similar: Let $B_r(0)$ denote the ball of radius r about 0 in $\boldsymbol{H}^2$. Then for any r and any $\epsilon > 0$ there exists $N = N(r, \epsilon)$ such that $B_r(0) \cap \Psi_n$ is within ϵ of $B_r(0) \cap \Psi_\infty$ in the Hausdorff metric. As $n \to \infty$ one needs to look "farther and farther out" in order to see the large-scale difference between Ψ_n and Ψ_∞.

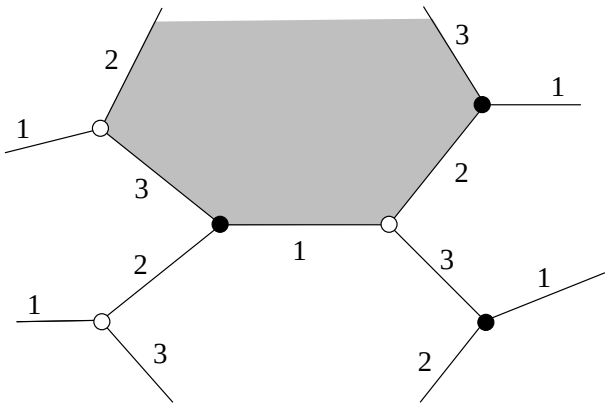

Figure 5.1: A tree with one complementary component shaded

Assume now that $n = 6m$ for some $m = 2, 3, 4,...$ Figure 5.1 shows a natural labeling of T by three numbers, so that all three edges incident to each vertex get a different label. The shaded region in the figure indicates an embedded disk. On the boundary of this disk the labels go $...1, 2, 3, 1, 2, 3,...$ Figure 5.1 also shows a 2-coloring of the vertices, so that adjacent vertices get opposite colors. Using the map ϕ we can transfer the decoration to Ψ_n when $n = 6m$, as stipulated. Also, we can transfer the labeling to Ψ_∞. Let Γ_n be the label-preserving symmetry group of Ψ_n. Likewise we define Γ_∞.

5.2.2 Comparing the Quotients

We want to compare $Q_\infty = \boldsymbol{H}^2/\Gamma_\infty$ to $Q_n = \boldsymbol{H}^2/\Gamma_n$. Let $\Gamma = \Gamma_\infty$ and $\widehat{\Gamma} = \Gamma_n$, etc.

To compute the topology of Q, we note that $\boldsymbol{H}^2 - \Psi$ consists of infinitely many horodisklike components, one per cusp of Γ. Let V_p be the closure of one of these components. The shaded region in Figure 5.1 gives an impression of what V_p looks like. Here are three facts.

- Γ acts transitively on the components of $\boldsymbol{H}^2 - \Psi$.
- The stabilizer Γ_p of V_p is isomorphic to $\boldsymbol{Z}$.
- $\partial V_p/\Gamma_p$ has 6 edges.

Hence, V_p/Γ_p is a punctured hexagon H. Other elements of Γ identify opposite sides of ∂H. Thus we can identify Q with $H/\sim$, where $\sim$ is a gluing of the opposite sides of ∂H. Hence, Q is a once-punctured torus—that is, a torus with a cone point of order ∞.

To compute the topology of $\widehat{Q}$, we note that $\boldsymbol{H}^2 - \widehat{\Psi}$ consists of infinitely many disklike components—n-gons actually—one per torsion point of $\widehat{\Gamma}$. Let $\widehat{V}_p$ be one of these components. We have the following parallel facts.

- $\widehat{\Gamma}$ acts transitively on the components of $\boldsymbol{H}^2 - \widehat{\Psi}$.
- The stabilizer $\widehat{\Gamma}_p$ of $\widehat{V}_p$ is isomorphic to $\boldsymbol{Z}/m$.
- $\partial\widehat{V}_p/\widehat{\Gamma}_p$ has 6 edges. (Recall that $n = 6m$.)

Hence, $\widehat{V}_p/\widehat{\Gamma}_p$ is a hexagon with an order-m cone point in the middle of it. Other elements of Γ identify opposite sides of ∂H. Thus, we can identify Q with $H/\sim$, where $\sim$ is a gluing of the opposite sides of ∂H. Hence, Q is a torus with a cone point of order $1/m$.

We think of the transition from Q to $\widehat{Q}$ as a kind of surgery. Let D^2_m denote the unit disk with its origin labeled by m. There is a subsurface $E \subset Q$ such that $E = S^1 \times [0, \infty)$. Here E is the end of Q. We choose a homeomorphism that identifies E with $D^2_m - \{0\}$. Then $\widehat{Q} = (Q \cup D^2_m)/\sim$ is the identification space. In Section 5.7 we will see that this example is really Dehn surgery in disguise.

A central feature of the toy example runs throughout this chapter. The topological difference between Q and $\widehat{Q}$ appears far away from the parts of these spaces that are covered by Ψ and $\widehat{\Psi}$.

5.3 HYPERBOLIC CASE: BEFORE FILLING

5.3.1 The Setup

Let M be a cusped hyperbolic 3-manifold. Let $G = \pi_1(M)$ be the fundamental group. Let $\rho : G \to \mathrm{Isom}(\boldsymbol{H}^3)$ be the representation of G as the universal covering group of M. Let $\Gamma = \rho(G)$. Then $M = \boldsymbol{H}^3/\Gamma$.

Each of the n cusps of M corresponds to a $\boldsymbol{Z}^2$-subgroup $\Gamma_j \subset \Gamma$ fixing a *cusp* $p_j \in S^2$ *of* Γ. Let $\mathrm{Cusp}(\Gamma)$ denote the set of all cusps of Γ. That is, $\mathrm{Cusp}(\Gamma)$ is the Γ-orbit of $p_1 \cup \cdots \cup p_n$.

5.3.2 The Busemann Complex

A general discussion of Busemann functions can be found in [**BH**]; however, we don't need this generality for our discussion. If we work in the upper half space model for $\boldsymbol{H}^3$, then all the Busemann functions are conjugate to the one in Equation 2.1.

To each $p \in \mathrm{Cusp}(\Gamma)$ we assign a Busemann function f_p. We can make the assignment Γ-equivariant in the sense that

$$f_p = f_{g(p)} \circ g \qquad \forall g \in \Gamma. \tag{5.1}$$

To do this, we first define the Busemann functions f_1,...,f_n for each of the cusps p_1,...,p_n, and then use the action of Γ and Equation 5.1 to define the rest of them.

We define the "Voronoi cell":

$$V_p = \{x \in \boldsymbol{H}^3 |\ f_p(x) \geq f_q(x)\ \forall q \in \mathrm{Cusp}(\Gamma)\}. \tag{5.2}$$

V_p is sometimes called the *Ford domain*. To imagine what V_p looks like, at least in the 2-dimensional case, one should think of the tree in Figure 5.1 embedded in $\boldsymbol{H}^2$. The shaded region in Figure 5.1 is then V_p for some choice of p. The unshaded regions represent other choices of V_p.

In the 3-dimensional case, V_p is a topological ball, which lies within a bounded distance from a horoball. ∂V_p is a polyhedral sphere, tiled by a pattern of geodesic polygons that is invariant under the $\boldsymbol{Z}^2$-stabilizer of p. We define

$$\Psi = \bigcup_{p \in \mathrm{Cusp}(\Gamma)} \partial V_p. \tag{5.3}$$

Then Ψ is a Γ-invariant piecewise geodesic 2-complex. We call Ψ the *Busemann complex*. Then Ψ is a Γ-invariant 2-complex embedded in $\boldsymbol{H}^3$.

For the discussion in the next section, it is useful to make a transversality assumption about the complex Ψ. We assume that no more than 3 faces

of each ∂V_p meet at a vertex, and no more than 4 of the V's intersect each other. In Chapter 12 we discuss the local model precisely. The transversality assumption guarantees that Ψ is stable under small perturbations, as we discuss below. Presumably this assumption usually holds true, though in our heuristic discussion we will not pause to verify this claim. We leave the rigorous justifications (for similar assumptions) for the spherical CR case.

5.3.3 From the Busemann Complex to Topology

Let $V_j = V_{p_j}$ for $j = 1,...,n$. Define

$$\Upsilon = \bigcup_{j=1}^{n} \Upsilon_j, \qquad \Upsilon_j = V_j/\Gamma_j. \tag{5.4}$$

By construction Γ_j stabilizes V_j, so the set Υ_j is well defined. The space $\partial\Upsilon_j$ is a torus, tiled by geodesic polygons, and moreover, Υ_j is homeomorphic to $\partial\Upsilon_j \times [0,\infty)$. In short, Υ_j has a torus end for $j = 1,...,n$.

To recover $M = \boldsymbol{H}^3/\Gamma$ from the structure of Ψ, we proceed as follows. Every point of $\boldsymbol{H}^3$ is Γ-equivalent to a point in $V_1 \cup \cdots \cup V_n$. Moreover, points in the interior of V_j are equivalent iff they belong to the same Γ_j-orbit. From these two facts, we deduce that $M = \Upsilon/\sim$, where $\sim$ is the pairing of the polygons of $\partial\Upsilon$ determined by the action of other elements of Γ. From this we recover the fact that M is a closed hyperbolic 3-manifold with n cusps.

5.4 HYPERBOLIC CASE: AFTER FILLING

5.4.1 The Setup

Suppose that the hyperbolic structure on M is perturbed slightly, so that it is incomplete. As in Section 5.1, let $\widehat{\rho} : G \to \mathrm{Isom}(\boldsymbol{H}^3)$ be the associated holonomy representation. Let $\widehat{\Gamma} = \widehat{\rho}(G)$. The phrase "if the structure on $\widehat{M}$ *is sufficiently close* to the structure on M" arises so often that we abbreviate it to *ISC*.

We have a surjective homomorphism $\eta : \Gamma \to \widehat{\Gamma}$ given by the equation

$$\eta = \widehat{\rho} \circ \rho^{-1}. \tag{5.5}$$

We define $\widehat{\Gamma}_j = \eta(\Gamma_j)$. Then $\widehat{\Gamma}_1,...,\widehat{\Gamma}_n$ are all abelian subgroups. The case corresponding to Dehn fillings on *all* the cusps of M corresponds to the case when each $\widehat{\Gamma}_j$ is a discrete $\boldsymbol{Z}$-subgroup consisting of loxodromic elements. For simplicity this is the case we consider. In this case, it turns out that $\widehat{\rho}$ is discrete ISC.

There is a unique geodesic α_j stabilized by $\widehat{\Gamma}_j$. ISC, then the endpoints of α_j are both close to p_j, and the action of $\widehat{\Gamma}_j$ on $\boldsymbol{H}^3$, far away from these endpoints, is close to the corresponding action of Γ_j on $\boldsymbol{H}^3$. Let $\mathrm{Axis}(\widehat{\Gamma})$ denote the $\widehat{\Gamma}$ orbit of $\alpha_1 \cup \cdots \cup \alpha_n$.

5.4.2 The Perturbed Complex

For each geodesic $\alpha \in \text{Axis}(\widehat{\Gamma})$, we choose a constant C_α and define a function $\widehat{f}_\alpha$ to have the form $\widehat{f}_\alpha(x) = (\text{distance to } \alpha) - C_\alpha$. Once we pick the constants $C_1,...,C_n$, so as to define $\widehat{f}_1,...,\widehat{f}_n$, the remaining constants are forced if we want the assignment to be $\widehat{\Gamma}$-invariant.

We can choose the constants $C_1,...,C_n$ so that, far away from $\widehat{\alpha}_j$ and p_j, the two functions $\widehat{f}_j$ and f_j are C^∞-close ISC. (All we need is C^1-close.) The geometric idea behind this observation is quite simple: The boundary of a fat tubular neighborhood of a geodesic is geometrically quite similar to a horosphere.

Now we define the sets $\widehat{V}_\alpha$ and $\widehat{\Psi}_j$ exactly as in equation 5.3. ISC, then $\widehat{\Psi}$ and Ψ have the same local combinatorial structure. Essentially, this is a consequence of the transversality and local finiteness properties of Ψ. The two complexes look the same in the midscale sense discussed in Section 5.2.1. What about the global difference? Whereas the connected components of $\boldsymbol{H}^3 - \Psi$ are horoball-like sets, the connected components of $\boldsymbol{H}^3 - \widehat{\Psi}$ are fat tubular neighborhoods of geodesics.

5.4.3 Comparing the Quotients

Let $\widehat{M} = \boldsymbol{H}^3/\widehat{\Gamma}$. We define $\widehat{\Upsilon}_j$ and $\widehat{\Upsilon}$ just as we defined Υ and Υ_j in Section 5.3.3. It turns out that all the sets $\widehat{V}_1,...,\widehat{V}_n$ are inequivalent ISC. From this it follows that $\widehat{M} = \widehat{\Upsilon}/\sim$, where $\sim$ is the pairing of the faces of $\partial\widehat{\Upsilon}$ induced from other elements of $\widehat{\Gamma}$.

$\partial\widehat{V}_j$ is a tiled version of a tubular neighborhood of α_j. In particular, $\partial\widehat{V}_j$ is an infinite cylinder. The $\boldsymbol{Z}$-subgroup acts on this cylinder freely and discretely. Hence, $\partial\widehat{\Upsilon}_j$, like $\partial\Upsilon_j$, is a torus. ISC, then the action of Γ_j on ∂V_j has the same local combinatorial behavior as the action of $\widehat{\Gamma}_j$ on $\partial\widehat{V}_j$. Hence, $\partial\Upsilon_j$ and $\partial\widehat{\Upsilon}_j$ are combinatorially identical.

The pairing of cells on $\partial\Upsilon$, which is determined by finitely many elements of Γ, is combinatorially identical to the pairing on $\partial\widehat{\Upsilon}$ ISC. Hence, $\partial\widehat{\Upsilon}$ is glued together in the same combinatorial way that $\partial\Upsilon$ is glued together. The only topological difference between M and $\widehat{M}$ takes place far away from $\partial\Upsilon/\sim$ and $\partial\widehat{\Upsilon}/\sim$. The difference is that Υ_j is homeomorphic to $T^2 \times [0,\infty)$, whereas $\widehat{\Upsilon}_j$ is a solid torus. Hence, the replacement of Υ_j by $\widehat{\Upsilon}_j$ amounts to Dehn filling. This happens for each $j = 1,...,n$ and explains why $\widehat{M}$ is obtained from M by filling all the cusps.

5.5 SPHERICAL CR CASE: BEFORE FILLING

5.5.1 The Setupp

Now we switch to the spherical CR setting. Let G, ρ, and $\Gamma = \rho(G)$ be as in §1.3. Let $\mathrm{Cusp}(\Gamma)$ denote the set of cusps of Γ. Let p_1,...,p_n be a minimal and complete list of representatives of the Γ-equivalence classes in $\mathrm{Cusp}(\Gamma)$. Let Γ_1,...,$\Gamma_n \subset \Gamma$ be the $\boldsymbol{Z}$-stabilizers of these cusps. Let Ω be the domain of discontinuity of Γ.

5.5.2 The Horotube Complex

For each cusp $p \in \mathrm{Cusp}(\Gamma)$, we will assign a horotube function f_p, as in Section 2.6.2. This assignment will be Γ-equivariant in the same sense as above. We define

$$V_p^\infty = \{x \in \Omega |\ f_p(x) \geq f_q(x)\ \forall q \in \mathrm{Cusp}(\Gamma)\}, \tag{5.6}$$

$$\Psi^\infty = \bigcup_{p \in \mathrm{Cusp}(\Gamma)} \partial V_p^\infty. \tag{5.7}$$

Compare Equations 5.2 and 5.3

Remark: In this section and the next, we prefer to use the notation V_j^∞ and Ψ^∞, because the corresponding sets live on the ideal boundary. We shall have occasion to consider the 4-dimensional objects that extend V^∞ and Ψ^∞, and we reserve the names V and Ψ for these extensions.

We call Ψ^∞ the *horotube complex*. By construction, Ψ^∞ is Γ-equivariant. We will take care to arrange that Ψ^∞ is a locally finite complex, tiled by compact smoothly embedded "polygons." (The polygons will actually be piecewise smooth compact surfaces with polygonal boundary.) This is where we need to use the hypothesis that Γ is a horotube group. In Chapters 10 and 11, we will see how to build the relevant transversality into this assignment.

5.5.3 From the Horotube Complex to Topology

We can deduce the structure of $Q^\infty = \Omega/\Gamma$ from the local structure of Ψ^∞, just as we did in the real hyperbolic case. For ease of exposition, we assume that ∂V_j^∞ is an infinite cylinder, for $j = 1,...,n$.

We define Υ_j^∞ and Υ as in the real hyperbolic case. For the same reason as in the real hyperbolic case, we have $Q^\infty = \Upsilon^\infty/\sim$, where $\sim$ is the equivalence relation on $\partial\Upsilon^\infty$ defined by the action of other elements of Γ.

Since ∂V^∞ is a cylinder and Γ_j acts freely on ∂V_j^∞, we see that $\partial\Upsilon^\infty$ is a torus and Υ^∞ is homeomorphic to $\partial\Upsilon^\infty \times [0,\infty)$. The easiest way to understand the topology of Υ^∞ is to note that the action of Γ_j on V_j^∞ is topologically equivalent to the action of the group generated by the map in

Equation 2.12 on the $\boldsymbol{C}$-horotube $\{(z,t)|\ |z| \geq 1\}$. The faces of $\partial\Upsilon^\infty$ are identified in some pattern, which results in a 3-manifold. In this way we recognize $\Upsilon^\infty/\sim$ as a manifold with n torus ends. The action of Γ_j on V_j^∞ gives us a canonical marking on the end of Υ_j^∞, as we will explain in Chapter 8. This induces a canonical marking for Ω/Γ.

5.6 SPHERICAL CR CASE: AFTER FILLING

5.6.1 The Setup

The setup is as described in Section 1.3. Let $\widehat{\rho} \in \mathrm{Rep}(G)$ be a representation that is far along in a sequence of representations converging nicely to ρ. Let $\widehat{\Gamma} = \widehat{\rho}(G)$. We define $\eta : \Gamma \to \widehat{\Gamma}$ as in Equation 5.5. We define $\widehat{\Gamma}_j = \eta(\Gamma_j)$. Here Γ_j is as in Section 5.5.1. We will consider the case when all the groups $\widehat{\Gamma}_j$ are finite groups generated by lens-elliptic elements as in Section 1.3.

5.6.2 The Perturbed Horotube Complex

We will show in Chapter 7 that the Γ_j-invariant horotube function f_j can be transplanted to a C^∞-close $\widehat{\Gamma}_j$-invariant function $\widehat{f_j}$. Using the action of $\widehat{\Gamma}$ we attach a perturbed horotube function to each $\widehat{\Gamma}$-conjugate of the groups $\widehat{\Gamma}_j$. We then define $\widehat{V}_j^\infty$ and $\widehat{\Psi}^\infty$ using the perturbed functions. (Actually, we will take a more roundabout approach, because we do not know a priori that $\widehat{\Gamma}$ is discrete.) It turns out that ISC the local combinatorial structure of $\widehat{V}^\infty$ and $\widehat{\Psi}^\infty$ is the same as the local combinatorial structure of V_j^∞ and Ψ^∞, and the midscale structure is the same. As in the real hyperbolic case, the stability of the combinatorics comes from the transversality and the local finiteness.

What about the global difference? Whereas $\Omega - \Psi^\infty$ consists of an infinite collection of horotubes. $\widehat{\Omega} - \widehat{\Psi}^\infty$ consists of an infinite collection of solid tori. It turns out that $\widehat{\Gamma}$ is discrete ISC. (Proving this assertion requires us to extend Ψ^∞ and $\widehat{\Psi}^\infty$ into $\boldsymbol{CH}^2$.) We then recognize $\widehat{\Omega}/\widehat{\Gamma}$ as $\widehat{\Upsilon}^\infty/\sim$, as in the real hyperbolic case.

5.6.3 Comparing the Quotients

How do Υ^∞ and $\widehat{\Upsilon}^\infty$ compare? The difference appears far away from $\partial\Upsilon^\infty/\sim$ and $\partial\widehat{\Upsilon}^\infty/\sim$. Let's first compare Υ_j with $\widehat{\Upsilon}_j$. Now, ∂V_j^∞ is an infinite cylinder, invariant under a $\boldsymbol{Z}$-action, whereas $\partial\widehat{V}_j^\infty$ is a surface with the same local structure but is invariant under a free $\boldsymbol{Z}/n$-action. The only possibility is that $\partial\widehat{V}_j^\infty$ is an unknotted torus, and hence, $\widehat{V}_j^\infty$ is a solid torus. The quotient of a solid torus by a free $\boldsymbol{Z}/n$-action is again a solid torus. Hence, $\widehat{\Upsilon}_j^\infty$ is a solid torus. We already know that Υ_j^∞ is homeomorphic to $T^2 \times [0,\infty)$. In short, $\widehat{\Upsilon}_j^\infty$ is a Dehn filling of Υ_j^∞. But then $\widehat{\Omega}/\widehat{\Gamma}$ is

obtained from Ω/Γ by filling all the cusps.

Remarks:
(i) The topology of a horotube is different from the topology of a horoball. On the other hand, the $\boldsymbol{Z}^2$-action on a horoball that arises in connection with a cusped hyperbolic manifold is different from the $\boldsymbol{Z}$-action on a horotube that arises in connection with a horotube group, and the transition from $\boldsymbol{Z}^2$ to $\boldsymbol{Z}$ in the real hyperbolic case is different from the transition from $\boldsymbol{Z}$ to $\boldsymbol{Z}/n$ in the spherical CR case. These differences cancel out, and the net result of Dehn filling is the same in both cases.
(ii) One should perhaps not be surprised by our parallel analysis. After all, Dehn surgery is one of the simplest kinds of 3-dimensional surgery. In the spherical CR setting, *some kind* of surgery ought to be taking place, and Dehn surgery is one of the simplest choices.
(iii) There is one difference between the hyperbolic case and the spherical CR case. In the spherical CR case, the topology of the domain of discontinuity $\widehat{\Omega}$ is frequently different than the topology of Ω. In the hyperbolic case, the underlying space $\boldsymbol{H}^3$ is the same for both group actions. This difference does not affect our general analysis, however.

5.7 THE TREE EXAMPLE REVISITED

We use the notation from Section 5.2. Rather than consider the action of Γ and $\widehat{\Gamma}$ on $\boldsymbol{H}^2$, we can consider the action of these groups on the unit tangent bundle $T_1\boldsymbol{H}^2$. This discussion is supposed to complement the discussion in Sections §4.2–4.3. If S is any subset of $\boldsymbol{H}^2$, then we let T_1S be the restriction of $T_1\boldsymbol{H}^2$ to S.

The quotient $T_1\boldsymbol{H}^2/\Gamma$ is a noncompact 3-manifold with a torus end. To see this, note that the end of $T_1\boldsymbol{H}^2/\Gamma$ has the form T_1H/Γ_p, where H is a horodisk contained in the interior of V_p. Furthermore, any oriented circle bundle over $S^1 \times [0,\infty)$ is trivial. Therefore,

$$T_1H/\Gamma_1 = (H/\Gamma_p) \times S^1 = (S^1 \times [0,\infty)) \times S^1 = T^2 \times [0,\infty). \tag{5.8}$$

The corresponding subset of $T_1\boldsymbol{H}^2/\widehat{\Gamma}$ has the form $T_1\widehat{H}/\widehat{\Gamma}_p$. Here $\widehat{H}$ is a disk centered at the fixed point of $\widehat{\Gamma}_p$. The finite group $\widehat{\Gamma}_p$ acts freely on the solid torus $T_1\widehat{H}$, and the quotient $T_1\widehat{H}/\widehat{\Gamma}_p$ is another solid torus. In this way we see $T_1\boldsymbol{H}^2/\widehat{\Gamma}$ as a Dehn filling of $T_1\boldsymbol{H}^2/\Gamma$.

To relate this picture to spherical CR geometry, we recall from Section 4.3 the isomorphism $I : T_1\boldsymbol{H}^2 \to S^3 - \boldsymbol{R}^2$, which intertwines the two group actions. I conjugates the action of Γ and $\widehat{\Gamma}$ to two subgroups of $PU(2,1)$, which we give the same name. Γ and $\widehat{\Gamma}$ both stabilize $\boldsymbol{R}\boldsymbol{H}^2$. Indeed $\Lambda = \widehat{\Lambda} = \boldsymbol{R}\boldsymbol{H}^2 \cap S^3$ and $\Omega = \widehat{\Omega} = S^3 - \boldsymbol{R}^2$. Hence, $\widehat{\Omega}/\widehat{\Gamma}$ is obtained from Ω/Γ by Dehn filling.

We can explicitly see the horotube complex I^∞ and the perturbed complex $\widehat{I}^\infty$. We can take $I^\infty = I(T_1\Psi)$. Each individual horotube has the form

$I(T_1V_p)$. As a check, note that $\partial(T_1V_p)/\Gamma_p$ is a circle bundle over a hexagon and, hence, a torus. The perturbed complex is $\widehat{I}^\infty = I(T_1\widehat{\Psi})$. Each set $I(T_1\widehat{V}_p)$ is a solid torus.

PART 2
Proof of the HST

Chapter 6 begins Part 2 of the monograph. Before we start Chapter 6, we will give an overview of this part of the monograph.

In Chapters 6–11, we present technical results that feed into the proof of the HST. To make the proof more "modular" and more checkable, we present the main results of each of these chapters in the first section. (The exception to this is Chapter 9, in which we use the results throughout the chapter.) For each of the chapters 6–8 and 9–10 the reader can, on first reading, just use the results in the first section as black boxes.

The material in Chapters 6–11 essentially breaks into two halves.

- Chapters 6–8 deal with basic properties of individual horotubes and the functions that define them. In particular, Chapter 8 gives a surgery formula in the toy case when the entire group is cyclic. This special case later feeds into the overall proof of the HST.

- Chapters 9–11 deal with collections of horotubes and the functions that define them. The main goal of these chapters is to set up the conditions necessary to make the heuristic discussion in Chapter 5 work out. That is, we want to define the horotube complex and control its geometry and topology.

Once we have assembled the technical results from Chapters 6–11, we present our general method for proving discreteness in Chapter 12. As the reader might expect, we use some version of the Poincaré fundamental polyhedron theorem. The reader who is familiar with geometric structures (e.g., see [**CEG**] or [**T0**]) might find our treatment in Chapter 12 a bit heavy-handed. However, we wanted to write things out in terms we like.

In Chapter 13 we present the overall proof of the HST, using the material from Chapters 6–12 essentially as a black box. At the end of Chapter 13 we prove Theorem 1.11.

Chapter Six

Extending Horotube Functions

6.1 STATEMENT OF RESULTS

Let P be a parabolic element, fixing a point $p \in S^3$. Let $f : S^3 - p \to \boldsymbol{R}$ be a horotube function, as in Section 2.6.2. We assume that $f(S^3 - p) \subset [0, 3]$. In particular, f is bounded. In this chapter, we will explain how we extend f to $\boldsymbol{CH}^2$ (and a bit outside $\boldsymbol{CH}^2$ as well). The simplest way to extend f would be to imitate the way that a function on the circle is extended to a harmonic function on the unit disk. We could integrate against a $PU(2,1)$-invariant kernel. This method is simple and intuitive but leaves us with a function that is hard to analyze. We will give a more hands-on extension that gives nearly the same result.

Let Z be the Siegel domain, as in Section 2.3. Let ∂Z be its ideal boundary in $\boldsymbol{C}^2$. If $p \in \partial \boldsymbol{CH}^2$ is any point, then there is some complex projective transformation T mapping $\boldsymbol{CH}^2$ to Z and p to ∞. We define

$$\boldsymbol{U}_p = T^{-1}(\boldsymbol{C}^2). \tag{6.1}$$

This set is a convenient enlargement of $\boldsymbol{CH}^2 \cup (S^3 - p)$, in which we extend our defining functions. The point is that we want to extend these functions so that they are smooth across $S^3 - p$.

As in Section 2.2.2, let $\mathrm{DIAM}_x(S)$ be the visual diameter of $S \subset S^3$ as seen from $x \in \boldsymbol{CH}^2$. In our setting we have $\langle f \rangle_t = f^{-1}[t, 3]$. Here is our result.

Lemma 6.1 (Extension) *f has a smooth extension $E : \boldsymbol{U}_p \to [0, 4]$ that satisfies the following for any sequence $\{x_n\}$ of points in $\boldsymbol{CH}^2$.*

- *If $E(x_n) \to 0$, then $\{\mathrm{DIAM}_{x_n}(\langle f \rangle_1)\} \to 0$.*
- *If $\{\mathrm{DIAM}_{x_n}(\langle f \rangle_0)\} \to 0$, then $E(x_n) \to 0$.*

At the end of the chapter, we will deduce, as a corollary of our proof, the following lemma.

Lemma 6.2 *Let T be a horotube based at $p \in S^3$. Let $\langle P \rangle$ be a $\boldsymbol{Z}$ parabolic subgroup that stabilizes T. Let $\{x_n\}$ be a sequence of points in $\boldsymbol{CH}^2$ converging to p. Suppose that $\mathrm{DIAM}_{x_n}(T) < 1/n$. Then there is a sequence $\{k_n\}$ such that $P^{k_n}(x_n)$ converges to a point of $S^3 - p$ on a subsequence.*

6.2 PROOF OF THE EXTENSION LEMMA

We work in the Siegel model Z, given in Equation 2.9:

$$Z = \{(z,w)|\ \mathrm{Re}(w) > |z|^2\} \subset \boldsymbol{C}^2 \subset \boldsymbol{CP}^2. \tag{6.2}$$

The action of P is conjugate either to the map in Equation 2.15 or Equation 2.16, and we normalize P so that it has one of these two forms. Either case has the same proof. We are working in $\mathcal{H}$ with $p = \infty$ and $\boldsymbol{U}_p = \boldsymbol{C}^2$. We have a projection $\pi : \boldsymbol{C}^2 \to \partial Z$ given by

$$\pi(z,w) = \big(z, |z|^2 + i\ \mathrm{Im}(w)\big). \tag{6.3}$$

Let $b : \boldsymbol{R} \to [0,1]$ be a smooth nondecreasing function such that $b(x) = 0$ iff $x \le 0$ and $\beta(x) = 1$ iff $x \ge 1$. On $\boldsymbol{C}^2$ we define the smooth function

$$\beta(z,w) = b\big(\mathrm{Re}(w) - |z|^2\big). \tag{6.4}$$

This function is positive on Z and 0 on ∂Z. Our extension is given by

$$E(x) = \beta(x) + f\big(\pi(x)\big). \tag{6.5}$$

Easy calculations show that $\beta \circ P = \beta$ and $\pi \circ P = P \circ \pi$. By hypothesis $f \circ P = f$. Combining all these facts gives us $E \circ P = E$.

6.2.1 Property 1

Suppose that $\{x_n\} \in Z$ with $E(x_n) \to 0$. Let's write $x_n = (z_n, w_n)$. We have $f(\pi(x_n)) \to 0$. In particular we have $\pi(x_n) \in \partial Z - \langle f \rangle_1$, a set that is compact mod P. Hence, there is a compact $K \subset \partial Z$ such that the P-orbit of $\pi(x_n)$ lies in K for all n. Replacing x_n by a P-equivalent point, if necessary, we can assume that $\pi(x_n) \in K$ for all n. Hence, $|z_n|$ and $\mathrm{Im}(w_n)$ are bounded. On the other hand, we have $\beta(x_n) \to 0$. This means that $\mathrm{Re}(w_n) - |z_n|^2 \to 0$. Hence, $\mathrm{Re}(w_n)$ is bounded. Hence, $|w_n|$ and $|z_n|$ are bounded. Hence, $\{x_n\}$ lies in a bounded subset of $\boldsymbol{C}^2$.

The fact that $\beta(x_n) \to 0$ forces x_n to exit every compact subset of Z. Given that $\{x_n\}$ lies in a bounded subset of Z, we can pass to a subsequence so that $\{x_n\}$ converges to a finite point $x \in \partial Z$. Since $f(\pi(x_n)) \to 0$, we have $x \notin \langle f \rangle_{1/2}$, say. All in all, x_n converges to a point $x \in \partial Z$ that is not contained in the closure of $\langle f \rangle_1$. But then $\mathrm{DIAM}_{x_n}(\langle f \rangle)_1 \to 0$.

6.2.2 Property 2

We suppose that $E(x_n) > s > 0$ for all n, and we show that $\mathrm{DIAM}_{x_n}(\langle f \rangle_0)$ is bounded away from 0. Let h_n be an automorphism of Z, fixing ∞, such that $h_n(x_n) = (0,1)$. It suffices to prove that the complements $\partial Z - h_n(\langle f \rangle_0)$ cannot Hausdorff-converge to all of ∂Z. Interpreting ∂Z as $\mathcal{H}$, the map h_n acts as a Carnot similarity with dilation factor λ_n. Taking a subsequence, we get the following three cases.

Case 1: Suppose that $\mathrm{Re}(w_n) - |z_n|^2 \to \infty$. This means that $\lambda_n \to 0$. Intuitively, these points are "rising away" from ∂Z, and h_n "pulls them back down." There is some D such that $\mathcal{H} - \langle f \rangle_0$ is within D units of a straight line in ∂Z. Hence, $\mathcal{H} - h_n(\langle f \rangle_0)$ is within $\lambda_n D$ units of some straight line in $\mathcal{H}$. Certainly such sets cannot Hausdorff-converge to all of $\mathcal{H}$.

Case 2: Suppose that $\mathrm{Re}(w_n) - |z_n|^2$ converges to a finite nonzero number. Using the notation from Case 1, the dilation factors λ_n converge to a nonzero finite number. But then there is some uniform D' such that $\mathcal{H} - h_n(\langle f \rangle_0)$ is within D' units of a straight line L_n in $\mathcal{H}$. Once again, it is not possible for $\mathcal{H} - h_n(\langle f \rangle_0)$ to Hausdorff-converge to all of $\mathcal{H}$.

Case 3: Suppose that $\mathrm{Re}(w_n) - |z_n|^2 \to 0$. Then $\beta(x_n) \to 0$. Hence, $f(\pi(x_n)) > s/2 > 0$ once n is large. We can write $h_n = e_n \circ p_n$, where p_n is a parabolic automorphism with $p_n(x_n) = (0, \epsilon_n)$ and e_n acts on $\mathcal{H}$ as a dilation that fixes the origin. Necessarily $\epsilon_n \to 0$. Since $f(x_n) > s/2$, there is a uniformly large Carnot ball about $\pi(x_n)$ contained in $\langle f \rangle_0$. Since p_n is a Carnot isometry, there is some uniformly large Carnot ball B centered at the origin of $\mathcal{H}$ and contained in $p_n(\langle f \rangle_0)$. But then $e_n(B)$ Hausdorff-converges to all of $\mathcal{H}$ and, hence, so does $h_n(\langle f \rangle_0)$. □

6.3 PROOF OF THE AUXILIARY LEMMA

Here we prove Lemma 6.2. We can choose a P-invariant horotube function f such that $\langle f \rangle_0 \subset T$. We can then extend f to E, as above. Our hypotheses on $\{x_n\}$ say that $\mathrm{DIAM}_{x_n}(T) \to 0$. Hence, $E(x_n) \to 0$, by Property 2. But now we just look at the proof of Property 1 above and see that it gives us exactly what we want—namely, we can replace each x_n by a P^{k_n}-translate that lies in a compact subset of $\boldsymbol{C}^2$ and, as $n \to \infty$, exits every compact subset of Z. □

Chapter Seven

Transplanting Horotube Functions

7.1 STATEMENT OF RESULTS

Let $P \in PU(2,1)$ be a parabolic element, fixing p. Let $\boldsymbol{U}_p$ be as in Equation 6.1. We use the Siegel model from Section 2.3, with $p = \infty$ and $\boldsymbol{U}_p = \boldsymbol{C}^2$. Let $\{P_n\}$ be a sequence of elements in $PU(2,1)$ with the property that that $P_n \to P$ geometrically, in the sense of Section 3.1.

Lemma 7.1 (Transplant) *Suppose that the sequence $\{P_n\}$ converges geometrically to P. Let $F : \boldsymbol{U}_p \to \boldsymbol{R}$ be a P-invariant smooth function. For each n there is a P_n-invariant smooth function $F_n : \boldsymbol{U}_p \to \boldsymbol{R}$. Moreover, $\{F_n\} \to F$ in $C^\infty(\boldsymbol{U}_p)$.*

For ease of exposition, we prove the Transplant Lemma when $\{P_n\}$ consists of elliptic elements. The other cases are similar and are actually easier.

7.2 A TOY CASE

Before proving the Transplant Lemma we treat a toy case, which is set in $\boldsymbol{H}^2$. The logic of the general case is the same as the logic in the toy case, except that the sets are higher dimensional.

We use the upper half-plane model of $\boldsymbol{H}^2$ and set $P(z) = z + 1$. Let $\{P_n\}$ be a sequence of elliptic isometries of $\boldsymbol{H}^2$ that converges geometrically to P.

Lemma 7.2 (Toy Case) *Suppose $F : \boldsymbol{C} \to \boldsymbol{R}$ is a smooth P-invariant function. For each n there is a P_n-invariant smooth function $F_n : \boldsymbol{C} \to \boldsymbol{R}$. Moreover, $F_n \to F$ in $C^\infty(\boldsymbol{C})$.*

7.2.1 Step 1

A fundamental domain for the action of P on $\boldsymbol{C}$ is given by the infinite vertical strip

$$V = \{z |\ \mathrm{Re}(z) \in [0,1]\} \tag{7.1}$$

with boundary components $V_j = \{z |\ \mathrm{Re}(z) = j\}$ for $j = 0, 1$. We can find a smooth function $\alpha : V \to [0,1]$ such that $\alpha \equiv j$ in a neighborhood of V_j. We do this by setting $\alpha(z) = a(\mathrm{Re}(z))$, where $a : \boldsymbol{R} \to \boldsymbol{R}$ is a smooth function equal to 0 in $(-\infty, 1/3)$ and 1 in $(2/3, \infty)$.

7.2.2 Step 2

Let $\{h_n\}$ be an unbounded sequence whose growth has yet to be determined. We shall want this sequence to converge to ∞ but to grow very slowly in comparison to the rate at which P_n converges to P. This condition on $\{h_n\}$ comes up often, and we will abbreviate it by saying that $\{h_n\}$ is *chosen wisely*. Let $R_n = [0,1] \times [-h_n, h_n]$. For later purposes we give an alternate definition of R_n. Let $\{P_r\}$ be the 1-parameter subgroup that contains P. That is, $P_r(z) = z + r$. Let $B(h_n)$ denote the vertical segment of length $2h_n$ centered at 0. Then

$$R_n = \bigcup_{r=0}^{1} P_r(B(h_n)). \tag{7.2}$$

R_n is the the shaded part of Figure 7.1 below.

If h_n is chosen wisely then the circle $C_n = P_n(V_0 \cup \infty)$ is transverse to ∂H_n and nearly parallel to V_1 along the right edge of R_n. Hence, there is a piecewise-analytic quadrilateral $\widehat{R}_n$, bounded by 3 line segments and an arc of C_n, which is Hausdorff close to R_n. For the purposes of generalization, we say more formally that

$$\widehat{R}_n = P_n(L_0) \cap \bigcup_{r=0}^{\infty} P_r\big(B(h_n)\big), \tag{7.3}$$

where L_0 is the left half-plane bounded by V_0. In Figure 7.1, the boundary $\partial\widehat{R}_n$ is drawn with thick lines. If $\{h_n\}$ is wisely chosen, then the Hausdorff distance between $\widehat{R}_n \cap K$ and $R_n \cap K$ tends to 0 if K is fixed and $n \to \infty$.

7.2.3 Step 3

Here we characterize how powers of P_n act on $\widehat{R}_n$. Let

$$I_n^k = P_n^k(\widehat{R}_n) \cap \widehat{R}_n. \tag{7.4}$$

For n large the two sets $\widehat{R}_n$ and $P_n(\widehat{R}_n)$ lie on opposite sides of C_n. Hence, I_n^1 amd I_n^{-1} are contained in the right and left "edges" of $\widehat{R}_n$, respectively.

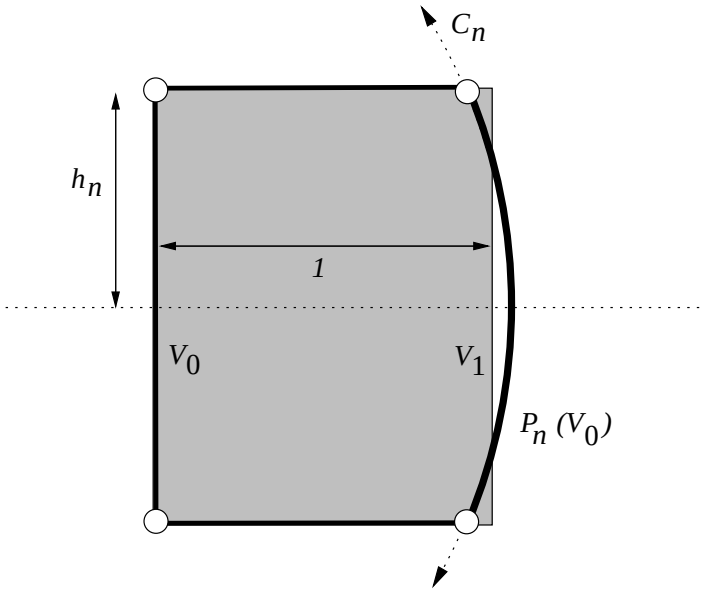

Figure 7.1: The quadrilaterals R_n and $\widehat{R}_n$

Lemma 7.3 *There is an unbounded sequence $\{h_n\}$ such that $I_n^k = \emptyset$ for all $|k| \geq 2$ and n large.*

Proof: If this lemma is false, then there is some h such that the choice $h_n \equiv h$ leads to an exponent $|k_n| \geq 2$ such that $I_n^{k_n} \neq \emptyset$. But then $\widehat{R}_n$ converges to a $1 \times 2h$ rectangle R_0. By construction some point of R_0 is within $1/2$ of some point of $P_n^{k_n}(R_0)$ for n large. Call this the *short translation property*. Since the fixed point of P_n tends to ∞, the short translation property forces $\{P_n^{k_n}\}$ to lie in a precompact set of maps. Since $P_n \to P$ geometrically, Lemma 3.1 forces a uniform bound on the true size of $\{k_n\}$. Passing to a subsequence, we can assume that $k_n \equiv k$. But then P_n^k converges to the translation P^k. However, $P^k(R_0)$ and R_0 are separated by at least one unit, contradicting the short translation property of P_n for n large. □

We include the above property in the wise choice property of $\{h_n\}$. In summary, if $\{h_n\}$ is wisely chosen and n is large, then I_n^k is contained in a side of $\widehat{R}_n$ for $|k| = 1$ and is empty otherwise.

7.2.4 Step 4

If h_n is wisely chosen, then the function α, from Step 1, will be 0 in a neighborhood of the left boundary component of $\widehat{R}_n$ and 1 in a neighborhood of the right boundary of $\widehat{R}_n$. On $\widehat{R}_n$ we define

$$F_n(z) = \big(1 - \alpha(z)\big)F(z) + \alpha(z)F\big(P_n^{-1}(z)\big). \tag{7.5}$$

Given z_0 in the left boundary component of $\widehat{R}_n$, we define $z_1 = P_n(z_0)$. Using the definitions, we compute

$$F_n(z_1) = 0 \times F(z_1) + 1 \times F\big(P_n^{-1}(z_1)\big) = F(z_0) = F_n(z_0).$$

No identifications are made on the top and bottom boundary components of $\widehat{R}_n$. Hence, the restriction of F_n to $\partial\widehat{R}_n$ is consistent with the action of P_n.

Let $A_n \subset \boldsymbol{C} \cup \infty$ denote the orbit of $\widehat{R}_n$ under the action of $\langle P_n \rangle$. We can then extend F_n to all of A_n by the action of $\langle P_n \rangle$. This extension can be made consistently because of Equation 7.5.

We extend F_n to all of $\boldsymbol{C}$ using a bump function. Since the sequence $\{A_n\}$ forms an exhaustion of $\boldsymbol{C}$, we do not care about the choice of bump function. The sequence $P \circ P_n^{-1}$ converges smoothly to the identity map on $\boldsymbol{C}$, uniformly on compact subsets of $\boldsymbol{C}$. Therefore, if we set

$$G = F \circ P_n^{-1} - F = F \circ (P \circ P_n^{-1}) - F,$$

then we see that $G \to 0$ in $C^\infty(\boldsymbol{C})$. Since $F_n - F = \alpha G$, we see that $F_n \to F$ in $C^\infty(\boldsymbol{C})$. □

7.3 PROOF OF THE TRANSPLANT LEMMA

We use the horospherical coordinates discussed in Section 2.3.4 and normalize so that P is either as in Equation 2.18 or Equation 2.19, namely,

$$\big((z,t),x\big) \to \big((uz,t+1),x\big). \tag{7.6}$$

$$\big((z,t),x\big) \to \big((1,s)\cdot(z,t),x\big). \tag{7.7}$$

In both cases we can include P in a 1-parameter family of maps $\{P_r\}$, which permute a parallel family $\{\Pi_r\}$ of 3-planes. Compare Lemma 2.7. In the case of Equation 2.18 we can take $\Pi_0 = \boldsymbol{C} \times \{0\} \times \boldsymbol{R}$, and in the case of Equation 2.19, we can take $\Pi_0 = i\boldsymbol{R} \times \boldsymbol{R} \times \boldsymbol{R}$. We set $\Pi_r = P_r(\Pi_0)$.

7.3.1 Step 1

Define

$$V = \bigcup_{r=0}^{1} \Pi_r. \tag{7.8}$$

Then V is a fundamental domain for the action of P on $\boldsymbol{C}^2$. Referring to Step 1 of the toy case, we define $\alpha(z,w) = a(r)$, where r is such that $(z,w) \in \Pi_r$. Then $\alpha = 0$ in a neighborhood of Π_0 and $\alpha = 1$ in a neighborhood of Π_1.

7.3.2 Step 2

For any $h \in \boldsymbol{R}^+$, let $B(h)$ denote the metric ball of radius h about the origin in Π_0. We choose a slowly growing sequence $\{h_n\}$ and define

$$R_n = \bigcup_{r=0}^{1} P_r\big(B(h_n)\big). \tag{7.9}$$

Compare Equation 7.2. Topologically R_n is the product of a 3-ball with $[0, 1]$.

If h_n is wisely chosen, then the set $\widehat{R}_n$, defined just as in Equation 7.3, will be topologically equivalent to R_n and geometrically close. This follows from transversality. We can think of $\widehat{R}^n$ as a set obtained by slightly warping the right boundary component of R_n.

7.3.3 Step 3

We define I_n^k as in Step 3 of the toy case. The obvious variant of Lemma 7.3 has essentially the same proof. Hence, $\{h_n\}$ can be wisely chosen so that, for n large, I_n^k is contained in a face of $\widehat{R}_n$ for $|k| = 1$ and otherwise empty.

7.3.4 Step 4

We define $A_n \subset \boldsymbol{CP}^2$ to be the $\langle P_n \rangle$-orbit of $\widehat{R}_n$. Now that all the relevant objects have been defined, this step is exactly the same as in the toy case. □

Chapter Eight

The Local Surgery Formula

8.1 STATEMENT OF RESULTS

Recall from Section 2.6.2 that a horotube is *nice* if its boundary is a smooth cylinder and the horotube is stabilized by a 1-parameter parabolic subgroup. In this chapter we will describe how the Dehn surgery works for individual nice horotubes. In Chapter 13 we will see this picture occur around each cusp of the manifold $M = \Omega/\Gamma$ from the HST.

Let T be a nice horotube, stabilized by a parabolic group $\langle P \rangle$. We normalize so that P is either as in Equation 2.12 or Equation 2.13. In the former case we also assume that $u \neq -1$. Let $\Omega = \mathcal{H}$ be the domain of discontinuity for the group $\langle P \rangle$. Let $\{P_n\}$ be a sequence of elements converging geometrically to P. We choose some large (but unspecified) n and let $\widehat{P} = P_n$. Let $\widehat{\Omega}$ be the domain of discontinuity for $\langle \widehat{P} \rangle$. Then $\widehat{\Omega}$ is S^3 minus 0, 1, or 2 points, depending on the type of $\widehat{P}$. In this chapter we will compare $\Upsilon = \Omega/\langle P \rangle$ to $\widehat{\Upsilon} = \widehat{\Omega}/\langle \widehat{P} \rangle$.

We define

$$R = \text{closure}(T)/\langle P \rangle, \qquad W = (\Omega - T)/\langle P \rangle. \tag{8.1}$$

Then

$$\Upsilon = R \cup W, \qquad \partial R = \partial W = R \cap W. \tag{8.2}$$

Note that $R = T^2 \times [0, \infty)$, where T^2 is a torus. R is the local model for a torus end of the manifold M.

Let f be a defining function for T, so that $T = \langle f \rangle_1$. Let $\widehat{f}$ be the C^∞-perturbation of f produced by the Transplant Lemma. Let

$$\widehat{T} = \langle \widehat{f} \rangle_1. \tag{8.3}$$

We define

$$\widehat{R} = \text{closure}(\widehat{T})/\langle \widehat{P} \rangle; \qquad \widehat{W} = (\widehat{\Omega} - \widehat{T})/\langle \widehat{P} \rangle. \tag{8.4}$$

Then

$$\widehat{\Upsilon} = \widehat{R} \cup \widehat{W}; \qquad \partial \widehat{R} = \partial \widehat{W} = \widehat{R} \cap \widehat{W}. \tag{8.5}$$

$\widehat{R}$ is the model for the end of $\widehat{M}$—i.e., the model for the surgered end of M. As in the unperturbed case, $\widehat{W}$ is far away from the surgered end. We think of W as being white and R as being red. All the action, so to speak, takes place in the red part. The purpose of this chapter is to prove the following lemma.

Lemma 8.1 (Local Surgery) *There is a canonical marking on ∂R with the following property. Relative to this marking $\widehat{\Upsilon}$ is obtained by performing a filling on Υ according to the following scheme.*

- *If $\widehat{P}$ is parabolic, then no filling is done.*
- *If $\widehat{P}$ is lens-elliptic of type (m,k) and sufficiently close to P, then the filling has type (m,k).*
- *If $\widehat{P}$ is loxodromic, then the filling has type $(0,1)$.*

In all cases, there is a homeomorphism from W to $\widetilde{W}$, canonical up to homotopy, that is C^0-close to the identity when lifted to a fundamental domain for ∂W.

Remark: The purpose of the homeomorphism between W and $\widehat{W}$ is to formalize the sense (in the nonparabolic case) in which $\widehat{\Upsilon}$ may be viewed as the space obtained by gluing the solid torus $\widehat{R}$ onto W, according to the above scheme. The "closeness" should be interpreted as follows. Our definition of $\widehat{P}$ depends on n, and as $n \to \infty$, the lift of our homeomorphism converges uniformly to the identity.

8.2 THE CANONICAL MARKING

To canonically identify $H_1(\partial R)$ with $\boldsymbol{Z}^2$, we need to choose a canonical basis $\{\alpha, \beta\}$ for $H_1(\partial R)$. We then identify α with $(0,1)$ and β with $(1,0)$.

Suppose first that P is as in Equation 2.12, with $u \neq -1$. We call this the $\boldsymbol{C}$-case. Since $u \neq -1$ there is a unique choice $u' \in \boldsymbol{R}$, of minimal length, such that $\exp(iu') = u$. (When $u = -1$ the two choices $u' = \pm\pi$ are the two choices of minimal length.) P includes, as the element P_1, in a 1-parameter subgroup $\{P_r\}$, which has the action

$$P_r(z,t) = \big(\exp(2\pi i u' r)z, t + r\big). \tag{8.6}$$

Here is our construction.

- $\partial T \cap (\boldsymbol{C} \times \{0\})$ is a closed circle on ∂T. We define $\alpha \in H_1(\partial R)$ as the homology class represented by this closed loop, given the clockwise orientation.
- We let $\beta \in H_1(\partial R)$ be the homology class represented by a quotient of the form $\{P_r(x)\}/\langle P\rangle$ for $x \in \partial T$. That is, we take an orbit of $\{P_r\}$ on ∂T and quotient out by the action of P_1 to obtain a closed loop. We orient β in the direction of increasing r.

The left-hand side of Figure 8.1 shows a picture. The two endpoints of β are identified in the quotient. Note that α is trivial in $H_1(W)$ and β generates $H_1(W) = \boldsymbol{Z}$.

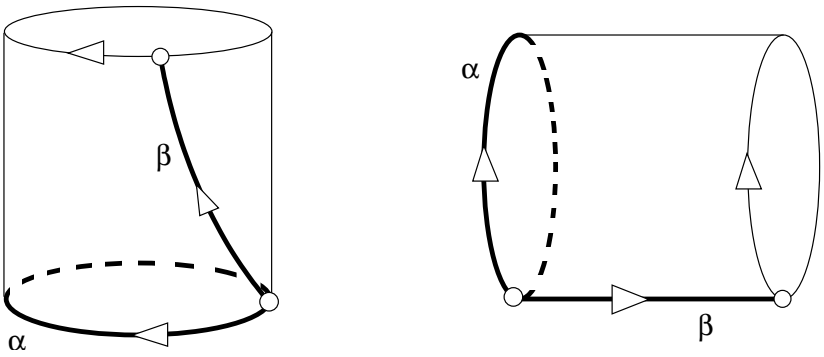

Figure 8.1: The canonical basis for homology

Suppose now that P is as in Equation 2.13. We call this the $\boldsymbol{R}$-case, even though P stabilizes an $\boldsymbol{R}$-circle only when $s = 0$. In any case, $P = P_1$ includes in the 1-parameter family $P_r(z,t) = (r, sr) \cdot (z,t)$.

- $\partial T \cap (\boldsymbol{R} \times i\boldsymbol{R})$ is a closed loop. We let $\alpha \in H_1(\partial R)$ be the homology class represented by one of the two possible orientations of α.
- We define β just as in the $\boldsymbol{C}$-case, with respect to $\{P_r\}$.

The right-hand side of Figure 8.1 shows a picure, when $s = 0$. When $s \neq 0$ one should imagine the picture "shearing" in the vertical direction. The two endpoints of α are identified in the quotient. We have yet to specify the orientation of α. We specify the orientation on α so that the triple $\alpha, \beta, \vec{v}$ gives the same orientation on ∂R in the $\boldsymbol{R}$-case as in the $\boldsymbol{C}$-case. Here $\vec{v}$ is the outward normal field. As in the $\boldsymbol{C}$-case, α is trivial in $H_1(W)$ and β generates $H_1(W) = \boldsymbol{Z}$.

Note that the right-hand side of Figure 8.1 is the same as the left-hand side but turned on its side. (We chose the orientations so as to get this property.) Our drawing coordinates are such that the $\boldsymbol{R}$-factor in $\mathcal{H} = \boldsymbol{C} \times \boldsymbol{R}$ is the vertical direction.

8.3 THE HOMEOMORPHISM

On any compact subset $K \subset \mathcal{H}$, the set $K \cap \partial\widehat{T}$ converges to $K \cap \partial T$ in the Hausdorff topology. We can choose K to contain a fundamental domain F' for ∂T. We can also choose F' to be an annulus, as in Figure 8.1. We can choose a nearby annular fundamental domain $\widehat{F}'$ for the action of $\widehat{P}$ on $\partial\widehat{T}$. Then each of the two loops comprising $\partial F'$ will be close to the corresponding loops bounding $\partial\widehat{F}'$.

Let F'_0 be the boundary component of F' so that $P(F'_0) = F'_1$. Let $\widehat{F}'_0$ be the corresponding boundary component of $\widehat{F}'$. We first map the loop $\widehat{F}'_0$ to the loop F'_0. Call this map h. We will promote h into a homeomorphism in a cell-by-cell way. Using the action of $\widehat{P}$ and P, we extend h so that it

maps F_1' to $\widehat{F}_1'$. Next, we extend h so that it maps the interior of F' to the interior of $\widehat{F}'$, in a way that is close to the identity.

F_0' bounds a disk D_0, which is contained in an affine plane. The interior of this disk is contained $\Omega - T$. We can find a disk $\widehat{D}^0$, bounded by $\widehat{F}_0'$, such that $\widehat{D}_0$ is C^∞-close to D_0 and the interior of $\widehat{D}_0$ is contained in $\widehat{\Omega} - \widehat{T}$. We set $\widehat{D}_1 = \widehat{P}(\widehat{D}_0)$. We extend h so that it maps D_0 to $\widehat{D}_0$. Using the action of P and $\widehat{P}$, we extend h so that h maps D_1 to $\widehat{D}_1$.

h now defines a homeomorphism from the boundary of a fundamental domain for $\Omega - T$ to the boundary of a fundamental domain for $\widehat{\Omega} - \widehat{T}$. These fundamental domains have the topological form $D^2 \times S^1$. In this situation, h clearly extends to the interiors. h descends to give the desired homeomorphism from W to $\widehat{W}$.

8.4 THE SURGERY FORMULA

In the parabolic case Υ and $\widehat{\Upsilon}$ are clearly homeomorphic. In the loxodromic case we have $\widehat{\Upsilon} = S^2 \times S^1$. The only surgery on a solid torus that yields $S^2 \times S^1$ is the one that kills α. This follows from Van Kampen's theorem. Hence, $\widehat{\Upsilon}$ is obtained from Υ by Dehn filling of type $(0, 1)$. Notice that the choice of β is irrelevant here. The rest of the section is devoted to the elliptic case. For ease of exposition, we will consider the $\boldsymbol{C}$-case. The $\boldsymbol{R}$-case is similar.

Given that the horotube function f has a standard form, we can obtain $\widehat{f}$ by a technique that is more concrete and explicit than the one given in the proof of the Transplant Lemma. Of course, the more explicit technique is closely related to the general technique.

We write

$$T = \bigcup_{s \in [0,m]} P_r(D), \tag{8.7}$$

where D is the complement of the unit disk in $\boldsymbol{C} \times \{0\}$. Rather than construct $\widehat{f}$ from the Transplant Lemma, we can define $\widehat{T}$ as the union

$$\widehat{T} = \bigcup_{s \in [0,m]} \widehat{P}_r(D). \tag{8.8}$$

We can then define $\widehat{f}$ in such a way that $\widehat{T} = \langle \widehat{f} \rangle_1$. (Compare Step 2 in the proof of the Transplant Lemma of Chapter 7.)
Here are some observations about Υ.

- $\Upsilon = R \cup W$ is homeomorphic to an open solid torus.
- $\partial R = \partial W$ is a torus because ∂T is a cylinder.
- W is a solid torus because $f^{-1}[0, 1]$ is a solid cylinder.
- $R = \partial R \times [1, \infty)$.

Here are the corresponding observations about $\widehat{\Upsilon}$.

- $\widehat{\Upsilon} = \widehat{R} \cup \widehat{W}$ is a lens space whose fundamental group is $\boldsymbol{Z}/m$.
- $\partial\widehat{R} = \partial\widehat{W}$ is a torus because $\widehat{f}^{-1}(1)$ is a torus.
- $\widehat{W}$ is a solid torus because $\widehat{f}^{-1}[0,1]$ is a solid torus.
- $\widehat{R}$ is a solid torus.

Looking at these observations side by side and using our homeomorphism h from the previous section, we can see that topologically $\widehat{\Upsilon}$ is obtained by attaching the solid torus $\widehat{R}$ onto W along ∂W. Here we are identifying W with $\widehat{W}$ *via* h. It only remains to uncover the type of surgery that is done in the elliptic case.

There is a canonical basis for $H_1(\partial\widehat{R})$. We define $\widehat{\alpha}$ to be homology class generated by ∂D. We define $\widehat{\beta}$ as we defined β in the previous section, using k/m in place of u'. The homeomorphism from ∂W to $\partial\widehat{W}$, being close to the identity, carries α to $\widehat{\alpha}$ and β to $\widehat{\beta}$.

Once we identify W with $\widehat{W}$, our Dehn filling has type (m', k'), where m' and k' are defined by the following property. The curve

$$\widehat{\gamma} = k'\widehat{\alpha} + m'\widehat{\beta} \tag{8.9}$$

is trivial in $H_1(\widehat{R})$. Using the fact that $\pi_1(S^3/\langle P\rangle) = \boldsymbol{Z}/m$, we see that $m' = m$.

We have an m-fold covering map $\pi : \partial\widehat{T} \to \partial\widehat{R}$, with the deck group given by $\langle\widehat{P}\rangle$. Let $\widehat{\alpha}_m$, $\widehat{\beta}_m$, and $\widehat{\gamma}_m$ be the lifts to $\partial\widehat{T}$. We think of these objects as actual curves on $\partial\widehat{T}$.

Since $\widehat{\alpha}$ is trivial in $H_1(\widehat{W})$, there are m distinct lifts of $\widehat{\alpha}$ to $\partial\widehat{T}$. Therefore, on the level of homology,

$$\pi_*(\widehat{\alpha}_m) = \widehat{\alpha}. \tag{8.10}$$

On the other hand, $\widehat{\beta}_m$ is just the orbit of $\{P_r|\ s \in [0,m]\}$. Therefore,

$$\pi_*(\widehat{\beta}_m) = m\widehat{\beta}. \tag{8.11}$$

Finally, $\widehat{\gamma}$ is trivial in $\widehat{R}$ in the same way that $\widehat{\alpha}$ is trivial in $\widehat{W}$. Therefore,

$$\pi_*(\widehat{\gamma}_m) = \widehat{\gamma}. \tag{8.12}$$

Note that $\widehat{\alpha}_m$ and $\widehat{\beta}_m$ serve as a basis for $H_1(\partial\widehat{T})$. We can write

$$\widehat{\gamma}_m = k''\widehat{\alpha}_m + m''\widehat{\beta}_m. \tag{8.13}$$

Applying π_* to both sides and using Equations 8.10, 8.11 and 8.12 we get

$$\widehat{\gamma} = k''\widehat{\alpha} + m''m\widehat{\beta}. \tag{8.14}$$

Combining Equations 8.9 and 8.14 we get $k'' = k'$ and $m'' = 1$. Therefore,

$$\widehat{\gamma}_n = k'\widehat{\alpha}_m + \widehat{\beta}_m. \tag{8.15}$$

The group $\widehat{P}_r$ twists k times as fast around the core curve

$$C_m = \{P_r(0,0)|s \in [0,m]\} \tag{8.16}$$

as it translates this curve. Hence, $\widehat{\beta}_m$ links C_m k times. On the other hand, α_m links C_m -1 times. Finally, γ_m links C_m 0 times. From this linking information and from Equation 8.15 we see that $k' = k$.

Chapter Nine

Horotube Assignments

9.1 BASIC DEFINITIONS

In the first three chapters of Part 2 we have considered individual horotubes and horotube functions. Now we thicken the plot and consider collections of these objects. In this chapter we do not assume that our group Γ is a horotube group. However, we assume that Γ shares many properties of a horotube group, namely the following.

- Γ is discrete.
- The limit set Λ is more than a single point.
- The regular set Ω is nonempty.
- Γ has a finite and nonzero number of conjugacy classes of rank one parabolic subgroups, and no rank-2 parabolic subgroups.

It would seem reasonable to allow for the possibility that Γ has some rank-2 parabolic subgroups but we forbid this to make the exposition simpler.

Let $\mathrm{Cusp}_p(\Gamma)$ denote the set of fixed points of parabolic elements. Each $p \in \mathrm{Cusp}(\Gamma)$ is stabilized by a unique maximal parabolic $\boldsymbol{Z}$ subgroup of Γ. We say that a *horotube assignment* is a Γ-equivariant assignment $p \to T_p$, where T_p is a horotube for each $p \in \mathrm{Cusp}(\Gamma)$. We require the following properties.

Property 1: The closure of T_p, in S^3, is contained in $\Omega \cup \{p\}$.

Property 2: Every point $q \in \Omega$ is contained in the closure of at least one horotube.

Property 3: Every compact subset of Ω intersects at most finitely many horotubes.

Property 4: For each $p \in \mathrm{Cusp}(\Gamma)$ there is a second horotube $T_p' \subset T_p$, such that $T_p' \cap T_q = \emptyset$ for all $q \neq p$. The assignment $p \to T_p'$ is also equivariant.

Below, whenever we speak of the "horotubes in our assignment," we refer to the horotubes of the form T_p and not the auxiliary horotubes of the form

T'_p. The smaller horotubes $\{T'_p\}$ are used only occasionally, for technical purposes.

9.2 THE MAIN RESULT

Lemma 9.1 *Let T_1 be one of the horotubes in our assignment. (And $T'_1 \subset T_1$ is the smaller horotube.) Let $\{g'_n\}$ be a sequence in $PU(2,1)$. Let Γ_1 denote the parabolic stabilizer of T_1. Suppose that the spherical diameter of $g'_n(\partial T_1)$ is uniformly bounded away from 0. Then there is a sequence $\{\gamma_n\} \in \Gamma_1$ such that the elements $g_n := g'_n \circ \gamma_n$ satisfy one of the following properties on a subsequence.*

1. *g_n converges in $PU(2,1)$.*
2. *$S^3 - g_n(T_1)$, $S^3 - g_n(T'_1)$, and $g_n(\Lambda)$ all converge to the same $\boldsymbol{F}$-circle. Hence, the spherical volume of $g_n(T_1)$ converges to* $\mathrm{vol}(S^3)$.
3. *There is a list T_2,...,T_M of horotubes and a compact*

$$K \subset \bigcup_{i=1}^{M} \overline{T}_i \cap \Omega$$

such that $S^3 - g_n(K)$ converges to a point. In particular, $g_n(\Lambda)$ converges to a point, and some horotube $g_n(T_i)$ has spherical volume converging to at least $\mathrm{vol}(S^3)/M$.

Proof: Let b_1 be the basepoint of T_1. Passing to a subsequence we can assume that $\{g'_n(b_1)\}$ converges to a point in S^3. There is therefore a convergence sequence $\{R_n\} \subset PU(2,1)$ of isometric rotations of S^3 so that $R_n \circ g'_n(b_1) = b_1$. If our theorem is true for $\{R_n \circ g'_n\}$, then it is also true for $\{g'_n\}$. So, without loss of generality we can assume that $g'_n(b_1) = b_1$ for all n.

Let $\boldsymbol{B}$ be a Heisenberg stereographic projection, which maps b_1 to ∞. There is some compact $S \subset \Omega$ such that the orbit $\Gamma_1 S$ contains ∂T_1. Hence, we can find $\gamma_n \in \Gamma_1$ such that $\boldsymbol{B}(g_n(S))$ contains a point of $\boldsymbol{B}(g'_n(\partial T_1))$, which is as close as possible to the origin in $\mathcal{H}$. Here we have set $g_n = g'_n \circ \gamma_n$. Note that $g_n(T_1) = g'_n(T_1)$, so $g_n(\partial T_1)$ has uniformly large spherical diameter.

Define

$$\tau_n = \boldsymbol{B}\big(g_n(T_1)\big), \qquad h_n = \boldsymbol{B} \circ g_n \circ \boldsymbol{B}^{-1}, \qquad S' = \boldsymbol{B}(S).$$

Then $h_n(\tau_1) = \tau_n$, and h_n is some Carnot similarity of $\mathcal{H}$. Finally, $h_n(S')$ contains a point of $\partial\tau_n$, which is as close as possible to the origin. Since $g_n(T_1)$ has uniformly large spherical diameter, we can find a uniformly large spherical ball B_n, centered on a point of $g_n(\partial T_1)$, which is disjoint from b_1. Passing to a subsequence we can assume that $B = B_n$ is independent of n.

Since b_1 is disjoint from B, there is some D such that $B' = \boldsymbol{B}(B)$ is within D units of the origin in $\mathcal{H}$. Hence, some point $y_n \in \partial\tau_n$ is within D units of the origin. But we can find some point $x_n \in S' \cap \partial\tau$ such that $h_n(x_n) = y_n$. In other words, we can find two points x_n and y_n uniformly close to the origin in $\mathcal{H}$ such that $h_n(x_n) = y_n$. We call this the *small action property*. We chose the set S with a view towards setting up the small action property, but now we can forget about S.

Let $|h_n|$ denote the dilation factor of h_n. Passing to a subsequence, we arrive at three cases.

Case 1: If $\{|h_n|\}$ is bounded away from both 0 and ∞, then the small action property implies that $\{h_n\}$ falls within a compact set of maps. Thus, we can pass to a subsequence so that $\{g_n\}$ converges in $PU(2,1)$.

Case 2: Suppose that $|h_n| \to 0$. Note that $\partial\tau_1$ is within a bounded neighborhood of some straight line L_1. Hence, every point of $\partial\tau_n$ is within a vanishingly small neighborhood of the straight line $L_n = h_n(L_1)$. Hence, $\mathcal{H} - \tau_n$ is contained within a vanishingly small tubular neighborhood of L_n. From the small action property and the fact that h_n is a contraction, we see that the lines $\{L_n\}$ fall within a compact set of straight lines of $\mathcal{H}$. Passing to a subsequence, we can assume L_n converges to a line L' in $\mathcal{H}$. If L_1 is a $\boldsymbol{C}$-circle, then so is L'. If L_1 is not a $\boldsymbol{C}$-circle, then h_n acts in such a way to decrease the "slope" of L_1 and L' must be an $\boldsymbol{R}$-circle. The point here is that h_n acts quadratically in the vertical direction and linearly in the horitonzal directions. Our analysis shows that $h_n(T_1)$ converges to the complement of an $\boldsymbol{F}$-circle. The same proof works for $g_n(T_1')$.

Since $\Lambda \subset S^3 - T_1$, we know that $g_n(\Lambda)$ converges to a set X contained in the limiting $\boldsymbol{F}$-circle. However, we can say more. $h_n(\boldsymbol{B}(\Lambda))$ is stabilized by the subgroup $h_n \circ \boldsymbol{B} \circ \Gamma_1 \circ \boldsymbol{B}^{-1} \circ h_n^{-1}$, a group whose generator acts on L_n with vanishingly small translation length. This shows that every point of L_n is within some η_n of $h_n(\boldsymbol{B}(\Lambda))$, where $\eta_n \to 0$. Hence, X equals our $\boldsymbol{F}$-circle.

Case 3: Suppose that $|h_n| \to \infty$. Then $|h_n^{-1}| \to 0$. The Carnot ball $h_n^{-1}(B')$ has a diameter that shrinks to 0 and a center that is uniformly close to the origin. (The second property comes from the small action property of h_n.) Finally, $h_n^{-1}(B')$ always intersects $\partial\tau_1$ because B' intersects τ_n.

Passing to a subsequence, we can assume that $h_n^{-1}(B)$ converges to a single point $x \in \partial\tau_1$. There is some $\delta > 0$ such that the Carnot ball K' of radius δ centered at x lies in $\boldsymbol{B}(\Omega)$ and only intersects M other horotubes. The finiteness property comes from Property 3 of our horotube assignment. Note that $\mathrm{diam}(h_n(K')) \to \infty$ and $B' \subset h_n(K')$. Note also that the distance from the $h_n(\partial K')$ to $\partial B'$ tends to ∞ and the center of B' is uniformly close to the origin. These properties imply that $h_n(K')$ converges to all of $\mathcal{H}$. Let $K = \boldsymbol{B}^{-1}(K')$. Then $g_n(K) = \boldsymbol{B}^{-1}(h_n(K'))$ converges to $S^3 - b_1$. By construction K is contained in the closure of M horotubes. This is our third alternative. □

9.3 COROLLARIES

Corollary 9.2 *Let T be a horotube from our assignment. Let $\epsilon > 0$ be given. There is some $\delta > 0$ with the following property. If $g \in PU(2,1)$ is such that $g(T)$ has spherical diameter at least ϵ, then there is some (possibly different) horotube U in our assignment such that $g(U)$ has spherical volume at least δ.*

Proof: Suppose we fix ϵ and find a sequence of counterexamples $\{g'_n\}$ and $\{T_n\}$ to this lemma. Passing to a subsequence, we can assume that all the T_n are Γ-equivalent. Replacing g'_n by $g'_n \circ y_n$ for suitable $y_n \in \Gamma$, we can assume that $T_n = T_1$ for all n. Then the horotubes $g'_n(T_1)$ all have spherical diameter at least ϵ. If the spherical diameter of $g'_n(\partial T_1)$ converges to 0, then $g'_n(T_1)$ converges to S^3 minus a point and we are done. Thus we need only consider the case when the spherical diameter of $g'_n(\partial T_n)$ is uniformly bounded away from 0. By Lemma 9.1 we can pass to a subsequence and find some horotube U such that $g_n(U)$ has uniformly large spherical volume. But then $U_n = \gamma_n(U)$ is such that $g'_n(U_n)$ has uniformly large spherical volume. This is a contradiction. □

Corollary 9.3 *Suppose that $\{g'_n\}$ is a sequence of elements in $PU(2,1)$ such that $g'_n(\Lambda)$ does not converge to a single point. Suppose also that $\{T_n\}$ is a sequence of horotubes from our assignment such that $g'_n(T_n)$ has uniformly large spherical diameter. Then $g'_n(T_n)$ also has uniformly large spherical volume.*

Proof: Applying the same trick as in the proof of Corollary 9.2, we can assume that $T_n = T_1$. If $g'_n(\partial T_1)$ has vanishingly small diameter, then on a subsequence $g'_n(T_1)$ converges to S^3 minus a point, and our result is obvious. Thus, we can assume that the diameter of $g_n(\partial T_1)$ is uniformly bounded away from 0. When we apply Lemma 9.1, we can't get Case 3 because $g_n(\Lambda) = g'_n(\Lambda)$ is more than a point. Hence, we have Case 1 or 2. In either case, $g'_n(T_1) = g_n(T_1)$ has uniformly large spherical volume. □

Corollary 9.4 *Suppose that Γ has a porous limit set. There is some $\epsilon > 0$ with the following property. For every $g \in PU(2,1)$, there is some horotube T_g such that $g(T_g)$ has spherical volume at least ϵ.*

Proof: There is some uniform $\epsilon' > 0$ with the following property. Some $x \in g(\Omega)$ is at least ϵ' from $g(\Lambda)$. But x lies in the closure of one of our

horotubes $g(T)$, and in particular, x is at least ϵ' from the basepoint of T. Hence, $g(T)$ has spherical diameter at least ϵ'. Now we apply Corollary 9.2. □

Corollary 9.5 *For any $\epsilon > 0$ there are only finitely many horotubes in our assignment that have spherical diameter greater than ϵ.*

Proof: Assume for the sake of contradiction that $\{T_n\}$ is an infinite sequence of distinct horotubes all having spherical diameter at least ϵ. If the diameter of ∂T_n converges to 0, then T_n converges to all of S^3 as $n \to \infty$. But this contradicts the fact that the limit set Λ is more than a single point. Thus the spherical diameter of ∂T_n is also bounded away from 0. Passing to a subsequence we can assume that $T_n = g_n'(T_1)$ for some $g_n' \in \Gamma$. Now we can apply Lemma 9.1. Note that $T_n = g_n(T_1)$ as well.

Case 1: Here $\{g_n\}$ is eventually constant because Γ is discrete. But then the horotubes $\{T_n\}$ are not all distinct. This is a contradiction.

Case 2: Recall that T_1' is the smaller horotube with the same basepoint as T_1. The horotubes $g_n(T_n')$ converge to the complement of an $\boldsymbol{F}$-circle in S^3. However, for different n these horotubes are all disjoint, by Property 4 of our horotube assignment. This is impossible.

Case 3: In this case $g_n(\Omega)$ converges to S^3 minus a single point. But $g_n(\Omega) = \Omega$ because $g_n \in \Gamma$. Hence, Λ is a single point. This is a contradiction. □

Recall from Section 2.2.2 that $\mathrm{DIAM}_x(S)$ is the visual diameter of a subset $S \subset S^3$ as seen from a point $x \in \boldsymbol{CH}^2$.

Corollary 9.6 *Let $p \in \mathrm{Cusp}(\Gamma)$ and $\{q_n\} \subset \mathrm{Cusp}(\Gamma)$. Suppose there is some $\epsilon > 0$ and a sequence $\{x_n\} \in \boldsymbol{CH}^2$ such that $\mathrm{DIAM}_{x_n}(T_p) > \epsilon$ and $\mathrm{DIAM}_{x_n}(T_{q_n}) > \epsilon$ for all n. Then the set $\{q_n\}/\Gamma_p$ is finite.*

Proof: Let $T_1 = T_p$ and let $T_n = T_{q_n}$. Let Γ_1 be the stabilizer of T_1. We will assume this result is false and derive a contradiction. Passing to a subsequence, we can assume that the points $\{q_n\}$ are pairwise inequivalent under the action of Γ_1. Passing to any further subsequence cannot destroy this property.

Let g_n' be an element such that $g_n'(x_n)$ is the origin in $\boldsymbol{CH}^2$. By construction the horotubes $g_n'(T_1)$ and $g_n'(T_n)$ all have spherical diameter at least ϵ. Suppose first that the spherical diameter of $g_n(\partial T_1)$ tends to 0. Since $g_n(T_1)$ has uniformly large spherical diameter, the only possibility is that $g_n(T_1)$ converges to S^3 minus a point. But then $g_n(T_n)$, a set whose diameter is uniformly large, intersects $g_n(T_1)$ for large n. Hence, T_1 intersects

T_n. We claim that there are only finitely many such horotubes with this property mod Γ_1. To see this, note that there is a compact subset $K \subset \Omega$ whose Γ_1-orbit contains $T_1 - T_1'$. Here T_1' is the smaller horotube mentioned in Property 4 for our horotube assignment. Since T_1' is disjoint from all the horotubes in our assignment except T_1, the Γ_1 orbit of K intersects all the horotubes that intersect T_1, including T_n for n large. But K only intersects finitely many horotubes from our assignment. Hence, every T_n is Γ_1-equivalent to one of finitely many horotubes.

Now we analyze the case where $g_n'(\partial T_1)$ has uniformly large spherical diameter. In this case we can apply Lemma 9.1 and use the notation from that result. Let $S_n = \gamma_n^{-1}(T_n)$. So, the basepoint of S_n is $\gamma_n^{-1}(q_n)$. Since $g_n = g_n' \circ \gamma_n$ we have $g_n(S_n) = g_n'(T_n)$. Hence, $g_n(S_n)$ has uniformly large spherical diameter. Also, note that the basepoint of S_n is never the same as the basepoint of T_1.

Case 1: If $\{g_n\}$ converges to a fixed map, then the list $\{S_n\}$ of horotubes must be finite, by Lemma 9.5. But then there are only finitely many points of the form $\gamma_n^{-1}(q_n)$. This contradicts the statement that the points in $\{q_n\}$ are pairwise distinct mod Γ_1.

Case 2: The spherical volume of $g_n(T_1)$ converges to the volume of S^3, and $g_n(\Lambda)$ converges to an $\boldsymbol{F}$-circle. In particular, $g_n(\Lambda)$ converges to a set that is more than a single point. Corollary 9.3 now applies, and we can say that $g_n(S_n)$ has uniformly large spherical volume. However, this contradicts the second alternative of Lemma 9.1. (In this alternative, $g_n(S_n)$, which lies in $g_n(S^3 - T_1')$, converges to the $\boldsymbol{F}$-circle.)

Case 3: If S_n does not belong to a certain list of M horotubes, then $g_n(S_n)$ converges to a single point. Hence, the sequence $\{S_n\}$ only has finitely many different horotubes. Hence, there are only finitely many points of the form $\gamma_n^{-1}(q_n)$. Hence, there are only finitely many q_n mod Γ_1. □

Chapter Ten

Constructing the Boundary Complex

10.1 STATEMENT OF RESULTS

Henceforth we assume that Γ is a horotube group. Our first goal is to prove the following lemma.

Lemma 10.1 (Structure) *Let E_1,...,E_n be the horocusps of Ω/Γ. There are horotubes $\widetilde{E}_1$,...,$\widetilde{E}_n$ and elements γ_1,...,$\gamma_n \in \Gamma$ such that $E_j = \widetilde{E}_j/\langle\gamma_j\rangle$. Furthermore every parabolic element of Γ is conjugate to a power of some γ_j. Thus, any maximal $\mathbf{Z}$ parabolic subgroup of Γ is conjugate in Γ to some $\langle\gamma_j\rangle$.*

Let Σ be the good spine for $M = \Omega/\Gamma$ guaranteed by Lemma 3.2. Then the horocusps E_1,...,E_n can be taken as the components of $M - \Sigma$. Let Ψ^∞ be the lift to Ω of Σ. According to Lemma 10.1, the cusps of Γ are in bijection with the components of $\Omega - \Psi^\infty$. Given a point $p \in \mathrm{Cusp}(\Gamma)$, let V_p^∞ be the component of $\Omega - \Psi^\infty$ corresponding to p. We would like to define Ψ^∞ in terms of horotube functions, as in Chapter 5.

We say that a *horotube function assignment* is an assignment $p \to f_p$, for each $p \in \mathrm{Cusp}(\Gamma)$. Here f_p is a horotube function. We require that our assignment is Γ-equivariant:

$$f_p = f_{g(p)} \circ g \qquad \forall g \in \Gamma. \tag{10.1}$$

We assume that the ranges of our horotube functions lie in the interval $[0, 3]$.

We say that our horotube function assignment is *adapted* to Ψ^∞ if the following conditions apply.

Condition 1: $f_p > 1 + \epsilon_0$ on V_p^∞. Here $\epsilon_0 > 0$ is some uniform constant.

Condition 2: V_p^∞ is the set of points $x \in \Omega$, where $f_p(x) > f_q(x)$ for all $p \neq q$.

Condition 3: Suppose x lies in the common boundary of the components V_1^∞,...,V_k^∞ of $\Omega - \Psi^\infty$. Then x is good with respect to the collection f_1,...,f_k, in the sense of Equation 3.3.

In Condition 3 we have $k \leq 4$ because Σ is a good spine. The point of Condition 3 is to make the local topology of Σ invariant under small

smooth perturbations of our horotube functions. Note that the definition for Condition 3 holds independent of the choice of Riemannian metric on S^3.

Lemma 10.2 (Horotube Assignment) *A horotube group has a horotube function assignment that is adapted to Ψ^∞ and has the following additional properties.*

(1) *The collections $\{\langle f_p\rangle_t\}$ are open covers of Ω for $t = 0, 1$. In particular $\langle f_p\rangle_t \subset \Omega$ for each $p \in \mathrm{Cusp}(\Gamma)$.*

(2) *Every compact $K \subset \Omega$ intersects only finitely many $\langle f_p\rangle_0$.*

(3) *$\langle f_p\rangle_2$ is a nice horotube on which the gradient ∇f_p does not vanish. Furthermore,* $\mathrm{closure}(\langle f_p\rangle_2) \cap \langle f_q\rangle_0 = \emptyset$ *if $p \neq q$.*

(4) *For any value of t, the closure of $\langle f_p\rangle_t$ is contained in $\Omega \cup \{p\}$.*

10.2 PROOF OF THE STRUCTURE LEMMA

We choose a good spine on $M = \Omega/\Gamma$ and a Riemannian metric adapted to it, as in Lemma 3.2. Let Ψ^∞ be the lift of Σ to Ω. We also lift the Riemannian metric on M to Ω.

Since Γ acts properly discontinuously on Ω and Ψ^∞ is the lift of a finite complex to Ω, the complex Ψ^∞ is locally finite.

Lemma 10.3 *Let $\widetilde{E}$ be one of the connected components of $\Omega - \Psi^\infty$. Then $\widetilde{E}$ is a horotube.*

Proof: The subtlety of this lemma is that we know E is covered by a horotube in S^3, but we don't (yet) know that $\widetilde{E}$ is a horotube. We do have a covering map $\pi : \widetilde{E} \to E$. We just have to show that the corresponding covering group is $\boldsymbol{Z}$ and generated by a parabolic element. By Lemma 2.7 E contains a nice horocusp. Let $\widetilde{F} \subset \widetilde{E}$ be the lift of F corresponding to the lift $\widetilde{E}$ of E. Then $\partial\widetilde{F}$ is a smooth surface in $\mathcal{H}$ locally stabilized by a 1-parameter parabolic subgroup, whose projection to ∂F is a torus. The only possibility is that $\partial\widetilde{F}$ is a cylinder and $\widetilde{F}$ is a nice horotube. Comparing $\widetilde{F}$ with F, we see that the covering group of $\widetilde{E}$ must be $\boldsymbol{Z}$ and generated by a parabolic. □

We say that the sets of the form $\widetilde{E}$, just discussed, are *spine horotubes*. The basepoint of any spine horotube is fixed by some parabolic element of Γ. However, we don't yet know that every parabolic element of Γ fixes the basepoint of some spine horotube.

Lemma 10.4 *Suppose that every parabolic $P \in \Gamma$ fixes the basepoint of some spine horotube. Then the Structure Lemma is true.*

Proof: Let $\widetilde{E}_1$,...,$\widetilde{E}_n$ be spine horotubes chosen so that $\widetilde{E}_j$ covers E_j for $j = 1,...,n$. Let $\langle \gamma_j \rangle \subset \Gamma$ be the $\boldsymbol{Z}$ parabolic covering group associated to $\widetilde{E}_j$. Let P be any parabolic of Γ. If P fixes the basepoint of some spine horotube, then some Γ-conjugate P' of P fixes the basepoint of one of the $\widetilde{E}_j$. Since $E_j = \widetilde{E}_j / \langle \gamma_j \rangle$ and E_j is necessarily a quotient of $\widetilde{E}_j / \langle \gamma_j, P' \rangle$, we must have $P' \in \langle \gamma_j \rangle$. □

Now we take the parabolic element $P \in \Gamma$ and try to show that it fixes the basepoint of a spine horotube.

Lemma 10.5 *Two spine horotubes A and B are disjoint iff they do not have a common basepoint.*

Proof: If A and B are not disjoint but do not have a common basepoint, then, after relabeling if necessary, some subset (more than one point) of the boundary of A is contained in B. But then some subset of Ψ^∞ is contained in B, which is absurd. □

Lemma 10.6 *For any $\epsilon, \delta \in (0,1)$ we can find $g \in PU(2,1)$ and some $k \neq 0$ such that $P' := gP^k g^{-1}$ maps any set of spherical volume $v > \epsilon$ to a set of spherical volume at least $(1-\delta)v$.*

Proof: We normalize so that P is as in Equation 2.12. With minor modifications, the same argument works if P is as in Equation 2.13. It suffices to construct sequences $\{g_n\} \in PU(2,1)$ and $\{k_n\} \in \boldsymbol{Z}$ such that $g_n P^{k_n} g_n^{-1}$ converges, uniformly on compacta, to the identity on $\mathcal{H}$. We choose g_n to be maps of the form given in Equation 2.14, with the dilation factor λ_n converging to 0. We first choose k_n so that the factor u^{k_n} in Equation 2.12 converges to 1. By construction, the maps $g_n P^{k_n} g_n^{-1}$ are Heisenberg isometries, which converge to the identity. □

Lemma 10.7 *Let T_p be the spine horotube based at p. Let $T'_p = T_p$. Then the horotube assignment $p \to T_p$ satisfies the four properties listed in Section 9.1*

Proof: Property 1 comes from the fact that $\overline{T}_p - p$ is the $\boldsymbol{Z}$-cover of a certain subset of the spine Σ. Property 2 follows from the fact that Ω is the union of Ψ^∞, a nowhere dense set, and the spine horotubes. Property 3 comes from the fact that Ψ^∞ is locally finite. Property 4 comes from the fact that T_p and T_q themselves are disjoint when $p \neq q$. This is why we can

set $T'_p = T_p$. □

Given the previous result, all the corollaries from Chapter 9 are available to us. So, let ϵ and $T = T_g$ be as in Lemma 9.4. Choose δ so that $(1-\delta)^{100/\epsilon} > 1/2$. Let P' be as in Lemma 10.6. Let $T = T_g$ be as in Corollary 9.4. So, $\tau = g(T)$ has spherical volume at least ϵ. But then $\tau_j = (P')^j(\tau)$ has spherical volume at least $\epsilon/2$ as long as $j < m := 100/\epsilon$. Since the sphere has total volume less than 50, there are indices $i, j < m$ such that $\tau_i \cap \tau_j \neq \emptyset$.

Let $T_j = P^{jk}(T)$. We compute that

$$\tau_j = (P')^j(\tau) = gP^{jk}g^{-1}\big(g(T)\big) = g\big(P^{jk}(T)\big) = g(T_j). \tag{10.2}$$

Hence, $g(T_i) \cap g(T_j) \neq \emptyset$. Hence, $T_i \cap T_j \neq \emptyset$. By Lemma 10.5, T_i and T_j have a common basepoint b. But then a power of P fixes b. Hence, P fixes b, and b is the basepoint of T. This completes our proof of Lemma 10.1. □

10.3 PROOF OF THE HOROTUBE ASSIGNMENT LEMMA

We will construct our horotube functions locally and then piece them together. First, we define an open cover of Ω, using four kinds of open sets.

(1) Cover the 0-skeleton of Ψ^∞ by small Riemannian balls centered at the 0-cells. We can choose this to be a Γ-invariant and disjoint collection of balls, which are so small that each one intersects Ψ^∞ in a set that is isometric to the intersection of a round ball centered at the origin with our third model.

(2) Cover the uncovered part of the 1-skeleton of Ψ^∞ by a locally finite Γ-invariant collection of Riemannian balls. We can choose these balls so small that we get the analogous Euclidean intersection property as in the 0-skeleton case. We also choose so that the Γ-orbit of each ball consists of a disjoint collection of balls.

(3) Cover the uncovered part of the 2-skeleton of Ψ^∞ by a locally finite and Γ-invariant collection of Riemannian balls, such that each ball intersects Ψ^∞ in a smooth open disk.

(4) Let V_p^∞ be the component of $\Omega - \Psi^\infty$ corresponding to the cusp p. Let U_p^∞ be an open subset of V_p^∞ whose boundary is so close to Ψ^∞ that U_p^∞ covers the previously uncovered part of V_p^∞.

We now attach some functions to each kind of set in our cover. First of all, for each ball U in our cover, we let U' be a slightly smaller concentric ball. By compactness, we can choose U' so that replacing U by U' everywhere still leaves us with a cover.

Let $V_1,...,V_k$ be the components of $U \cap \Psi$. We think of $U \cap \Psi^\infty$ as a subset of

one of our models $M(k,3)$ for $k = 2,3,4$, discussed in Section 3.2. Recall that $M(k,3)$ is the subset of the 3-dimensional hyperplane consisting of points whose first k coordinates sum to 0. Then V_j is a subset of the component of this hyperplane, where the ith coordinate is largest. We define $f_j : \Pi \to \boldsymbol{R}$ by the function

$$f_j(x) = \beta(x)(1 + x_j), \tag{10.3}$$

where $\beta(x)$ is a radially symmetric bump function. By construction the functions $f_1,...,f_k$ define our model in a neighborhood of the origin. At any point $x \in \partial V_j$ the gradient $\nabla\beta(x)$ points along ∂V_j by symmetry and $\nabla(x_j)$ points into V_j. Hence, ∇f_j points into V_j at any point of ∂V_j, where β is nonzero. Call this the *bump construction*.

If U is any of the metric balls in our cover, thenlet $V_1,...,V_k$ be the components of $U - \Psi^\infty$. Using the bump construction we can define functions $f_{U,j}$ such that $f_{U,j}(x) > f_{U,i}(x)$ if $x \in V_j$ and $i \neq j$ as long as $x \in U'$. We can also guarantee that $\nabla f_{U,j}$ points into V_j for every point of $U' \cap \partial V_j$. If p is a cusp corresponding to the component V_j, then we define $f_{U,p} = f_{U,j}$. Otherwise we define $f_{U,p} = 0$. Once we make the construction in one of our balls, we use the Γ-action to propagate the construction to the Γ-orbit of the given ball. We can do this without any conflict because the Γ-orbit consists of disjoint balls.

For the (big) sets $U = U_p^\infty$ of type 4, we proceed as follows. We let $f_{U,p}$ be a Γ_p-invariant bump function, which is identically 1 on U_p^∞ and vanishes outside a small neighborhood of U_p^∞. For $p \neq q$ we define $f_{U,q} = 0$. Once we make this construction for a complete list of inequivalent cusps, we do the remainder of the sets using the action of Γ.

Now we set $f_p = \sum_U f_{U,p}$. This is a locally finite sum. Equation 10.1 is true by construction. We adjust the range of f_p so that $f_p > 1 + \epsilon_0$ on V_p^∞ and the range of f_p lies in $[0,3]$. By construction V_p^∞ consists precisely of those points where $f_p(x) > f_q(x)$ for all $q \neq p$. Finally, suppose x lies in the common boundary of components $V_1^\infty,...,V_k^\infty$. Then $\nabla f_j(x)$ points into V_j^∞ because it is a finite and nontrivial sum of vectors, which have this property by construction. This forces Equation 3.3.

By compactness there is some small $\delta > 0$ such that $f_p(x)$ vanishes outside the δ neighborhood of V_p^∞. Hence, if $\{x_n\}$ is any sequence of points in Ω converging to $\Lambda - \{p\}$, then $f_p(x_n) = 0$ for large n. This is easiest to see in Heisenberg space: The support of f is the component of (roughly) a large cylinder, and $\Lambda - p$ runs through the interior of this (rough) cylinder. Thus we can make f smooth on $S^3 - p$ by defining $f = 0$ on $\Lambda - p$. From everything we have said, our horotube function assignment $p \to f_p$ is adapted to Ψ^∞.

Now we verify the remaining properties. By construction $\langle f_p \rangle_1$ covers the closure of V_p^∞ in Ω. Hence, the union $\bigcup \langle f_p \rangle_1$ covers Ω. This implies the first part of Lemma 10.2 (1). The second part of Lemma 10.2 (1) follows from the fact, already mentioned, that f_p vanishes outside a δ-neighborhood, in the complete Riemannian metric, of V_p^∞.

For Lemma 10.2 (2), it suffices to show that every ball B of radius δ

intersects at most finitely many sets of the form $\langle f \rangle$. Every set $\langle f \rangle_0$ is within δ of the corresponding spine horotube. If B intersects infinitely many sets $\langle f \rangle_0$, then a concentric ball of radius 2δ intersects infinitely many spine horotubes, contradicting the local finiteness of Ψ^∞.

To guarantee Lemma 10.2 (3), we modify our function f_p on U_p^∞ so that the level set $\langle f_p \rangle_2$ is a nice horotube so far inside U_p^∞ that it is disjoint from the rest of our cover. We can also arrange that ∇f is transverse to this horotube. One way to make this construction is to do it on the relevant end of M and then lift. Once we do this construction for a complete collection of inequivalent cusps, we use the action of Γ to handle all the cusps.

Let $X = \partial \langle f_p \rangle_t \cap \Omega$. Since the interior of $\langle f_p \rangle_t$ is contained in Ω, the accumulation set of $\langle f_p \rangle_t$ in $S^3 - \Omega$ coincides with the accumulation set of X in $S^3 - \Omega$. Note that X is the $\boldsymbol{Z}$-fold cover of a compact subset of Ω/Γ and the covering group of X is precisely Γ_p, the stabilizer subgroup of p. Any accumulation point of X in $S^3 - \Omega$ is the limit of a sequence of points $\{x_k\}$ in X that exits every compact subset of Ω. Since X/Γ_p is compact, there are elements $\gamma_k \in \Gamma_p$ such that $y_k = \gamma_k^{-1}(x_k)$ stays in a compact subset $K \subset X$. But then γ_k is an increasingly high power of the generator of Γ_p, and $\gamma_k(K)$ converges to p. Hence, p is the limit of $\{x_k\}$. Hence, the only accumulation point of $\langle f \rangle_t$ in $S^3 - \Omega$ is p. This is Lemma 10.2 (4).

Chapter Eleven

Extending to the Inside

11.1 STATEMENT OF RESULTS

In this chapter we will assume for ease of exposition that Γ is a horotube group with no elliptic elements. At the end of the proof of the HST in Chapter 13, we will discuss how to modify our arguments when there exist elliptic elements.

We have already constructed a horotube function assignment $p \to f_p$, which defines the complex Ψ^∞. In this chapter we will extend the function f_p so that the extension F_p is smooth on the set $\boldsymbol{U}_p$ from Chapter 6 and has the kind of transversality guaranteed by Equation 3.3, at least for "relevant" sets of defining functions.

Once we have our function assignment, we will define

$$\Psi = \bigcup_p \partial V_p, \qquad V_p = \{x \in \boldsymbol{CH}^2 : \ F_p(x) > F_q(x) \ \forall q \neq p\}. \tag{11.1}$$

Then Ψ is a 3-complex that extends Ψ^∞. In terms of our function assignment, we have

$$\Psi^\infty = \bigcup_p \partial V_p^\infty, \qquad V_p^\infty = \{x \in \Omega : \ f_p(x) > f_q(x) \ \forall q \neq p\}. \tag{11.2}$$

Let $p \to E_p$ be the function assignment we get from the Extension Lemma. Various points of $\boldsymbol{CH}^2$ might not be good with respect to various "relevant" collections of our functions $\{E_p\}$, in the sense of Equation 3.3 from Section 3.3. So, we need to perturb slightly.

Lemma 11.1 *There is some ϵ_0 with the following property. For any $x \in \boldsymbol{CH}^2$ there is some $p \in \mathrm{Cusp}(\Gamma)$ such that $E_p(x) > 100\epsilon_0$.*

Proof: Suppose this is false. By the Extension Lemma we would have $\mathrm{DIAM}_{x_n}(\langle f_q \rangle_1) \to 0$ for all $q \in \mathrm{Cusp}(\Gamma)$. Let $g_n \in PU(2,1)$ be such that $g_n(x_n) = 0$. Then the spherical diameter of $g_n(\langle f_q \rangle_1)$ tends to 0 as $n \to \infty$. Hence, the spherical volume of every single set $g_n(\langle f_q \rangle_1)$ tends to 0. Given that Λ is porous, this contradicts Lemma 9.4. $\square$

Given $x \in \boldsymbol{CH}^2$ we say that a finite collection of functions $F_1,...,F_k$ is *relevant for* x if $\min F_i(x) > \epsilon_0$.

We say that the function assignment $p \to F_p$ is a *good extension* if the following conditions are met.

- F_p is smooth on $\boldsymbol{U}_p$.
- $F_p(x)/E_p(x) \in [1/2, 2]$ for all $x \in \boldsymbol{CH}^2$ and all $p \in \mathrm{Cusp}(\Gamma)$.
- F_p agrees with f_p on $S^3 - p$.
- Any $x \in \boldsymbol{CH}^2$ is good with respect to any relevant finite collection of defining functions.

The first main result in this chapter is the following lemma.

Lemma 11.2 (Transversality) *A good extension exists.*

Here are several additional properties of our complex Ψ. Several of these results use the porous limit set condition in an essential way.

Lemma 11.3 (Local Structure) *Every point $x \in \Psi$ has a neighborhood U, together with a diffeomorphism $U \to \boldsymbol{R}^4$, that carries $U \cap \Psi$ either to the product of $\boldsymbol{R}$ with one of the local models for Ψ^∞, or else to the cone on the 2-skeleton of a regular tetrahedron in $\boldsymbol{R}^4$.*

Lemma 11.4 (Compatibility) *Ψ extends to a Γ-invariant neighborhood $U \subset \boldsymbol{CP}^2$ of $\boldsymbol{CH}^2 \cup \Omega$ such that $\Psi \cap U$ is transverse to S^3 and $\Psi^\infty = \Psi \cap \Omega$. In particular, Ψ has no vertices in U, and for $k = \{1, 2, 3, 4\}$ each k-cell C of $\Psi \cap U$ canonically corresponds to the $(k-1)$-cell $C \cap \Omega$ of Ψ^∞. Finally, the unbounded cells of Ψ are in bijection with the cells of Ψ^∞. The unbounded cell σ of Ψ corresponds to the cell $\sigma^\infty := \sigma \cap \Omega$.*

Lemma 11.5 (Finiteness) *There is a finite subset $S \subset \mathrm{Cusp}(\Gamma)$ with the following property. For each $x \in \Psi$ there is some $\gamma = \gamma_x \in \Gamma$ such that $\gamma(S_x) \subset S$. Also, there is a compact subset $K_0 \subset \boldsymbol{CH}^2 \cup \Omega$ such that every cell of $\Psi \cup \Psi^\infty$ is Γ-equivalent to a cell contained entirely in K_0.*

11.2 PROOF OF THE TRANSVERSALITY LEMMA

Say that a closed metric ball $B \subset \boldsymbol{CH}^2$ is *small* if $2B$ injects into $\boldsymbol{CH}^2/\Gamma$ under the quotient map. Here $2B$ is the ball concentric with B and with twice the radius. Let $\mathcal{B}$ denote the Γ-orbit of the set of all small balls with rational center and radius. Then $\mathcal{B}$ is a countable cover of $\boldsymbol{CH}^2$, and any small ball is approximated by members of $\mathcal{B}$. Let $\mathcal{S}$ denote the set of finite subsets of $\mathrm{Cusp}(\Gamma)$. The group Γ acts diagonally on the product set $\mathcal{B} \times \mathcal{S}$. Let $X_1, X_2, \ldots$ be an enumeration of representatives of the equivalence classes such that that every pair in $\mathcal{B} \times \mathcal{S}$ is Γ-equivalent to a unique X_j.

Say that $X_j = (B_j, S_j)$ is *noticeable* if every function associated to S_j is everywhere greater than, say, $\epsilon_0/100$ on $2B_j$. Say that $X_i = (B_j, S_j)$ is *good* if every point of B_j is good with respect to the collection of functions named

by S_j. Both concepts are Γ-invariant. Since B_j is compact, the goodness of X_j is stable under small perturbations of our functions. From Lemma 3.3 there are arbitrarily small smooth perturbations on B_j of the functions associated to S_j, which make X_j good. We make such a perturbation provided that X_j is noticeable. Using a bump function, we make these perturbations global but supported in $2B_j$. After we make the perturbation for X_j, we use the action of Γ to extend the perturbation in an equivariant way to the orbit $\Gamma(2B_j)$. We can do this because the orbit $\Gamma(2B_j)$ consists of disjoint balls, and the support of our perturbation is contained in the union of these balls.

Once we have finished with the index j, we have some new function assignment $p \to E_p^{(j)}$. Let c_j denote the maximum amount (in the sup norm) that any partial derivative of any function is changed on $2B_j$ during the jth perturbation. If we make $\{c_j\}$ decay rapidly enough, then we can guarantee the following.

- The limit $F_p = \lim_{j\to\infty} E_p^{(j)}$ is smooth and $E_p(x)/F_p(x) \in [1/2, 2]$ for all $x \in \boldsymbol{CH}^2$ and all cusps p.

- X_j is good with respect to the assignment $p \to F_p$ for all j. Hence, the whole assignment is good.

What makes this work is that any compact subset of $\boldsymbol{CH}^2$ only intersects finitely many balls in the orbit $\Gamma(2B_j)$, for any fixed j.

We extend F_p to all of $\boldsymbol{U}_p$ by setting $F_p = E_p$ on $\boldsymbol{U}_p - \boldsymbol{CH}^2$. So far we have not tried to control what happens as we approach Ω. We now add in this control. Let $\{K_j\}$ be a compact exhaustion of $\boldsymbol{CH}^2 \cup \Omega$. Say that a ball in the orbit $\Gamma(2B_j)$ is *j-important* if it intersects K_j. Given the properness of the Γ-action, there are only finitely many j-important balls for each j. Say that the jth perturbation has *important size* ϵ if each partial derivative of each of the involved functions is changed by at most ϵ on a j-important ball. Let $c_j' \geq c_j$ denote the important size of the jth perturbation.

Let d_j denote the minimal Euclidean distance between a point in S^3 and a point in a j-important ball. We now require $c_j' < 2^{-j}\exp(-1/d_j^2)$. For any $x \in \Omega$ there is an integer k and a neighborhood U of x such that $U \subset K_k$ and U is disjoint from all the j-important balls if $j < k$. If $y \in \boldsymbol{CH}^2$ lies in U and is within d of x, then $F_p(y) - E_p(y)$ has all partial derivatives less than $\exp(-1/d^2)$ at x. This estimate shows that $F_p - E_p$ vanishes outside of $\boldsymbol{CH}^2$ and is smooth in a neighborhood of $\boldsymbol{CH}^2 \cup \Omega$. It only remains to show that E_p is smooth at a point $x \in S^3 - \Omega - p$. By construction $f_p(x) = 0$. So $F_p < \epsilon_0/100$ in a small neighborhood of x. But then $\widehat{F}_p = F_p$ in this neighborhood.

11.3 PROOF OF THE LOCAL STRUCTURE LEMMA

We begin with a few general results that help pin down some of the local structure.

Lemma 11.6 *Given any $\epsilon > 0$ there is some $\delta > 0$ with the following property. If $p \in \mathrm{Cusp}(\Gamma)$ and $x \in \boldsymbol{CH}^2$ then $F_p(x) > \epsilon$ implies that $\mathrm{DIAM}_x(\langle f_p \rangle_0) > \delta$. Here δ is independent of p.*

Proof: Since each of our functions F_p satisfies the conclusions of the Extension Lemma, we know that there is some δ_p, dependent on p, that has the properties of this lemma. However, from the Γ-equivariance, we have $\delta_p = \delta_q$ if p and q are Γ-equivalent cusps. Since there are only finitely many cusps mod Γ, the infimum of all the δ_p is positive. □

Corollary 11.7 *Let K be a compact subset of $\boldsymbol{CH}^2$. Then there are only finitely many functions F_p such that the maximum value of F_p on K exceeds ϵ_0.*

Proof: If this was false, then by Lemma 11.6 and compactness there would be infinitely many horotubes of the form $\langle f_p \rangle_0$ having spherical diameter greater than some δ. But this contradicts Lemma 9.5. □

Let $x \in \Psi$ be a point. Let S_x denote the set of cusps p such that $x \in \partial V_p$. From Lemma 10.2 and the Transversality Lemma, we have $F_p(x) \geq 50\epsilon_0$ for any $p \in S_x$. Corollary 11.7 now says that S_x is a finite set. Let $p_1,...,p_k$ be the members of S_x. Let $F_j = F_{p_j}$. Given that we only have to worry about finitely many other functions in a neighborhood of x, we can take a small neighborhood U of x such that

$$\sup_{p \notin S_x} \sup_{y \in U} \widehat{F}_p(y) < \min_j \inf_{y \in U} \widehat{F}_j(y).$$

Hence, the components of $U - \Psi^\infty$ are precisely $V_1 \cap U,...,V_k \cap U$. The entire structure of $\Psi \cap U$ is determined by the functions $F_1,...,F_k$.

We think of U as a subset of $\boldsymbol{R}^4$ and x the origin. We know that the functions $F_1,...,F_k$ are smooth functions on U, and the endpoints of their gradients are the vertices of a nondegenerate $(k-1)$-dimensional simplex. (This is an immediate consequence of Equation 3.3.) Define

$$G_j = F_j - \frac{1}{k} \sum_{i=1}^{k} F_i.$$

Since we are subtracting the same function from each F_j to create G_j, the functions G_j also define $\Psi \cap U$. Note that $\sum G_j = 0$.

Suppose first that $k = 5$. The smooth map $\phi(x) = (G_1(x),...,G_k(x))$ carries $\Psi \cap U$ into the model $M(5,4)$. The vectors $\nabla G_1(0),...,\nabla G_k(0)$ are

obtained from the vectors $\nabla F_1(0),...,\nabla F_k(0)$ simply by subtracting off the center of mass of the simplex by these vectors. Hence, $\nabla G_1(0),...,\nabla G_k(0)$ form the vertices of a simplex whose center of mass is the origin. In particular, each vector $\nabla G_j(0)$ is separated from the other vectors by some hyperplane through the origin. But this shows that the linear differential $d\phi$ maps $\nabla G_j(0)$ to a nontrivial vector, which points into the jth component of $M(5,4)$. Hence, $d\phi(0)$ is nondegenerate. Hence, ϕ is a diffeomorphism in a neighborhood of 0. Evidently ϕ carries each component of $U - \Psi$ into the corresponding component of $M(5,4)$.

When $k = 2,3,4$, we let $A_{k+1},...,A_5$ be a basis for the orthogonal complement of the subspace containing $\nabla G_1(0),...,\nabla G_k(0)$. We define $G_m(x) = x \cdot A_m$ for $m = k+1,...,5$. We define ϕ as above. An argument very similar to what we did above shows that ϕ is a diffeomorphism near the origin and maps $\Psi \cap U$ into the model $M(k,4)$. This completes the proof of the Local Structure Lemma. □

11.4 PROOF OF THE COMPATIBILITY LEMMA

Compactness on the Boundary: Since Ψ^∞/Γ is compact, there is some compact set $K \subset \Omega$ such that Ψ^∞ is contained in the Γ-orbit of K. From the Horotube Assignment Lemma there is a finite subset $S_K \subset \mathrm{Cusp}(\Gamma)$ such that f_p vanishes identically on K unless $p \in S_K$. Let $f_1,...,f_k$ be the complete list of functions that do not identically vanish on K. Then the local combinatorial structure of Ψ^∞ is entirely determined by the functions $f_1,...,f_k$.

No Spurious Vertices: Now we consider what Ψ looks like in a neighborhood of K. We claim that there is no sequence of points $\{x_n\} \in \boldsymbol{CH}^2$ converging to a point in Ω such that the defining set S_{x_n} consists of 5 elements. If this is false, then we can use the Γ-action to arrange that $x_n \to x \in K$ and S_x consists of 5 elements. This contradicts the fact that $|S_x| \le 4$, which follows from the condition that our function assignment $p \to f_p$ is adapted to Ψ^∞.

The Product Structure near Infinity: Working in the ball model of $\boldsymbol{CH}^2$, we let V denote a small neighborhood of the compact set K in $\boldsymbol{C}^2$. If we choose V small enough, then no point in $\Psi \cap V$ is a vertex. (A vertex would have a 5-element defining set.) Moreover, if we choose V small enough, then all the extended functions $F_1,...,F_k$ are defined and smooth throughout V. We can extend Ψ to V just by using the finitely many functions $F_1,...,F_k$.

For any $x \in K$ the gradients of the defining functions $f_1,...,f_a$ associated to x are in general position, when measured in $T_x(S^3)$. But this means that none of these gradients is perpendicular to $T_x(S^3)$. In other words, the extension of Ψ we have defined is transverse to S^3 at x. If we choose V small

enough, then this transversality persists throughout $V \cap S^3$, by continuity. Thus we can extend Ψ to all of V (namely, somewhat outside $\boldsymbol{CH}^2 \cup \Omega$). Finally, we can extend Ψ to the set ΓV using the group action Γ and the equivariance of our function assignment.

By construction Ψ is defined in ΓV and transverse to S^3 on all of $\Gamma K \subset \Omega$. Since $\Psi \cap S^3 \subset \Gamma K$, we see that Ψ is transverse to S^3 on all of Ω. (This statement has no content for points not in ΓK.) The transversality properties of Ψ in ΓV guarantee that $\Psi \cap \Omega = \Psi^\infty$ and that the k-cells of $\Psi \cap U$ are naturally in bijection with the $(k-1)$-cells of Ψ^∞.

Extending the Neighborhood: Now we extend ΓV to be a Γ-invariant neighborhood of Ω in such a way as to guarantee that $\Psi \cap U = \Psi \cap \Gamma V$. We choose a locally finite and Γ-invariant open cover of $\Omega - \Gamma V$ by sets whose closure lies in $\Omega - \Psi^\infty$. Let U_1, U_2,... be an enumeration (without redundancy) of of Γ-equivalence classes of our sets. By compactness, we can certainly extend U_j to an open subset U_j' of $\boldsymbol{C}^2$, which avoids both $\Psi \cap V$ and Λ. The desired open set is then $V \cup \bigcup_j \Gamma(U_j')$.

The Bijection between the Cells: Let σ be an unbounded cell of Ψ. Let $A(\sigma)$ be the accumulation set of σ on S^3. At least two horotube functions are positive on σ. Hence,

$$A(\sigma) \subset I := \langle f_p \rangle_0 \cap \langle f_q \rangle_0, \qquad p \neq q.$$

But $I \subset \Omega$ by Lemma 10.2 (4). Hence, $A(\sigma) \subset \Omega$. Given our product structure in a neighborhood of Ω, we must have $\sigma \cap S^3 = \sigma^\infty$, a cell of Ψ^∞, and $A(\sigma)$ must be the closure of σ^∞. Conversely, any cell of Ψ^∞ extends to an unbounded cell of Ψ, by the construction above. □

11.5 PROOF OF THE FINITENESS LEMMA

The First Statement: Consider the set

$$S(p) = \bigcup_{x \in \partial V_p} S_x.$$

Lemma 11.8 *$S(p)/\Gamma_p$ is a finite set.*

Proof: From Lemma 11.1 and the definition of V_p, we have $F_p(x) > 50\epsilon_0$. Hence, $E_p(x) > 25\epsilon_0$. But if $q \in S_x$, then again we must have $E_q(x) > 25\epsilon_0$. But then, by Lemma 11.6, we have $\mathrm{DIAM}_x(\langle f_q \rangle_0) > \delta_0$. Lemma 9.6 (which requires the porous limit set condition) now tells us that $S(p)/\Gamma_p$ is a finite set. □

Let $S'(p)$ denote a large enough finite set so that each $q \in S(p)$ is Γ_p-equivalent to a cusp in $S(p)$. For any $q \in S'(p)$, let $S'(p,q)$ denote those

cusps r such that p, q, and r all belong to some defining set S_y for some $y \in \Psi$. Given this definition, we have $S_y \subset S'(p,q)$.

Lemma 11.9 *$S'(p,q)$ is a finite set.*

Proof: Let α be a generator for Γ_p, and let β be a generator for Γ_q. Let $\{r_n\}$ be a supposedly infinite collection of elements of $S'(p,q)$. Then $r_n \in S(p) \cap S(q)$. Given that $S'(p)$ and $S'(q)$ are both finite, we can pass to a subsequence so that

$$r_n = \alpha^{a_n}(s_1) = \beta^{b_n}(s_2).$$

Here $s_1 \in S'(p)$ and $s_2 \in S'(q)$ and a_n and b_n are appropriately chosen exponents, which tend to ∞. But $\alpha^{a_n}(s)$ converges to p, and $\beta^{b_n}(s)$ converges to the fixed point of q. This is a contradiction. □

Define the finite set

$$S''(p) = \bigcup_{q \in S'(p)} S'(p,q).$$

First, suppose that $x \in \partial V_p$. Then there is some γ such that $p, q \in \gamma(S_x)$ for some $q \in S'(p)$. But then $\gamma(S_x) \subset S'(p,q)$ by definition. Hence, $\gamma(S_x) \subset S''(p)$. More generally, we get $\gamma(S_x) \subset S''(p)$, provided that $x \in V_q$, where q is Γ-equivalent to p. Let p_1,...,p_k be a complete list of equivalence classes of cusps. Then the set $S = S''(p_1) \cup \cdots \cup S''(p_k)$ satisfies the first statement of the Finiteness Lemma.

The Second Statement: Note that we can attach a defining set to each open cell of Ψ. The defining set for a cell is the same as the defining set for any point on that cell. Hence, if σ is a cell of Ψ, then there is some γ so that the defining set for $\gamma(\sigma)$ is contained in our finite set S. Now we take K to be the closure of the union of all the cells whose defining sets are subsets of S. By construction K is the closure of a finite union of cells. Hence, K could equally well be described as the union of the closures of finitely many cells. By the Compatibility Lemma, each cell of Ψ has compact closure in $\boldsymbol{CH}^2 \cup \Omega$. Hence, K is a compact subset of $\boldsymbol{CH}^2 \cup \Omega$. By construction, every cell of Ψ is equivalent to a cell of K. □

Chapter Twelve

Machinery for Proving Discreteness

12.1 CHAPTER OVERVIEW

The method we present bears some resemblance to the Poincaré fundamental polyhedron theorem (see [**FZ**], [**M1**]), but is really a reformulation of Thurston's holonomy theorem (see [**T0**], [**CEG**]).

- In Section 12.2, we will define the notion of a *simple complex*. This is just a higher dimensional version of the concept discussed in [**Mat**] for 3-manifolds. The complexes Ψ and Ψ^∞ are both simple complexes.

- In Section 12.3, we will use simple complexes to define the basic object that we call a (G, X)-*chunk*. Here (G, X) refers to the data for a geometric structure. Intuitively, (G, X)-chunks are the pieces one gets after cutting open a (G, X)-manifold along a simple complex. However, this description is the reverse of what we want to do. We want to put together a bunch of (G, X)-chunks and show that they produce a (G, X)-manifold (or -orbifold.)

- In Section 12.4, we will explain what it means to glue together a finite number of (G, X)-chunks based on a *geometric equivalence relation* derived from a simple complex. This procedure sometimes leads to a (G, X)-manifold or -orbifold and sometimes it doesn't.

- In Section 12.5, we introduce the idea of *aligning by a simple complex*. The idea is that we have some auxiliary simple complex that provides local models for the results of all the gluings made by a geometric equivalence relation. When our geometric equivalence relation is aligned by a simple complex, the quotient is a (G, X)-manifold or -orbifold.

Our method is supposed to be a generalization of the straightforward method that beginning topology students are often taught to use when they verify that, for example, a square with opposite sides identified really is a manifold. We just go around, cell by cell, checking that the pieces fit together in the right way.

12.2 SIMPLE COMPLEXES

Let M be a smooth n-manifold. We say that a *simple complex* is a closed subset $\Psi \subset M$ such that every point $x \in \Psi$ has attached to it a pair (f, U), where the following apply.

- U is a neighborhood of x in M.
- $f : U \to \boldsymbol{R}^n$ is a diffeomorphism to $\boldsymbol{R}^n$ that carries $U \cap \Psi$ to one of our models $M(k, n)$.

We call the pair (f, U) a *simple chart* and U a *simple neighborhood*.

We have already shown that our complexes Ψ^∞ and Ψ from Chapters 10–11 are simple. Below we will use the symbol Ψ to refer to a general example of a simple complex. We hope that this doesn't cause confusion.

A simple complex naturally has a subdivision into smooth open submanifolds, which we call "cells," even though they are not necessarily topological balls. We define a one-sided neighborhood of a cell σ of Ψ to be a component of $V \cap \Psi$, where V is a small neighborhood of σ, which deformation retracts to σ. The lightly shaded sets in Figure 12.1 are one-sided neighborhoods of points in the hexagonal tiling, and the darkly shaded set is a one-sided neighborhood of a 1-cell.

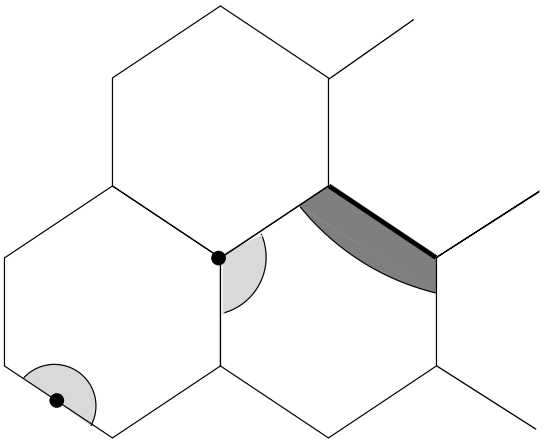

Figure 12.1: Some one-sided neighborhoods in the hexagonal grid

12.3 CHUNKS

Let (G, X) be as in Section 3.5. Let $\Psi \subset X$ be a simple complex. We say that a (G, X)-*manifold–chunk derived from* $\Psi \subset X$ is a compact Hausdorff space $\Upsilon = M \cup H$ such that M is a (G, X)-manifold and H is partitioned into finitely many subsets, which we call *cells*. To each cell C of H there is attached a pair (U, f), such that we have the following.

- U is an open neighborhood of closure(C) in Υ.
- $C' = f(C)$ is a cell of Ψ.
- $f : U \to U' \subset X$ is a homeomorphism from U to a one-sided neighborhood U' of C'.
- $f : U \cap M \to U' - \Psi$ is a (G, X)-isomorphism.

Here is some terminology associated to these objects. We call the neighborhood U, considered above, a *cell neighborhood*. We call the map f a *cell map*. We write $\partial\Upsilon = H$.

When the pair (G, X) is understood, we will call Υ a *manifold chunk*. As a variant of the definition above, we can require that M is a complex hyperbolic orbifold with isolated singularities, and in this case we require that Υ is metrically complete rather than compact. We call M an *orbifold chunk* for short. When the same remarks apply to both manifold and orbifold chunks, we will just use the word *chunk*. Here are some examples.

- The spaces Υ and $\widehat{\Upsilon}$ from Sections 5.3 and 5.4 are real hyperbolic manifold chunks derived from the complexes Ψ and $\widehat{\Psi}$, at least assuming the transversality assumptions there hold.
- The space Υ^∞, considered in Section 5.5, is a spherical CR manifold chunk derived from the horotube complex Ψ^∞. We will consider this more formally in Chapter 12.
- Let Ψ be the object from Chapter 11. Let V_p be a component of $\boldsymbol{CH}^2 - \Psi$. Let Γ_p be the stabilizer of V_p. Then $\overline{V}_p/\Gamma_p$ is a complex hyperbolic manifold chunk.

12.4 GEOMETRIC EQUIVALENCE RELATIONS

Suppose now that Υ is a chunk, not necessarily connected. Let C_1 and C_2 be two distinct cells of Υ that have the same dimension. Let $\overline{C}_j$ be the closure of C_j in Υ_j. Then $\overline{C}_j$ is a finite union of cells, with C_j itself being the top-dimensional cell in the cluster.

We say that a *geometric pairing* between C_1 and C_2 is a homeomorphism $\psi : \overline{C}_1 \to \overline{C}_2$ that is compatible with the two structures. More precisely, we mean the following.

- There are cell neighborhoods U_1 and U_2 of C_1 and C_2.
- There is a cell C with one-sided cell neighborhoods U_1' and U_2'.
- There are cell maps $f_1 : U_1 \to U_1'$ and $f_2 : U_2 \to U_2'$ with the property that $f_1(x) = f_2 \circ \phi(x)$ for all $x \in \overline{C}_1$.

- When C_1 and C_2 are top-dimensional cells, we require that $U_1' \cup U_2'$ is an open neighborhood of C, so that C divides the interiors of these sets from each other.

Let Υ be a chunk. Let $\sim$ be an equivalence relation on $\partial\Upsilon$. We say that the restriction of $\sim$ to the union $\overline{C}_1 \cup \overline{C}_2$ is *realized* by a geometric pairing ψ of C_1 and C_2 if, for all $x_1 \in \overline{C}_1$ and $x_2 \in \overline{C}_2$, we have

$$x_1 \sim x_2 \qquad \Longleftrightarrow \qquad \psi(x_1) = x_2. \tag{12.1}$$

We call $\sim$ a *geometric equivalence relation* (or geometric relation for short) on $\partial\Psi$ if the following conditions hold.

Condition 1: For every two distinct cells C_1 and C_2, the restriction of $\sim$ to $\overline{C}_1 \cup \overline{C}_2$ is realized by a geometric pairing. In particular, a point in the interior of an m_1-cell is equivalent to a point in the interior of an m_2-cell iff $m_1 = m_2$.

Condition 2: Every top-dimensional cell of $\partial\Upsilon$ is equivalent under $\sim$ to exactly one other top-dimensional cell. (In light of Condition 1 above, it makes sense to talk about an equivalence between the cells of $\partial\Upsilon$.)

Condition 3: No two points on the same cell of $\partial\Upsilon$ are equivalent to each other.

A geometric relation need not lead to a (G, X)-manifold or -orbifold. Here is an example. Suppose our chunks are right-angled regular hyperbolic hexagons. We could glue finitely many of these hexagons together, 5 per vertex, to produce a topological surface. However, the surface we produce would have a geometric cone point at each vertex.

On the other hand, we have a tiling of $\boldsymbol{H}^2$ by right-angled hexagons, with 4 fitting around each vertex. Using this tiling as a guide, we could take a finite union of right-angled regular hexagons and glue them together, 4 per vertex, to produce a hyperbolic surface. This time, the result would be a bona fide hyperbolic surface. The tiling provides all the local models for how our hexagons fit together. In this way, the tiling aligns our gluing pattern for the surface. The next section makes this idea more formal.

12.5 ALIGNMENT BY A SIMPLE COMPLEX

12.5.1 Basic Definition

We continue with the same notation from the previous section. Suppose that Υ is a (G, X)-chunk derived from a simple complex $\Psi \subset X$. Suppose also that $\sim$ is a geometric relation on $\partial\Upsilon$. We can partition the cells of $\partial\Upsilon$ into equivalence classes. Each equivalence class $\mathcal{C}$ consists of cells that are all equivalent to each other and to no other cell. We say that the equivalence class $\mathcal{C} = \{C_1,...,C_k\}$ is *aligned by* Ψ if the following holds.

- There are cell neighborhoods U_1,...,U_k of C_1,...,C_k, respectively, together with cell maps $f_j : U_j \to U_j'$.
- The union $U' = U_1' \cup \cdots \cup U_k'$ is an open neighborhood of a cell C of Ψ and $f_j(\overline{C}_j) = \overline{C}$.
- A point $x_i \in U_i$ is equivalent to a point $x_j \in U_j$ iff $f_i(x_i) = f_j(x_j)$.

In other words, the sets $U_j' \cap \Psi$ intersect each other in precisely the same way that the sets $U_j \cap \partial\Upsilon$ are paired together for $\sim$. Notice that the maps f_1,...,f_k piece together canonically to give a homeomorphism

$$f_1 \cup \cdots \cup f_k : (U_1 \cup \cdots \cup U_k)/\sim \quad \to \quad U'. \tag{12.2}$$

By definition $\mathcal{C}$ consists of 2 members if $\mathcal{C}$ is a top-dimensional equivalence class. In this case Ψ automatically aligns $\mathcal{C}$. We say that Ψ *aligns* $\sim$ if Ψ aligns every equivalence class of $\sim$.

12.5.2 The Main Results

We say that the pair (G, X) is *hypersurface rigid* if an element of G is determined by its action on any smooth codimension-1 surface in X. For instance $(SL_2(\boldsymbol{R}), \boldsymbol{R}^2)$ is not hypersurface rigid. On the other hand, both $(PU(2,1), S^3)$ and $(PU(2,1), \boldsymbol{CH}^2)$ are easily seen to be hypersurface rigid. Now for the main result of this section.

Lemma 12.1 *Suppose (G, X) is hypersurface rigid and Ψ is a simple complex in X. Let Υ be a (G, X)-manifold–chunk derived from Ψ. Let $\sim$ be a geometric pairing on $\partial\Upsilon$ that is aligned by Ψ. Then $\Upsilon/\sim$ has a canonical (G, X)-structure that extends the (G, X)-structure already on $\Upsilon - \partial\Upsilon$.*

Proof: For each $x \in \partial\Upsilon$ we construct a coordinate chart into X, as follows. x belongs to some cell C_1. Let $\mathcal{C} = \{C_1,...,C_k\}$ be the equivalence class of C_1. Equation 12.2 gives the coordinate chart from a neighborhood U of x in $\Upsilon/\sim$ into X. Say that a *preferred chart about x* is the restriction of the map defined in Equation 12.2 to a neighborhood $W \subset U$ of x such that W does not intersect any cells in $\overline{C}_j - C_j$, for $j = 1,...,k$. (If these cells are 0-cells, then the condition is vacuous.) Say that a *preferred chart about $y \in \Upsilon - \partial\Upsilon$* is an open coordinate chart that is disjoint from $\partial\Upsilon$.

To complete our proof, we just need to see that the overlap functions defined by pairs of preferred charts are all locally in G. Let (W_1, f_1) and (W_2, f_2) be two preferred charts with $W_1 \cap W_2 \neq \emptyset$.

Case 1: Suppose W_1 and W_2 are both disjoint from $\partial\Upsilon/\sim$. Then the result follows just by virtue of the fact that these charts already come from a (G, X) structure.

Case 2: Suppose W_1 is disjoint from $\partial\Upsilon/\sim$ but W_2 is not. Then $W_1 \cap W_2$

is contained in a single connected component of Υ, say, Υ_1. In this case, the restriction f_2 to W_2 is a local (G, X)-isomorphism, by definition of cell maps. The same is true, tautologically, of the restriction of f_1 to W_1. Hence, the restrictions of f_1 and f_2 to each component of $W_1 \cap W_2$ are both local (G, X)-isomorphisms. Hence, $f_1 \circ f_2^{-1}$ is the restriction of an element of G on $f_2(W)$.

Case 3: Suppose that neither W_1 nor W_2 is disjoint from $\partial\Upsilon/\sim$. This is the interesting case. Let W be a connected component of $W_1 \cap W_2$. Let $\phi = f_1 \circ f_2^{-1}$. We want to show that the restriction of ϕ to $f_2(W)$ coincides with an element of G. By continuity, it suffices to show that the restriction of ϕ to a dense subset of $f_2(W)$ coincides with an element of G. Now, the restriction of ϕ to $f_2(W) - \Psi$ coincides, on each connected component, with an element of G. Any two connected components of $f_2(W) - \Psi$ can be joined by a finite string of connected components, each of which accumulates on a common top-dimensional cell of Ψ. But the restriction of ϕ to such adjacent components must be the same element of G, due to the hypersurface ridigity property. This means that the restriction of ϕ to $f_2(W) - \Psi'$ coincides with a single element of G. Here Ψ' is the complex obtained from Ψ by deleting all the top-dimensional faces. Since $\dim(\Psi') < \dim(f_2(W))$, the set $f_2(W) - \Psi'$ is dense in $f_2(W)$, and we are done. □

An orbifold chunk comes equipped with a metric that extends the complex hyperbolic metric defined on an open dense set. We say that an orbifold chunk is *complete* if this metric is complete.

Lemma 12.2 *Suppose (G, X) is hypersurface rigid and Ψ is a simple complex in X. Let Υ be a complex hyperbolic orbifold chunk that is complete as a metric space. Let $\sim$ be a geometric pairing on $\partial\Upsilon$ that is aligned by Ψ. Then $\Upsilon/\sim$ is a complex hyperbolic orbifold with isolated singularities.*

Proof: Nothing at all happens in neighborhoods of the singularities, so these neighborhoods inject into $\Upsilon/\sim$. Since the singular points in Υ are disjoint from $\partial\Upsilon$, the same analysis in Lemma 12.1 shows that $\Upsilon/\sim$ admits a complex hyperbolic metric in the complement of its finitely many singularities. Finally, the completeness of $\Upsilon/\sim$ follows from the fact that the chunk itself is complete. Putting all these facts together establishes the lemma. □

Chapter Thirteen

Proof of the HST

In this chapter we will assemble the ingredients from the previous chapters to prove the Horotube Surgery Theorem (HST). Our proof has two themes. One theme is essentially a detailed working out of the material in Sections 5.5–5.6. The other theme is a discreteness proof using the machinery from Chapter 12. Until Section 13.9 we make the blanket assumption that Γ has no elliptic elements. At the end of this chapter we will explain how to deal with elliptic elements.

We will use the notation from the statement of the HST. As in Chapter 5, we will use the notation ISC to denote that $\widehat{\rho}$ is sufficiently far along in a sequence of representations that converges nicely to ρ. (We think of ISC as standing for "if sufficiently close".)

13.1 THE UNPERTURBED CASE

By construction Ψ^∞ is a simple complex modeled on $(PU(2,1), S^2)$. Each 1-cell of Ψ^∞ is an open arc. Each 2-cell is an open disk whose boundary is a finite polygon. The complement $\Omega - \Psi^\infty$ consists of the pieces of the form V_p^∞, one per cusp. The deformation retraction from Ω/Γ onto the good spine Σ lifts to give, in particular, a deformation retraction from V_p^∞ onto ∂V_p^∞. This situation forces ∂V_p^∞ to be a cylinder and V_p^∞ to be a torus cross a ray.

The complex Ψ is a simple complex modeled on $(PU(2,1), \boldsymbol{CH}^2)$. From the Compatibility Lemma of Chapter 11, the unbounded cells of Ψ are canonically bijective with the cells of Ψ^∞.

Let $p_1,...,p_k \in \mathrm{Cusp}(\Gamma)$ be a list of pairwise Γ-inequivalent cusps. For ease of notation we set $V_j = V_{p_j}$, etc. Let Γ_j be the stabilizer of p_j in Γ. Define

$$\Upsilon = \bigcup_{j=1}^{k} \Upsilon_j, \qquad \Upsilon_j = V_j/\Gamma_j. \tag{13.1}$$

Since every point of $\boldsymbol{CH}^2$ is equivalent to some point on V_j, we see that $\boldsymbol{CH}^2/\Gamma = \Upsilon/\sim$, where $\sim$ is a geometric relation aligned by Ψ, in the sense of Section 12.5. The story on the boundary is the same. The gluing relations on the boundary pieces Υ_j^∞ are just the extensions of the ones on the pieces Υ_j.

We want to identify this same kind of structure for the group $\widehat{\Gamma}$.

13.2 THE PERTURBED CASE

Let K_0 be the compact set from the Finiteness Lemma of Chapter 11. We choose some extremely large but as yet unspecified pair of integers n_1 and n_2. We think of n_2 as being much larger than n_1, say, $n_2 = \exp(100n_1)$. For $m = 1, 2$ we let K_m denote the union $\bigcup_\gamma \gamma(K_0)$, where the union is taken over all the words in Γ having length less than n_m relative to some generating set. We call n_1 and n_2 the *master constants*.

Recall that every point $x \in \Psi$ has a defining set $S_x \subset \mathrm{Cusp}(\Gamma)$. The structure of Ψ in a neighborhood of x is determined entirely by the $k \leq 5$ functions associated to this defining set.

For $m = 1, 2$ let U_m be an open subset of $\boldsymbol{C}^2$, which contains K_m. We require that the closure $\overline{U}_m$ is contained in the neighborhood of $\boldsymbol{CH}^2 \cup \Omega$, guaranteed by the Compatibility Lemma, on which Ψ exists. In particular $\overline{U}_m \cap \Omega$ is compact. We think of U_m as a very small tubular neighborhood of K_m. By the Finiteness Lemma there is a finite set $S_m \subset \mathrm{Cusp}(\Gamma)$ such that $S_x \subset S_m$ for all $x \in U_m \cap \Psi$. The point here is that $U_m \cap (\boldsymbol{CH}^2 \cup \Omega)$ is contained in a finite union of translates of the compact set K from the Finiteness Lemma, and slightly outside of $\boldsymbol{CH}^2 \cup \Omega$, our product structure guarantees that no new defining functions are needed.

Naive Definition: Now we are going to use the Transplant Lemma from Chapter 6. First, we will make the most naive construction, and then we will improve it. Each $p \in S_m$ is stabilized by some parabolic group $\Gamma_p = \rho(G_p)$. We can apply the Transplant Lemma to F_p relative to the group $\widehat{\rho}(G_p)$. Call the resulting function $\widehat{F}_p$.

ISC the function $\widehat{F}_p$ is defined throughout U_m. We set

$$\widehat{\Psi}_m = \bigcup_{p \in S_m} \partial \widehat{V}_{p,m}, \quad \widehat{V}_{p,m} = \{x \in U_m |\ \widehat{F}_p(x) > \widehat{F}_q(x)\ \ \forall q \in S_m - p\}. \quad (13.2)$$

In other words, we adjust the functions associated to S_m to define $\widehat{\Psi}_m$ throughout U_m. We set $\widehat{\Psi}_m^\infty = \widehat{\Psi}_m \cap \Omega$.

Given the transversality we have built into Ψ, namely, the fact that every point of Ψ is good with respect to its defining set, the cells of Ψ are stable under sufficiently small perturbations. ISC there will be a bijection between the cells of Ψ entirely contained in U_m and the cells of $\widehat{\Psi}_m$ entirely contained in U_m. The corresponding cells σ and $\widehat{\sigma}$ will be close in the sense that there is a smooth map between them that is C^∞-close to the identity. Working skeleton by skeleton, we can arrange these smooth maps so that they are compatible and give a cellular homeomorphism between the corresponding unions of cells.

The Improvement: The construction we have made isn't quite what we want because the resulting sets are not in any sense equivariant under the action of $\widehat{\Gamma}$. Now we will improve the situation.

Given a pair of cusps $p, q \in \Gamma$, let $H_{p,q} \subset G$ denote the set of elements g

such that $\gamma = \rho(g)$ conjugates the stabilizer Γ_p to Γ_q. That is, $\gamma\Gamma_p\gamma^{-1} = \Gamma_q$. Necessarily we have $\gamma(p) = q$. Hence, $H_{p,q}$ is a double coset of the form $G_q g G_p$, where $\rho(G_p) = \Gamma_p$ and $\rho(G_q) = \Gamma_q$. There is a finite collection $\mathcal{D}$ of such double cosets having the form $H_{p,q}$, where $p, q \in S_2$.

Choose a list $p_1,...,p_k$ of pairwise inequivalent Γ-equivalence classes of cusps of S_m. By construction (or the Finiteness Lemma) S_2 is contained in the Γ-orbit of S_1, and so we can take the same list for $m = 1, 2$. We define $\widehat{F}_j = \widehat{F}_{p_j}$ for $j = 1,...,k$. For each $p \in S_2 - \{p_1,...,p_k\}$ choose some element g such that $g(q) = p_j$. Then g represents some member $\mathcal{D}$. We redefine

$$\widehat{F}_q = \widehat{F}_j \circ \widehat{\gamma}. \tag{13.3}$$

Here we have set $\widehat{\gamma} = \widehat{\rho}(g)$. We claim that $\widehat{F}_q$ is equivariant with respect to the group $\widehat{\Gamma}_q$. To see this we choose $a \in G_q$ and note that there is some $b \in G_{p_j}$ such that $gag^{-1} = b$. In other words $ga = bg$. We compute

$$\widehat{F}_q \circ \widehat{\alpha} = \widehat{F}_j \circ \widehat{\gamma} \circ \alpha = \widehat{F}_j \circ \widehat{\beta} \circ \widehat{\gamma} =^* \widehat{F}_j \circ \widehat{\gamma} = \widehat{F}_q.$$

Here we have set $\widehat{\alpha} = \widehat{\rho}(a)$, etc. The starred equality comes from the fact that $\widehat{F}_j$ is invariant under $\widehat{G}_{p_j}$, by the Transplant Lemma. Suppose that $g' = dga$ for $a \in G_q$ and $d \in G_{p_j}$. Then

$$\widehat{F}_j \circ \widehat{\gamma}' = \widehat{F}_j \circ \widehat{\delta} \circ \widehat{\gamma} \circ \widehat{\alpha} = \widehat{F}_j \circ \widehat{\gamma} \circ \widehat{\alpha} = \widehat{F}_q \circ \widehat{\alpha} = \widehat{F}_q.$$

In other words, our definition of $\widehat{F}_q$ is independent of a coset representative.

Suppose now that $q, r \in S_2$ are such that $g(q) = r$ for some $g \in G$. Then both cusps are equivalent to the same p_j. We can write $g = g_r^{-1} g_q$, where $g_q(q) = p_j = q_r(r)$. We compute that

$$\widehat{F}_r \circ \widehat{\gamma} = \widehat{F}_r \circ \widehat{\gamma}_r^{-1} \circ \widehat{\gamma}_q = \widehat{F}_j \circ \widehat{\gamma}_q = \widehat{F}_q.$$

This computation makes sense wherever the functions are defined and, in particular, in the set S_2. Letting D denote the union of all the elements in $\mathcal{D}$, we have established the equivariance

$$\widehat{F}_p = \widehat{F}_{\widehat{\gamma}(p)} \circ \gamma, \qquad \forall p \in S_2; \quad \forall g \in D. \tag{13.4}$$

As above we have set $\widehat{\gamma} = \widehat{\rho}(g)$. Our equation holds as long as all evaluations involved lie in U_2. Our complexes are as symmetric as possible.

We have shown that our definition of $\widehat{F}_q$ only depends on our double coset, but we can choose representatives of these cosets from a fixed finite list. ISC the images of each member of this fixed list under ρ and $\widehat{\rho}$ will be very close in $PU(2,1)$. Thus, our refined definition of $\widehat{\Psi}_m$ does not lose the original approximation properties it had.

The Engulfing Property: Our constructions depend on the master constants n_1 and n_2. Given that $S_1 \subset S_2$ and all the functions F_p corresponding to $p \in S_2 - S_1$ vanish on the compact set $\overline{U}_1$, any cell of $\widehat{\Psi}_1^\infty$ is also a cell of $\widehat{\Psi}_2^\infty$. In short $\widehat{\Psi}_1^\infty \subset \widehat{\Psi}_2^\infty$. Given Equation 13.4 we have ISC

$$\widehat{\gamma}(\widehat{\Psi}_1^\infty) \subset \widehat{\Psi}_2^\infty, \qquad |\widehat{\gamma}| < \sqrt{n_2}. \tag{13.5}$$

Here $|\widehat{\gamma}|$ is the length of $\widehat{\gamma}$ in the generators. Technically, we are measuring the shortest possible length of g in the generators, where g is such that $\widehat{\gamma} = \widehat{\rho}(g)$. There is nothing significant about $\sqrt{n_2}$, except that it is much smaller than $n_2 - n_1$ and much larger than n_1.

Moreover, we have a homeomorphism, near the identity, carrying cells of $\widehat{\Psi}^\infty$ to cells of $\widehat{\Psi}$. By Equation 13.4 this homeomorphism is compatible with the actions of Γ and $\widehat{\Gamma}$, at least for words of length less than, say, $\sqrt{n_2}$. Thus we can see combinatorially how $\widehat{\gamma}(\widehat{\Psi}_1)$ sits inside $\widehat{\Psi}_2$ by looking at the corresponding cells of Ψ.

13.3 DEFINING THE CHUNKS

In this section we describe the chunks $\widehat{\Upsilon}_j$, which (later on) will piece together to form $\boldsymbol{CH}^2/\widehat{\Gamma}$ in the same way that the chunks Υ_j defined above piece together to form $\boldsymbol{CH}^2/\Gamma$. Our construction uses in a vital way the fact that $\widehat{\rho}(G_p)$ converges geometrically to $\rho(G_p)$ for each cusp p of Γ. We will see below how this assumption is used.

On the Boundary: We will first describe the construction on the boundary. We let $X^\infty = X \cap S^3$ for any of the objects we have considered above. (All the intersections actually take place in Ω.) Given any $p \in S_1$ (the smaller of the two collections of cusps), we let $\widehat{V}^\infty_{p,1}$ be the (boundary version of the) set from Equation 13.2. We let $\widehat{B}^\infty_p$ denote the union of cells of $\partial\widehat{V}^\infty_{p,1}$ of the form $\widehat{\sigma}$, corresponding to cells σ of ∂V^∞_p under the cellular bijection we have arranged above. We have the corresponding nearby set $B^\infty_p \subset \Psi^\infty$. We define

$$\text{Extend}(\widehat{B}^\infty_p) = \bigcup_{\widehat{\gamma} \in \widehat{\Gamma}_p} \widehat{\gamma}(\widehat{B}^\infty_p). \tag{13.6}$$

Let's explain what is going on, starting with the unperturbed group. The set ∂V^∞_p is a cylinder partitioned into smooth cells. The set B^∞_p is a large finite subset of this cylinder. One should picture this set as a big annulus that contains many translates of a fundamental domain for the action of Γ_p on ∂V^∞_p. The set $\widehat{B}^\infty_p$ is the image of B^∞_p under a tiny perturbation. One should picture $\widehat{B}^\infty_p$ as essentially the same big annulus but slightly wiggled.

Let g be a generator of G_p. Let $\gamma = \rho(g)$ and $\widehat{\gamma} = \widehat{\rho}(g)$. Consider first the action of γ on B^∞_p. There is plenty of overlap between B^∞_p and $\gamma(B^\infty_p)$. However, the overlap is seamless because we know that globally γ just acts so as to permute ∂V^∞_p. Indeed the set $\text{Extend}(B^\infty_p)$ is precisely ∂V^∞_p. We recover the original cylinder.

Now consider the situation for $\widehat{\gamma}$ acting on $\widehat{B}^\infty_p$. We will concentrate on the most interesting case, when $\widehat{\Gamma}_p$ is a finite elliptic subgroup. Recall that the master constants n_1 and n_2 determine our constructions above. ISC we can assume that the order N of Γ_p is much larger than n_2, say,

$n > \exp(\exp(100n_2))$. There is some smallish integer k such that $\widehat{\gamma}^j(\widehat{B}_p^\infty) \cap \widehat{B}_p^\infty = \emptyset$ unless j mod N (the representative with the smallest absolute value) lies in $[-k, k]$. We can choose n_2 so as to guarantee that $k < \sqrt{n_2}$.

Given Equation 13.5 and the intertwining properties of our homeomorphism from $\widehat{\Psi}_2$ to Ψ_2, we see that

$$\bigcup_{j=1}^{k} \widehat{\gamma}_j(\widehat{B}_p^\infty)$$

is just a finite union of cells of $\widehat{\Psi}_2$, each one close to the corresponding cell of ∂V_p^∞. Every point on Extend(B_p^∞) is equivalent to one in our finite union. Hence, the whole object is a locally embedded combinatorial surface.

If Extend$(\widehat{B}_p^\infty)$ is not embedded, then $\widehat{B}_p^\infty$ intersects $\widehat{\gamma}^d(\widehat{B}_p^\infty)$ for some exponent d whose true size is large in the sense of Section 3.1. But this contradicts the fact that $\widehat{\Gamma}_p$ converges geometrically to Γ_p, as shown in Lemma 3.1. Without the geometric convergence, our cylinder really would wrap around. With the geometric convergence, the wrapping does not occur, and Extend$(\widehat{B}_p^\infty)$ is an embedded polyhedral surface and, in fact, a torus.

Now we define $\widehat{V}_p^\infty$ to be the component of $S^3 - \text{Extend}(\widehat{B}_p^\infty)$ that lies on the same side of Extend$(\widehat{B}_p^\infty)$ as V_p^∞ lies on ∂V_p^∞. One can line up the sides easily because, at least in the large compact set K_1, the cells of the one object line up almost exactly with the cells of the other object.

By construction $\widehat{\Gamma}$ acts on $\widehat{V}_p^\infty$, and we define $\Upsilon_p^\infty = \widehat{V}_p^\infty/\widehat{\Gamma}_p$. This solid torus is an $(S^3, PU(2,1))$-chunk. Let $\widehat{\Upsilon}_1^\infty$,...,$\widehat{\Upsilon}_k^\infty$ be the chunks corresponding to a complete list of pairwise distinct Γ-equivalence classes of cusps. We have concentrated on the elliptic case. When $\widehat{\Gamma}_p$ is parabolic or loxodromic, the same construction is essentially identical.

On the Inside: Our constructions in the previous section go through essentially word for word. We just take the ∞ superscript off all the symbols. Here are the minor differences. Given the local singular structure of Ψ and $\widehat{\Psi}$, the set ∂V_p is a 3-manifold partitioned into cells. The same goes for Extend$(\widehat{B}_p)$. The so-called cells of ∂V_p and Extend$(\widehat{B}_p)$ need not be open balls. ∂V_p and Extend$(\widehat{B}_p)$ might have more than one boundary component. However, it is possible to choose consistent and corresponding normal vector fields on every boundary component. The point is that we can line everything up almost exactly in the compact set K_1. Thus we define $\widehat{V}_p$ to be the set obtained by filling in the boundary components of Extend$(\widehat{B}_p)$ so as to be consistent with how V_p is filled in from ∂V_p. Given the Compatibility Lemma and the fact that we have perturbed everything smoothly, we get the following property. If $U \subset \boldsymbol{CH}^2$ is a small neighborhood of a point $x \in \partial \widehat{V}_p^\infty$, then the closure of $U \cap \widehat{V}_p^\infty$ is the accumulation set on S^3 of $U \cap \widehat{V}_p^\infty$. Call this the *one-side property*.

We define $\widehat{\Upsilon}_p = \widehat{V}_p/\widehat{\Gamma}_p$. Again, we only need this for a complete list p_1,...,p_k of pairwise distinct Γ-equivalence classes of cusps. We let $\widehat{\Upsilon}_1$,...,$\widehat{\Upsilon}_k$

be the corresponding complex hyperbolic orbifold chunks.

13.4 THE DISCRETENESS PROOF

By construction, the boundary of $\widehat{\Upsilon}_j$ is combinatorially identical to the boundary of Υ_j. Just as the action of Γ gives a geometric equivalence relation to the cells on the boundary of $\Upsilon = \bigcup_{i=1}^k \Upsilon_k$, the action of (a finite subset of) $\widehat{\Gamma}$ gives a geometric equivalence relation to the cells on the boundary of $\widehat{\Upsilon} = \bigcup \widehat{\Upsilon}_j$. Equation 13.4 guarantees that this works, as long as the master constants are large enough. ISC there is a natural bijection between these cells, which respects the equivalence relation. In short, the two spaces are glued together in combinatorially identical ways.

The geometric equivalence relation on Υ is aligned by the simple complex Ψ. We can choose the master constants so that the compact set $\overline{U}_1$ contains a finite subset of Ψ, which has all the local gluing models. Given Equations 13.4 and 13.5, the equivalence relation on $\widehat{\Upsilon}_j$ is aligned by the simple complex $\widehat{\Upsilon}_2$, considered as a subset of the complex hyperbolic manifold U_2. Intuitively $\widehat{\Psi}_2$ is a model for how various translates of $\widehat{\Psi}_1$ fit together. From Lemma 12.2, the quotient space $\widehat{Q} = \widehat{\Upsilon}/\sim$ is a complex hyperbolic orbifold with isolated singularities. Likewise, $\widehat{Q}^\infty$ is a spherical CR manifold. Indeed, given our product structure in a neighborhood of Ω, the quotient $\widehat{Q}$ is a capped complex hyperbolic orbifold, in the sense of Chapter 3, and the ideal boundary is $\widehat{Q}^\infty$.

Let $\widehat{\Gamma}^*$ be the orbifold universal covering group of $\widehat{Q}$, guaranteed by Lemma 3.5. Given the way we have constructed $\widehat{Q}$, we see that $\widehat{\Gamma}^*$ contains a generating set for $\widehat{\Gamma}$ and vice versa, at least if we choose our master constants large. Hence, $\widehat{\Gamma} = \widehat{\Gamma}^*$. This proves that $\widehat{\Gamma}$ is discrete. Moreover, $\boldsymbol{CH}^2/\widehat{\Gamma} = \widehat{Q}$. Since $\widehat{Q} \cup \widehat{Q}^\infty$ is a capped orbifold, we see by Lemma 3.7 that $\widehat{\Omega}/\widehat{\Gamma} = \widehat{Q}^\infty$.

The object

$$[\widehat{\Psi}] = (\partial \widehat{\Upsilon})/\sim \tag{13.7}$$

has the same local structure as $\widehat{\Psi}_1$ and, hence, is a simple complex in $\widehat{Q}$. We can lift $[\widehat{\Psi}]$ to a $\widehat{\Gamma}$-invariant simple complex $\widehat{\Psi} \subset \boldsymbol{CH}^2$. ISC we have $\widehat{\Psi}_2 \subset \widehat{\Psi}$. The same remarks apply to the boundary, and we have a simple complex $\widehat{\Psi}^\infty$, which extends $\widehat{\Psi}_2^\infty$ ISC. By construction, the components of $\widehat{\Omega} - \widehat{\Psi}^\infty$ are lifts of the interiors of the sets $\widehat{\Upsilon}_j^\infty$. Hence, by symmetry every component of $\widehat{\Omega} - \widehat{\Psi}$ is a $\widehat{\Gamma}$-translate of one of the sets $\widehat{V}_j^\infty$ we constructed above.

13.5 THE SURGERY FORMULA

We have a description of Ω/Γ as a gluing together of chunks Υ_1^∞,...,Υ_k^∞ according to a geometric equivalence relation. Here $\Upsilon_j^\infty = V_j^\infty/\Gamma_j$ is a torus

cross a ray. We have the analogous setup for the perturbed group, where now $\widehat{\Upsilon}_j^\infty$ is a 3-manifold with boundary. The boundaries of Υ_j^∞ and $\widehat{\Upsilon}_j^\infty$ have combinatorially identical cell decompositions, and there is a cellular homeomorphism between them.

We have a nice horotube $T_j = \langle f_j \rangle_2 \subset V_j^\infty$ and a corresponding smooth surface $\widehat{T}_j := \langle f_j \rangle_2 \subset \widehat{\Upsilon}_j$. Fix some index j. For ease of exposition we will suppose that $\widehat{\Gamma}_j$, the stabilizer of $\widehat{V}_j^\infty$, is generated by a lens-elliptic element of type (m, k). The Local Surgery Lemma gives us partitions

$$(S^3 - p)/\Gamma_j = W \cup (T_j/\Gamma_j), \qquad S^3/\widehat{\Gamma}_j = \widehat{W} \cup (\widehat{T}_j/\Gamma_j), \tag{13.8}$$

together with a homeomorphism $h : W \to \widehat{W}$, which is close to the identity in the sense of Chapter 8: We can lift h to a fundamental domain for W, and it will be close to the identity as a map on a subset of S^3. Perturbing h slightly we can arrange that h restricts to the cellular homeomorphism we already have from $\partial\Upsilon_j^\infty$ to $\partial\widehat{\Upsilon}_j^\infty$.

Let $R = T_j/\Gamma_j$ and $\widehat{R} = \widehat{T}_j/\widehat{\Gamma}_j$. The Local Surgery Lemma gives us a description of $\widehat{W} \cup \widehat{R}$ as being obtained by performing (m, k) Dehn filling on R according to the canonical marking on ∂R. Here we use our homeomorphism h to identity W with $\widehat{W}$. But h also identifies $\Upsilon_j^\infty - R$ with $\widehat{\Upsilon}_j^\infty - \widehat{R}$. Hence, $\widehat{\Upsilon}_j^\infty$ is obtained from Υ_j^∞ by performing a Dehn filling of type (m, k). All this filling takes place away from $\partial\Upsilon_j^\infty$ and is not affected by the way we glue the various pieces together to make Ω/Γ. Hence, the same filling that happens on the individual Υ_j^∞ also happens inside Ω/Γ. This gives the surgery formula in the HST.

13.6 HOROTUBE GROUP STRUCTURE

Suppose that the quotient $\widehat{\Omega}/\widehat{\Gamma}$ is noncompact. This is the case in the HST where at least one cusp is not filled. From the analysis in the previous section, $\widehat{\Omega}/\widehat{\Gamma}$ can be written as the union of a compact set and a finite disjoint union of horocusps. From the analysis above, $\boldsymbol{CH}^2/\widehat{\Gamma}$ is a complex hyperbolic orbifold with isolated singularities. Hence, $\widehat{\Gamma}$ is a discrete group of isolated type. It only remains to verify that $\widehat{\Lambda}$ is porous.

Suppose, for the sake of getting a contradiction, that $\{g_n\} \in PU(2,1)$ is a sequence such that $g_n(\widehat{\Lambda})$ Hausdorff-converges to all of S^3. Let x_n be such that $g_n(x_n) = (0,0)$. For any point $x \in \boldsymbol{CH}^2$, let $\mathrm{VIS}(x)$ denote the largest distance d such that $g_x(\Omega)$ contains a ball of spherical diameter at least d. By construction we have $\mathrm{VIS}(x_n) \to 0$. If $\gamma \in \widehat{\Gamma}$, then by symmetry we have $\mathrm{VIS}(x) = \mathrm{VIS}(\gamma(x))$. Since there are only finitely many cusps mod Γ, we can use the above symmetry and pass to a subsequence so that $x_n \in V_1$ for all n. Here V_1 is particular component of $\boldsymbol{CH}^2 - \widehat{\Psi}$ that is independent of n. If $\{x_n\}$ lies in a compact subset of $\boldsymbol{CH}^2$, then obviously $\mathrm{VIS}(x_n)$ does not converge to 0. Hence, $x_n \to x \in S^3$ on a subsequence. Recall that $\widehat{V}_1^\infty$

is an open component of $\widehat{\Omega} - \widehat{\Psi}^\infty$. Let

$$Y = \widehat{V}_1^\infty \cup \partial \widehat{V}_1^\infty \cup \mathrm{Fix}(\widehat{\Gamma}_1). \tag{13.9}$$

$$\partial Y = \partial \widehat{V}_1^\infty \cup \mathrm{Fix}(\widehat{\Gamma}) \tag{13.10}$$

is either a torus, a cylinder compactified at one point, or a sphere, depending on whether $\widehat{\Gamma}_1$ is elliptic, parabolic, or loxodromic.

Lemma 13.1 *$x \in Y$.*

Proof: Given a set $S \subset \boldsymbol{CH}^2$ let $A(S)$ denote the set of accumulation points of S on S^3. It suffices to show that $A(\widehat{V}_1) \subset Y$, and this is what we will do.

ISC the set $\widehat{V}_1$ has the same local combinatorial structure as V_1. In particular, the compact cells of $\partial \widehat{V}_1$ lie in $\boldsymbol{CH}^2$, and every noncompact cell $\widehat{\sigma}$ of $\partial \widehat{V}_1$ is such that $A(\widehat{\sigma}) \subset \widehat{\Omega}$ is a cell of $\partial \widehat{V}_1^\infty$. Moreover, there are only finitely many cells of $\widehat{V}_1/\widehat{\Gamma}_1$. From this we conclude that $A(\partial \widehat{V}_1) = \partial Y$. Now $S^3 - \partial Y$ consists of two components, one of which is the interior of Y and the other of which is $S^3 - Y$.

If some sequence of $\widehat{V}_1$ converges to a point $x' \in S^3 - Y$, then some other sequence of $\widehat{V}_1$ converges to a point $x'' \in S^3 - Y$ contained in some small neighborhood U of a cell $\widehat{\sigma}^\infty$ of ∂Y. This contradicts the one-side property mentioned at the end of Section 13.3. Hence, x' does not exist. □

Lemma 13.2 *$\widehat{\Gamma}_1$ is a parabolic subgroup, and $x = \widehat{p}_1$, the fixed point of $\widehat{\Gamma}_1$.*

Proof: Note that $Y - \mathrm{Fix}(\widehat{\Gamma}_1) \subset \widehat{\Omega}$. So, if $x \notin \mathrm{Fix}(\widehat{\Gamma}_1)$, then $x \in \widehat{\Omega}$. If $x \in \widehat{\Omega}$, then $\mathrm{VIS}(x_n) \to 1$, the spherical diameter of S^3. Hence, $x \in \mathrm{Fix}(\widehat{\Gamma}_1)$. If $\widehat{\Gamma}_1$ is an elliptic subgroup, then $\mathrm{Fix}(\widehat{\Gamma}_1) \cap S^3 = \emptyset$. Hence, there is no location at all for x. This case is impossible. Suppose that $\widehat{\Gamma}_1$ is a loxodromic subgroup generated by some loxodromic element $\widehat{h}$. Then we can choose a sequence $\{h_n\} \in \widehat{\Gamma}_1$ such that $y_n = h_n(x_n)$ converges to neither fixed point of $\widehat{\Gamma}_1$. The idea here is that the generator of $\widehat{\Gamma}_1$ attracts about one fixed point and repels around the other. By symmetry we have $\mathrm{VIS}(y_n) = \mathrm{VIS}(x_n)$. But y_n converges to a point in $\widehat{\Omega}$, and hence, $\mathrm{VIS}(y_n) \to 1$. The only case we haven't ruled out is the case where $\widehat{\Gamma}_1$ is parabolic and $x = \widehat{p}_1$, the fixed point. □

Let $\widetilde{T}_1 = \langle \widehat{f}_2 \rangle$. Recall that $\mathrm{DIAM}_{x_n}(\widehat{T}_1)$ is the visual diameter of $\widehat{T}_1$.

Lemma 13.3 $\mathrm{DIAM}_{x_n}(\widehat{T}_1)$ *is bounded away from* 0.

Proof: Suppose that $\mathrm{DIAM}_{x_n}(\widehat{T}_1) \to 0$. Lemma 6.2 then says that there are elements $h_n \in \widehat{\Gamma}_1$ such that $h_n(x_n)$ converges to a point $y \in S^3 - \widehat{p}_1$. Since

h_n stabilizes $\widehat{V}_1$, we could replace our original sequence $\{x_n\}$ by $\{h_n(x_n)\}$ and rerun the previous two lemmas. This would tell us that $y = \widehat{p}_1$, a contradiction. □

Lemma 13.4 $\mathrm{DIAM}_{x_n}(\partial \widehat{T}_1)$ *is bounded away from* 0.

Proof: Suppose, on the contrary, that $\mathrm{DIAM}_{x_n}(\partial \widehat{T}_1) \to 0$. Recall that $g_n(x_n) = (0,0)$. The spherical diameter of $g_n(\widehat{T}_1)$ is bounded away from 0, but the diameter of $g_n(\partial \widehat{T}_1)$ converges to 0. This is only possible if $g_n(\widehat{T}_1)$ is converging to all of S^3. But then $\mathrm{VIS}(x_n) \to 1$. This is a contradiction. □

Now we are in a position to apply Lemma 9.1. Note that $\widehat{\Omega}/\widehat{\Gamma}$ is the union of a compact set together with a finite union of horocusps, one of which can be taken to coincide with $\widehat{T}_1/\widehat{\Gamma}_1$. Thus we can produce a horotube assignment for $\widehat{\Gamma}$, which includes $\widehat{T}_1$ as one of the horotubes and satisfies all the conditions needed for Lemma 9.1, a result that does not use the porous limit set condition. Recall that $x_n = g_n^{-1}(0,0)$. Adjusting g_n as in Lemma 9.1 we can assume that $\{g_n\}$ satisfies the conclusions of Lemma 9.1. None of the alternatives in Lemma 9.1 leads to $g_n(\widehat{\Lambda})$ filling up all of S^3. This contradiction finishes our proof that $\widehat{\Lambda}$ is porous.

13.7 PROOF OF THEOREM 1.11

We summarize the information we get from the proof of the HST.

- There are locally isomorphic complexes $\Psi^\infty \subset \Omega$ and $\widehat{\Psi}^\infty \subset \widehat{\Omega}$. The individual cells inject into Ω/Γ and $\widehat{\Omega}/\widehat{\Gamma}$, respectively.
- When no filling occurs, we have a homeomorphism $h : \Omega/\Gamma \to \widehat{\Omega}/\widehat{\Gamma}$, which is a cellular homeomorphism from Ψ^∞/Γ to $\widehat{\Psi}^\infty/\Gamma$.
- For each cusp p of Ω, there is a set V_p^∞ whose boundary lies in Ψ^∞. These *pieces* partition Ω. The same goes for $\widehat{\Omega}$.

Since ρ and $\widehat{\rho}$ are horotube representations of the same group, with the same set of parabolic elements, the cusps of Γ are canonically bijective with the cusps of $\widehat{\Gamma}$. This means that each cell of Ψ^∞ can be matched up canonically with a cell of $\widehat{\Psi}^\infty$. The cell σ of Ψ^∞ gets matched to the cell $\widehat{\sigma}$ of $\widehat{\Psi}^\infty$, which is defined by the corresponding collection of cusps. Moreover, each cell injects into the quotient. Thus we can canonically lift $h : \Psi^\infty/\Gamma \to \widehat{\Psi}^\infty/\widehat{\Gamma}$ cell by cell to a homeomorphism $\widetilde{h} : \Psi^\infty \to \widehat{\Psi}^\infty$.

The pieces of Ω are in bijection with the cusps of Γ, and the pieces of $\widehat{\Omega}$ are in bijection with the cusps of $\widehat{\Gamma}$. Hence, the pieces of Ω are naturally in bijection with the pieces of $\widehat{\Omega}$. Let V be a piece of Ω, and let $\widehat{V}$ be the

corresponding piece of $\widehat{\Omega}$. Let π be the projection from V into Ω/Γ. We would like to lift h to a homeomorphism $V \to \widehat{V}$, given that we already have a lift $\widetilde{h} : \partial V \to \partial \widehat{V}$.

Choose some $x_0 \in \partial V$. Given any $y \in V$, let α be a path in V joining x_0 to y. That is, $\alpha(0) = x_0$ and $\alpha(1) = y$. Let $[\alpha]$ denote the projection of α to V/Γ. Let β be the unique lift to $\widehat{\Omega}$ of the path $h([\alpha])$ such that $\beta(0) = \widetilde{h}(x_0)$. We define $\widetilde{h}(x) = \beta(1)$.

Lemma 13.5 *Our extension is well defined.*

Proof: Given any set $X \subset \Omega$, let $[X]$ denote the corresponding set in Ω/Γ. Note that the covering map $\Omega \to \Omega/\Gamma$ is normal. Every closed loop in Ω/Γ either has all closed lifts or no closed lifts. The normality comes from the presence of the group action. The same goes for the covering of $\widehat{\Omega}$ over $\widehat{\Omega}/\widehat{\Gamma}$. We also note that V deformation retracts to ∂V.

Suppose that α_1 and α_2 are two paths in V that join x_0 to y. Let β_1 and β_2 be the corresponding paths in $\widehat{\Omega}$. We need to show that $\beta_1(1) = \beta_2(1)$. Let A be the loop $\alpha_1 * \alpha_2^{-1}$, where $*$ means concatenation. Let B be any lift of the closed loop $h([A])$, a loop based at $[x_0]$. Then $\beta_1(x) = \beta_2(x)$ iff B is closed. To show that B is closed, it suffices to find a loop χ, based at $[x_0]$, which is isotopic (rel $[x_0]$) to $[A]$ such that $h(\chi)$ has a closed lift. Since V deformation retracts to ∂V, the loop A is isotopic (rel x_0) to a loop $A' \subset \partial V$. We let $\chi = [A']$. One of the lifts of $h(\chi)$ is $\widetilde{h}(A')$, a closed loop. We get an isotopy from $[A]$ to χ by pushing down the isotopy between A and A'. Hence, B is closed. □

Our extension, being well defined, is certainly continuous. We can make the same construction reversing the roles of V and $\widehat{V}$. Hence, $\widetilde{h}$ has a continuous inverse. Now we know that $\widetilde{h}$ extends to give a homeomorphism from V to $\widetilde{V}$ that is compatible with the group actions. All these extensions fit together across the different pieces, giving us our global homeomorphism.

13.8 DEALING WITH ELLIPTICS

Here we explain how to modify our constructions to handle the case where Γ has some elliptic elements of isolated type. The difficulty presented by the elliptic points is that the Transversality Lemma of Chapter 11 requires a free action of Γ on $\boldsymbol{CH}^2$. There is a nice fix for this problem.

- We redefine Cusp(Γ) to be the union of all the parabolic fixed points of Γ, together with all the torsion points—i.e., fixed points of elliptic elements of Γ.
- We delete from $\boldsymbol{CH}^2$ all the torsion points and make the same construction in the proof of the Transversality Lemma but relative to the

"punctured space." We don't care how these functions interact very near the deleted points because of the construction to follow.

- In the case where q is a torsion point, we define a function F_q that has a pole at q and vanishes outside a small neighborhood of q. This function then dominates all the others in a small neighborhood of q.

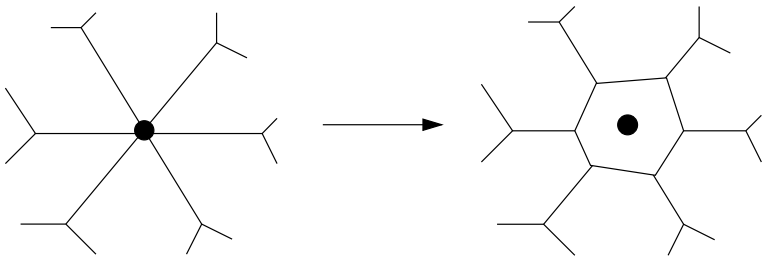

Figure 13.1: Resolving the singularities

With these changes made, we get the same result as in the Transversality Lemma, relative to the augmented list of functions. We then use these functions to define Ψ. By construction Ψ does not intersect the torsion points, and each torsion point is surrounded by a bounded component of $\boldsymbol{CH}^2 - \Psi$. Figure 13.1 shows a caricature of the situation.

All the constructions in this chapter now go through, pretty much word for word. The extra little components we introduce by our modified construction play no role in the analysis other than accounting for the isolated singular points in $\boldsymbol{CH}^2/\widehat{\Gamma}$, which are already present in $\boldsymbol{CH}^2/\Gamma$.

PART 3
The Applications

In this part of the monograph we establish all the applications of the HST, using Theorems 1.3 and 1.4 as black boxes. In Part 4 we will prove Theorems 1.3 and 1.4.

Chapter Fourteen

The Convergence Lemmas

14.1 STATEMENT OF RESULTS

In this chapter we investigate the interplay between the ordinary convergence of sequence of elements $\{P_n\} \in PU(2,1)$ to some $P \in PU(2,1)$ and the stronger geometric convergence defined in Section 3.1. Recall that a parabolic element is $\boldsymbol{R}$-parabolic iff it stabilizes an $\boldsymbol{R}$-slice and is $\boldsymbol{C}$-parabolic iff it stabilizes a unique $\boldsymbol{C}$-slice. In the $\boldsymbol{R}$-parabolic case, the map is conjugate to the map in Equation 2.13 with $s = 0$. In the $\boldsymbol{C}$-parabolic case, the map is conjugate to the one in Equation 2.12 with $u \neq 1$. We say that a parabolic element P is *irregular* if it is conjugate to the one in Equation 2.12 with $u = 1$ and otherwise *regular*. Even though the irregular parabolics are conjugate to a very simple map on Heisenberg space, they are hard for us to deal with.

Lemma 14.1 (Convergence I) *Let P be a regular parabolic element. Suppose that $\{P_n\}$ is a sequence of parabolic elements such that $P_n \to P$ algebraically. Then $\langle P_n \rangle$ converges geometrically to $\langle P \rangle$.*

For elliptic elements the situation is trickier and our results are more limited. Lens-elliptic elements of type (m,k) are defined in Equation 2.7. Say that an elliptic element P is *weak lens-elliptic* if it has the same trace as a lens-elliptic element. The reader familiar with Theorem 6.2.4 in [**G**] might be puzzled about this definition. In [**G**, Theorem 6.2.4] it is claimed that the trace of an such an element determines its conjugacy class; but this is not quite right. If we take a diagonal matrix in $SU(2,1)$ and suitably permute the diagonal entries, then we obtain another matrix with the same trace which might not be conjugate.

Lemma 14.2 (Convergence II) *Let P be a $\boldsymbol{C}$-parabolic element. Suppose that $\{P_n\}$ is a sequence of weak lens-elliptic elements converging algebraically to P. Then $\langle P_n \rangle$ converges geometrically to $\langle P \rangle$. If P_n' is a sequence of lens-elliptic elements converging to P, and P_n and P_n' have the same trace, then P_n and P_n' are conjugate for large n.*

Our last result is obviously tailored to a specific application.

Lemma 14.3 (Convergence III) *Let P be an $\boldsymbol{R}$-parabolic element. Suppose that $p_0, p_1, p_2 \in S^3$ are 3 distinct points. For $j = 0, 1, 2$ let $\{p_{j,n}\}$ be a*

sequence of points in $\boldsymbol{CH}^2$ *such that* $\lim_{n\to\infty} p_{j,n} = p_j$. *For* $j = 1, 2$ *let* J_j *be the complex reflection fixing* p_0 *and* p_j, *and let* $J_{j,n}$ *be the complex reflection fixing* $p_{0,n}$ *and* $p_{j,n}$. *Suppose that* $P_n = J_{1,n}J_{2,n}$ *stabilizes an* $\boldsymbol{R}$*-slice* Π_n. *Then* P_n *converges geometrically to* P.

14.2 PRELIMINARY LEMMAS

Here we establish a few easy results that help us with the proofs of Convergence Lemmas II and III.

Lemma 14.4 *Let* $\{P_n\}$ *be a sequence of elliptic elements in* $PSL_2(\boldsymbol{R})$ *that converges algebraically to a parabolic element* $P \in PSL_2(\boldsymbol{R})$. *Suppose, for each* n, *that* P_n *is conjugate to a rotation through an angle of* $2\pi/m_n$ *for some* $m_n \in \boldsymbol{N}$. *Then* P_n *converges geometrically to* P.

Proof: Let Λ_n be a family of m_n evenly spaced geodesic rays emanating from the fixed point of P_n. Let F_n be the wedge-shaped region bounded by a pair of successive rays. Then F_n is a fundamental domain for $\langle P_n \rangle$, and evidently F_n converges to a fundamental domain for $\langle P \rangle$. This shows that $\langle P_n \rangle$ converges in the Hausdorff topology to $\langle P \rangle$. □

Remark: The condition on the rotation angle is necessary. For instance, if $\{P_n\}$ rotates by an angle of $4\pi/n$, with n odd, then a suitable sequence of the form $\{P_n^{k_n}\}$ converges to the square root of P.

Corollary 14.5 *Let* P *be an* $\boldsymbol{R}$*-parabolic element. Suppose that* $\{P_n\}$ *is a sequence of elliptic elements in* $PU(2,1)$ *that converges algebraically to a parabolic element* $P \in PU(2,1)$. *Suppose that there is an* $\boldsymbol{R}$*-slice* Π *such that* P_n *stabilizes* Π *for all* n *and* P_n *rotates* Π *through an angle of* $2\pi/m_n$ *for some* $m_n \in \boldsymbol{N}$. *Then* P_n *converges geometrically to* P.

Proof: We will first deal with the question of Hausdorff convergence. An element of $PU(2,1)$ that stabilizes an $\boldsymbol{R}$-slice is determined by its action on this slice. The Hausdorff convergence now follows from Lemma 14.4.

We still need to show that $\langle P_n \rangle$ acts freely on S^3. To see this, note that the first $m_n - 1$ powers of P_n rotate both Π and $\Pi_n^{\perp}$ nontrivially. Here $\Pi_n^{\perp}$ is the $\boldsymbol{R}$-slice perpendicular to Π through the fixed point of P_n. From this property it is easy to see that none of the first $m_n - 1$ powers of P_n fixes a point on S^3, and the m_nth power is trivial. □

Corollary 14.6 *Let* P *be a* $\boldsymbol{C}$*-parabolic element. Suppose that* $\{P_n\} \in PU(2,1)$ *is a sequence of lens-elliptic elements that converges algebraically to a parabolic element* $P \in PU(2,1)$. *Suppose that there is a* $\boldsymbol{C}$*-slice* Π *such that* P_n *and* P *stabilize* Π *for all* n. *Then* P_n *converges geometrically to* P.

Proof: Let m_n be the order of P_n. If m_n does not tend to ∞, then the sequence $\{P_n\}$ falls within a finite number of conjugacy classes of elliptic elements. In this case it is impossible for P_n to converge to P. Therefore, we have $m_n \to \infty$. Now P_n rotates Π through an angle of either $2\pi/m_n$ or $2k_n\pi/m_n$. Note that $m_n \to \infty$ by hypothesis. If P_n rotates Π by an angle of $2k_n\pi/m_n$ for $k_n > 1$, then P_n rotates the normal bundle $N(\Pi)$ of Π by $2\pi/m_n$, an amount that tends to 0 as $n \to \infty$. On the other hand, P rotates $N(\Pi)$ by a nonzero angle. This is where we use the fact that P has a nontrivial twist. Hence, it is impossible for P_n to rotate $N(\Pi)$ by an angle of $2\pi/m_n$ once n is sufficiently large. Hence, P_n rotates $N(\Pi)$ by an angle of $2k_n\pi/m_n$ and $k_n \to \infty$. By Lemma 14.4, $P_n|_\Pi$ converges geometrically to $P|_\Pi$. Now let $\pi : \boldsymbol{CH}^2 \to \Pi$ be an orthogonal projection. Then $\pi^{-1}(F_n)$ is a fundamental domain for the action of P_n that converges to the fundamental domain $\pi^{-1}(F)$ for P. Here F_n and F are as in the proof of Lemma 14.4. This picture implies that $\langle P_n \rangle$ Hausdorff-converges to $\langle P \rangle$. Also $\langle P_n \rangle$ acts freely because (by definition of lens-elliptic elements) k_n is relatively prime to m_n. □

14.3 PROOF OF THE CONVERGENCE LEMMA I

We work in $\mathcal{H}$ so that P is one of the maps in either Equation 2.12 or 2.13. The fixed point of P_n necessarily converges to ∞ because otherwise P would fix some point in $\mathcal{H}$. Hence, there is some sequence $\{\epsilon_n\} \in PU(2,1)$ converging to the identity, so that $\epsilon_n \circ P_n \circ \epsilon_n^{-1}$ fixes ∞ for all n. But P_n converges geometrically to P iff $\epsilon_n P_n \epsilon_n^{-1}$ converges geometrically to P. Hence, without loss of generality, we can assume that P_n fixes ∞ for all n.

The irregular parabolic elements all have trivial actions on $\boldsymbol{C}$. Hence, a sequence of irregular parabolic elements cannot converge to a regular parabolic element. So P_n is a regular parabolic element for large n.

If P_n does not converge geometrically to P, then we can find a sequence of exponents e_n such that $P_n^{e_n}$ converges to some parabolic element not in $\langle P \rangle$. If $\{e_n\}$ is bounded, then this is impossible because any given finite power of P_n converges to the same finite power of P. Hence, $\{e_n\}$ is unbounded. For any Heisenberg automorphism Q, let $[Q]$ denote the induced action of Q on $\boldsymbol{C}$.

Case 1: Suppose that P is not $\boldsymbol{C}$-parabolic and P_n is not $\boldsymbol{C}$-parabolic. Then we can normalize so that P is as in Equation 2.13. Then $[P](z) = z+1$ and $[P_n](z) = z + t_n$. We must have $t_n \to 1$. But $[P_n^{e_n}](z) = z + t_n e_n$, and this sequence of maps cannot converge if $\{e_n\}$ is unbounded. Hence, $P_n^{e_n}$ cannot converge.

Case 2: Suppose that P is $\boldsymbol{C}$-parabolic. Then P is as in Equation 2.12 with $u \neq 1$. We claim that P_n is also $\boldsymbol{C}$-parabolic for n large. Otherwise,

$[P_n]$ converges to either the identity or a translation. This is not possible because $[P]$ is a nontrivial rotation and $[P_n]$ converges to $[P]$. (We could also use Lemma 2.1 here.) Since P_n is a $\boldsymbol{C}$-parabolic, there is a unique $z_n \in \boldsymbol{C}$ such that P_n stabilizes $z_n \times \boldsymbol{R}$. Note that $[P_n]$ is a nontrivial rotation about z_n. Since these rotations converge to rotation about 0, we must have $z_n \to 0$. Hence, there is a nearly horizontal plane Π_n such that P_n maps Π_n parallel to itself. The height difference between Π_n and $P_n(\Pi_n)$ tends to 1. Hence, for any compact K, there is a uniformly bounded number of P_n-translates of Π_n, which intersect K. This makes it impossible for $P_n^{e_n}$ to converge to any element if e_n is unbounded.

Case 3: Now we have the subtle case. Suppose that P is not $\boldsymbol{C}$-parabolic and P_n is $\boldsymbol{C}$-parabolic. Since the twist of P_n is nontrivial, there is a unique $z_n \in \boldsymbol{C}$ such that P_n stabilizes the line $z_n \times \boldsymbol{R}$. Since $[P_n]$ converges to a translation and $[P_n]$ acts as a rotation, we must have $|z_n| \to \infty$.

For any complex number z, let $\theta(z)$ denote the small angle that the line $\overline{0z}$ makes with $\boldsymbol{R}$. We claim that $\theta(z_n)$ converges to $\pi/2$. To see this, note that $[P_n]$ is nearly a translation along the line tangent to the circle centered at z_n and containing 0. Since $[P_n]$ converges to the map $z \to z + 1$, our tangent line must converge to $\boldsymbol{R}$. Hence, $\overline{0z}$ must converge to $i\boldsymbol{R}$. Now we know that $\theta(z_n)$ converges to $\pi/2$.

There is a plane Π_n such that Π_n and $P_n(\Pi_n)$ are parallel. In fact Π_n can be taken as any spinal sphere with one pole on $z_n \times \boldsymbol{R}$ and one pole at ∞. Indeed, Π_n is just the image of $\boldsymbol{C} \times \{0\}$ under some Heisenberg automorphism. We can adjust the finite pole of Π_n such that the origin of $\mathcal{H}$ lies in Π_n. Now, Π_n is tangent to the contact plane at its pole and, hence, is nearly vertical at the origin. Since $\theta(z_n) \to \pi/2$, the action of P_n near 0 is nearly perpendicular to Π_n. Hence, near 0, the planes Π_n and $P_n(\Pi_n)$ are a uniformly bounded distance from each other. Now we get the same contradiction as in Case 2.

14.4 PROOF OF THE CONVERGENCE LEMMA II

14.4.1 Convergence for Lens-Elliptic Elements

We first prove the Convergence Lemma II under the assumption that P_n is actually lens-elliptic. Since P_n is lens-elliptic, we know that P_n stabilizes two $\boldsymbol{C}$-circles C_n and C_n'. (These $\boldsymbol{C}$-circles bound perpendicular $\boldsymbol{C}$-slices, which correspond to the eigenspaces of P_n acting on the tangent space of its fixed point.) We label so that P_n rotates C_n by $2\pi/m_n$ and C_n' by $2\pi k_n/m_n$. This is to say that P_n rotates by $2\pi k_n/m_n$ in the direction normal to C_n.

Suppose we knew that the collection $\{C_n\}$ fell within a compact subset of $\boldsymbol{C}$-slices. Then we could conjugate P_n by arbitrarily small elements of $PU(2,1)$ so that P_n stabilizes a $\boldsymbol{C}$-slice Π independent of n. But then Corollary 14.6 applies, and we are done with the convergence part of the

Convergence Lemma II. Thus, our goal is to show that $\{C_n\}$ cannot exit every compact subset of S^3.

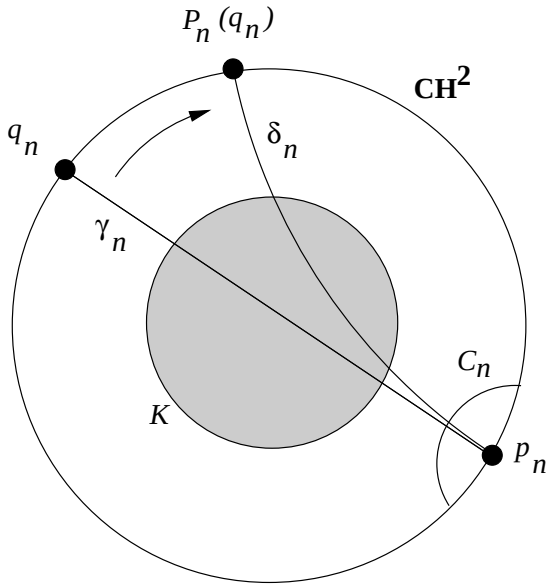

Figure 14.1: Geometric construction of geodesics

Suppose, on the contrary, that C_n exits every compact subset of the space of $\boldsymbol{C}$-circles. We can take a subsequence so that C_n converges to the fixed point $p \in S^3$ of P. There is a geodesic γ_n, through the origin of $\boldsymbol{CH}^2$, that is perpendicular to C_n. Let γ_n' be the tangent vector to γ_n at the point $\gamma_n \cap C_n$. Let $q_n \in S^3$ be the endpoint of γ_n that is far away from p. Let $r_n = P_n(p_n)$. Then $q_n \to q$, the point antipodal to p, and $r_n \to r = P(q)$. Let $\delta_n = P_n(\gamma_n)$. Note that δ_n is also perpendicular to C_n. Let δ_n' denote the tangent vector to δ_n at $\delta_n \cap C_n$. Figure 14.1 shows a schematic picture.

Since P_n moves the origin of $\boldsymbol{CH}^2$ a uniformly bounded amount, there is some compact $K \subset \boldsymbol{CH}^2$ such that γ_n and δ_n both intersect K for all n. Let π_n be an orthogonal projection onto C_n. Then $\gamma_n \cap C_n = \pi_n(\gamma_n)$ and $\delta_n \cap C_n = \pi_n(\delta_n)$ are both subsets of $\pi_n(K)$, a subset of vanishingly small diameter.

The tangent vectors γ_n' and δ_n' are not based at the same point. However, their basepoints are vanishingly close together, in the complex hyperbolic metric. (Both the basepoints lie in $\pi_n(K)$.) Thus we can measure the angle between the two vectors by parallel-translating γ_n to δ_n along the geodesic joining their basepoints. Call this angle θ_n. Note that θ_n must converge to the nonzero limit of $2\pi k_n/m_n$. (Since the twist of P is nontrivial, the quantity $2\pi k_n/m_n$ converges to a nonzero number.) On the other hand, due to the divergence properties of geodesics in pinched negative curvature, γ_n and δ_n must become increasingly parallel to each other in a small neighborhood of $\pi_n(K)$, in order that they both extend way out and pass through

K. This implies that $\theta_n \to 0$, and we have a contradiction. This contradiction establishes the Convergence Lemma II under the assumption that P_n is lens-elliptic.

14.4.2 Weak Versus Strong Lens-Elliptics

Now we suppose that $\{P_n\}$ is weak lens-elliptic and $P_n \to P$. We will show that P_n is actually lens-elliptic for n large. Our previous result then finishes the proof of the Convergence Lemma II. Let P_n' be a lens-elliptic element with the same trace as P_n. (We don't assume that $P_n' \to P$.)

The element P_n' is represented by a matrix in $SU(2,1)$ which is conjugate to a diagonal matrix with diagonal entries $\omega_1', \omega_2', \omega_3'$. Solving the equations

$$\omega_1'\omega_2'\omega_3' = 1, \qquad \frac{\omega_1'}{\omega_3'} = \exp\left(\frac{2\pi i}{m_n}\right), \qquad \frac{\omega_2'}{\omega_3'} = \exp\left(\frac{2\pi i k_n}{m_n}\right),$$

we get

$$\omega_1' = \exp\left(2\pi i \frac{-k_n + 4}{6m_n}\right),$$

$$\omega_2' = \exp\left(2\pi i \frac{2k_n - 2}{6m_n}\right),$$

$$\omega_3' = \exp\left(2\pi i \frac{-k_n - 2}{6m_n}\right).$$

Likewise, P_n is represented by a matrix in $SU(2,1)$ that is conjugate to a diagonal matrix with diagonal entries $\omega_1, \omega_2, \omega_3$. The fact that P_n and P_n' have the same traces forces $(\omega_1, \omega_2, \omega_3)$ to be a permutation of $(\omega_1', \omega_2', \omega_3')$. If the permutation is trivial, then we are done. It remains to rule out the other 5 possibilities. Using the symmetry $(z, w) \to (w, z)$, we can reduce to 2 cases.

Case 1: First, suppose $\omega_3 = \omega_2'$ and $\omega_2 = \omega_3'$. Then P_n is conjugate to the map

$$(z, w) \to \left(\frac{\omega_1'}{\omega_2'}z, \quad \frac{\omega_3'}{\omega_2'}w\right) = \left(\exp\left(2\pi i \frac{-k_n + 2}{2m_n}\right)z, \exp\left(2\pi i \frac{-k_n}{m_n}\right)w\right). \tag{14.1}$$

Since $-k_n$ is relatively prime to m_n, some power of P_n is lens-elliptic in the case under consideration.

Since some power of P_n is lens-elliptic, this power of P_n stabilizes a unique pair of $\boldsymbol{C}$-circles. But then P_n stabilizes the same pair of $\boldsymbol{C}$-circles C_n and C_n'. The quantity k_n/m_n cannot converge to 0, for otherwise the trace of P_n converges to 3, rather than to the trace of P. Hence, P rotates *both* C_n and C_n' by an amount that does not converge to 0. But then the same argument as in the previous section applies to both C_n and C_n', showing that neither C_n nor C_n' exits a certain compact subset of $\boldsymbol{C}$-circles. Passing to a subsequence, we can therefore assume that C_n and C_n' converge. But then

P_n converges to an elliptic element. This is a contradiction.

Case 2: Second, suppose that $\omega_3 = \omega_1'$ and $\omega_1 = \omega_3'$. Then P_n is conjugate to the map

$$(z,w) \to \Big(\frac{\omega_3'}{\omega_1'}z, \quad \frac{\omega_2'}{\omega_1'}w\Big) = \Big(\exp\Big(2\pi i\frac{-1}{m_n}\Big)z, \exp\Big(2\pi i\frac{k_n-2}{2m_n}\Big)w\Big). \quad (14.2)$$

In this case P_n^{-1} is lens-elliptic. Hence, P_n^{-1} converges geometrically to P^{-1} by our previous result. Hence, P_n converges geometrically to P. This finishes Case 2.

If P_n' happens to converge to P, then P_n and P_n' both converge geometrically to P. This rules out Case 2 above, because P_n rotates the normal direction to the plane Π_n by about half as much as P_n' rotates the normal direction to Π_n'. Here Π_n and Π_n' are $\boldsymbol{C}$-slices that converge to the $\boldsymbol{C}$-slice stabilized by P. Having ruled out all the nonconjugate cases, we see that P_n and P_n' must be conjugate.

This completes the proof of the Convergence Lemma II. □

14.5 PROOF OF THE CONVERGENCE LEMMA III

Evidently P_n converges to P in the ordinary sense. Since P_n stabilizes an $\boldsymbol{R}$-slice and rotates it by a nontrivial amount, P_n acts freely on S^3. We just need to show that $\langle P_n \rangle$ Hausdorff-converges to $\langle P \rangle$.

We will show that P_n stabilizes an $\boldsymbol{R}$-slice Π_n whose Euclidean diameter is bounded away from 0. Assume this for the moment. Then, by taking a subsequence, we can assume that Π_n converges to an $\boldsymbol{R}$-slice Π. Then we conjugate by vanishingly small elements to arrange that $\Pi_n = \Pi$ for all n. Lemma 14.5 now finishes the proof. So, to finish our proof we have to establish our claim that P_n stabilizes a uniformly large $\boldsymbol{R}$-slice Π_n.

For ease of notation we drop the subscript n. We write $\Pi' = \Pi_n$ and $P' = P_n$ to distinguish them from their parabolic counterparts. Let C_j be the $\boldsymbol{C}$-slice fixed by J_j. If p_0, p_1, p_2 already lie in an $\boldsymbol{R}$-slice, then we let Π' be this $\boldsymbol{R}$-slice and we are done. Henceforth, we assume that p_0, p_1, p_2 do not all lie in an $\boldsymbol{R}$-slice. If p_0, p_1, p_2 all lie in a $\boldsymbol{C}$-slice, then $J_1 = J_2$ and P' is trivial. Hence, p_0, p_1, p_2 do not lie in a $\boldsymbol{C}$-slice.

Lemma 14.7 *There exists $q_2 \in C_1$, not equal to any p_i, such that p_0, p_1, q_2 all lie in an $\boldsymbol{R}$-slice.*

Proof: This is most easily seen by normalizing so that

$$p_0 = (0,0), \qquad p_1 = (r,0), \qquad p_2 = (z,s),$$

for some $r, s \in \boldsymbol{R} - \{0\}$ and some $z \in \boldsymbol{C} - \boldsymbol{R}$. Consider the point

$$q_2 = (|z|^2, \overline{z}s).$$

Since $|z|^2 + s^2 \leq 1$, the distance from q_2 to 0 is $|z| < 1$. Since $\overline{z}p_2 = q_2$ and multiplication by any complex number stabilizes C_1, we see that $q_2 \in C_1$. Applying the map $(u, v) \to (u, zv)$ to the points p_0, p_1, q_2, we move these points to $\boldsymbol{R}^2$. Hence, p_0, p_1, q_2 all lie in a common $\boldsymbol{R}$-slice. □

Let Π' be the $\boldsymbol{R}$-slice containing the points (p_0, p_1, q_2). There is a uniform lower bound to the Euclidean distance between p_0 and p_1, both of which are contained in Π'. Hence, there is a uniform lower bound to the Euclidean diameter of Π'. We need to show that P' stabilizes Π'.

Lemma 14.8 *Suppose that J is a complex reflection fixing a $\boldsymbol{C}$-slice C. Suppose that R is an $\boldsymbol{R}$-slice that intersects C in a real geodesic. Then $J(R) = R$.*

Proof: We normalize so that $C = \boldsymbol{C} \times \{0\} \cap \boldsymbol{CH}^2$ and $J(z, w) = (z, -w)$ and $R \cap \boldsymbol{C} = (-1, 1) \times \{0\}$. Then R has the form $\{(r, e^{i\theta}s) |\ r, s \in \boldsymbol{R}\} \cap \boldsymbol{CH}^2$. But then clearly $J(R) = R$. □

The $\boldsymbol{C}$-slice C_1 contains p_0 and p_2. Hence, it also contains q_2. Hence, $J_1(q_2) = q_2$. The intersection $\gamma_1 = \Pi' \cap C_1$ is the real geodesic γ_1, containing the two points p_0 and q_2. Hence, $\gamma_1 \in C_1$. Hence, $\Pi' \cap C_1$ is a real geodesic. Our lemma now shows that $J_1(\Pi') = \Pi'$.

Since J_2 fixes p_0 and p_1, we see that the real geodesic γ_2 containing the two points p_0 and p_1 is the intersection $C_2 \cap \Pi'$. Our lemma now says that $J_2(\Pi') = \Pi'$. Hence, $P'(\Pi') = \Pi'$, as desired.

Chapter Fifteen

Cusp Flexibility

15.1 STATEMENT OF RESULTS

We want to apply the HST to the group $\mathbf{\Gamma_3}$ from Theorem 1.4, so we need some perturbations. In this chapter we construct the basic examples.

Let $\mathcal{R}$ denote the space of representations of $\boldsymbol{Z}_3 * \boldsymbol{Z}_3$ into $SU(2,1)$ in which the two generators map to order-3 elliptics acting freely on S^3. As is typical, we consider two representations the same if they are conjugate in $SU(2,1)$. The dimension count we give below shows that $\mathcal{R}$ is a 4-dimensional analytic manifold, at least in the region of interest to us.

We say that a *diamond group* is a representation $x \in \mathcal{R}$ such that $A_x B_x$ and $A_x B_x^{-1}$ are both parabolic. Let $\mathcal{D}$ be the set of diamond groups. In Section 4.6.3 we explained how to view $\mathbf{\Gamma_3}$ as a point $x_0 \in \mathcal{R}$. We have $x_0 \in \mathcal{D}$ by Equation 4.25. Thus $\mathbf{\Gamma_3}$ is both the golden group and a diamond group. We hope the reader forgives us this terminology. The reason for the term *diamond group* is that these groups are constructed using a kind of diamond of points in S^3. See below.

There is a natural *trace map* $F : \mathcal{R} \to \boldsymbol{C}^2$ given by

$$F(x) = \big(\mathrm{Tr}(A_x B_x), \mathrm{Tr}(A_x B_x^{-1})\big). \tag{15.1}$$

This chapter is devoted to proving the following result.

Lemma 15.1 (Cusp Flexibility) *There are points $x \in \mathcal{D}$ arbitrarily close to x_0 with the following properties.*

(1) *Both $A_x B_x$ and $A_x B_x^{-1}$ are $\boldsymbol{C}$-parabolic with irrational twist.*

(2) *F is an open mapping in a neighborhood of x.*

(3) *dF_x maps the tangent space at x onto $\{0\} \times \boldsymbol{C}$.*

The open mapping statement means that F maps small open sets containing x to open sets containing $F(x)$. Numerical calculations suggest that dF is nonsingular on an open dense subset of $\mathcal{D}$, but we don't need this stronger result.

We will use the first two items of the Cusp Flexibility Lemma to prove Theorem 1.5 and the last item to prove Theorem 1.7. The (dis)interested reader might want to skip ahead to the next chapter to see how the Cusp Flexibility Lemma is used before getting into the details of its proof.

15.2 A QUICK DIMENSION COUNT

Here is a dimension count that shows that $\mathcal{R}$ is 4-dimensional. Each $\boldsymbol{R}$-elliptic order-3 element fixes a pair of complex slices. We assign to each such element T a pair (p, Π), where $p \in \boldsymbol{CH}^2$ is the fixed point of T and Π is one of the two complex lines stabilized by T. To be definite, we always take the eigenspace whose eigenvector is $\exp(2\pi i/3)$. The space of $\boldsymbol{C}$-slices in $\boldsymbol{CH}^2$ is 4-dimensional because it is an open subset of $(\boldsymbol{CP}^2)^*$, the space of complex lines in $\boldsymbol{CP}^2$. Once we choose the $\boldsymbol{C}$-slice, there is a 2-dimensional family of choices for the point. All in all, we get a 6-parameter family of choices for each element. Together we have a 12-parameter family of choices, but $SU(2,1)$ is 8-dimensional and permutes these choices. Since there are no relations amongst the generators and (generically) no common stabilizers of our pair of flags, the dimension count is right.

Remark: See [**FP**] for a classification of the representations of $\boldsymbol{Z}_3 * \boldsymbol{Z}_2$ (the modular group) into $SU(2,1)$. The components of that representation variety are all 1-dimensional, and two of them appear as subsets of the larger family we consider here.

15.3 CONSTRUCTING THE DIAMOND GROUPS

Let $\mathcal{S} \subset \mathcal{R}$ denote those representations σ such that $A_\sigma B_\sigma^{-1}$ is parabolic. Here we will define a smooth map that bijectively maps an open subset of $\boldsymbol{R}^3$ onto a neighborhood of x_0 in $\mathcal{S}$. Given

$$(t,u,v) \in (0,1) \times (-1,0) \times \left(\frac{-\pi}{2}, \frac{\pi}{2}\right)$$

the following points in $\mathcal{H}$.

$$a_1 = (0,1), \quad a_2 = (0,-1), \quad b_1 = (t,0), \quad b_2 = \big(\exp(iv)\ u, 0\big). \tag{15.2}$$

We take $t \in (0,1)$ because the $\boldsymbol{C}$-reflection in the circle $S^1 \times \{0\}$ fixes a_1 and a_2 and acts as the map $t \to 1/t$ when restricted to $\boldsymbol{R} \times \{0\}$. The case $t = 1$ corresponds to the situation where a_1, a_2, b_1 all lie in the same $\boldsymbol{R}$-circle, and this is far from x_0.

Any ideal triangle has a 3-fold symmetry, and we let A and B be the order-3 elements whose orbits are

$$A: \quad a_2 \to a_1 \to b_1, \qquad\qquad B: \quad a_2 \to a_1 \to b_2. \tag{15.3}$$

(We are slightly abusing notation here and having $SU(2,1)$ act on $\mathcal{H}$.) There is some t_0 such that $(t_0, -t_0, 0)$ corresponds to x_0. The point here is that the complex reflection $(z,t) \to (z,-t)$ lies in the group $\boldsymbol{\Gamma}'$ and swaps b_1 and b_2.

Lemma 15.2 *For (s,t,u) as above, the element AB^{-1} is parabolic.*

Proof: Consider first the analogous picture in $\boldsymbol{H}^2$. In this case we can place in each triangle a triple of horocircles, mutually tangent at the centers of the edges. The 3-fold symmetry of each triangle permutes each horocircle triple. Suppose that A permutes some triangle τ, B permutes the adjacent triangle, and AB^{-1} fixes a common vertex x. Then AB^{-1} maps the horocircle h based at x back to some other horocircle $AB^{-1}(h)$ based at x. So AB^{-1} is parabolic iff $h = AB^{-1}(h)$, which is true iff the centers of symmetry of the corresponding triangles match.

One can play a similar game in $\boldsymbol{CH}^2$. To each ideal triangle we may associate a maximally symmetric triple of horospheres based at the vertices and mutually tangent at the centers of symmetry of the edges. Once again, if AB^{-1} fixes a common vertex, then AB^{-1} is parabolic iff it maps the corresponding horosphere back to itself. Again, this is to say that the centers of symmetry match up at the common edge.

The pair of points a_1, a_2 in the triangle (a_1, a_2, b_1) determines a $\boldsymbol{C}$-slice A_1. Let $\gamma_1 \subset A_1$ be the geodesic in $\boldsymbol{CH}^2$ joining a_1 to b_1. The horospheres about a_1 and a_2, relative to the triangle (a_1, a_2, b_1), are tangent along a point x_1 of γ_1. There is an anti-holomorphic isometry I_1 that stabilizes A_1, fixes x_1, and interchanges a_1 with a_2. This isometry acts on $\mathcal{H}$ as $(z, t) \to (\overline{z}, -t)$. The fixed point set of I_1 intersects A_1 in the geodesic γ_1', which is perpendicular to γ_1 at the above-mentioned tangency point. γ_1' intersects $\mathcal{H}$ in the two points 0 and ∞. All the same remarks can be made with b_2 in place of b_1, except that the isometry I_2 is a conjugate of I_1 by the rotation $(z, t) \to (\exp(iv)\ z, t)$. In particular, γ_2' intersects $\mathcal{H}$ at 0 and ∞. Hence, $\gamma_1' = \gamma_2'$ and $x_1 = x_2$. Our two sets of horospheres match along a common edge. □

Given any element of $\mathcal{S}$ fairly near x_0, we let a_2 be the fixed point of AB^{-1}. We then let $a_1 = A(a_2)$ and $b_1 = A(a_1)$. We must have $B^{-1}(a_1) = a_2$. Then we let $b_2 = B^{-1}(a_2)$. This gives us our points (a_1, a_2, b_1, b_2). If two representations σ and σ' are conjugate, then the conjugating element in $SU(2, 1)$ carries the points (a_1, a_2, b_1, b_2) to the points (a_1', a_2', b_1', b_2'). Hence, our smooth map gives a bijection from a neighborhood of $(t_0, -t_0, 0)$ onto a neighborhood of x_0.

15.4 THE ANALYTIC DISK

Let $\mathcal{D}$ denote the subset of $\mathcal{R}$ consisting of the diamond groups, as in the Cusp Flexibility Lemma. Since we have coordinatized $\mathcal{S}$ in a neighborhood of x_0 as an open set in $\boldsymbol{R}^3$, it makes sense to talk about analytic sets and functions on $\mathcal{S}$.

Lemma 15.3 *$\mathcal{D}$ intersects a neighborhood of x_0 in an analytically embedded 2-dimensional subdisk of $\mathcal{S}$.*

Proof: We let δ be the function from Lemma 2.1. We define $D : \mathcal{S} \to \boldsymbol{R}$ by the equation $D(t,u,v) = \delta(AB)$, where A and B depend on the parameters (t,u,v). We have $\mathcal{D} = D^{-1}(0)$. To prove Lemma 15.3, we just have to show that the differential dD is nonzero at x_0. Let $v = (1,-1,0)$ be the tangent vector at x_0 corresponding to the variation $t \to (t,-t,0)$. It suffices to show that $dD_{x_0}(v) \neq 0$.

Now, the groups $(t,-t,0)$ are conjugate in $PU(2,1)$ to the groups constructed in Section 4.6.3. The point here is that both groups have a complex reflection conjugating A to B. Letting τ be the parameter in Section 4.6.3, suppose we knew that the correspondence $t \leftrightarrow \tau$ was smooth and regular in a neighborhood of $t_0 \leftrightarrow \tau_0$. Then equation 4.24 would imply $dD_{x_0}(v) \neq 0$.

The calculation in Section 4.6.3 shows that the argument of the triple product X in Equation 4.19 is a smooth regular function of τ in a neighborhood of τ_0. To compute the corresponding triple product for $(a_1, a_2, b_1(t))$, points that live in Heisenberg space, we follow [**FP**]. The point $(z,t) \in \mathcal{H}$ corresponds to the vector $(|z|^2 + it, \sqrt{2}z, 1)$ relative to the Hermitian form

$$\langle Z, W \rangle' = Z_1 \overline{W}_3 + Z_2 \overline{W}_2 + Z_1 \overline{W}_3.$$

Using this form and Equation 15.2, we compute

$$\langle a_1, a_2 \rangle' \langle a_2, b_1(x) \rangle' \langle b_1(t), a_1 \rangle' = 2i(t^2+i)^2.$$

This triple product equals the corresponding triple product for the corresponding points in S^3. The last equation shows that the argument of the triple product is a smooth regular function of t in a neighborhood of t_0. This shows that our correspondence $t \leftrightarrow \tau$ is smooth and regular in a neighborhood of $t_0 \leftrightarrow a_0$. □

15.5 PROOF OF THE CUSP FLEXIBILITY LEMMA

Lemma 15.4 *The map F is at most 2 to 1 in a neighborhood of $x_0 \in \mathcal{R}$.*

Proof: This result follows from the main result in Sean Lawton's beautiful thesis [**La**] on $SL_3(\boldsymbol{C})$ representations of free groups. There are two main points to this deduction. First, the representations in $\mathcal{R}$ near x_0 are *closed points* in the $SL_3(\boldsymbol{C})$ character variety. Second, an $SL_3(\boldsymbol{C})$ conjugacy between two of our representations is actually an $SU(2,1)$ conjugacy because our representations are *Zariski dense* in $SU(2,1)$. □

Remark: In the next section we will give an elementary and self-contained proof of Lemma 15.4.

Henceforth, we restrict our attention to a neighborhood of x_0 in which F is at most 2 to 1. In particular, we replace $\mathcal{D}$ by a smaller disk that is entirely contained in this small neighborhood of x_0. We say that an open set $U \subset \mathcal{R}$ is *good* if F is injective on U.

Lemma 15.5 *Any point $x = (t, u, v)$, with $v \neq 0$, has a good neighborhood, provided it is sufficiently close to x_0.*

Proof: There is a canonical and continuous involution of $\mathcal{R}$ that commutes with F. We choose any anti-holomorphic isometry ψ of $\boldsymbol{CH}^2$ and consider the involution $(A, B) \to (A', B')$, where in general $X' = \psi^{-1} X^{-1} \psi$. This works because

$$\mathrm{Tr}(A'B') = \overline{\mathrm{Tr}(A^{-1}B^{-1})} = \overline{\mathrm{Tr}((BA)^{-1})} = \mathrm{Tr}(BA) = \mathrm{Tr}(AB).$$

Similarly $\mathrm{Tr}(A'(B')^{-1}) = \mathrm{Tr}(AB^{-1})$. Given that F is at most 2 to 1 in a neighborhood of x_0, we can prove this result simply by showing that the point in $\mathcal{R}$ corresponding to (t, u, v) is not a fixed point of our involution when $v \neq 0$. Our involution is independent of the choice of ψ, and we can choose ψ so that the corresponding map on $\mathcal{H}$ is given by $\psi(z, t) = (\overline{z}, -t)$. But then a direct calculation shows that our involution maps the point in $\mathcal{R}$ corresponding to (t, u, v) to the point corresponding to $(t, u, -v)$. □

Lemma 15.6 *There is an open dense subset $\mathcal{D}_1$ of $\mathcal{D}$ with the property that each point in $\mathcal{D}'$ has a good neighborhood.*

Proof: Let $\pi : \mathcal{D} \to \boldsymbol{R}$ be the map $(t, u, v) \to v$. Let $\mathcal{X}$ denote the set of points of the form $(t, u, 0)$. The variation used in the proof of Lemma 15.3 is tangent to $\mathcal{X}$. Therefore π is nonconstant on $\mathcal{D}$. Being analytic, π maps an open dense set onto the complement of 0. Now we apply the previous result. □

Properties (1) and (2): In light of Lemma 15.4, the set $F(\mathcal{D})$ is 2-dimensional. Hence, there are points in $\mathcal{D}$ for which neither AB nor AB^{-1} have trace 3. But then an open dense set of points has this property because F is analytic on $\mathcal{D}$. Hence, there is an open dense subset $\mathcal{D}_2$ of $\mathcal{D}$ for which AB and AB^{-1} are both $\boldsymbol{C}$-parabolic. If $x \in \mathcal{D}_3 = \mathcal{D}_1 \cap \mathcal{D}_2$, then x has a good neighborhood. But then, in a neighborhood of x, the map F is an injective and continuous map between 4-dimensional spaces. Hence, F is an open mapping in a neighborhood of x by the invariance of domain theorem.

Property (3): Let $\mathcal{U}$ be a component of $\mathcal{D}_3$. Then F maps $\mathcal{U}$ into a region of $\boldsymbol{C}^2$ consisting of pairs (z_1, z_2), where both z_1 and z_2 are traces of $\boldsymbol{C}$-parabolic elements. But the set $\mathcal{E}$ of these pairs of complex numbers is a smooth embedded disk $\mathcal{E}$. The point is that the projection $\pi(\mathcal{E})$ is an arc of an analytic curve. See [**G**, pp. 204–205].

By Sard's theorem, there is a point $x \in \mathcal{U}$, where dF_x maps the tangent space to $\mathcal{U}$ surjectively to the tangent space to $\mathcal{E}$. But then an open dense subset of $\mathcal{U}$ has this property, by analyticity. Hence, there is an open dense subset $\mathcal{D}_4 \subset \mathcal{D}_3$ such that dF_x maps the tangent space to $\mathcal{D}$ onto the tangent space to $\mathcal{E}$.

Let $\pi_2 : \boldsymbol{C}^2 \to \boldsymbol{C}$ be the projection onto the second factor, and let $x \in \mathcal{D}_4$. Given that dF_x is surjective from the tangent space of $\mathcal{D}$ to the tangent space of $\mathcal{E}$, we can certainly find $v_1 \in T_x$ such that $dF_x(v_1) = (0, z_1)$, with $z_1 \neq 0$ being tangent to $\pi_2(\mathcal{E})$. Also, we can choose a vector $v_2 \in V_x$ tangent to $\mathcal{S}$ and nearly parallel to the vector $(1, -1, 0)$ used in the proof of Lemma 15.3. If x is sufficiently close to x_0, then we will have $dF_x(v_2) = (0, z_2)$ with z_2 transverse to $\pi_2(\mathcal{E})$. But then dF_x maps the span of $\{v_1, v_2\}$ onto $\{0\} \times \boldsymbol{C}$.

This completes the proof of the Cusp Flexibility Lemma. □

15.6 THE MULTIPLICITY OF THE TRACE MAP

Here we give an elementary and self-contained proof of Lemma 15.4.

Let $\omega = \exp(2\pi i/3)$. The matrices A and B are conjugate, so there is some $M \in SU(2,1)$ such that $B = MAM^{-1}$. After conjugating in $SU(2,1)$ we can assume that

$$A = \begin{bmatrix} \omega & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad M = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \tag{15.4}$$

For the representation $x_0 \in \mathcal{R}$, we can take $M = I_0$, the complex reflection from Section 4.6.3. All the entries in this matrix are nonzero. (See the comment at the end of Section 4.6.3.) Hence, for all representations sufficiently close to x_0, all the entries in M are nonzero.

Let $\Gamma(M)$ denote the group generated by A and B. We want to study the extent to which $\mathrm{Tr}(AB)$ and $\mathrm{Tr}(AB^{-1})$ determine M. Given $u = (u_1, u_2)$, a vector of unit complex numbers, the matrix

$$G_u = \begin{bmatrix} u_1 & 0 & 0 \\ 0 & u_2 & 0 \\ 0 & 0 & \overline{u_1 u_2} \end{bmatrix}$$

commutes with A. But then the group $\Gamma(M')$ is conjugate to $\Gamma(M)$, where

$$M' = \Gamma(G_u M G_v) = \begin{bmatrix} a_{11}u_1v_1 & a_{12}u_1v_2 & a_{13}u_1\overline{v_1v_2} \\ a_{12}u_2v_1 & a_{22}u_2v_2 & a_{23}u_2\overline{v_1v_2} \\ a_{31}v_1\overline{u_1u_2} & a_{32}v_2\overline{u_1u_2} & a_{33}\overline{u_1u_2v_1v_2} \end{bmatrix}.$$

Looking carefully at how the u's and v's are distributed in the last equation, we see that we can choose these unit complex numbers such that

$$a_{11}, a_{12}, a_{13}, a_{21} \in \boldsymbol{R}^+. \tag{15.5}$$

It is convenient to define the *auxiliary cubics*

$$X_0 = a_{11}a_{22}a_{33}, \qquad X_2 = a_{13}a_{21}a_{32}, \qquad X_4 = a_{12}a_{23}a_{31},$$

$$X_1 = a_{11}a_{23}a_{32}, \qquad X_3 = a_{12}a_{21}a_{33}, \qquad X_5 = a_{13}a_{22}a_{31}. \tag{15.6}$$

Note that

$$\det(M) = \sum_{i=0}^{5} (-1)^i X_i = 1. \tag{15.7}$$

Let $\lambda = \exp(\pi i/3)$ be the usual 6th root of unity. We compute that

$$\frac{1}{3}\mathrm{Tr}(AB) = \lambda^1 X_1 + \lambda^3 X_3 + \lambda^5 X_5, \quad \frac{1}{3}\overline{\mathrm{Tr}(B^{-1}A^{-1})} = \lambda^1 \overline{X}_1 + \lambda^3 \overline{X}_3 + \lambda^5 \overline{X}_5, \tag{15.8}$$

$$\frac{1}{3}\mathrm{Tr}(AB^{-1}) = \lambda^0 X_0 + \lambda^2 X_2 + \lambda^4 X_4, \quad \frac{1}{3}\overline{\mathrm{Tr}(BA^{-1})} = \lambda^0 X_0 + \lambda^2 \overline{X}_2 + \lambda^4 \overline{X}_4. \tag{15.9}$$

For any matrix $X \in SU(2,1)$, the traces of X and X^{-1} are conjugate. Thus the two expressions in Equation 15.8 are actually equal. Subtracting one from the other and using the fact that $\omega = \lambda^2$ we find that

$$\mathrm{Im}(X_1) + \omega \mathrm{Im}(X_3) + \omega^2 \mathrm{Im}(X_5) = 0.$$

But this is only possible if

$$\mathrm{Im}(X_1) = \mathrm{Im}(X_3) = \mathrm{Im}(X_5). \tag{15.10}$$

Using Equation 15.9 in the same way, we have

$$\mathrm{Im}(X_0) = \mathrm{Im}(X_2) = \mathrm{Im}(X_4). \tag{15.11}$$

Averaging the two expressions in Equation 15.8 we find that

$$\frac{\lambda^{-1}}{3}\mathrm{Tr}(AB) = \mathrm{Re}(X_1) + \omega \mathrm{Re}(X_3) + \omega^2 \mathrm{Re}(X_5). \tag{15.12}$$

Likewise,

$$\frac{1}{3}\mathrm{Tr}(AB^{-1}) = \mathrm{Re}(X_0) + \omega \mathrm{Re}(X_2) + \omega^2 \mathrm{Re}(X_4). \tag{15.13}$$

Equation 15.12 determines $\mathrm{Re}(X_i)$ up to a global constant for $i = 1, 3, 5$. Likewise, Equation 15.13 determines $\mathrm{Re}(X_i)$ up to a potentially different global constant for $i = 0, 2, 4$. All in all, Equations 15.10-15.13 determine X_0, X_2, X_4 up to a complex constant and X_1, X_3, X_5 up to a potentially different complex constant.

Let's put this another way. Suppose $\widehat{M} \in SU(2,1)$ is such that $\Gamma(M)$ and $\Gamma(\widehat{M})$ have the same image under F. Then there are complex numbers z_1, z_2 such that $\widehat{X}_i = X_i + z_1$ for $i = 0, 2, 4$ and $\widehat{X}_i = X_i + z_2$ for $i = 1, 3, 5$. Expanding Equation 15.7 we get

$$1 = \sum_{i=0}^{5} (-1)^i \widehat{X}_i = \sum_{i=0}^{5} (-1)^i X_i + 3z_1 - 3z_2 = 1 + 3z_1 - 3z_2.$$

Hence, $z_1 = z_2 = z$. In summary,

$$\widehat{X}_i = \widehat{X}_i + z, \qquad i = 0, 1, 2, 3, 4, 5. \tag{15.14}$$

To determine z we use the fact that $X_0X_2X_4 = X_1X_3X_5$, and likewise for the $\widehat{X}_i$. Expanding, we get

$$0 = \widehat{X}_1\widehat{X}_3\widehat{X}_5 - \widehat{X}_0\widehat{X}_2\widehat{X}_4$$

$$= (X_1 - z)(X_3 - z)(X_5 - z) - (X_0 - z)(X_2 - z)(X_4 - z) = Cz - z^2,$$

where

$$C = X_1X_3 + X_3X_5 + X_5X_1 - X_0X_2 - X_2X_4 - X_4X_0.$$

The only values of z that make this last equation true are $z = 0$ and $z = C$. Hence, there are at most two choices of $\{X_i\}$ that lead to the same F value. To finish the proof of Lemma 15.4, we just have to see that the values of X_i and the constraints in Equation 15.5 determine M.

We have normalized so that a_{ij} and $\widehat{a}_{ij}$ are real multiples of each other for $(i, j) = (1, 1), (1, 2), (1, 3), (2, 1)$. The equality of our various cubics forces this to be true for all indices. For instance, from $a_{11}a_{22}a_{33} = \widehat{a}_{11}\widehat{a}_{22}\widehat{a}_{33}$ we see that a_{33} and $\widehat{a}_{33}$ are positive real multiples. In short, $\widehat{a}_{ij} = r_{ij}a_{ij}$ for positive r_{ij}. The equations $r_{11}r_{22}r_{33} = 1 = r_{12}r_{21}r_{33}$ lead to $r_{21}/r_{11} = r_{22}/r_{12}$. Using the other equations in a similar way, we see that there are $\lambda_2, \lambda_3 \in \boldsymbol{R}^+$ such that

$$(r_{k1}, r_{k2}, r_{k3}) = \lambda_k(r_{11}, r_{12}, r_{13}), \qquad k = 2, 3. \tag{15.15}$$

We pick $\zeta \in \boldsymbol{R}$, and for $k = 2, 3$, we define vectors $s_k = (s_{k1}, s_{k2}, s_{k3})$, where

$$s_{kj} = \cos(\zeta)\mathrm{Re}(a_{1j}\overline{a}_{kj}) + \sin(\zeta)\mathrm{Im}(a_{1j}\overline{a}_{kj}). \tag{15.16}$$

Since M is nonsingular, almost every choice of ζ leads to vectors s_2 and s_3 which are linearly independent. We make such a choice. Since $M, \widehat{M} \in SU(2, 1)$, we have the equations

$$s_k \cdot (1, 1, -1) = 0, \qquad \lambda_k s_k \cdot (r_{11}^2, r_{12}^2, -r_{13}^2) = 0, \qquad k = 2, 3. \tag{15.17}$$

Since s_2 and s_3 are linearly independent, $(1, 1, -1)$ and $(r_{11}^2, r_{12}^2, -r_{13}^2)$ are multiples of the ordinary cross product $s_2 \times s_3$ and, hence, multiples of each other. Since the r_{ij} are all positive, we get that $r_{11} = r_{12} = r_{13}$.

We repeat the same argument as above, using the transpose of M in place of M, with the result that $r_{11} = r_{21} = r_{31}$. From here it is easy to see that $r_{ij} = 1$ for all i, j. Hence, $M = \widehat{M}$.

Chapter Sixteen

CR Surgery on the Whitehead Link Complement

16.1 TRACE NEIGHBORHOODS

The goal of this chapter is to prove Theorem 1.5.

We fix a left-invariant Riemannian metric on $SU(2,1)$ just for the convenience of being able to give names to open sets. The choice of metric doesn't really matter. Let $P \in SU(2,1)$ be an element, and let $B_\epsilon(P)$ be the ϵ-ball about P. We say that a *trace neighborhood* of P is a subset $Y \subset SU(2,1)$ such that $P \in Y$, and for any $\epsilon > 0$ the set

$$X_\epsilon := \{\mathrm{Tr}(h) |\ h \in Y \cap B_\epsilon(P)\} \tag{16.1}$$

contains an open neighborhood of $\mathrm{Tr}(P)$ in $\boldsymbol{C}$.

Here is an obvious consequence of this definition.

Lemma 16.1 *Let Y be a trace neighborhood of P, and suppose that $\{\tau_n\}$ is a sequence of complex numbers converging to $\mathrm{Tr}(P)$. Then for n sufficiently large there is some $P_n \in Y$ whose trace is τ_n.*

Proof: For n sufficiently large we can find some ϵ_n so that the set X_{ϵ_n} contains τ_n. By making the sequence decay slowly enough, we can arrange that $\epsilon_n \to 0$. But then we can find some element P_n, within ϵ_n of P, whose trace is τ_n. □

Say that the twist u of a $\boldsymbol{C}$-parabolic P element is *irrational* if

$$u = \exp(2\pi i\theta), \qquad \theta \in \boldsymbol{R} - \boldsymbol{Q}. \tag{16.2}$$

Geometrically, this means that every power of P has a nontrivial twist.

Lemma 16.2 *Let P be a $\boldsymbol{C}$-parabolic element with an irrational twist u. Let $\{(m_n, k_n)\}$ be a sequence of relatively prime pairs such that $|k_n/m_n| < 1/2$ and $\exp(2\pi i k_n/m_n) \to u$. Then there is a sequence $\{P'_n\}$ of lens-elliptic elements of type (m_n, k_n) converging to P.*

Proof: Note that $\lim m_n \to \infty$ because k_n/m_n converges to an irrational number. Let Π be the complex slice stabilized by P. We can choose P'_n so that P'_n stabilizes Π and $P'_n|_\Pi$ converges geometrically to $P|_\Pi$. We then adjust P'_n so that P'_n acts in the normal direction as multiplication by $u_n = \exp(2\pi i k_n/m_n)$. If $|u_n - u|$ is small, then P and P'_n have nearly

the same action both on Π and $N(\Pi)$. But then P'_n and P are close in $SU(2,1)$. $\square$

Corollary 16.3 *Suppose that P is a $\boldsymbol{C}$-parabolic element with an irrational twist u. Let Y be a trace neighborhood of P. Let $\{(m_n, k_n)\}$ be a sequence of relatively prime pairs with the property that $|k_n/m_n| < 1/2$ and $\exp(2\pi i k_n/m_n) \to u$. Then for n large there is some lens-elliptic element $P_n \in Y$ of type (m_n, k_n), and the sequence P_n converges geometrically to P.*

Proof: From Lemma 16.2 we can find a sequence $\{P'_n\}$ of lens-elliptic elements converging to P such that P'_n has type (m_k, n_k). We don't know that $P'_n \subset Y$. Let τ_n be the trace of P'_n. Since $P'_n \to P$, the traces τ_n converge to the trace of τ. By Lemma 16.1 we can find for large n a parabolic element $P_n \in Y$ such that $\mathrm{Tr}(P_n) = \tau_n$. Moreover, $P_n \to P$ algebraically. By definition P_n is weak lens-elliptic. The first statement of the Convergence Lemma II now says that $P_n \to P$ geometrically. The second statement of the Convergence Lemma II says that P_n and P'_n are conjugate once n is large. Hence, P_n is lens-elliptic for n large. $\square$

16.2 APPLYING THE HST

Let $\mathcal{R}$ be the space of representations of $\boldsymbol{Z}_3 * \boldsymbol{Z}_3$ considered in the previous chapter. If $\sigma \in \mathcal{R}$, then we let A_σ and B_σ be the two generators of $\Gamma_\sigma = \sigma(\boldsymbol{Z}_3 * \boldsymbol{Z}_3)$. We think of $\boldsymbol{\Gamma_3}$ as being a particular representation $\rho \in \mathcal{R}$. By the Cusp Flexibility Lemma we can find arbitrarily nearby representations σ such that $A_\sigma B_\sigma$ and $A_\sigma B_\sigma^{-1}$ are both $\boldsymbol{C}$-parabolic with irrational twist and the trace map F from Equation 15.1 is an open mapping in a neighborhood of σ. In the terminology above we have the following.

- The set $\{A_\tau B_\tau |\ \tau \in \mathcal{R}\}$ is a trace neighborhood of $A_\sigma B_\sigma$.
- The set $\{A_\tau B_\tau^{-1} |\ \tau \in \mathcal{R}\}$ is a trace neighborhood of $A_\sigma B_\sigma^{-1}$.

Now we use the Convergence Lemma I to see that the parabolic subgroups $\langle A_\sigma B_\sigma \rangle$ and $\langle A_\sigma B_\sigma^{-1} \rangle$ converge geometrically to the parabolic subgroups $\langle A_\rho B_\rho \rangle$ and $\langle A_\rho B_\rho^{-1} \rangle$ as $\sigma \to \rho$. By Theorem 1.4 the group $\boldsymbol{\Gamma_3}$ is a horotube group. As long as σ is close enough to ρ, then the HST tells us that σ is a horotube representation of $G_3/\ker(\sigma)$ and M_σ is homeomorphic to $\mathbf{Wh}$.

Lemma 16.4 $\ker(\sigma)$ *is trivial.*

Proof: This result is an immediate consequence of Theorem 1.11. Here we give an alternate proof for the sake of exposition. From our description of the space $\mathcal{D}$ in the previous chapter, we can find a continuous 1-parameter

family of representations $\{\rho_t\} \in \mathcal{D}$ with $\rho_0 = \rho$ and $\rho_1 = \sigma$. The HST applies to all these groups, and so they are all discrete. If $g \in \mathrm{Ker}(\sigma)$, then there are parameters $t \in (0,1)$, where $\rho_t(g)$ is arbitrarily close to the identity. But this contradicts the Margulis lemma. □

Now we know that σ is a horotube representation of G_3. Let α_1 and α_2 be the twists of $A_\sigma B_\sigma$ and $A_\sigma B_\sigma^{-1}$, respectively. For any $\epsilon > 0$ let $C(\epsilon) \subset \boldsymbol{Z}^4$ denote those 4-tuples (m_1, k_1, m_2, k_2) such that

- $|k_j/m_j| < 1/2$ and $(m_j, k_j) = 1$ for $j = 1, 2$.
- $\min(m_1, k_1, m_2, k_2) > 1/\epsilon$ and $|\alpha_j - \exp(2\pi i k_j/m_j)| < \epsilon$ for $j = 1, 2$.

If ϵ is sufficiently small and $(m_1, k_1, m_2, k_2) \in C(\epsilon)$, then by Corollary 16.3 we can find some $\tau \in \mathcal{R}$ such that $A_\tau B_\tau$ is a lens-elliptic element of type (m_1, k_1) and $A_\tau B_\tau^{-1}$ is a lens-elliptic element of type (m_2, k_2) *and* the representation τ is close enough to satisfy the hypotheses of the HST with respect to the horotube representation σ.

By the HST, the representation τ is discrete and $M_\tau = \Omega_\tau/\tau(G_3)$ is the (m_1, k_1, m_2, k_2) filling of $M_\sigma = \Omega_\sigma/\sigma(G_3)$. These coordinates are taken with respect to the homology bases given in the statement of the HST. Hence, the (m_1, k_1, m_2, k_2) filling of **Wh** admits a complete spherical CR structure, namely, $\Omega_\tau/\tau(G_3)$. By construction of $C(\epsilon)$, there is some open convex cone C such that $C(\epsilon)$ contains all but finitely many points of $C \cap \mathcal{D}(\mathbf{Wh})$. This completes the proof of Theorem 1.5. □

Chapter Seventeen

Covers of the Whitehead Link Complement

17.1 POLYGONS AND ALTERNATING PATHS

In this chapter we will prove Theorem 1.7. Our first goal is to construct a finite index subgroup of $\mathbf{\Gamma_3}$ based on a finite trivalent tree Y. There is a tiling $\mathcal{T}$ of $\boldsymbol{H}^2$ by ideal triangles, that is invariant under the action of the group G_3 considered in the previous chapter. Indeed, if τ is a single ideal triangle, then the tiling is created by considering the orbit of τ under the ideal triangle group generated by reflections in the sides of τ.

Given any ideal polygon I in $\boldsymbol{H}^2$ that is tiled by a finite number of ideal triangles from our tiling, we have the *dual tree* Y_I of I. The tree has one vertex per ideal triangle in the triangulation, and two vertices are joined by an edge iff the corresponding triangles share an edge. Conversely, if we start with a tree, then we can build the polygon for which the given tree is the dual tree. In this way we let $I = I_Y$ be the ideal polygon associated to our tree Y. We always take Y to have an odd number of vertices, so that I has an odd number n of ideal triangles in its triangulation. We also assume that at least one terminal vertex of Y is incident to a vertex of valence 2.

Let G_I' be the group generated by reflections in the sides of I. Let G_I be the even subgroup. The quotient $\widetilde{I} = \boldsymbol{H}^2/G_I$ is the double of I. The cusps of G_I are bijective with the vertices of I, and there are $n+2$ such vertices. Corresponding to G_I there is a subgroup Γ_I of finite index in $\mathbf{\Gamma}$, the golden triangle group. It turns out that $\Omega/\mathbf{\Gamma}_I$ has one additional cusp, corresponding to the $\boldsymbol{C}$-parabolic cusps of the Whitehead link complement. As preparation for identifying this cusp, we consider certain special paths in our polygon and its double.

We have a tower of maps

$$\widetilde{I} \to I \to \tau. \tag{17.1}$$

We say that a closed geodesic γ in $\widetilde{I}$ is *primitive* if it does not trace around a shorter closed geodesic with the same image. Likewise, we say that a closed curve in I is *primitive* if it doesn't trace around a shorter closed curve with the same image. There is a canonical closed path in τ, namely, the geodesic triangle that connects the midpoints of τ. This path is shown schematically in Figure 17.1. Note that this path is a billiard path in the classical sense. It bounces off the sides of τ according to the law of geometric optics; the angle of incidence equals the angle of reflection.

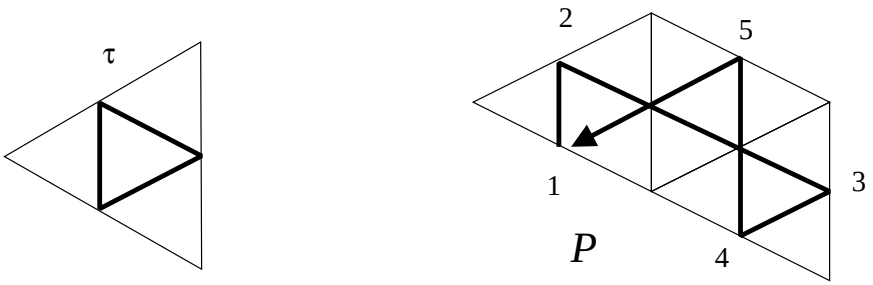

Figure 17.1: Examples of alternating paths

We say that an alternating path in I is one that covers the triangular path in τ and forms a billiard path on I. Figure 17.1 shows an example. The numbers in Figure 17.1 indicate the order in which the path visits the sides of I.

Lemma 17.1 *I has a unique primitive alternating billiard path.*

Proof: This result is proved by induction. Certainly the result is true for τ. For the induction step, suppose we create I from I' by adding a triangle T to one of the edges e of I'. To get the primitive alternating path on I, we follow the primitive alternating path on I' until we reach e. We make a detour and bounce off the two far edges of T, then we return to our original route. This is the induction step. □

We say that a primitive alternating geodesic on $\widetilde{I}$ is a closed geodesic that covers the primitive alternating billiard path γ on I. We create such a path by lifting γ to $\widetilde{I}$. Those readers familiar with billiards will recognize this construction as *doubling* a billiard path. At any rate, γ has one or two lifts, depending on the parity of the number of triangles in I. In the even case, γ has two lifts, both having the same length as γ. In the odd case, the case of interest to us, γ has a unique lift, and this lift has twice the length of γ. Henceforth, we assume that I has an odd number of triangles, and we call the unique primitive lift of γ the *canonical alternating path* on $\widetilde{I}$.

17.2 IDENTIFYING THE CUSPS

Let G_I be the group constructed in the previous section, from the polygon I. As mentioned above, we assume that I is odd. Let Γ_I be the image of G_I under the representation ρ from Theorem 1.4. Then Γ_I is a subgroup of finite index in $\mathbf{\Gamma}$ and, hence, is a horotube group. To apply the HST to Γ_I we first need to identify its parabolic subgroups. Since Γ_I is a subgroup of $\mathbf{\Gamma}$, we know from Theorem 1.4 that every parabolic element of Γ_I is either conjugate to $(I_iI_j)^k$ or else $(I_0I_1I_2)^{2k}$. In the former case the element is $\boldsymbol{R}$-parabolic, and in the latter case the element is $\boldsymbol{C}$-parabolic.

Lemma 17.2 *Any two maximal $\boldsymbol{C}$-parabolic subgroups of Γ_I are conjugate in Γ_I. Hence, Ω/Γ_I has one $\boldsymbol{C}$-parabolic horocusp.*

Proof: A loxodromic element g of G_I stabilizes a geodesic, and this geodesic projects to a closed geodesic $[g]$ in $\widetilde{I}$. Let's call g_1 and g_2 *weakly conjugate* iff $[g_1] = [g_2]$. Let $\lambda(g)$ denote the hyperbolic translation length of g along its geodesic axis.

We say that two $\boldsymbol{C}$-parabolic elements γ_1 and γ_2 are *weakly conjugate* if the corresponding elements $g_1, g_2 \in G_I$ are weakly conjugate. In this case we set $\lambda(\gamma_j) := \lambda(g_j)$. This definition makes sense because g_1 and g_2 are loxodromic in G_I. Note that g_1 and g_2 are powers of a common element of G'. Hence, $[g_1] = [g_2]$. Indeed, both these geodesics must be multiples of the canonical alternating path on $\widetilde{I}$. One can easily check this statement for the group G, and then the general case is a tautology based on the fact that $\widetilde{I}$ covers $\boldsymbol{H}^2/G$ in a way that maps the canonical alternating path on $\widetilde{I}$ to the canonical alternating path on $\boldsymbol{H}^2/G$.

Let $\gamma_1 \in \Gamma_I$ be a $\boldsymbol{C}$-parabolic element that minimizes λ amongst all $\boldsymbol{C}$-parabolic elements. Let Γ_1 be the cyclic group generated by γ_1. Let G_1 be the corresponding subgroup of G_I, generated by g_1. We claim that Γ_1 is a maximal $\boldsymbol{C}$-parabolic subgroup of Γ_I.

Choose some other $\boldsymbol{C}$-parabolic element $\mu \in \Gamma_I$. Let $m \in G_I$ be the corresponding element. Since the geodesic stabilized by m projects to a multiple of the canonical alternating geodesic, we see that there is some element $u \in G_I$ such that umu^{-1} stabilizes the same geodesic stabilized by G_1. If $umu^{-1} \notin G_1$, then some element of the larger subgroup $\langle G_1, umu^{-1}\rangle$ has shorter translation length than does the generator g_1 of G_1. But then γ_1 does not minimize λ amongst all $\boldsymbol{C}$-parabolic elements. This is a contradiction. Hence, μ is Γ_I-conjugate to an element of Γ_1. Hence, any maximal $\boldsymbol{C}$-parabolic subgroup in Γ_I is conjugate in Γ_I to Γ_1. □

17.3 TRACEFUL ELEMENTS

Here we gather some more information about traces of elements in $SU(2,1)$.

We will denote the Lie algebra of $SU(2,1)$ by L. Here L is an 8-dimensional real vector space consisting of matrices of the form

$$\frac{dg(t)}{dt}|_I := \lim_{t\to 0} \frac{g(t) - I}{t}, \tag{17.2}$$

where $g(t) \in SU(2,1)$ is a 1-parameter family of matrices converging to the identity matrix I. We will denote elements of L by λ_1, λ_2, etc.

Here are two examples. If

$$g_1(t) = \begin{bmatrix} e^{it} & 0 & 0 \\ 0 & e^{it} & 0 \\ 0 & 0 & e^{-2it} \end{bmatrix}, \qquad g_2(t) = \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{17.3}$$

then the corresponding derivatives in λ are

$$\lambda_1 = \begin{bmatrix} i & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -2i \end{bmatrix}, \qquad \lambda_2 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{17.4}$$

For any $h \in SU(2,1)$ and any λ, we have the *adjoint action*. The matrix $h\lambda h^{-1}$ lies in λ as well. In particular, we can take $h = I_0 I_2$, one of the matrices from Equation 4.13, and then set

$$\lambda_3 = h\lambda_1 h^{-1}, \qquad \lambda_4 = h\lambda_2 h^{-1}. \tag{17.5}$$

Given any $A \in SU(2,1)$, we have a linear map

$$f(\lambda) = \mathrm{Tr}(\lambda A). \tag{17.6}$$

We say that A is *traceful* if f is onto $\boldsymbol{C}$. This is an extremely mild condition. It is easy to see that this is a conjugacy invariant notion. If A is traceful, then so is hAh^{-1}. We check explicitly that the $\boldsymbol{R}$-parabolic element $I_0 I_1$ of the golden triangle group is traceful. Hence, all $\boldsymbol{R}$-parabolic elements are traceful, as well as all sufficiently nearby parabolic elements.

More generally, we say that a pair of elements $(A, B) \in SU(2,1)^2$ is a *traceful pair* if that map $f(\lambda) = (\mathrm{Tr}(\lambda A), \mathrm{Tr}(\lambda B))$ is onto $\boldsymbol{C}^2$. Again this is a very mild condition, true on an open dense set of pairs.

Lemma 17.3 *Let I_0, I_1, I_2 be the generators of the golden triangle group. Let $P_i = I_{i-1} I_{i+1}$, with indices taken mod 3. For any pair of indices $i \neq j$ and any exponents $a = \pm 1$ and $b = \pm 2$, the pair (P_i^a, P_j^b) is traceful.*

Proof: We compute explicitly in all cases that f maps the 4 Lie algebra vectors λ_1,...,λ_4 onto $\boldsymbol{C}^2$, when f is defined relative to any pair (P_i, P_j). We omit the calculation because practically any random choice of 4 Lie algebra vectors would give the same result. □

17.4 TAKING ROOTS

Let k be some positive integer. Here we gather some information about taking kth roots of elements in $SU(2,1)$. The following result is certainly not the most general one possible, but it is what we need.

Lemma 17.4 *Let P be a $\boldsymbol{C}$-parabolic element, and let U be a neighborhood of P in $SU(2,1)$. Suppose that $P^{1/k}$ is some kth root of P. If U is sufficiently small, then there is a continuous way to choose a kth root P' for all $P' \in U$.*

Proof: P stabilizes some fixed slice Π and rotates through some angle θ in the direction normal to Π. The element $P^{1/k}$ stabilizes Π and rotates the normal direction by $\theta/k + 2\pi n/k$ for some integer n. Let $P' \in U$ be some

other element. Suppose we know that P' stabilizes some $\boldsymbol{C}$-slice Π' close to Π. Then any choice of $(P')^{1/k}$ would stabilize Π' and rotate the normal direction by $\theta'/k + 2\pi n'/k$, where θ' and θ are close. The continuous choice comes from taking $n' = n$.

Now we explain why P' must stabilize a $\boldsymbol{C}$-slice Π' that is close to Π. When P' is loxodromic or parabolic, we know that P' stabilizes a unique $\boldsymbol{C}$-slice Π' and rotates the normal direction by an amount that is uniformly bounded away from 0. The same argument as in Section 14.4.1 says that Π' cannot exit every compact subset of $\boldsymbol{CH}^2$. But then Π' must converge to Π as $P' \to P$.

Suppose now that P' is elliptic. The fixed point x' of P' tends to the fixed point of P. The same argument as in Section 14.4.1 shows that there are some tangent vectors based at x' that are moved through a very small angle by P'. But this implies that P' must stabilize some $\boldsymbol{C}$-slice Π' through x' and rotate it by a vanishingly small amount. But then the amount that P' rotates normal to Π' is bounded away from 0, and the argument in Section 14.4.1 applies in the same way as in the loxodromic and parabolic cases. □

Suppose now that P is a $\boldsymbol{C}$-parabolic element and we have a continuous family $\{P_\tau\}$ of elements, which includes P. By the previous result we can define the kth root $P_\tau^{1/k}$, provided that all elements of P_τ are sufficiently close to P.

Lemma 17.5 *If the set $\{P_\tau^{1/k}\}$ forms a trace neighborhood of $P^{1/k}$, then the set $\{P_\tau\}$ forms a trace neighborhood of P.*

Proof: The important point here is that the trace of A^k is determined by the trace of A. In fact, the real and imaginary parts of the trace of A^k are polynomials in the real and imaginary parts of the trace of A. If $\widehat{P}$ is any element of $SU(2,1)$ sufficiently close to P (but not necessarily in our family), then we can choose $\widehat{P}^{1/k}$ close to P. If $\widehat{P}$ is close enough to P, then $\mathrm{Tr}(\widehat{P}^{1/k})$ will agree with $\mathrm{Tr}(P_\tau^{1/k})$ for some τ. But then $\mathrm{Tr}(\widehat{P})$ agrees with $\mathrm{Tr}(P_\tau)$. Since $\widehat{P}$ was arbitrary, we see that the set $\{P_\tau\}$ forms a trace neighborhood of P. □

17.5 APPLYING THE HST

From what we have seen above, the group Γ_I has $m = n + 3$ equivalence classes of maximal parabolic subgroups. Let $\mathcal{S}$ denote the set of representations of G_I. The group Γ_I is the image of G_I under a horotube representation $\rho \in \mathcal{R}$. Let $P_{1,\rho}, ..., P_{m,\rho}$ be a list of equivalence classes of maximal parabolic subgroups of $\rho(G_I)$. We let $P_{1,\rho}$ be the $\boldsymbol{C}$-parabolic element.

Lemma 17.6 *After relabeling our elements, we can arrange that the pair $(P_{2,\rho}, P_{3,\rho})$ is traceful.*

Proof: This result uses the side condition on our tree. Our tree has at least one univalent vertex, which is incident to a vertex of valence 2.

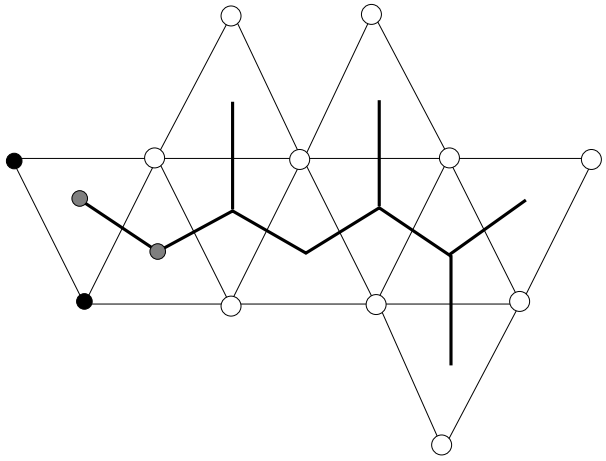

Figure 17.2: The special part of the polygon.

There is some pair of vertices on our polygon I such that the parabolic elements fixing these elements are simultaneously conjugate to a pair from Lemma 17.3. Figure 17.2 shows a representative example of the general situation. □

There is some integer k so that $P_{1,\rho}$ is conjugate in $SU(2,1)$ to the kth power of the element $A_\rho B_\rho$ considered in the previous two chapters. We might as well conjugate so that some kth root of $P_{1,\rho}$ is exactly $A_\rho B_\rho$. As long as $P_{1,\sigma}$ is sufficiently close to $P_{1,\rho}$, we can take a unique kth root $P_{1,\sigma}^{1/k}$ of $P_{1,\sigma}$, which is close to $A_\rho B_\rho$. This follows from Lemma 17.4.

We define the *trace map*

$$F(\sigma) = \big(\mathrm{Tr}(P_{1,\sigma}^{1/k}), ..., \mathrm{Tr}(P_{m,\sigma})\big), \tag{17.7}$$

which maps $\sigma \in \mathcal{S}$ to $\boldsymbol{C}^m$. (We use $P_{1,\sigma}^{1/k}$ so as to fit our situation into the Cusp Flexibility Lemma.)

We claim that there is a representation σ arbitrarily close to our original representation ρ in which dF is nonsingular and all the parabolic elements are $\boldsymbol{C}$-parabolic with irrational twist. Once we have this situation, the end of the proof here is identical to what we did in the previous chapter. Lemma 17.5 tells us that our trace neighborhood of $P_{1,\sigma}^{1/k}$ gives us a trace neighborhood of $P_{1,\sigma}$.

Let $\mathcal{R}$ be the representation space considered in the previous chapter. Given any element of $\mathcal{R}$, we can simply pass to finite index to get an element in $\mathcal{S}$. Thus we think of $\mathcal{R}$ as being a subset of $\mathcal{S}$. From the Cusp Flexibility Lemma, we can find a representation σ, arbitrarily close to ρ, for which all

the parabolics are $\boldsymbol{C}$-parabolic with irrational twist, and the image of the trace map dF_σ contains any vector of the form $(z_1, 0,...,0)$ with $z_1 \in \boldsymbol{C}$.

We have a projection $\pi : \boldsymbol{C}^m \to \boldsymbol{C}^{m-1}$ obtained by forgetting about the first coordinate. In light of what we have just said above, it suffices to show that $\pi \circ dF_\sigma$ is onto $\boldsymbol{C}^{m-1}$. We know that the pair $(P_{2,\rho}, P_{2,\rho})$ is traceful, and so the same result holds for σ, provided that it is sufficiently close to ρ.

Note that the group G_I, being the fundamental group of a punctured sphere, is freely generated by the parabolic elements corresponding to all but one of the punctures. The same argument as in the proof of Theorem 1.5 shows that σ is injective, provided that it is close enough to ρ. Hence, our group $\sigma(G_P)$ is freely generated by the parabolic elements $P_{3,\sigma},...,P_{m,\sigma}$, with

$$P_{2,\sigma} = P_{3,\sigma} \cdots P_{m,\sigma} \tag{17.8}$$

the product. We can choose a representation nearby σ simply by replacing these elements with $g_3 P_{3,\sigma},...,g_m P_{m,\sigma}$, where $g_3,...,g_m$ are arbitrarily chosen small elements in $SU(2,1)$.

Let λ be a Lie algebra element that generates some element g. If we set $g_3 = g$ and make the other g_j the identity, then $P_{j,\sigma}$ is replaced by $gP_{j,\sigma}$ for $j = 2, 3$, and otherwise $P_{j,\sigma}$ is untouched. Since the trace is a linear map, we have

$$\pi \circ df(\lambda, 0, \cdots, 0) = \big(\mathrm{Tr}\big(\lambda P_{2,\sigma}\big), \mathrm{Tr}(\lambda P_{3,\sigma}), 0, \cdots, 0\big). \tag{17.9}$$

But we know that the pair $(P_{2,\sigma}, P_{3,\sigma})$ is traceful. Therefore, the image of $\pi \circ dF$ contains any vector of the form $(z_2, z_3, 0,...,0)$. If we just vary $P_{4,\sigma}$ and leave $P_{j,\sigma}$ untouched for $j = 2, 4,...,m$, then we can produce any vector of the form $(*, 0, z_4, 0,...,0)$, where we have no control over the number in the first position. However, we can kill the first entry by subtracting the vector $(*, 0,...,0)$, which lies in our image. Hence, the vector $(0, 0, 0, z_4, 0,...,0)$ lies in the image of $\pi \circ dF$. The same goes for the remaining coordinate vectors like this. Hence, $\pi \circ dF$ is onto $\boldsymbol{C}^{m-1}$.

This completes our proof of Theorem 1.7. □

Chapter Eighteen

Small-Angle Triangle Groups

18.1 CHARACTERIZING THE REPRESENTATION SPACE

The goal of this chapter is to prove Theorem 1.10. We will use the notation from Chapter 4. Recall that $\diamond\text{Rep}(\zeta)$ consists of those complex hyperbolic ζ-triangle group representations such that the product $I_0I_1I_2$ is loxodromic. First, we will pin down the structure of $\diamond\text{Rep}(\zeta)$. Let $\alpha(\infty)$ denote the parameter in $\diamond\text{Rep}(\infty,\infty,\infty)$ corresponding to the golden triangle group.

Lemma 18.1 *If $|\zeta|$ is sufficiently large, then $\diamond\text{Rep}(\zeta)$ is a proper subinterval of $\text{Rep}(\zeta)$, and its critical endpoint $\alpha(\zeta)$ converges to $\alpha(\infty)$.*

Proof: If α is much larger than $\alpha(\infty)$ and $\min|\zeta|$ is large, then $I_0I_1I_2$ is loxodromic by continuity. We just have to understand what happens when $|\zeta|$ is large and α is near $\alpha(\infty)$.

Let $S \subset \boldsymbol{R}^4$ denote the set of the form $(\zeta_1^{-1},\zeta_2^{-1},\zeta_3^{-1},\alpha)$, for α close to $\alpha(\infty)$ and $|\zeta|$ large. Consider the map $f: S \to \boldsymbol{R}$ given by $\delta(I_0I_1I_2)$. Here δ is as in Equation 2.5. This map is analytic in the interior of S and extends by analytic continuation to a neighborhood of S. (Geometrically, we can just consider ultraideal triangles as well as the ζ-triangles.) Let $p_0 \in S$ be the point $(0,0,0,\alpha(\infty))$.

Looking at Equation 4.14 we see that $\partial f/\partial s$ is nonzero at p_0. The coordinate change from s to α is a diffeomorphism in a neighborhood of $\alpha(\infty)$. Hence, $\partial f/\partial\alpha$ is nonzero at p_0. By the implicit function theorem we can take S and U small enough so that $f^{-1}(0)\cap U$ is a smooth hypersurface transverse to $(0,0,0)\times\boldsymbol{R}$ and the gradient ∇f has a nonzero component in the α direction throughout U. Our result follows immediately. □

Remark: In [**P**] the explicit formula

$$\text{trace}(I_0I_1I_2) = 8r_0r_1r_2\exp(i\alpha) - \big(4(r_0^2+r_1^2+r_2^2)-3\big)$$

is derived. We could have equally well based our proof of Lemma 18.1 on this formula and Lemma 2.1. We chose the more abstract approach above because it is self-contained.

18.2 DISCRETENESS

Now we prove that every representation in $\diamond\mathrm{Rep}(\zeta)$ is discrete, provided that $|\zeta| = \min(\zeta_i)$ is sufficiently large.

Recall that each complex hyperbolic triangle group corresponds to a pair (ζ, α), where α is the polar angular invariant of the triangle on which the group is based. Suppose that Theorem 1.10 is false. Then we can find a sequence (ζ_n, α_n) such that $|\zeta_n| \to \infty$ but the corresponding group $\Gamma(\zeta_n, \alpha_n)$ is indiscrete. It follows from Lemma 18.1 that we can pass to a subsequence so that α_n converges to some α in the closure of $\diamond\mathrm{Rep}(\infty, \infty, \infty)$. Thus, we can choose our triangle $T(\zeta_n, \alpha_n)$ so that it converges to the ideal triangle corresponding to the limit point in the closure of $\diamond\mathrm{Rep}(\infty, \infty, \infty)$. Let Γ be the corresponding complex hyperbolic ideal triangle group.

Let G be the even subgroup of the real hyperbolic ideal triangle group. By either Theorem 1.3 or 1.4, Γ is the image of a horotube representation ρ of G. We interpret $\Gamma(\zeta_n, \alpha_n)$ as the image of an (unfaithful) representation $\rho_n : G \to PU(2,1)$. Evidently ρ_n converges algebraically to ρ.

If Γ is not the even subgroup of the golden triangle group, then the only peripheral subgroups in Γ are the ones generated by elements conjugate to I_iI_j. Looking at our construction of the triangle groups, we see that the element $\rho_n(I_iI_j)$ converges to $\rho(I_iI_j)$ precisely as hypothesized in the Convergence Lemma III. Hence, $\rho_n(I_iI_j)$ converges geometrically to $\rho(I_iI_j)$. By symmetry, the same result holds for any element in G that is conjugate to I_iI_j. Note also that the elliptic subgroups of $\rho_n(G)$ act freely on S^3 because they are lens-elliptic of type $(-1, \zeta_i)$ for $i = 0, 1, 2$. All in all, ρ_n converges nicely to ρ in the sense of the HST. The HST now applies, showing that ρ_n is discrete for n large. This is a contradiction.

We still have to deal with the possibility that Γ is the even subgroup of the golden triangle group. In this case, ρ is a horotube group, and the only new additional peripheral subgroups are the ones conjugate to $\langle\rho((I_0I_1I_2)^2)\rangle$. But by hypothesis the corresponding elements in $\rho_n(G)$ are parabolic. Hence, the corresponding subgroups converge geometrically by the Convergence Lemma I. Once again, $\{\rho_n\}$ converges nicely to ρ, and the HST shows that ρ_n is discrete for n large. We have the same contradiction.

Now we know that $\Gamma(\zeta, \alpha)$ is discrete if ζ is sufficiently large and $\alpha \in \diamond\mathrm{Rep}(\zeta)$.

18.3 HOROTUBE GROUP STRUCTURE

We continue the notation from the previous section and also set $\Gamma_n = \rho_n(G)$. The HST gives more information than we have used in our proof above. Since not all the cusps get filled when $\rho_n \to \rho$, we know that Γ_n is again a horotube group when n is sufficiently large. More precisely, $\Gamma_n = \widehat{\rho}_n(G_n)$, where $\widehat{\rho}_n$ is a horotube representation of the quotient group $\widehat{G}/\ker(\rho_n)$. Therefore, we

have the following corollary.

Corollary 18.2 *If $|\zeta|$ is sufficiently large and $\alpha \in \diamond\mathrm{Rep}(\zeta)$, then there is a group $G(\zeta,\alpha)$ and a horotube representation $\rho(\zeta,\alpha) : G(\zeta,\alpha) \to PU(2,1)$ such that $\Gamma(\zeta,\alpha)$ is the image of $G(\zeta,\alpha)$ under $\rho(\zeta,\alpha)$. The group $G(\zeta,\alpha)$ is the quotient of G by the kernal of $\rho \in \diamond\mathrm{Rep}(\zeta)$ corresponding to α.*

The obvious guess about these groups and representations is correct.

Lemma 18.3 *The representation $\rho \in \diamond\mathrm{Rep}(\zeta)$ is the horotube representation mentioned in Corollary 18.2, and the group $G(\zeta)$, the even subgroup of the real hyperbolic ζ-triangle group, is the abstract group mentioned in Corollary 18.2.*

Proof: Note that $G(\zeta,\alpha)$ is always a quotient of $G(\zeta)$ because our representation ρ does respect the relations of $G(\zeta)$. We just have to show that ρ does not act trivially on an element of G whose image in $G(\zeta)$ is nontrivial. Let $g \in G$ be such a candidate element. We will fix ζ and vary α. When $\alpha = \pi$ we know everything there is to know about ρ. It is just the Fuchsian representation of $G(\zeta)$ stabilizing an $\boldsymbol{R}$-slice. Hence, $\rho(g)$ is nontrivial when $\alpha = \pi$. As α decreases from π, the group $\rho(G)$ remains discrete. Hence, by the Margulis lemma, $\rho(g)$ remains uniformly bounded away from the identity element. This shows that $\rho(g)$ cannot converge to the identity element as α decreases from π to the other endpoint of $\diamond\mathrm{Rep}(\zeta)$. □

This completes the proof of Theorem 1.10. We note also that Theorem 1.11 applies to every sufficiently close pair of parameters. Then we can compose finitely many of the homeomorphisms produced by Theorem 1.11 to get a topological conjugacy, on the regular sets, for any two groups $\Gamma(\zeta,\alpha_1)$ and $\Gamma(\zeta,\alpha_2)$ for any two parameters $\alpha_1, \alpha_2 \in \diamond\mathrm{Rep}(\zeta)$. Below we will deduce Corollary 1.12 by extending our conjugacy across the limit set.

18.4 TOPOLOGICAL CONJUGACY

Let Γ be the Fuchsian representation of the group $G(\zeta)$, and let $\widehat{\Gamma}$ be some other representation corresponding to a parameter in $\diamond\mathrm{Rep}(\zeta)$. To prove Theorem 1.12 it suffices to consider the case of Γ and $\widehat{\Gamma}$.

We have a homeomorphism $h : \Omega \to \widehat{\Omega}$, which conjugates Γ to $\widehat{\Gamma}$. It remains to extend h to a homeomorphism from Λ to $\widehat{\Lambda}$ in a way that makes the result a self-homeomorphism of S^3 conjugating Γ to $\widehat{\Gamma}$. We can identify Ω with $T_1\boldsymbol{H}^2$, the unit tangent bundle of $\boldsymbol{H}^2$, as we discussed in Chapter 4. We can identify Λ with the real circle $S^1 = \boldsymbol{R}^2 \cap S^3$.

The construction we give can be based on any finite collection of elements $S = \{g_1,...,g_k\} \subset \Gamma$ such that $\widehat{\rho}(g_i)$ is loxodromic for all i. For the sake of concreteness we let S denote the set of words having the form $\iota_i\iota_j\iota_i\iota_k$, with

i, j, k distinct. Let $\mathcal{G} \subset \Gamma$ denote those elements of Γ that are conjugate to S. Let $\mathcal{A}$ denote the set of geodesics α such that α is the translation axis of some $g \in \mathcal{G}$. Note that g and g^{-1} do not both belong to $\mathcal{G}$, at least when $\min(\zeta_1, \zeta_2, \zeta_3) \geq 4$. Hence, the elements of $\mathcal{G}$ are in bijection with the elements of $\mathcal{A}$. Choose some point z, not contained in any α belonging to $\mathcal{A}$; the precise choice doesn't matter. Let $\mathcal{H}$ denote the set of half-spaces H such that H is bounded by $\alpha \in \mathcal{A}$ and disjoint from z. We set $\alpha_H = \alpha$, the geodesic bounding H. Note that $H_1 \subset H_2$ implies that α_1 and α_2 do not share a common endpoint. We call this the *strict nesting property*.

We have the fibration $\pi : T_1\boldsymbol{H}^2 \to \boldsymbol{H}^2$. For any $H \in \mathcal{H}$ we define

$$[H] = \pi^{-1}(H) \cup \{p\} \cup \{q\}. \tag{18.1}$$

Here p and q are the endpoints of the geodesic α_H. Then $[H]$ is a topological ball, stabilized by the loxodromic element $g_H \in \mathcal{G}$, which stabilizes α_H. The balls $[H]$ and $[H']$ are nested in the same way that the subspaces $[H]$ and $[H']$ are nested. The boundary of $[H]$ is $[\alpha_H] = \pi^{-1}(\alpha_H) \cup \{p\} \cup \{q\}$. We would like to define $[\widehat{H}] = \omega([H])$, but we don't yet know that ω extends continuously to S^1. This is actually what we are trying to prove. To define $[\widehat{H}]$ we will take an indirect approach that sidesteps the difficulty of extension by using the behavior of loxodromic elements.

For each $\alpha \in \mathcal{A}$ we define

$$[\widehat{\alpha}] = \text{closure}\big(\omega([\alpha] - \{p\} - \{q\})\big). \tag{18.2}$$

The closure is taken in S^3. Here p and q are the fixed points of g_α, the element of $\mathcal{G}$ stabilizing α. Let $\widehat{g}_\alpha$ be the corresponding element of $\widehat{\Gamma}$.

Lemma 18.4 *$[\widehat{\alpha}] \subset S^3$ is a $\widehat{g}_\alpha$-invariant embedded sphere, and furthermore, $[\widehat{\alpha}] \cap \widehat{\Lambda} = \{\widehat{p}\} \cup \{\widehat{q}\}$. Here $\widehat{p}$ and $\widehat{q}$ are the fixed points of $\widehat{g}_\alpha$.*

Proof: Since ω is a homeomorphism, $[\widetilde{\alpha}] = \omega([\alpha] - \{p\} - \{q\})$ is homeomorphic to a twice-punctured sphere. By definition, the closure of this twice-punctured sphere is $[\widehat{\alpha}]$. Let $\{x_n\}$ be any sequence of points that exits every compact subset of $[\widetilde{\alpha}]$. We claim that $\{x_n\}$ converges to either $\widehat{p}$ or $\widehat{q}$ on a subsequence. Given the claim, we have $[\widehat{\alpha}] = [\widetilde{\alpha}] \cup \{\widehat{p}\} \cup \{\widehat{q}\}$. Hence, $[\widehat{\alpha}]$ is an embedded sphere that intersects $\widehat{\Lambda}$ precisely in the set $\{\widehat{p}\} \cup \{\widehat{q}\}$. The invariance is immediate from the equivariance of ω.

Now for the claim. Set $g = g_\alpha$. There is a compact $K \subset T_1\boldsymbol{H}^2$, which serves as a fundamental domain for the action of $\langle g \rangle$ on $([\alpha] - \{p\} - \{q\})$. Hence, there is a compact subset $\widehat{K} \subset \widehat{\Omega}$, which serves as a fundamental domain for the action of $\langle \widehat{g} \rangle$ on $[\widetilde{\alpha}]$. This means that there is an unbounded sequence of exponents $\{k_n\}$ such that $\widehat{g}^{k_n}(x_n) \in \widehat{K}$ for all n. Passing to a subsequence we can assume that either $k_n \to \infty$ or $k_n \to -\infty$. In the former case $\widehat{g}^{-n_k}(\widehat{K}) \to \widehat{p}$, and in the latter case $\widehat{g}^{-n_k}(\widehat{K}) \to \widehat{q}$. Hence, $\{x_n\}$ converges, on a subsequence, to either $\{p\}$ or $\{q\}$. $\square$

Now let $H \in \mathcal{H}$ be some half-space. Being an embedded sphere, $[\widehat{\alpha}_H]$ separates S^3 into two components, one of which contains $\omega([H]-\{p\}-\{q\})$. We let $[\widehat{H}]$ be the closure of this component.

Lemma 18.5 *For any $\epsilon > 0$ there are only finitely many sets of the form $[\widehat{H}]$ that have spherical diameter greater than ϵ.*

Proof: This follows immediately from Lemma 2.6 applied to the sequence $\{[H_k]\}$. □

Definition of the Map: Now we define a map $\lambda : S^1 \to \widehat{\Lambda}$ as follows. From our choice of $\mathcal{H}$ we see that each x is the accumulation set of a nested sequence $\{H_n\} \in \mathcal{H}$. We define $\lambda(x) = \bigcap[\widehat{H}_n]$. This is a single point by the preceding result. From the nesting properties of elements of $\mathcal{H}$, we see that $\lambda(x)$ is independent of the chosen nesting sequence. Note that $\bigcap[\widehat{H}_n]$ is an accumulation point of $\widehat{\Lambda}$. Since $\widehat{\Lambda}$ is closed, we have $\lambda(x) \in \widehat{\Lambda}$.

Continuity: If $x, y \in S^1$ are close together, then x and y will have defining sequences that agree for a large number of terms. But then $\widehat{x}$ and $\widehat{y}$ are close together, by Lemma 18.5. This shows that λ is continuous.

Bijectivity: If $x \neq y$, then we can find disjoint H_x and H_y such that $x \in \beta_{H_x}$ and $y \in \beta_{H_y}$. Then $\widehat{x}$ and $\widehat{y}$ are contained in disjoint subsets of S^3, namely, $[\widehat{H}_x]$ and $[\widehat{H}_y]$. This shows that λ is injective. Let $\widehat{\mathcal{G}} \subset \widehat{\Gamma}$ denote the set of elements corresponding to $\mathcal{G} \subset \Gamma$. The set of fixed points of elements of $\widehat{\mathcal{G}}$ is dense in $\widehat{\Lambda}$. If x is a fixed point of an element $g \in \mathcal{G}$, then we can take our sequence $\{H_n\}$ so that $g(H_n) = H_{n+1}$ for all n. From this we see that $\lambda(x)$ is the corresponding fixed point of $\widehat{g}$. Hence, $\lambda(S^1)$ is dense in $\widehat{\Lambda}$. Since λ is continuous and S^1 is compact, $\lambda(S^1)$ is a compact subset of $\widehat{\Lambda}$. Being compact and dense, $\lambda(S^1)$ must be all of $\widehat{\Lambda}$.

Homeomorphism: The map $\sigma = \omega \cup \lambda : (T_1\boldsymbol{H}^2 \cup S^1) \to S^3$ is a bijection, separately continuous on $T_1\boldsymbol{H}^2$ and S^1, that conjugates the action of Γ to the action of $\widehat{\Gamma}$. To show that σ is continuous on $T_1\boldsymbol{H}^2 \cup S_1 = S^3$, it suffices to check the following compatibility. If $x \in S^1$ and $\{x_n\}$ is a sequence of points in $T_1\boldsymbol{H}^2$ with $x_n \to x$, then $\sigma(x_n) \to \sigma(x)$. Let $y_n = \pi(x_n) \subset \boldsymbol{H}^2$. Then $y_n \to x$. Therefore, there is a sequence of subspaces $\{H_n\}$ of $\mathcal{H}$ such that $x = \bigcap H_n$ and $y_{n'} \subset H_n$ for all n' sufficiently large. But then $x_{n'} \subset [H_n]$ and $\sigma(x_{n'}) \subset [\widehat{H}_n]$. Since $\sigma(x) = \bigcap[\widehat{H}_n]$ we see that $\sigma(x_n) \to \sigma(x)$, as desired. Hence, σ is everywhere continuous. A continuous bijection from a closed space to a Hausforff space is a homeomorphism.

PART 4
Structure of Ideal Triangle Groups

Chapter Nineteen

Some Spherical CR Geometry

This chapter' begins Part 4 of the monograph. In this part we prove Theorems 1.4 and 1.3.

19.1 PARABOLIC *R*-CONES

We introduced the *hybrid cone* construction in [**S1**] and used it in [**S0**] to define *hybrid spheres*. In [**FP**], hybrid spheres were renamed $\boldsymbol{R}$*-spheres*. To be consistent, we rename the hybrid cone the $\boldsymbol{R}$*-cone*.

For $\boldsymbol{F} \in \{\boldsymbol{C}, \boldsymbol{R}\}$ recall that an $\boldsymbol{F}$*-circle* is the accumulation set, in S^3, of an $\boldsymbol{F}$-slice. Say that an $\boldsymbol{F}$*-arc* is a connected arc of an $\boldsymbol{F}$-circle. Given a curve $C \subset S^3$ and a point $q \in S^3 - C$, we want to produce a surface by coning C to q using $\boldsymbol{R}$-arcs. Since there are infinitely many $\boldsymbol{R}$-arcs that join two points, we need some extra data to select one. Say that a *flag* is a pair (E, q), where E is a $\boldsymbol{C}$-circle that contains q.

Lemma 19.1 *Let $x \in S^3 - E$. There is a unique $\boldsymbol{R}$-circle $\gamma = \gamma(E, q; x)$ such that $x \in \gamma$, $q \in \gamma$, and $\gamma \cap (E - q) \neq \emptyset$.*

Proof: We normalize by a Heisenberg stereographic projection so that $E = E_0 = \{0\} \times \boldsymbol{R}$, the *center* of $\mathcal{H}$, and $q = \infty$ and $x = (1, 0)$. Then $(\boldsymbol{R} \times \{0\}) \cup \infty$ is the only $\boldsymbol{R}$-circle that does the job. □

Definition: We call (E, q) the *standard flag* when it is as in the previous proof.

Let $\Sigma(E, q; x)$ be the portion of γ that connects q to x but avoids $E - q$. Given a set $S \subset S^3 - E$, we define

$$\Sigma(E, q; S) = \bigcup_{x \in S} \Sigma(E, q; x). \tag{19.1}$$

We call Σ an $\boldsymbol{R}$*-cone*. Our construction is natural. The $PU(2, 1)$-image of an $\boldsymbol{R}$-cone is again an $\boldsymbol{R}$-cone.

19.2 PARABOLIC *R*-SPHERES

Let C_1 be a $\boldsymbol{C}$-circle that links E_0, and let (E_0, q_0) be a flag. We write $I_1 = I_{C_1}$, the complex reflection fixing C_1, and $(E_2, q_2) = I_1(E_0, q_0)$. We

define the ***R**-sphere*

$$\Sigma = \Sigma(E_0, q_0; C_1) \cup \Sigma(E_2, q_2; C_1). \tag{19.2}$$

In analogy with spinal spheres, we call C_1 the *equator* and (E_j, q_j) the *poles*. We call each component of $\Sigma - C_1$ a *hemisphere*. Σ is determined by its equator and poles. Σ is a spinal sphere iff E_0 and C_1 bound perpendicular $\boldsymbol{C}$-slices. In general we have the following lemma.

Lemma 19.2 *Σ is a piecewise analytic sphere. Moreover, the $\boldsymbol{C}$-circle connecting the poles of Σ only intersects Σ at the poles.*

Proof: We normalize so that (E_0, q_0) is the standard flag. Then C_1 is an ellipse that links E_0. The map $(z,t) \to (z/|z|, t)$ is injective and analytic on C_1 and is a fibration from $(\boldsymbol{C} - \{0\}) \times \boldsymbol{R}$ onto an infinite cylinder. $\Sigma(E_0, q_0; C_1)$ is obtained from C_1 by gluing on rays, which are subsets of the fibers of the fibration. From this description it is clear that $\Sigma(E_0, q_0; C_1)$ is an embedded disk and analytic away from $\{q_0\} \cup C_1$. The same argument works for $\Sigma(E_2, q_2; C_1)$, by symmetry. Let $\pi : \mathcal{H} \to \boldsymbol{C}$ be projection. By symmetry, I_1 maps the exterior of the cylinder $\Lambda = \pi^{-1}(C_1)$ into the interior. Thus $\Sigma(E_2, q_2; C_1) - C_1$ and $\Sigma(E_0, q_0; C_1) - C_1$ lie in different components of of Λ. Hence, the two equators intersect exactly along C_1.

The $\boldsymbol{C}$-circle γ connecting the poles is a vertical line that links C_1. But then $\gamma \cap \Sigma(E_0, p_0; C_1)$ is just ∞, one of the poles. But $I_1(\gamma) = \gamma$, and I_1 swaps the two poles. □

If Σ is not a spinal sphere, then there is a unique $\boldsymbol{R}$-circle contained in Σ. We call this $\boldsymbol{R}$-circle the $\boldsymbol{R}$-axis. The $\boldsymbol{R}$-axis intersects C_1 and E_0 each in two points. To see this, we normalize so that (E_0, p_0) is the standard flag and the center of C_1 lies in $\boldsymbol{R}^+ \times \{0\}$. (When Σ is a spinal sphere, the center of C_1 would lie at the origin.) Then $\boldsymbol{R} \times \{0\}$ is the $\boldsymbol{R}$-axis. Under this normalization the $\boldsymbol{R}$-reflection $(z,t) \to (\overline{z}, t)$ is a symmetry of Σ fixing the $\boldsymbol{R}$-axis. We call this the ***R**-symmetry* of Σ and sometimes denote it by J_R. Thus Σ has an $\boldsymbol{R}$-symmetry and also a $\boldsymbol{C}$-symmetry coming from inversion in C_1.

19.3 PARABOLIC ELEVATION MAPS

Here we will analyze the map from Lemma 19.2 in more detail. Given a flag (E, q), let $\boldsymbol{R}(E, q)$ denote the set of $\boldsymbol{R}$-circles that contain q and intersect E at some other point. Then $\boldsymbol{R}(E, q)$ is naturally a cylinder. We can identify $\boldsymbol{R}(E, q)$ with the set of pairs (x, v), where $x \in E - q$ is a point and v is a tangent vector based at x and contained in the contact plane at x. Alternatively, when (E, q) is the standard flag, we can identify $\boldsymbol{R}(E, q)$ with $S^1 \times \boldsymbol{R}$. The identification sends the element of $\boldsymbol{R}(E, q)$ to its point of intersection with $S^1 \times \boldsymbol{R}$.

Given any $x \in S^3 - E$, we define $\Psi(x)$ to be the element of $\boldsymbol{R}(E,q)$ that contains x. If we normalize so that (E,q) is the standard flag and identify $\boldsymbol{R}(E,q)$ with $S^1 \times \boldsymbol{R}$, then Ψ is as in Lemma 19.2:

$$\Psi(z,t) = \Big(\frac{z}{|z|}, t\Big). \tag{19.3}$$

We always identify the $\boldsymbol{R}$-factor with the vertical direction. We call Ψ the *parabolic elevation map*. This map, of course, depends on the flag (E,q). The map Ψ conjugates a parabolic element stabilizing (E,q) to an isometry of $\boldsymbol{R}(E,q)$; when it is given, the flat metric comes from its identification with $S^1 \times \boldsymbol{R}$.

Remark: At this point the reader might want to skip ahead to the next chapter to see our main construction and take a long look at Figure 20.1. The rest of this chapter consists of the technical results needed to show that Figure 20.1 really has the properties that are obvious from its appearance.

19.4 A NORMALITY CONDITION

Say that an $\boldsymbol{R}$-circle γ is *affiliated* to $\Sigma(E_j, q_j; C_1)$ if γ is the extension of an $\boldsymbol{R}$-arc of $\Sigma(E_j, q_j; C_1)$. The $\boldsymbol{R}$-axis is the unique $\boldsymbol{R}$-circle affiliated to both $\Sigma(E_0, q_0; C_1)$ and $\Sigma(E_2, q_2; C_1)$. We say that Σ is *normal* if every $\boldsymbol{R}$-circle affiliated to $\Sigma(E_0, q_0; C_1)$, other than R, links E_2. This is equivalent to the statement that every affiliate of $\Sigma(E_2, q_2; C_1)$, other than R, links E_0.

Let Σ be the $\boldsymbol{R}$-sphere from Equation 19.2. Let Ψ be the elevation map associated to (E_0, q_0). We normalize so that (E_0, q_0) is the standard flag, as above. Let R be the $\boldsymbol{R}$-axis of Σ. Let C_1^* denote the $\boldsymbol{C}$-circle obtained by rotating C_1 isometrically by 180 degrees about E_0.

Lemma 19.3 (Normality Criterion) *Σ is normal provided the following.*

(1) *E_0 and E_2 are linked.*

(2) *$\Psi(E_2) \cap \Psi(C_1)$ is two points, and $\Psi(E_2) \cap \Psi(C_1^*)$ is two points.*

(3) *Some $\boldsymbol{R}$-circle γ_0 affiliated with $\Sigma(E_2, q_2; C_1)$ links E_0.*

Proof: Our proof has four steps.

Step 1: By construction $A := \Psi(R - E_0)$ consists of a pair of "antipodal" points on $S^1 \times \boldsymbol{R}$. We claim $\Psi(E_2) \cap \Psi(C_1) = A$. By symmetry $E_2 \cap R$ consists of two points, and (since E_2 links E_0) these two points of R lie on either side of $R - E_0$. Hence, $A \subset \Psi(E_2)$. Likewise, $A \subset \Psi(C_1)$. Hence, $A \subset \Psi(E_2) \cap \Psi(C_1)$. If $\Psi(E_2) \cap \Psi(C_1)$ is exactly two points, then we must

have $A = \Psi(E_2) \cap \Psi(C_1)$. By symmetry $A \subset \Psi(C_1^*)$. Hence, the same argument as above shows that $\Psi(E_2) \cap \Psi(C_1^*) = A$.

Step 2: Suppose that $\gamma \neq R$ is an affiliate of $\Sigma(E_2, q_2; C_1)$. Then $\gamma' = I_1(\gamma)$ is an affiliate of $\Sigma(E_0, q_0; C_1)$. The image $\Psi(\gamma' - E_0)$ consists of two points, one of which lies in $\Psi(C_1)$ and one of which lies in $\Psi(C_1^*)$. The point is that $\gamma' - E_0$ is the union of two fibers of Ψ. One of these fibers intersects C_1, and the other one intersects C_1^*.

Step 3: Let $B = \Psi(\gamma' - E_0)$. We claim that $A \cap B = \emptyset$. Let $\mu : S^1 \times \boldsymbol{R} \to S^1$ denote projection. Then $\mu(A)$ is a pair of antipodal points on S^1, and $\mu(B)$ is a pair of antipodal points on S^1. Since C_1 links E_0, the map $\mu \circ \Psi$ is injective on C_1. Hence, $\mu(A)$ and $\mu(B)$ are different pairs of antipodal points. Hence, $A \cap B = \emptyset$.

Step 4: From Step 1 we get $\Psi(E_2) \cap (\Psi(C_1) \cup \Psi(C_1^*)) = A$. From Step 2 we get $\Psi(\gamma' - E_0) \subset \Psi(C_1) \cup \Psi(C_1^*)$. Hence, $\Psi(E_2) \cap \Psi(\gamma' - E_0) \subset A$. From Step 3 we know that $\Psi(\gamma' - E_0) \cap A = \emptyset$. Hence, $\Psi(E_2) \cap \Psi(\gamma' - E_0) = \emptyset$. Hence, $E_2 \cap (\gamma' - E_0) = \emptyset$. But $E_2 \cap E_0 = \emptyset$ because these circles are linked. Hence, $E_2 \cap \gamma' = \emptyset$. Hence, $E_0 \cap \gamma = \emptyset$. Now we know that the affiliates of $\Sigma(E_2, q_2; C_1)$, other than R, either all link or all fail to link. Since one links, they all do. □

19.5 USING NORMALITY

Say that a *parabolic cospinal family* is a collection of spinal spheres that all share a common pole and a common spine. We can always normalize to get the *standard parabolic cospinal family*, consisting of spinal spheres $\{\Sigma_t' |\ t \in \boldsymbol{R}\}$ in $\mathcal{H}$, where the poles of Σ_t' are $(0, t)$ and ∞. We say that an $\boldsymbol{R}$-circle links a parabolic cospinal family if it links the common spine.

Lemma 19.4 (Cospinal Intersection) *Suppose that $\mathcal{S}$ is some parabolic cospinal family. Let γ be an $\boldsymbol{R}$-circle that links $\mathcal{S}$. Then γ is tangent to exactly two spinal spheres in $\mathcal{S}$.*

Proof: We normalize so that $\mathcal{S}$ is the standard parabolic family. Let $\pi : \mathcal{H} \to \boldsymbol{C}$ be the projection map. If γ is tangent to a spinal sphere Σ in $\mathcal{S}$ at p, then γ is tangent to the $\boldsymbol{R}$-circle L that contains p. Here L is unique because p is not a pole. The point here is that both γ and L are also tangent to the contact plane through p, and this plane is transverse to Σ. Now, $\pi(\gamma)$ is a lemniscate. One the lobes of $\pi(\gamma)$ surrounds 0, and $\pi(L)$ is a line containing 0, that is tangent to $\pi(\gamma)$ at $\pi(p) \neq 0$. Given the nature of lemniscates, there are exactly 2 such lines. See Figure 19.1. □

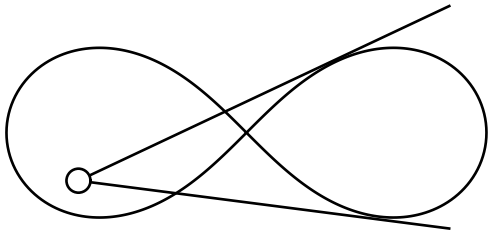

Figure 19.1: A lemniscate and 2 tangent lines

Now we examine how our elevation maps interact with $\boldsymbol{R}$-circles. Say that a curve $\alpha \in S^1 \times \boldsymbol{R}$ is a *monotone arc* if it is smooth and regular on its interior and has positive slope. Say that α is *major* if α winds more than 2π times around in the S^1 factor and otherwise is *minor*. Figure 20.2 shows a nice example of two monotone arcs, one major and one minor.

Lemma 19.5 (Elevation Image) *Let Ψ be an elevation map based on a flag (E, q). Let γ be an $\boldsymbol{R}$-circle that links E. Then $\Psi(\gamma)$ is the union of two monotone arcs, one major and one minor.*

Proof: We normalize so that (E, q) is the standard flag. By definition γ links the standard parabolic cospinal family. Let $s \to \gamma(s)$ be a smooth parametrization of γ. Let $\alpha(s) = (z(s), t(s)) = \Psi(\gamma(s))$. By the Cospinal Intersection Lemma, there are two parameters s_1 and s_2 such that $\gamma(s)$ is not tangent to any fiber of Ψ when $s \neq s_1, s_2$. The kernel of $d\Psi_x$ is precisely the tangent vector to the element of $\boldsymbol{R}(E, q)$, which contains x, at x. Thus, when $\gamma(s)$ is not tangent to one of the spinal spheres in the standard parabolic cospinal family, $\alpha'(s)$ is nonzero.

Let C_s denote the contact plane at $x = \gamma(s)$. Then $d\Psi_x(C_s)$ is a line of positive slope. This follows from the description of the contact distribution in $\mathcal{H}$ as the plane field annihilated by the 1-form $dt - xdy - ydx$. This shows that α has positive slope where it is regular. Since γ links E, the total argument of α varies by more than 2π. Hence, one of the monotone arcs is major. The other monotone arc projects to $\boldsymbol{C}$ in such a way as to be contained entirely in one of the lobes of the leminscate $\pi(\gamma_s)$, as shown in Figure 19.1. Hence, this arc is minor. □

Chapter Twenty

The Golden Triangle Group

20.1 MAIN CONSTRUCTION

The goal of this chapter is to prove everything in Theorem 1.4 except the statement about $\Omega/\mathbf{\Gamma_3}$. We will deal with this last statement in Chapter 21.

Let I_0, I_1, I_2 be the generators of $\mathbf{\Gamma}'$, the golden triangle group—see Equation 4.13. The element I_j fixes pointwise a $\boldsymbol{C}$-circle C_j. The $\boldsymbol{C}$-circles C_{i-1} and C_{i+1} intersect pairwise in a point v_i. The element $I_iI_jI_k$ fixes a point q_j and stabilizes a $\boldsymbol{C}$-circle E_j. Since $I_kI_jI_i$ and $I_iI_jI_k$ are inverses of each other, they both stabilize the flag (E_j, q_j). A routine calculation shows that C_i links $E_{j\pm1}$. Here indices are taken mod 3. Define

$$\Sigma_j = \Sigma(E_{j-1}, q_{j-1}; C_j) \cup \Sigma(E_{j+1}, q_{j+1}; C_j). \tag{20.1}$$

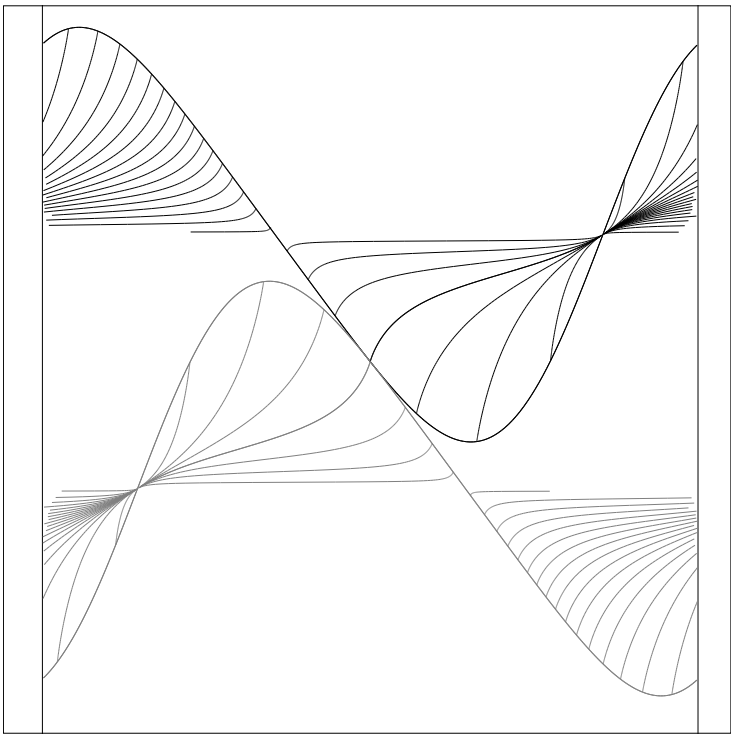

Figure 20.1: Elevation projection of Σ_1 and Σ_2

Note that Σ_j is an $\boldsymbol{R}$-sphere because $I_j(E_{j-1}, q_{j-1}) = (E_{j+1}, q_{j+1})$. Figure 20.1 shows a plot of $\Psi(\Sigma_1)$ in black and $\Psi(\Sigma_2)$ in gray. The sinusoidal-like curves are $\Psi(C_1)$ and $\Psi(C_2)$. (Incidentally, this picture indicates that C_1 and C_2 both link E_0 because $\Psi(C_1)$ and $\Psi(C_2)$ wind once around the cylinder.) The $\boldsymbol{R}$-arcs foliating $\Sigma(E_0, q_0; C_1)$ are mapped to points in Figure 20.1. Likewise, the $\boldsymbol{R}$-arcs foliating $\Sigma(E_0, q_0; C_2)$ are mapped to points. Figure 20.1 indicates that $\Psi(C_1)$ lies above $\Psi(C_2)$ and that the two curves intersect only at the one point $\Psi(v_0)$ where they must intersect.

Below we will prove the following lemma.

Lemma 20.1 (Intersection) $\Psi(C_1) \cap \Psi(C_2) = \Psi(v_0)$.

The arcs in Figure 20.1 appear to be monotone arcs, in the sense of Section 19.4. The black arcs of $\Psi(\Sigma(E_0, q_0; C_1))$ have both endpoints on $\Psi(C_1)$. One of the endpoints varies with the arc and the other endpoint does not. In [**S0**, §4.5] we compute that the unvarying endpoint lies above the highest point of $\Psi(C_2)$. This is obvious from the picture. Likewise, the unvarying endpoint of the gray arcs lies below the lowest point of $\Psi(C_1)$.

Below we will prove the following lemma.

Lemma 20.2 (Monotone) *The $\boldsymbol{R}$-arcs of $\Psi(\Sigma(E_1, q_1; C_2))$ are all monotone.*

The same result holds for the arcs of $\Psi(\Sigma(E_2, q_2; C_1))$, by symmetry. Assuming both the Intersection Lemma and the Monotone Lemma, we get that $\Psi(\Sigma_1) \cap \Psi(\Sigma_2)$ is a single point. But this means that $\Sigma_1 \cap \Sigma_2$ is a single $\boldsymbol{R}$-arc, namely, $\Sigma(E_0, q_0; v_0)$. But this means that Σ_1 and Σ_2 bound closed balls B_1 and B_2 in S^3, which have disjoint interiors. By symmetry the same statement can be made for each distinct pair (Σ_i, Σ_j). Moreover, the three $\boldsymbol{R}$-spheres are permuted by an order-3 symmetry. From this we see that there are three balls $B_0, B_1, B_2 \subset S^3$, with disjoint interiors, such that I_j preserves $\partial B_j = \Sigma_j$ and interchanges the two components of $S^3 - B_j$. From this picture, which looks quite a bit like a classical Schottky group, we see that $\boldsymbol{\Gamma}$ is discrete and that the corresponding representation is an embedding.

One more ingredient goes into the proof of Theorem 1.4. Let $\mathcal{S}$ denote the $\boldsymbol{\Gamma}'$-orbit of $\Sigma_0 \cup \Sigma_1 \cup \Sigma_2$.

Lemma 20.3 (Shrinking) *For any $\epsilon > 0$ there are only finitely many elements of $\mathcal{S}$ that have diameter greater than ϵ.*

As we will see, the Shrinking Lemma lets us identify Ω, the domain of discontinuity, and also prove that Λ, the limit set, is porous.

20.2 THE PROOF MODULO TECHNICAL LEMMAS

We will spend most of our time dealing with the subgroup $\boldsymbol{\Gamma}$, and then we will bring in $\boldsymbol{\Gamma_3}$ at the end. We have already mentioned that $\boldsymbol{\Gamma}'$ is a discrete

embedding. The point here is that the set $U = S^3 - (B_0 \cup B_1 \cup B_2)$ is freely permuted by the elements of $\mathbf{\Gamma}'$. Since $\mathbf{\Gamma}'$ is a discrete embedding, so is $\mathbf{\Gamma}$.

Next we need a technical result.

Lemma 20.4 *Let $\mathcal{B}$ denote the Γ-orbit of the balls $B_0 \cup B_1 \cup B_2$. Suppose that $\{\beta_j\}$ is a maximal and infinite collection of balls of $\mathcal{B}$ with the property that $X = \bigcap \partial\beta_i \neq \emptyset$. Then $X \in \{v_0, v_1, v_2, q_0, q_1, q_2\}$.*

Proof: Applying the group action and the 3-fold symmetry of the picture, we can assume that $\beta_1 = B_0$ and $\beta_2 = I_0(B_1)$. Thus

$$\partial\beta_1 \cap \partial\beta_2 = I_0(\Sigma_0 \cap \Sigma_1) = I_0\big(\Sigma(E_2, q_2; v_2)\big). \tag{20.2}$$

Given the maximality condition, β_3 is one of two balls, namely, $\beta_3 = I_0 I_1(B_2)$ or $\beta_3 = I_0 I_1(B_0)$.

Suppose that $\beta_3 = I_0 I_1(B_2)$. Then

$$\partial\beta_2 \cap \partial\beta_3 = I_0 I_1(\Sigma_1 \cap \Sigma_2) = I_0 I_1\big(\Sigma(E_0, q_0; v_0)\big) = I_0\big(\Sigma(E_2, q_2; v_0)\big).$$

Using the above equations we get

$$\partial\beta_1 \cap \partial\beta_2 \cap \partial\beta_3 = I_0\big(\Sigma(E_2, q_2; v_0) \cap \Sigma(E_2, q_2; v_2)\big) = I_0(q_2) = q_1.$$

Hence, $\bigcap \beta_i = q_1$.

Suppose that $\beta_3 = I_0 I_1(B_0)$. Then

$$\partial\beta_2 \cap \partial\beta_3 = I_0 I_1(\Sigma_1 \cap \Sigma_0) = I_0 I_1\big(\Sigma(E_2, q_2; v_2)\big) = I_0\big(\Sigma(E_0, q_0; v_2)\big).$$

Hence,

$$\partial\beta_1 \cap \partial\beta_2 \cap \partial\beta_3 = I_0\big(\Sigma(E_2, q_2; v_2)\big) \cap I_0\big(\Sigma(E_0, q_0, v_2)\big) = I_0(v_2) = v_2.$$

Hence, after using symmetry, we get either $X = q_1$ or $X = v_2$. □

Define

$$F = S^3 - (B_0^o \cup B_1^o \cup B_2^o) - (q_0 \cup q_1 \cup q_2) - (v_0 \cup v_1 \cup v_2). \tag{20.3}$$

Say that the points $v_0, v_1, v_2, q_0, q_1, q_2$ are the *ideal vertices* of F.

Corollary 20.5 *$F \subset \Omega$.*

Proof: First, we will show that $F \subset \Omega$. Since $\mathbf{\Gamma}'$ freely permutes the orbit of U, we know that $U \subset \Omega$. So, it suffices to consider a point $x \in F - U$. Our analysis in the previous lemma shows that, with finitely many exceptions, $g(B_i) \cap \{x\} = \emptyset$. Here $g \in \mathbf{\Gamma}'$. Hence, there is a neighborhood V of x such that V only intersects finitely many translates of F. Hence, $V \subset \Omega$. Hence, $x \subset \Omega$. This shows that $F \subset \Omega$. □

Lemma 20.6 *F is a fundamental domain for the action of $\mathbf{\Gamma}'$ on Ω.*

Proof: Let Λ be the limit set of $\mathbf{\Gamma}'$. Let $\Lambda' \subset S^3$ denote the set of points p such that every neighborhood of p intersects infinitely many elements of $\mathcal{B}$. For instance $v_0 \in \Lambda'$. By construction Λ' is nonempty, closed, and $\mathbf{\Gamma}'$-invariant. Hence, $\Lambda \subset \Lambda'$. From the Shrinking Lemma we see that any $p \in \Lambda'$ is an accumulation point of a sequence of the form $\{q_j\}$, where q_j is a pole point of an $\boldsymbol{R}$-sphere bounding some element of $\mathcal{B}$. But such poles belong to Λ. Hence, $\Lambda' = \Lambda$. But the orbit of F is precisely the complement of Λ'. Hence, Ω is the orbit of F. Hence, F is a fundamental domain for the action of $\mathbf{\Gamma}'$ on Ω. □

Below we will prove the following lemma.

Lemma 20.7 (Horocusp) *The following quotients are horocusps:*

$$V_1/\mathbf{\Gamma}; \qquad \big(V_0 \cup I_0(V_0)\big)/\mathbf{\Gamma}; \qquad \big(Q_0 \cup Q_1 \cup Q_2 \cup I_0(Q_0) \cup I_0(Q_1) \cup I_0(Q_2)\big)/\mathbf{\Gamma}.$$

By symmetry $V_2/\mathbf{\Gamma}$ is also a horocusp. This fact, together with the facts from the Horocusp Lemma, takes care of all the neighborhoods of the vertices in $F \cup I_0(F)$. Thus, once we delete the neighborhoods V_i, Q_i, $I_0(V_i)$, and $I_0(Q_i)$, we are left with a compact subset of F. Thus we see that $\Omega/\mathbf{\Gamma}$ is the union of a compact set and 4 horocusps. It remains to show that Λ is porous. We will prove this result in two steps.

Lemma 20.8 *Suppose that $g \in PU(2,1)$ is such that $g(\Omega)$ contains no ball of spherical radius $1/100$. Then there is some ball $\beta \in \mathcal{B}$ such that $g(\partial\beta)$ has spherical diameter at least $1/100$.*

Proof: We have $\Lambda \subset B_0 \cup B_1 \cup B_2$. Each B_j contains two smaller spheres, $I_j(B_{j-1})$ and $I_j(B_{j+1})$. The set $\Lambda \cap B_j$ is contained in $I_j(B_{j-1}) \cap I_j(B_{j+1})$. Continuing in this way, each of the two little balls just mentioned contains two little balls and so on. By the Shrinking Lemma, every nested sequence shrinks to a point.

We will suppose our claim is false and derive a contradiction. Each of the three spheres $g(\Sigma_0)$, $g(\Sigma_1)$, $g(\Sigma_2)$ has diameter less than η. This means that each $g(B_j)$ is either very small or very large. If all three are very small, then $g(\Lambda)$, which is contained in the union of these balls, is not η-dense. Hence, one of the $g(B_j)$, let's say $g(B_0)$, is very large. Consider the two balls $g(B_{01})$ and $g(B_{02}) \subset g(B_0)$. If both of these balls are small, then Λ is not η-dense. Hence, one of the balls, say, $g(B_{01})$, is very large. Continuing in this way, we produce an infinite nested sequence of balls, each of which contains a hemisphere. Hence, the infinite intersection of these balls contains a hemisphere and does not shrink to a point. This contradiction establishes our claim. □

We ought to be able to extract the statement that Λ is porous directly from Lemma 20.8, but as it is we need to put in a bit more work. Using the structure of Ω/Γ, we can create a horotube assignment $p \to T_p$ for Γ that satisfies all the properties given in Section 9.1. We can choose our horotube

assignment so that F only intersects 4 horotubes and the basepoints of these horotubes are the vertices of F. (This point will become much clearer in the next chapter, where we analyze F in detail.)

For the purposes of contradiction we will suppose that $\{g_n\} \in PU(2,1)$ is a sequence of elements for which $g_n(\Lambda)$ is becoming dense. In particular, we can assume that Lemma 20.8 holds. Let $\beta_n \in \mathcal{B}$ be such that $g_n(\partial\beta_n)$ has spherical diameter at least $1/100$. Note that $\partial\beta_n \cap \Lambda$ consists of 4 points, the vertices of one of the fundamental domain translates adjacent to $\partial\beta_n$. Also, $\partial\beta_n - \Lambda$ is covered by 4 horotubes T_{nj} for $j = 1,2,3,4$. Since $g_n(\partial\beta_n)$ has diameter at least $1/100$, there is some point

$$x \in g_n(T_{jn} \cap \partial\beta_n)$$

that is at least, say, $1/1000$ from the basepoint of $g_n(T_{jn})$. Here j is one of $1,2,3,4$. Hence, there is some horotube T_n from our assignment such that $g_n(T_n)$ has uniformly large spherical diameter. Looking at the options in Lemma 9.1, we see that $g_n(\Lambda)$ cannot converge to S^3, which is a contradiction. Hence, Λ is porous.

Now we know that $\mathbf{\Gamma}$ is a horotube group. It remains to analyze the types of elements in $\mathbf{\Gamma}$. All the elements of $\mathbf{\Gamma}$ are either loxodromic or parabolic. Hence, $\mathbf{\Gamma}$ trivially has isolated type. Finally, the Structure Lemma tells us that any parabolic of $\mathbf{\Gamma}$ is conjugate to one that fixes a vertex of $F \cup I_0(F)$. Hence, every parabolic element of $\mathbf{\Gamma}$ is conjugate to either $(I_iI_j)^a$ for some $a \in \boldsymbol{Z}$ or $(I_0I_1I_2)^b$ for some $b \in 2\boldsymbol{Z}$.

Now we consider $\mathbf{\Gamma_3}$. Since the order-3 element of this group acts with an isolated fixed point, $\mathbf{\Gamma_3}$ has isolated type. All the other horotube group properties are automatically preserved in going from $\mathbf{\Gamma}$ to $\mathbf{\Gamma_3}$. Given what we already know about the parabolic elements of $\mathbf{\Gamma}$ and the fact that $\mathbf{\Gamma}$ has index 3 in $\mathbf{\Gamma_3}$, we see that any parabolic element of $\mathbf{\Gamma_3}$ is conjugate in $\mathbf{\Gamma_3}$ to either $(I_0I_1)^a$ for some $a \in \boldsymbol{Z}$ or $(I_0I_1I_2)^b$ for some $b \in 2\boldsymbol{Z}$. This completes the proof of Theorem 1.4 (except the last statement) modulo the several lemmas mentioned above.

20.3 PROOF OF THE HOROCUSP LEMMA

20.3.1 Some Topological Models

First, we are going to build topological models for the quotients we are taking. In the next chapter we will build more global versions of these models and actually give a second proof of the Horocusp Lemma.

Let $\Delta \subset \boldsymbol{R}^2$ be the closed unit disk. Let $f_1 : \Delta \to [0,1]$ be a smooth function with the property that $f(x,y) = 0$ iff $y = 0$ and $x \geq 0$. Let

$$N_1 = \{(x,y,z) \in \Delta \times \boldsymbol{R} |\ |z| \leq f_1(x,y)\} - \{(0,0,0)\}. \tag{20.4}$$

We call $(0,0,0)$ the *puncture* of N. We call the segment $\{(t,0,0)|\ t \in (0,1]\}$ the *singular arc* of N_1. Note that a small closed ball around $(0,0,0)$ in $\boldsymbol{R}^3$ intersects $\boldsymbol{R}^3 - \text{interior}(N_1)$ in a pair of closed cells, which intersect along

the singular arc. When we speak of a *boundary component* of N_1, we always mean a component that contains $(0,0,0)$ in its closure. Thus, we mean to refer to the "top" and "bottom" of N_1 when we refer to the boundary components of N_1.

Now we want to see how copies of N_1 fit together. Let $f_2 : \Delta \to [0,1]$ be a smooth function such that $f_2(x,y) = 0$ iff $x = 0$ and $y = 0$. We define N_2 exactly as N_1 but with f_2 in place of f_1. Note that that N_2 is homeomorphic to the union of two copies of N_1, joined along a common boundary in such a way that their punctures match and their singular arcs do not match. The same picture obtains if we glue together $n > 2$ copies of N_1, in parallel, in such a way that the singular arcs of at least one consecutive pair do not coincide.

To really get a clear picture of N_2, we note that N_2 is homeomorphic to the set

$$\big(\boldsymbol{C} - \text{interior}(\Delta)\big) \times [0,1].$$

In this case the puncture point is ∞. Here we are abusing notation slightly and thinking of Δ as the closed unit disk in $\boldsymbol{C}$ rather than in $\boldsymbol{R}^2$. The map $(z,t) \to (z,t+1)$ (or any other map from Equation 2.12) identifies the top and bottom of this version of N_2, and the quotient is a horocusp. Indeed, the orbit of this version of N_2 under the iterates of the map in Equation 2.12 is the standard horocusp we discussed in Chapter 2.

We might have built a similar model of N_2 based on the maps in Equation 2.13. In this case N_2 would be a suitably chosen unbounded region in $\mathcal{H}$ between the two planes $i\boldsymbol{R} \times \boldsymbol{R}$ and $(1 + i\boldsymbol{R}) \times \boldsymbol{R}$.

20.3.2 The Side Pairings

We claim that we can take the sets $V_i \cap F$ and $Q_i \cap F$ to be homeomorphic to N_1. Consider the set $V_0 \cap F$. The point v_0 is the intersection of two balls B_1 and B_2 whose interiors are disjoint. The intersection $\partial B_1 \cap \partial B_2 = \Sigma_1 \cap \Sigma_2$ is the arc $\Sigma(E_0, q_0; v_0)$. The point v_0 is one of the endpoints of this arc. From this we see that a small neighborhood of v_0 in F is homeomorphic to N_1. Essentially the same argument works for $Q_0 \cap F$.

Now let's consider the set V_1. We have $V_1 = (V_1 \cap F) \cup (V_1 \cap I_0(F))$. These sets are obtained by sticking two copies of N_1 together along a boundary component, in such a way that the punctures match and the singular arcs do not match. The point here is that

$$I_0\big(\Sigma(E_1, q_1; v_1) \cap \Sigma(E_1, q_1; v_1)\big) = \{v_1\}.$$

Thus we see that V_1 is homeomorphic to N_2. The action of $I_0 I_2$ on V_1 pairs one boundary component to the other, in the same topological way that the map $(z,t) \to (z,t+1)$ pairs the top and bottom of our unbounded version of N_2. Thus $V_1/\boldsymbol{\Gamma}$ is a horocusp. (It would be better here to use a model of N_2 based on Equation 2.13, but we hope our argument is clear enough as is.)

The analysis of $(V_0 \cup I_0(V_0))/\mathbf{\Gamma}$ is similar. Rather than repeat the details, we will give an argument from symmetry. Our analysis of $V_1/\mathbf{\Gamma}$ shows that there is a horotube in S^3 stabilized by I_1I_2 and based at v_0 contained entirely in Ω. By symmetry, there must be a similar horotube stabilized by I_0I_2 and based at v_1. The quotient of this horotube by the group generated by I_0I_1 is the horocusp $(V_0 \cup I_0(V_0))/\mathbf{\Gamma}$. The idea here is that the picture in the quotient is independent of the choice of fundamental domain.

Finally, consider the quotient

$$\big(Q_0 \cup Q_1 \cup Q_2 \cup I_0(Q_0) \cup I_0(Q_1) \cup I_0(Q_2)\big)/\mathbf{\Gamma}.$$

This quotient coincides with the quotient $(Q_0 \cup Q_1 \cup Q_2)/\mathbf{\Gamma}'$. Note that no point q_i lies in the $\boldsymbol{C}$-circle C_0 fixed by I_0. Hence, $Q_i = Q_i \cap F$. The map I_1 conjugates $I_1I_0I_2$ to $I_0I_2I_1$ and, hence, maps q_2 to q_0. Thus I_1 pairs one boundary component of Q_0 to a boundary component of Q_2. Also I_2 conjugates $I_1I_0I_2$ to $I_2I_1I_0$. Thus I_2 pairs the other boundary component of Q_0 to a boundary component of Q_1. The union $Q_0 \cup I_1(Q_2) \cup I_2(Q_1)$ is homeomorphic to N_2. The map $I_0I_1I_2$ pairs the top and bottom components of our union, and the quotient is again a horocusp.

This completes the proof of the Horocusp Lemma. □

20.4 PROOF OF THE INTERSECTION LEMMA

Here we prove that $\Psi(C_1)$ and $\Psi(C_2)$ only intersect in a single point. We first need some preliminary results.

Suppose that C is a $\boldsymbol{C}$-circle in $\mathcal{H}$ that links $(\{0\} \times \boldsymbol{R})$. Let $\pi(z,t) = z$ be the projection to $\boldsymbol{C}$. Then $\pi(C)$ is a circle in $\boldsymbol{C}$ that surrounds 0. Let r be the radius of $\pi(C)$, and let d be the distance from the center of $\pi(C)$ to the origin. Then $d < r$. If $d > 0$, then we define the *aspect* $A(C) = (r/d)^2$. Here is a version of [**S0**, Lemma 2.1].

Lemma 20.9 *Let A be the aspect of C. Let Ψ be the map given by $\Psi(z,t) = (\arg z, t)$. Up to scaling and rotation $\Psi(C)$ is the graph of*

$$f_A(t) = \sin(t)\big(\cos(t) + \sqrt{A - \sin^2(t)}\big) \tag{20.5}$$

Proof: We normalize so that $(1,0)$ is the center of mass of C. Then $d = 1$, and C is contained in the contact plane through $(1,0)$. This plane is spanned by $(1,0)$ and $(i,2)$. Let C_θ be the point on C such that the line through 0 and $\pi(C_\theta)$ makes an angle of θ with the x-axis. Then $\Psi(C)$ is the graph of the function $\theta \to \text{height}(C_\theta) = 2y$, where $C_\theta = (x,y)$. Our formula comes from solving the equations $(x-1)^2+y^2 = r^2$ and $x = y\cot(\theta)$ in terms of y. □

Lemma 20.10 *If $A \geq 9$, then f_A'' is negative on $(0,\pi)$ and positive on $(\pi, 2\pi)$.*

Proof: This is an exercise in calculus, which we do in [**S0**]. □

We normalize so that (E_0, q_0) is the standard pair and $R_0 = (\{\boldsymbol{R} \times \{0\}) \cup \infty$. Then the $\boldsymbol{R}$-reflection in R_0 is given by the isometric rotation $(z,t) \to (\overline{z}, -t)$. This shows that $A(C_1) = A(C_2)$. A direct calculation shows that $A(C) = 10$. See [**S0**] for details and also see Figure 20.4. Let ψ_j be the function whose graph is $\Psi(C_j)$. Up to rotation and scaling, ψ_1 and ψ_2 coincide with f_{10}, the function in Equation 20.5. Hence, by Lemma 20.10, the function ψ_j is convex on an interval of length π and concave on the complementary interval of length π. By symmetry we have $\psi_j'' = 0$ only at the values corresponding to $\Psi(C_j) \cap L_j$.

We have $\Psi(v_0) = (0,0)$. By symmetry we get

$$\psi_1(0) = \psi_2(0), \qquad \psi_1'(0) = \psi_2'(0), \qquad \psi_1''(0) = -\psi_2''(0). \tag{20.6}$$

We compute that $\psi_1''(0) > 0$. Hence, $\psi_2''(0) < 0$. To finish the proof it suffices to show that $\psi_1(t) > \psi_2(t)$ for $t \in (0, \pi]$. There are values $t_1, t_2 \in (0, \pi)$ such that $\psi_j''(t_j) = 0$. The point $(t_j, \psi_j''(t_j))$ is one of the points of $\Psi(C_j) \cap L_j$. We suppose $t_2 \le t_1$, as suggested by Figure 20.1. The other case is similiar. Then on $(0, t_2)$ we have $\psi_1 > \psi_2$ because of our initial conditions at 0 and the fact that $\psi_1'' > 0$ on $(0, t_1)$ and $\psi_2'' < 0$ on $(0, t_1)$. For $t \in [t_2, \pi)$ the curve $\Psi(C_2)$ lies below L_2, and L_2 lies completely below $\Psi(C_2)$.

This completes the proof of the Intersection Lemma. □

20.5 PROOF OF THE MONOTONE LEMMA

We index the $\boldsymbol{R}$-arcs of $\Sigma(E_2, q_2; C_1)$ by points $\theta \in C_1$. That is,

$$\gamma_\theta = \Sigma(E_2, q_2; \theta). \tag{20.7}$$

Let $\widehat{\gamma}_\theta$ be the $\boldsymbol{R}$-circle that contains γ_θ. Thus the curve $\widehat{\gamma}_\theta$ is affiliated with $\Sigma(E_2, q_2; C_1)$, as in Section 19.4. Let $\alpha_\theta = \Psi(\gamma_\theta)$, and let $\widehat{\alpha}_\theta = \Psi(\widehat{\gamma}_\theta)$. Figure 20.2 shows a plot of some α_θ, shown in black, and $\widehat{\alpha}_\theta - \alpha_\theta$, shown in dark gray. $\Psi(C_1)$ is shown in light gray.

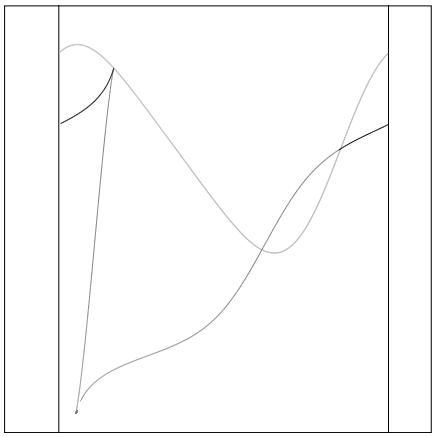

Figure 20.2: Elevation projection of C_1 and an affiliated $\boldsymbol{R}$-circle

Since $\widehat{\alpha}_\theta$ winds once around $S^1 \times \boldsymbol{R}$, we see that $\widehat{\gamma}_\theta$ links E_0. According to the Elevation Image Lemma, $\widehat{\alpha}_\theta$ is the union of a major monotone arc and a minor monotone arc. Clearly, α_θ is contained in the major monotone arc. It appears that α_θ shares an endpoint with the major monotone arc, and this is not an accident.

Lemma 20.11 *The endpoint $\Psi(\theta)$ of α_θ is a singular point of $\widehat{\alpha}$.*

Proof: Let C_θ be the contact plane at θ. Then I_1 rotates C_θ by 180 degrees. This means that γ_θ and $I_1(\gamma_\theta)$ are tangent at θ. But $I_1(\gamma_\theta) = \Sigma(E_0, q_0; \theta)$ is a fiber of the elevation map Ψ. Hence, γ_θ is tangent to a fiber of Ψ at θ. □

Suppose that Σ_1 is normal. Then every affiliate of $\Sigma(E_2, q_2; C_1)$, except the axis R_1, links E_0. But then the Elevation Image Lemma applies to all these arcs. As θ varies away from the one parameter we have checked, the following three things persist.

- The decomposition into major and minor monotone arcs varies continuously. This follows from the uniqueness of the decomposition.
- α_θ always shares one endpoint with the major monotone arc.
- If α_θ is contained entirely inside the major monotone arc of $\widehat{\alpha}_\theta$ then α_θ is a minor monotone arc. The point here is that α_θ just moves directly from one of its endpoints to the other, and this other endpoint is less than 2π around the cylinder.

From this structure we see that α_θ *cannot cease* from being a minor subarc of the major monotone arc of $\widehat{\alpha}_\theta$ until θ reaches one of the parameters of $R_1 \cap C_1$. The linking fails at these two parameters. Thus, assuming Σ_1 is normal, we see that half the arcs of $\Psi(\Sigma(E_2, q_2; C_1))$ are monotone arcs—namely, the half that are indexed by parameters in the same component of

$C_1 - R_1$ as the one parameter we plotted. However, the other half of the arcs are monotone by symmetry. It remains only to prove the following.

Lemma 20.12 Σ_1 *is normal.*

Proof: We will verify the 3 conditions of our normality criterion from Lemma 19.3.

A direct calculation, which we omit, shows that E_2 and E_0 are linked. One way to see this pictorially is to look at the gray curve in Figure 20.3. This curve is $\Psi(E_2)$. (The aspect ratio is different from Figures 21.1 and 21.2 because we wanted to fit all of $\Psi(E_2)$ on the screen.) Note that this curve winds once around the cylinder $S^1 \times \boldsymbol{R}$. This implies that E_2 links E_0, the axis of the cylinder. This is Lemma 19.3 (1).

The two black curves in Figure 20.3 are $\Psi(C_1)$ and $\Psi(C_1^*)$, with $\Psi(C_1)$ being the lower of the two curves. (These curves are isometric, with one being obtained from the other by rotating the cylinder by π.) All the curves are graphs of functions similar to the one given by Lemma 20.10. Each function has one max and one min. From the nesting pattern of the extrema we see immediately that $\Psi(C_1) \cap \Psi(E_2)$ consists of a pair of points. Lemma 20.13 shows the more delicate result that $\Psi(C_1^*) \cap \Psi(E_2)$ consists of a pair of points. Lemma 19.3 (2).

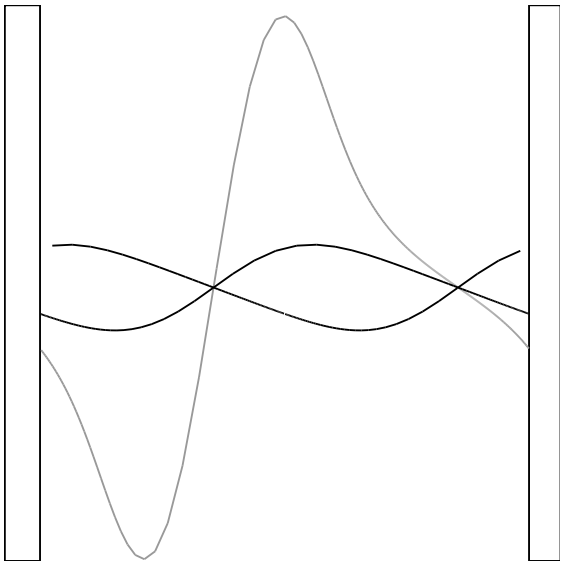

Figure 20.3: Elevation projection of C_1, C_1^* (black), and E_2 (gray)

The affiliate drawn in Figure 20.2 winds once around the cylinder. Hence, this affiliate $\Sigma(E_2, q_2; C_1)$ links E_0. This is Lemma 19.3 (3). □

Lemma 20.13 $\Psi(C_1^*) \cap \Psi(E_2)$ *consists of a pair of points.*

Proof: Given $p \in \mathcal{H}$ we write $p = (\pi(p), v(p))$. Recall that R_1 is the $\boldsymbol{R}$-axis of Σ_1. We know that $\Psi(C_1^*) \cap \Psi(E_2)$ contains the two points of $\Psi(R_1 - E_0)$. Compare Lemma 19.3 (1). Here we rule out another intersection point.

We normalize so that (E_0, q_0) is in standard position. Let e denote the center of $\pi(E_2)$, and let c denote the center of $\pi(C_1^*)$. Both c and e lie in the line $r = \pi(R_1)$. Figure 20.4 shows $\pi(C_1^*)$ in black and $\pi(E_2)$ in gray. Let $r' \not\subset r$ be a ray through the origin. A direct calculation–see Section 22.4.2—shows that

$$|c| < \frac{2}{7}|e|, \qquad |r' \cap \pi(C_1^*)| < 3|r' \cap \pi(E_2)|. \tag{20.8}$$

Let Δ_e denote the area of the triangle with vertices $0, e, r' \cap \pi(E_2)$. Likewise, let Δ_c denote the area of the triangle with vertices $0, e, r' \cap \pi(C_1^*)$. The origin is a common vertex of these triangles; one side of each triangle lies in r, and one side of each triangle lies in r. From Equation 20.8 we can map Δ_c properly into Δ_e using an element of $SL_2(\boldsymbol{R})$ with eigenvectors r and r' and eigenvalues 3 and 1/3. Hence, $\Delta_c < \Delta_e$.

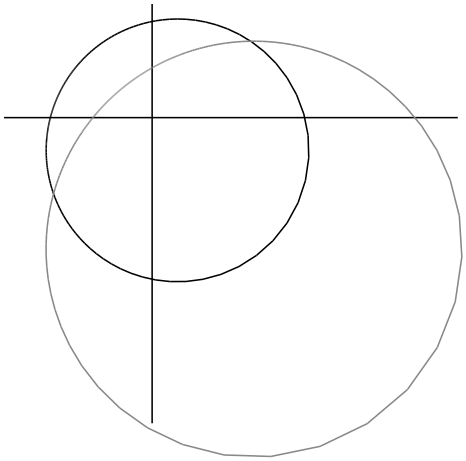

Figure 20.4: The projection to $\boldsymbol{C}$ of C_1^* (black) and E_2 (gray)

If $\Psi(C_1^*) \cap \Psi(E_2)$ had a third intersection point, then we could find a fiber R' of Ψ, which intersects both C_1^* and E_2. Let $r' = \pi(R')$. Then r' is a ray through the origin. The Lift Principle II from Section 2.4 gives

$$v(R' \cap E_2) = 4\Delta_e > 4\Delta_c = v(R' \cap C_1^*).$$

This is a contradiction.

This completes the proof of the Monotone Lemma. □

20.6 PROOF OF THE SHRINKING LEMMA

Let $\mathcal{B}$ denote the Γ-orbit of the balls $B_0 \cup B_1 \cup B_2$. It suffices to show that every nested intersection of the form $\bigcap \beta_n$, with $\beta_n \in \mathcal{B}$, is a single point.

Suppose, for the sake of contradiction, that this fails for some sequence. Using Lemma 20.4 and the fact that $\mathcal{B}$ is the orbit of finitely many balls, we boil the problem down to three cases.

Case 1: There are infinitely many pairs $(\beta_{n_k}, \beta_{n_k+1})$ such that $\partial\beta_{n_k} \cap \partial\beta_{n_k+1} = \emptyset$. Moreover, the collection of our chosen pairs is finite modulo the action of $\mathbf{\Gamma}$. This case follows immediately from Lemma 20.6.

Case 2: Letting $g = I_0 I_1 I_2$, we have $\partial\beta_1 = \Sigma_1$ and $g(\beta_i) = \beta_{i+3}$. Since the sets $g^n(\beta_1)$ are nested, it suffices to show that the boundary sets $g^n(\partial\beta_1) = g^n(\Sigma_1)$ shrink to points. We normalize by a Heisenberg stereographic projection so that g is as in Equation 2.12. Then (E_1, q_1) is the standard flag, and Σ_1 is contained between two horizontal planes in $\mathcal{H}$. Hence, $g^n(\partial\Sigma_1)$ exits every compact subset of $\mathcal{H}$. This does it.

Case 3: Letting $g = I_0 I_1$, we have $\partial\beta_1 = \Sigma_1$ and $g(\beta_i) = \beta_{i+2}$. Here we improve on the argument in [**S0**]. Again, it suffices to show that $\{g^n(\Sigma_1)\}$ shrinks to a point. The map $g = I_0 I_1$ is conjugate to the map in Equation 2.13, with $s = 0$. Let Π denote the contact plane at v_2. There is a tangent line L in Π such that g stabilizes an $\boldsymbol{R}$-circle through v_2 iff this $\boldsymbol{R}$-circle is tangent to L. Compare the proof of Lemma 2.3.

There are 2 $\boldsymbol{R}$-arcs of Σ_1 emanating from v_2. Both of these $\boldsymbol{R}$-circles are tangent to some line $L' \subset \Pi$. Suppose $\{g_n(\Sigma_1)\}$ fails to shrink to a point. Then there is some $\boldsymbol{R}$-circle λ, tangent to L, with the following property. The intersection $\lambda \cap \Sigma_1$ is such that $g^k(\lambda \cap \Sigma_1)$ lies a uniformly bounded distance from v_2 in S^3, for infinitely many exponents k. But this happens iff λ intersects Σ_1 in points arbitrarily close to v_2. Note that all the $\boldsymbol{R}$-circles of Σ_1 near v_2 are nearly parallel to L'. Thus, our intersection property implies that $L = L'$.

To finish our proof, we just have to show that $L' \neq L$. If $L' = L$, then each of the two $\boldsymbol{R}$-arcs of Σ_1 through v_2 is actually stabilized by g. But then the intersection $\bigcap_{i=1}^{n} g^i(\Sigma_1)$ would contain an arc for any n, and we know that this intersection is just a single point for $n > 1$.

This completes the proof of the Shrinking Lemma. □

Chapter Twenty One

The Manifold at Infinity

In this chapter we will prove the last statement of Theorem 1.4—$\Omega/\mathbf{\Gamma_3}$ is homeomorphic to the Whitehead link complement. We already have a fundamental domain for the action of $\mathbf{\Gamma}$, namely, $F \cup I_0(F)$, where F is as in Equation 20.3. We really just have to analyze the topology of this domain and its side pairings. This is the approach taken in [**S0**]. Here we will take a more combinatorial (and long-winded) approach because it gives us a global understanding of the way $\mathbf{\Gamma_3}$ acts on Ω. Also, the combinatorial approach shows off some of the beauty of $\mathbf{\Gamma_3}$.

In the first section we will describe the topology of F, since this is necessary for any further progress. Describing the topology of F amounts to building a combinatorial model $\widetilde{F}$ for F. After we build $\widetilde{F}$ we will build a combinatorial model $\widetilde{\Omega}$ for Ω, tiled by copies of $\widetilde{F}$. There is an isomorphic copy $\widetilde{\mathbf{\Gamma_3}}$, which acts on $\widetilde{\Omega}$. After we describe this combinatorial action, we will analyze the quotient $\widetilde{\Omega}/\widetilde{\mathbf{\Gamma_3}}$ and show that it is homeomorphic to the Whitehead link complement. Finally, we will identify our model with the actual group action of $\mathbf{\Gamma_3}$ on Ω.

21.1 A MODEL FOR THE FUNDAMENTAL DOMAIN

21.1.1 Everything but the Null Curve

Almost all the information about F can be gleaned from the extensive markings on ∂F.

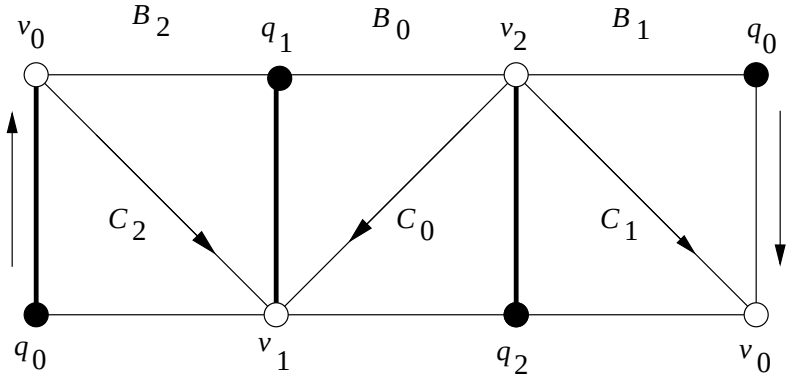

Figure 21.1 Three marked balls pairwise intersecting along edges

In Figure 21.1 we represent a solid ball as a "pillow," a thin neighborhood of a square. Each pillow intersects each other one along an arc—we are tempted to say a "seam." Figure 21.1 shows schematically how 3 balls fit together so that their boundaries have the same markings as our balls B_0, B_1, B_2. Note that the $\boldsymbol{C}$-circle C_i drawn on Σ_i is only half visible. The other half goes "around the back." Figure 21.1 exactly captures the markings and the intersection pattern of our three balls. There is a homeomorphism from $B_0 \cup B_1 \cup B_2$ that, when restricted to ∂F, respects all the markings. It remains to embed Figure 21.1 in space as a set $\widetilde{F}$ so that the homeomorphism from ∂F to $\partial \widetilde{F}$ extends to a homeomorphism from F to $\widetilde{F}$.

Before we make the embedding, we need to talk about some subtle points of the diagram. Each $\boldsymbol{C}$-circle in S^3 has a preferred orientation. This comes from the fact that complex lines are oriented in $\boldsymbol{C}^2$. We choose our orientation so that compact $\boldsymbol{C}$-circles in $\mathcal{H}$ project to clockwise oriented circles. The arrows in Figure 21.1 illustrate the orientations on the $\boldsymbol{C}$-circles. We let C_i^+ denote the portion of C_i that connects v_{i-2} to v_{i-1}. Here indices are taken mod 3. We let $C_i^- = C_i - C_i^-$. Figure 21.1 shows C_2^+ and C_0^- and C_1^+. The order-3 symmetry S of $\mathbf{\Gamma_3}$ permutes the 3 "pillowcases" in such a way as to preserve the orientations. In particular $S(C_i^+) = C_{i+1}^+$.

Now we explain how our picture is embedded in S^3. For this purpose it is useful to fatten up $B_0 \cup B_1 \cup B_2$ a bit. Let N_i denote a very thin and pointy topological ball–a "banana"–whose two cone points are the endpoints of $B_{i-i} \cup B_{i+1}$ and whose interior is a neighborhood of $B_{i-1} \cup B_{i+1}$. Then

$$T = (B_0 \cup B_1 \cup B_2) \cup (N_0 \cup N_1 \cup N_2)$$

is a solid torus that contains the curves C_0, C_1, C_2 on its boundary. Furthermore, T deformation retracts to $B_0 \cup B_1 \cup B_2$. To understand how $B_0 \cup B_1 \cup B_2$ is embedded in S^3, it suffices to find a curve on ∂T that is 0 in $H_1(S^3 - T)$. This tells us how to "twist" Figure 20.1 when we embed it into S^3.

Below we will prove the following lemma.

Lemma 21.1 (Null) $\gamma = C_2^+ \cup C_0^- \cup C_1^+$ *is null in* $H_1(S^3 - T)$.

Assuming the Null Lemma, we can embed Figure 21.1 in space. After we embed Figure 21.1 in $S^3 = \boldsymbol{R}^3 \cup \infty$, we take the closure of the complement minus the vertices. This is $\widetilde{F}$. Due to the Null Lemma, our homeomorphism extends to a homeomorphism from F to $\widetilde{F}$. Figure 21.2 shows a union of 3 tetrahedra T_0, T_1, T_2. The two tetrahedra on the right T_1, T_2 are supposed to fit inside T_0, the tetrahedron on the left, so that the relevant vertices match.

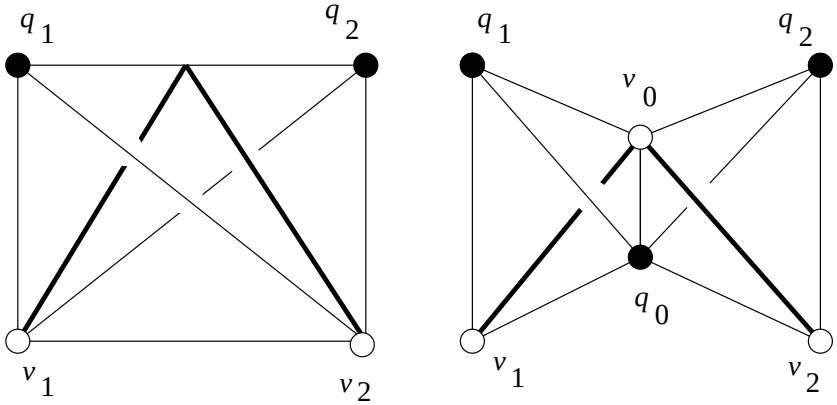

Figure 21.2: Two small tetrahedra fitting inside a large one

Remark: After some effort you can convince yourself that the union of tetrahedra $T_0 \cup T_1 \cup T_2$ can actually be built. That is, it is not an "impossible projection" of tetrahedra. To see that the union can be built, picture q_1 and v_2 lying in front of the picture plane—i.e. closer to your nose—q_2 and v_1 lying behind the picture plane, and v_0 and q_0 lying in the picture plane. Now take the convex hulls of the relevant 4-tuples of vertices.

To get our embedding, we identify B_0 with the *outside* of T_0, and we identify B_1 and B_2 with the inside of T_1 and T_2, respectively. We have labeled the vertices of $\widetilde{F}$ just as we labeled the vertices of F. We have drawn the curve $\widetilde{\gamma}$, the analogue of γ, using dark lines. The dark curve on the left is $\widetilde{C}_0^-$. The two dark curves on the right are $\widetilde{C}_1^+$ and $\widetilde{C}_2^+$. Our convention is that $\widetilde{C}_j^+$ is always an edge of the tetrahedron and $\widetilde{C}_j^-$ lies in the union of two faces. The whole curve $\widetilde{C}_j$ is represented by a triangle.

We can see that $\widetilde{\gamma}$ is isotopic to the triangle $\widetilde{\gamma}'$, which is the boundary of the convex hull of the points q_1, q_2, v_0. The convex hull itself, which serves as a bounding disk, has an interior disjoint from our three embedded balls. This shows that $\widetilde{\gamma}'$ and, hence, $\widetilde{\gamma}$ are null homotopic. Hence, the obvious homeomorphism from ∂F (the set in S^3) to $\partial\widetilde{F}$ extends to give a homeomorphism from F to $\widetilde{F}$.

For later purposes it is worth pointing out some hidden symmetry of $\widetilde{F}$. It can be decomposed into 6 tetrahedra. To make the pattern more clear, we let V_i denote the set $v_{i-1} \cup v_{i+1}$. Likewise, we define $Q_i = q_{i-1} \cup q_{i+1}$. Then $\widetilde{F}$ is triangulated by the 6 tetrahedra $\lambda(V_i \cup Q_j)$, where $i \neq j$ and λ stands for the convex hull. In the sections that follow this one, we construct this same combinatorial structure in a much more canonical way.

21.1.2 Proof of the Null Lemma

To finish our proof that $\widetilde{F}$ really is homeomorphic to F, we need to prove the Null Lemma. We remark that our analysis in [**S0**] took the Null Lemma

for granted, without giving a proof. In hindsight, we think that this was too large a jump in the argument, and we take the opportunity to fill in the details here.

It turns out that it is relatively easy to see which curve is null when we work in $\mathcal{H}$ and normalize using the map B from Equation 4.15. Our plan is to exhibit a curve $\beta \subset B_0 \cup B_1 \cup B_2$ that is a deformation retract of $B_0 \cup B_1 \cup B_2$. Then we will show that our chosen γ does not link β. It is convenient to set $\tau = \sqrt{5/3}$. We have the equations

$$v_0 = (1,0), \qquad v_1 = \Big(\frac{-1+3i\tau}{4}, -\tau\Big), \qquad v_2 = \Big(\frac{-1-3i\tau}{4}, \tau\Big),$$

$$q_0 = \infty, \qquad q_1 = \Big(\frac{-1-i\tau}{4}, \frac{-\tau}{2}\Big), \qquad q_2 = \Big(\frac{1+i\tau}{4}, \frac{\tau}{2}\Big),$$

Figure 21.3 shows the approximate locations of these points, as well as the approximate locations of C_0, C_1, C_2.

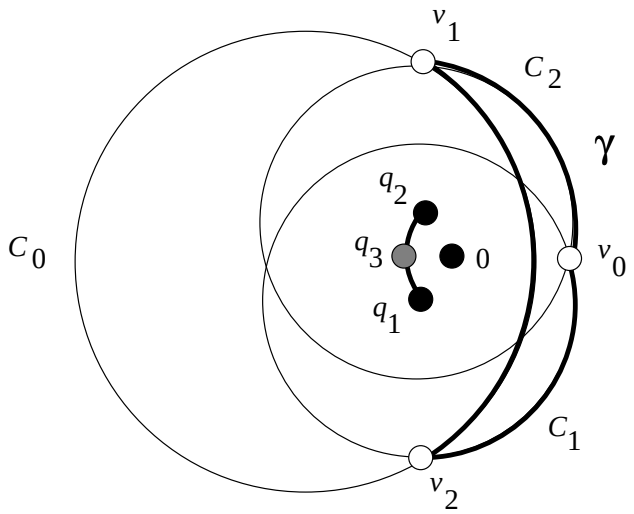

Figure 21.3: Heisenberg projection of vertices and unlinked loops

By Lemma 19.2 the upward vertical $\boldsymbol{C}$-arc β_1 connecting q_2 to ∞ lies entirely inside B_1, except for the endpoints $q_2 \cup \infty$. Likewise, the downward vertical $\boldsymbol{C}$-arc β_2 connecting q_1 to ∞ lies entirely in B_2, except for the endpoints $q_1 \cup \infty$. Finally, the short $\boldsymbol{C}$-arc β_0 joining q_1 to q_2 lies entirely inside B_0, except for the endpoints. We compute that β_0 intersects $\boldsymbol{R}\times\{0\}$ at the point $q_3 = (1-\tau, 0)$, which lies to the left of $(0,0)$. Hence, $\pi(\beta_0)\cap\pi(\gamma) = \emptyset$. Hence, $\pi(\beta) \cap \pi(\gamma) = \emptyset$. Hence, β and γ are unlinked. Here π is the projection from $\mathcal{H}$ to $\boldsymbol{C}$. By construction $B_0 \cup B_1 \cup B_2$ deformation retracts to β, a curve that really runs right through this set. Hence, γ is null.

This completes the proof of the Null Lemma. □

21.2 A MODEL FOR THE REGULAR SET

Let G' be the ordinary ideal triangle group acting on $\boldsymbol{H}^2$ generated by reflections $\iota_0, \iota_1, \iota_2$. Let $\mathcal{T}$ be the tiling of $\boldsymbol{H}^2$ by ideal triangles generated by the action of G'. Let $\mathcal{G}$ denote the union of geodesics stabilized by words that are conjugate to $\iota_0\iota_1\iota_2$.

Let V denote the union of the ideal vertices of $\mathcal{T}$ together with the geodesics in $\mathcal{G}$. In other words, each object in V is either a geodesic of $\mathcal{G}$ or an ideal vertex of $\mathcal{T}$. Note that the elements of V are in canonical bijection with the parabolic fixed points of $\mathbf{\Gamma}'$. We will now assign 6 labelle simplices to each ideal triangle τ of $\mathcal{T}$.

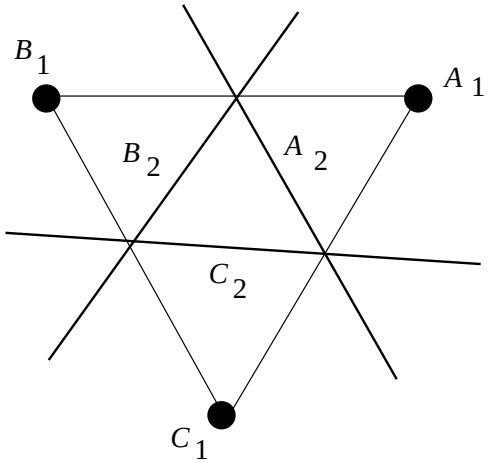

Figure 21.4: A schematic picture of an ideal triangle and 3 geodesics

Given an ideal triangle τ of $\mathcal{T}$, let $V_\tau \subset V$ denote the set of geodesics of V that intersect τ together with the 3 ideal vertices of τ. We label the vertices A_1, B_1, C_1 and the geodesics A_2, B_2, C_2, as shown in Figure 21.4. To τ we associate the following 4-tuples:

$(A_1B_1B_2C_2)$,

$(A_1B_1C_2A_2)$,

$(B_1C_1C_2A_2)$,

$(B_1C_1A_2B_2)$,

$(C_1A_1A_2B_2)$,

$(C_1A_1B_2C_2)$.

We let $\widetilde{F}_\tau$ denote the union of the corresponding tetrahedra. We delete

the 6 vertices from $\widetilde{F}$. We will see that $\widetilde{F}$ and F are homeomorphic in a way that naturally respects the combinatorial structures of these objects. Also, looking at the decomposition into tetrahedra, we see that $\widetilde{F}_\tau$ is combinatorially isomorphic to the set $\widetilde{F}$ constructed in the previous section.

We let $\widetilde{\Omega}$ be the union of all the simplices $\widetilde{F}_\tau$ as τ ranges over all ideal triangles in $\mathcal{T}$. By construction, the modular group and, hence, G' act naturally on $\widetilde{\Omega}$. Let $\widetilde{\mathbf{\Gamma}}'$ denote this representation of G' as a group action on $\widetilde{\Omega}$. Let $\widetilde{\mathbf{\Gamma}}$ be the even subgroup. $\widetilde{F}_\tau$ is by construction a fundamental domain for the action of $\widetilde{\mathbf{\Gamma}}'$ on $\widetilde{\Omega}$. We let τ be the ideal triangle whose sides are fixed by the three generators $\iota_0, \iota_1, \iota_2$, and then we set $\widetilde{F} = \widetilde{F}_\tau$.

Lemma 21.2 *$\widetilde{\Omega}$ is a combinatorial 3-manifold.*

Proof: To show that $\widetilde{\Omega}$ is a combinatorial manifold, we need to establish two things.

- Every 2-cell in $\widetilde{\Omega}$ includes in exactly 2 tetrahedra—i.e., the link of a 2-simplex is a 0-sphere. (The 2-cells are all triangles, but we prefer to use the word *2-cell* to distinguish it from the ideal triangles of $\mathcal{T}$.)
- The link of every 1-simplex in $\widetilde{\Omega}$ is a finite polygon—i.e., a combinatorial 1-sphere.

We don't need to analyze the links of the vertices because we have deleted them from $\widetilde{\Omega}$.

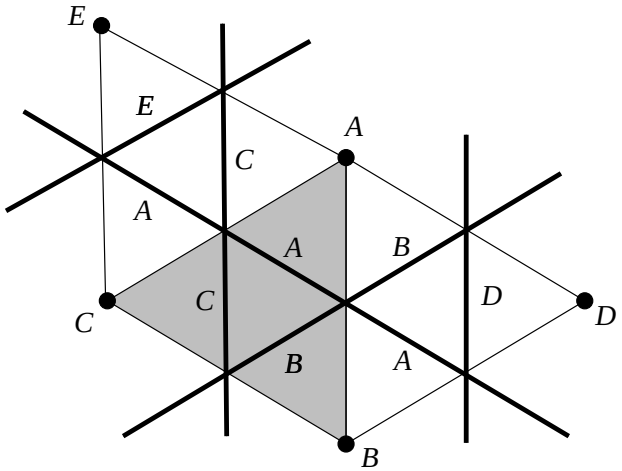

Figure 21.5: A schematic picture of three adjacent ideal triangles

Figure 21.5 shows (schematically) three adjacent triangles of $\mathcal{T}$, as well as the $\mathcal{G}$-objects associated to these triangles. Here A_1,...,E_1 are ideal vertices of $\mathcal{T}$ and A_2,...,E_2 are geodesics in $\mathcal{G}$. (To make Figure 21.5 simpler, we leave off the subscripts.) Using G_6 we can normalize so that a given 2-cell

is labeled by objects associated to the shaded ideal triangle and that one of the labels of the 2-cell is A_1. Indeed, modulo the action of G_6 there are four 2-cells in $\widetilde{\Omega}$. These include in 2 tetrahedra as follows. Here is s list that indicates the 0-sphere links.

$(A_1A_2B_2) \hookrightarrow (A_1A_2B_2\boldsymbol{C}_1), (A_1A_2B_2\boldsymbol{D}_1)$,

$(A_1B_2C_2) \hookrightarrow (A_1B_2C_2\boldsymbol{B}_1), (A_1B_2C_2\boldsymbol{C}_1)$,

$(A_1B_1A_2) \hookrightarrow (A_1B_1A_2\boldsymbol{C}_2), (A_1B_1A_2\boldsymbol{D}_2)$,

$(A_1B_1C_2) \hookrightarrow (A_1B_1C_2\boldsymbol{A}_2), (A_1B_1C_2\boldsymbol{B}_2)$.

Modulo the action of G_6 there are three 1-cells of $\widetilde{\Omega}$. Here is a list of representative 1-cells and their links:

(A_1B_1) link: $\langle A_2C_2B_2D_2\rangle$,

(A_2B_2) link: $\langle A_1C_1B_1D_1\rangle$,

(A_1A_2) link: $\langle B_2C_1E_2E_1C_2B_1D_2D_1\rangle$.

In the first item, the successive tetrahedra containing (A_1B_1) are (cyclically) $(A_1B_1\boldsymbol{A}_2\boldsymbol{C}_2), (A_1B_1\boldsymbol{C}_2\boldsymbol{B}_2), (A_1B_1\boldsymbol{B}_2\boldsymbol{D}_2), (A_1B_1\boldsymbol{D}_2\boldsymbol{C}_2)$. The other 2 items have a similar interpretation. In all cases we have finite polygon links. □

Let $\widetilde{\boldsymbol{\Gamma_3}}$ denote the index-3 supergroup of $\widetilde{\boldsymbol{\Gamma}}$ obtained by adjoining the order-3 symmetry of any triangle of $\mathcal{T}$. Note that $\widetilde{\boldsymbol{\Gamma_3}}$ freely and transitively permutes the edges of $\mathcal{T}$.

21.3 A MODEL FOR THE QUOTIENT

Lemma 21.3 *$\widetilde{\Omega}/\widetilde{\boldsymbol{\Gamma_3}}$ is homeomorphic to the Whitehead link complement.*

Proof: To each edge e of the modular tiling we can associate 4 tetrahedra of $\widetilde{\Omega}$. For instance, to the edge e whose vertices are A_1, B_1 we associate the tetrahedra comprising the link of the edge labeled by A_1 and B_1, namely,

$$\begin{bmatrix} A_1 \\ B_1 \\ A_2 \\ C_2 \end{bmatrix}, \quad \begin{bmatrix} A_1 \\ B_1 \\ C_2 \\ B_2 \end{bmatrix}, \quad \begin{bmatrix} A_1 \\ B_1 \\ B_2 \\ D_2 \end{bmatrix}, \quad \begin{bmatrix} A_1 \\ B_1 \\ C_2 \\ A_2 \end{bmatrix}. \tag{21.1}$$

These 4 tetrahedra together make an octahedron, as shown in Figure 21.6.

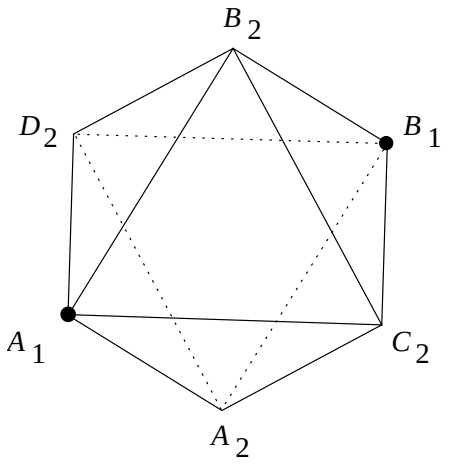

Figure 21.6: A combinatorial octahedron

Note that $\widetilde{\mathbf{\Gamma_3}}$ transitively permutes the edges of $\mathcal{T}$. Hence, a single octahedron—for instance, the one associated to e above—serves as a fundamental domain. It remains to figure out the face pairings. There are 4 pairings, which we call A, B, C, D. There is an order-3 symmetry of the ABC triangle, which induces the pairing

$$A: \qquad \begin{bmatrix} A_1 \\ B_2 \\ C_2 \end{bmatrix} \leftrightarrow \begin{bmatrix} B_1 \\ C_2 \\ A_2 \end{bmatrix}. \tag{21.2}$$

The generator of the parabolic subgroup fixing B_1 has the action $B_2 \to D_2$ and $C_2 \to A_2$. Thus we have the pairing

$$B: \qquad \begin{bmatrix} B_1 \\ B_2 \\ C_2 \end{bmatrix} \leftrightarrow \begin{bmatrix} B_1 \\ D_2 \\ A_2 \end{bmatrix}. \tag{21.3}$$

There is an order 3 symmetry of the ABD triangle, which induces

$$C: \qquad \begin{bmatrix} A_1 \\ A_2 \\ D_2 \end{bmatrix} \leftrightarrow \begin{bmatrix} B_1 \\ D_2 \\ B_2 \end{bmatrix}. \tag{21.4}$$

Looking at the generator of the parabolic subgroup fixing A_1, we get

$$D: \qquad \begin{bmatrix} A_1 \\ A_2 \\ C_2 \end{bmatrix} \leftrightarrow \begin{bmatrix} A_1 \\ D_2 \\ B_2 \end{bmatrix}. \tag{21.5}$$

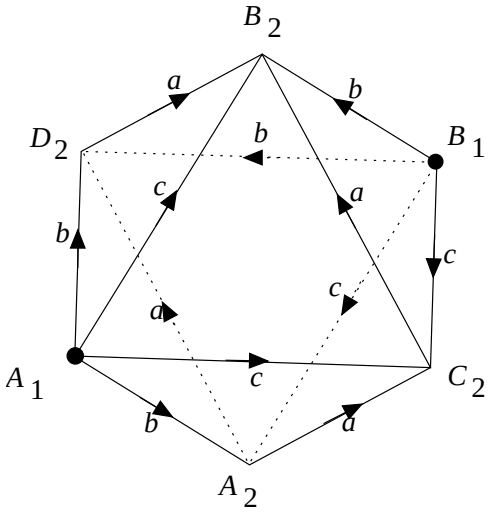

Figure 21.7: The face pairings of the combinatorial octahedron

These pairings give rise to the gluing diagram in Figure 21.7, which coincides with the one for the Whitehead link complement in [**R**, p. 545]. □

Remark: In the octahedral tiling we constructed, every edge includes in 4 octahedra. Hence, if these octahedra are taken to be regular ideal hyperbolic octahedra, then they fit together to put a hyperbolic structure on $\widetilde{\Omega}$. In other words, without even analyzing the face pairings, we can see that our quotient is a hyperbolic manifold.

21.4 IDENTIFICATION WITH THE MODEL

To finish our proof of Theorem 1.4 we need to construct a homeomorphism from $\widetilde{\Omega}$ to Ω that conjugates the action of $\widetilde{\mathbf{\Gamma_3}}$ on $\widetilde{\Omega}$ with the action of $\mathbf{\Gamma_3}$ on Ω. As we remarked above, our sets $\widetilde{F}_\tau$ are combinatorially identical to the model $\widetilde{F}$ constructed in Section 21.1. Thus we can fix the triangle τ whose vertices are v_0, v_1, v_2 and identify $\widetilde{F}$ with $\widetilde{F}_\tau$. As in Section 21.1 we have a homeomorphism from $\widetilde{F}$ to F that respects all the markings. This homeomorphism also respects the action of the side pairings of F and $\widetilde{F}$ by the generators of $\mathbf{\Gamma}'$ and $\widetilde{\mathbf{\Gamma}}'$, respectively. Given the treelike nature of the way the translates of these fundamental domains fit together, we now see that our homeomorphism from $\widetilde{F}$ to F extends to a homeomorphism from $\widetilde{\Omega}$ to Ω that conjugates the one group action to the other. The order-3 symmetry acts the same on each object, and so our homeomorphism conjugates the action of $\widetilde{\mathbf{\Gamma_3}}$ on $\widetilde{\Omega}$ to the action of $\mathbf{\Gamma_3}$ on Ω.

This completes the proof of the last statement in Theorem 1.4. □

Chapter Twenty Two

The Groups near the Critical Value

In this chapter we prove Theorem 1.3 for parameter values $s \in (\underline{s} - \epsilon, \overline{s})$. Here $\overline{s} = \sqrt{125/3}$ is the critical parameter, $\underline{s} = \sqrt{35}$ is the "Goldman-Parker parameter", and ϵ is some small but unspecified value. As we mentioned in Chapter 1, our proof relies on some technical details from [**S5**].

22.1 MORE SPHERICAL CR GEOMETRY

The constructions we make here are "loxodromic analogues" of the constructions we made in Chapters 18–19, which were based on the parabolic element $I_1 I_0 I_2$. Here will explain enough about these constructions at least to show two computer plots that give evidence for the main result, which is proved in [**S5**].

22.1.1 Loxodromic $\boldsymbol{R}$-Cones

This section parallels Section 19.1. Let E be a $\boldsymbol{C}$-circle, and let $Q \subset E$ be a proper $\boldsymbol{C}$-arc. E bounds a $\boldsymbol{C}$-slice $\boldsymbol{H}^2_E$. The points $y \in Q$ and $z \in E - Q$ are said to be *harmonic conjugates* if the geodesic γ_Q in $\boldsymbol{H}^2_E$, which joins the endpoints of Q, is perpendicular to the geodesic in $\boldsymbol{H}^2_E$, which joins y to z. For instance:

- If E is normalized to be $(\{0\} \times \boldsymbol{R}) \cup \infty \subset \mathcal{H}$ and Q is taken to be $(\{0\} \times \boldsymbol{R}^+) \cup \infty$, then $(0, -r)$ and $(0, r)$ are harmonic conjugates.
- If E is as above but the endpoints of Q are $(0, \pm u)$, then the points $(0, v_1)$ and $(0, v_2)$ are harmonic conjugates iff $v_1 v_2 = u^2$.

Up to $PU(2,1)$ either special case we just described is the only case.

Lemma 22.1 *Given $X \in S^3 - E$, there is a unique $\boldsymbol{R}$-circle $\gamma(E, Q; X)$, which contains X and intersects E in a harmonic pair of points.*

Proof: We normalize so that $E = (\{0\} \times \boldsymbol{R}) \cup \infty \subset \mathcal{H}$ and $Q = \{0\} \times \boldsymbol{R}^+$. Let Σ_r denote the spinal sphere whose poles are $(0, -r)$ and $(0, r)$. Then $\mathcal{H} - \{0\}$ is foliated by the pairwise cospinal spinal spheres $\{\Sigma_r |\ r \in \boldsymbol{R}^+\}$. One of these spinal spheres, Σ_x, contains x. Since x is not a pole of Σ_x, exactly one of the leaves of the $\boldsymbol{R}$-foliation of Σ_x contains x. $\square$

Let $\Sigma(E,Q;X)$ be the portion of $\gamma(E,Q;X)$ that joins X to Q but avoids $E-Q$. Given a subset $S \subset S^3 - E$, we define the *loxodromic hybrid cone* $\Sigma(E,Q;S)$ as in the parabolic case. Here we are coning not to a point q but to points in an arc Q.

22.1.2 Loxodromic *R*-Spheres

This section parallels Section 19.2. Let C_1 be a $\boldsymbol{C}$-circle that links another $\boldsymbol{C}$-circle E_0. Let us also assume that C_1 and E_0 do not bound perpendicular $\boldsymbol{C}$-slices. In the language of [**S5**] we would say that C_1 and E_0 are normally linked. An easy calculation in [**S5**] shows that all the pairs of interest to us are normally linked. Indeed, the cases we consider are just tiny perturbations of the parabolic case.

Let (E_0,Q_0) be a pair, with Q_0 a nontrivial arc of E_0. Let

$$(E_2,Q_2) = I_1(E_0,Q_0).$$

Here $I_1 = I_{C_1}$. We define the *loxodromic hybrid sphere*

$$\Sigma = \Sigma(E_0,Q_0;C_1) \cup \Sigma(E_2,Q_2;C_1) \tag{22.1}$$

exactly as in the parabolic case. As in the parabolic case, we call C_1 the *equator* of Σ. We call (E_0,Q_0) and (E_2,Q_2) the *poles* of Σ. The object $\Sigma(E_0,Q_0;C_1)$ is an embedded topological disk, analytic away from $C_1 \cup Q_0$. See [**S5**, §3] for the short proof.

22.1.3 Loxodromic Elevation

This section parallels Section 19.3. Let $\boldsymbol{R}(E,Q)$ denote the set of $\boldsymbol{R}$-arcs that intersect E as in Lemma 22.1. As in the parabolic case we can identify the cylinder $\boldsymbol{R}(E,Q)$ with the set of pairs of the form (x,v), where $x \in E-Q$ and v is a tangent vector at x, contained in the contact plane at x. Given $y \in S^3 - E$ we define $\Psi(y)$ to be the element of $\boldsymbol{R}(E,Q)$ that contains y.

The formula for Ψ is more complicated in the loxodromic case. Let Q_1 and Q_2 be the two endpoints of Q. Let $\widehat{Q}_1$ and $\widehat{Q}_2$ be lifts of Q_1 and Q_2, respectively. Let $\widehat{Q}$ be a vector such that $\langle \widehat{Q}, \widehat{Q}_j \rangle = 0$. Our map depends on our choice of lifts but only up to an affine homeomorphism of the range—i.e., scaling/translating the vertical $\boldsymbol{R}$ direction and rotating in the S^1 direction. Given $X \in S^3 - E$ we let $\widehat{X}$ be a lift of $\widehat{X}$. We define

$$\Psi(X) = \left(\left[\frac{\langle \widehat{X}, \widehat{Q} \rangle}{\sqrt{\langle \widehat{X}, \widehat{Q}_1 \rangle \langle \widehat{X}, \widehat{Q}_2 \rangle}} \right], \log \left| \frac{\langle \widehat{X}, \widehat{Q}_1 \rangle}{\langle \widehat{X}, \widehat{Q}_2 \rangle} \right| \right). \tag{22.2}$$

Here $[z] = z/|z|$. It is possible to take a branch of the square root function, which makes the above equation well defined. This is a basic topological fact, which is easiest to see when Q is a very short arc, so that $\langle \widehat{X}, \widehat{Q}_1 \rangle$ and $\langle \widehat{X}, \widehat{Q}_2 \rangle$ always have nearly the same argument. The range of Ψ is $S^1 \times \boldsymbol{R}$, as in the parabolic case.

Just as we had a parabolic cospinal family in Section 18.5, here we have a *loxodromic cospinal family*. The loxodromic cospinal family associated to the pair (E, Q) is the collection of cospinal spinal spheres whose poles lie on E and are harmonic with respect to (E, Q). The $\boldsymbol{R}$-arcs foliating these spinal spheres are the fibers of the loxodromic elevation map. This is just as in the parabolic case.

Versions of the Cospinal Intersection Lemma and the Elevation Image Lemma hold in the loxodromic case, but these are somewhat more technical to state. We refer the reader to [**S5**] for complete details. We also mention that [**S5**, Lemma 3.15], which is a version of Lemma 19.3, has a glitch in it. We take the opportunity to fix this in Section 22.4.

22.2 MAIN CONSTRUCTION

Let Γ' be one of the complex hyperbolic triangle groups for $s \in (\underline{s} - \epsilon, \overline{s})$. Let I_0, I_1, I_2 be the standard generators. The element I_j is a complex reflection in a $\boldsymbol{C}$-circle C_j. The $\boldsymbol{C}$-circles C_i and C_j intersect pairwise in a point p_{ij}. The element $I_i I_j I_k$ and its inverse $I_k I_j I_i$ are loxodromic. They stabilize a $\boldsymbol{C}$-circle E_j and a $\boldsymbol{C}$-arc $Q_j \subset E_j$. The endpoints of Q_j are the two fixed points of $I_i I_j I_k$. We choose Q_j so that it varies continuously with s and shrinks to a point as $s \to \overline{s}$. In analogy with the main construction of Chapter 20, we define

$$\Sigma_j = \Sigma(E_{j-1}, Q_{j-1}; C_j) \cup \Sigma(E_{j+1}, Q_{j+1}; C_j). \tag{22.3}$$

Note that Σ_j is a loxodromic $\boldsymbol{R}$-sphere because

$$I_j(E_{j\pm 1}, Q_{j\pm 1}) = (E_{j,\mp 1}, Q_{j\mp 1}). \tag{22.4}$$

The point is that I_j conjugates $I_j I_{j+1} I_{j+2}$ to $I_{j+1} I_{j+2} I_j$, with indices (as always) taken mod 3.

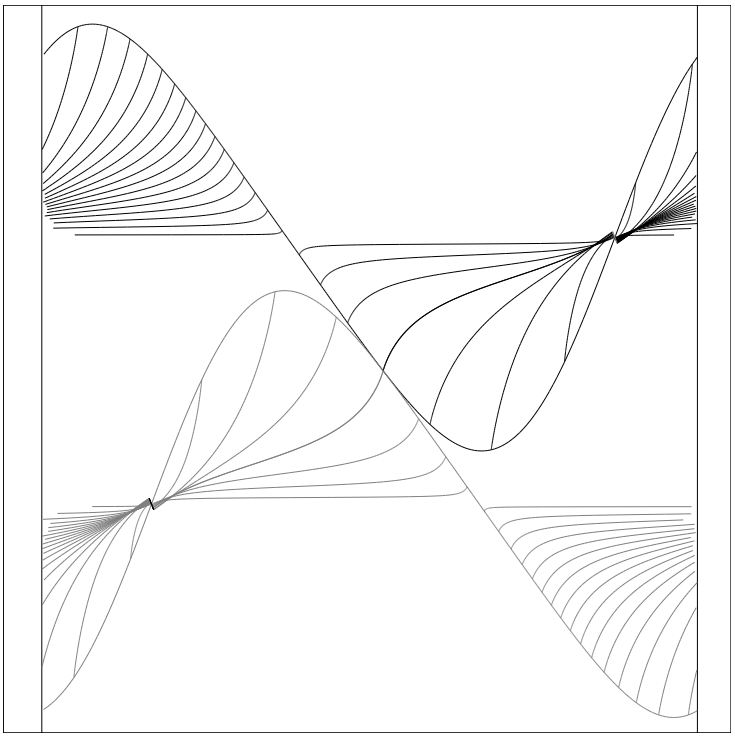

Figure 22.1: An elevation projection of two $\boldsymbol{R}$-spheres at parameter $s = \underline{s}$

Figure 22.1 shows a plot of $\Psi(\Sigma_1 - E_0)$ in black and $\Psi(\Sigma_2 - E_0)$ in gray for the parameter $s = \underline{s}$. The $\boldsymbol{R}$-arcs foliating $\Sigma(E_0, Q_0; C_1) - E_0$ are mapped to points in Figure 22.1. Likewise, the $\boldsymbol{R}$-arcs foliating $\Sigma(E_0, Q_0; C_2) - E_0$ are mapped to points. Notice the similarity to Figure 20.1, the picture at the parameter $\overline{s}$. Indeed, one has to look closely at the pictures to see the difference. (We tuned the aspect ratio to nearly match the one in Figure 19.1.) For $s \in [\underline{s}, \overline{s})$ the pictures (properly scaled) all look just about the same, and interpolate between Figure 22.1 and Figure 19.1. For comparison, we include a plot for the much smaller parameter $s = 1$. Computer evidence suggests that our construction works just fine for all parameters in $[0, \overline{s})$. However, certain technical details prevent us from easily dealing with parameters significantly less than $\underline{s}$. Fortunately, the groups for $s < \underline{s}$ are well understood, thanks to the work of Goldman and Parker.

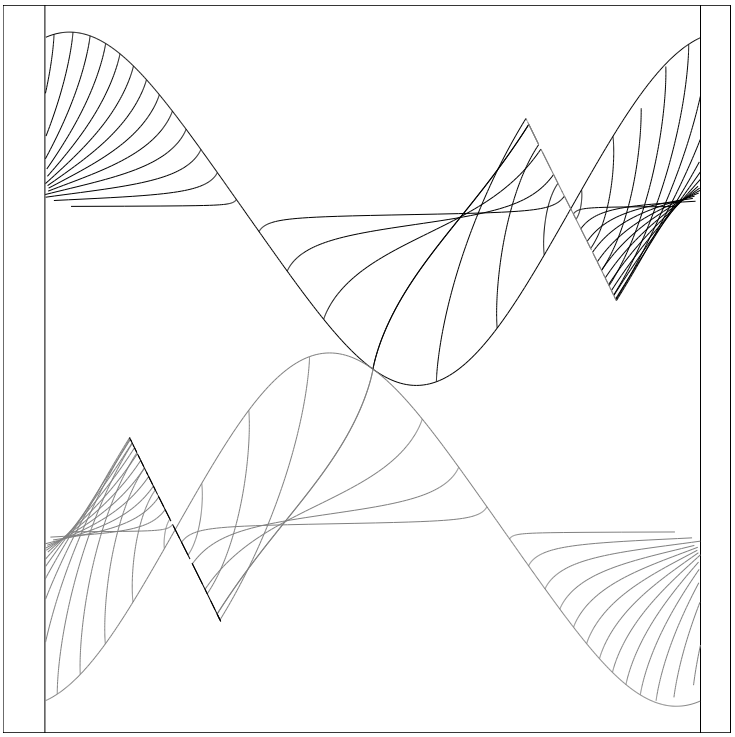

Figure 22.2: An elevation projection of two $\boldsymbol{R}$-spheres at parameter $s = 1$

In [**S5**] we show for $s \in (\underline{s}-\epsilon, \overline{s})$ that $\Psi(\Sigma_1 - E_0)$ and $\Psi(\Sigma_2 - E_0)$ intersect only at one point. Hence, $(\Sigma_1 - E_0) \cap (\Sigma_2 - E_0)$ is an $\boldsymbol{R}$-arc. On the other hand, $\Sigma_1 \cap E_0$ and $\Sigma_2 \cap E_0$ are both arcs. These arcs overlap. Thus $\Sigma_1 \cap \Sigma_2$ is the union of 2 arcs, which meet to make a T. In particular $\Sigma_1 \cap \Sigma_2$ is a contractible set. By symmetry, the same holds for all pairs (Σ_i, Σ_j).

We also show in [**S5**] that Σ_j is an embedded ball. Given the nature of the intersections, we have essentially the same topological picture as we developed in the parabolic case. That is, we have 3 balls B_0, B_1, B_2, with disjoint interiors, such that I_j exchanges the two components of B_j. This topological picture implies that Γ' is discrete and thereby proves the Goldman-Parker conjecture. The topological picture also serves as a basis for our proof in Section 22.3 that Γ is a horotube group.

22.3 HOROTUBE GROUP STRUCTURE

Our proof of the Shrinking Lemma, given in Section 20.6, goes through here almost word for word, with only one change. Case 2 does not occur because the element $g = I_0 I_1 I_2$ is not parabolic. Indeed, g would have to fix the endpoints of the (point or interval) $\bigcap \partial\beta_n$, a proper set of Q_2. But g fixes the endpoints of Q_2 and no other points.

Once we have the Shrinking Lemma, we can identify a fundamental domain for the action of Γ' as follows. We take the closure of $S^3 - B_0 \cup B_1 \cup B_2$ and delete the parabolic fixed points. Whereas we deleted 6 vertices in the parabolic case, here we only delete 3, because some of the parabolic elements are now loxodromic. The same horocusp picture obtains in a neighborhood of the 3 vertices of the fundamental domain. Thus we see that Ω/Γ is the union of a compact set K and 3 horocusps.

As in the critical case, the limit set Λ is precisely the nested intersection of all the balls in the orbit $\Gamma(B_0 \cup B_1 \cup B_2)$. The same argument as in Section 20.2 shows that Λ is porous. Now we know that Γ is a horotube group for $s \in (\underline{s} - \epsilon, \overline{s})$.

22.4 THE LOXODROMIC NORMALITY CONDITION

In this section we take the opportunity to fix a glitch in [**S5**, Lemma 3.15], the loxodromic version of Lemma 19.3. In stating [**S5**, Lemma 3.15], we left out the hypothesis that $\Psi(C_1^*) \cap \Psi(E_2)$ consists of two points. We didn't notice this omission because, in the part of the proof analogous to Lemma 19.3 (2), we wrote that $\Psi(\gamma' - E_0)$ is a single point of $\Psi(C_1)$. Actually $\Psi(\gamma' - E_0)$ consists of two points, one in $\Psi(C_1)$ and one in $\Psi(C_1^*)$. This is clear from our computer pictures, but somehow we missed it when writing [**S5**]. (In [**S0**], dealing with the parabolic case, we took an approach not involving the normality condition, so the issue didn't come up there.)

To fix [**S5**, Lemma 3.15] we need to prove the loxodromic version of Lemma 20.13. The result is clear from figures like Figure 22.4, which look almost the same independent of parameter $s \in [\underline{s}, \overline{s}]$, but it certainly requires proof. Though we state our argument for the interval $[\underline{s}, \overline{s})$, the fact that we always get inequalities shows that the argument actually works for $(\underline{s} - \epsilon, \overline{s})$ as long as ϵ is small enough. Incidentally, our proof should give the reader an ample dose of the kind of nitty-gritty details contained in [**S5**].

22.4.1 An Estimate for the Lemniscate

We include an estimate for lemniscates, which we use in Section 22.4. The *standard lemniscate* has the polar coordinate equation $r^2 = \cos(2\theta)$. One of the lobes of this lemniscate consists of points having a nonnegative real part. We let L denote this lobe. Say that a *radiant* of L is a line segment that joins the origin to another point of L. For any $s \in (0,1)$ there is one radiant R_s of L, which lies in the sector bounded by the lines $\theta = 0$ and $\theta = \pi/4$. The radiant R_s divides the region bounded by L into two pieces, the smaller of which has area $A(s)$.

Lemma 22.2 *If $s < 4/5$, then $A(s) < s^4/7$.*

Proof: Let $\alpha \in (0, \pi/4)$ be the angle such that R_s is contained in the line $\theta = \alpha$. The region of interest lies in the sector bounded by the lines $\theta = \alpha$

and $\theta = \pi/4$. Using the formula for area in polar coordinates, we have

$$A(s) = \int_\alpha^{\pi/4} \frac{r^2\ d\theta}{2} = \int_\alpha^{\pi/4} \frac{\cos(2\theta)d\theta}{2} = \frac{1}{4}\big(1 - \sin(2\alpha)\big)$$

$$= \frac{1}{4}\big(1 - \sqrt{1 - \cos^2(2\alpha)}\big) = \frac{1}{4}(1 - \sqrt{1 - s^4}) \leq \frac{s^4}{7}.$$

It is a calculus exercise to show that the last inequality holds for $s \in [0, 4/5]$. □

We can get a more general result by scaling.

Corollary 22.3 (Lemniscate Estimate) *Suppose that L is a lobe of a lemniscate having radius R. Suppose $\lambda/R < 4/5$. A length λ radiant of L divides the region bounded by L into two pieces, the smaller of which has area at most $\lambda^4/(7R^2)$.*

22.4.2 Rough Geometric Features

Here we recall some estimates from [**S5**]. (For the parameter $\overline{s}$ these estimates are just direct calculations.) We normalize so that we have the following conditions.

- $E_0 = (\{0\} \times \boldsymbol{R}) \cup \infty$.
- $\pi(C_1)$ has radius 1 and is centered at a point $t \in (0, \infty)$. In this case $\pi(C_1^*)$ is centered at $c = -t$. In [**S5**, Section 5.3] we prove that $t^{-2} \in [9.5, 10]$. Hence,

  $$t \in \left[\frac{1}{\sqrt{10}}, \frac{1}{3}\right] \tag{22.5}$$

  for all parameters $s \in [\underline{s}, \overline{s}]$.
- $Q_0 \subset E_0$ is the unbounded interval whose endpoints have the form $(0, \pm u)$. In [**S5**, Lemma 6.2] we prove that

  $$u > 4.2 \tag{22.6}$$

  for all parameters $s \in [\underline{s}, \overline{s}]$.

As in the parabolic case, let r' be a ray through the origin. Let c be the center of $\pi(C_1^*)$, and let e be the center of $\pi(E_2)$.

Lemma 22.4 *Equation 20.8 is true for all parameters in $[\underline{s}, \overline{s}]$. That is,*

$$|c| < \frac{2}{7}|e|, \qquad |r' \cap \pi(C_1^*)| < 3|r' \cap \pi(E_2)|.$$

Proof: Using the fact that the complex reflection I_1 maps E_0 to E_2 and acts as ordinary inversion about t when restricted to $\boldsymbol{R}$, we get that $\pi(E_2)$ intersects $\boldsymbol{R}$ in the points $t = I_1(\infty)$ and $-1/t + t = I_1(0)$. Hence, $\pi(E_2)$ has center $e = t - 1/(2t) < -7/6$. Since C_1^* is just the rotated version of C_1, we see that C_1 has center $c = -t \in (-1/3, 0)$ and radius 1. Clearly, $|e| > 2|c|/7$. This is the first half of Equation 20.8. Let (*) denote the second half of Equation 20.8. Note that $\pi(C_1^*)$ intersects $\boldsymbol{R}$ in the two points $-1 - t$ and $1 - t$. The worst case of (*) occurs when the ray r' converges to the positive real axis. This limit is $(1-t)/t < 3$ for t as in Equation 22.5. □

Let F be any fiber of the map Ψ. Then the projection $\pi(F)$ is one lobe of a lemniscate whose vertex lies at the origin. Recall from Section 22.4.1 that the radius of $\pi(F)$ is the length of the longest radiant of $\pi(F)$.

Lemma 22.5 *Let F be a fiber of Ψ that contains a point $p = (z, v)$ of C_1^*, with $t > 0$. Then $\pi(F)$ has radius at least $\sqrt{8.5/v}$. In particular $\pi(F)$ has radius at least 3.57.*

Proof: Let r denote the radius of $\pi(F)$. One lobe of the standard lemniscate has area $1/2$, and so $\pi(F)$ has area $r^2/2$. Now F intersects E_0 in the two points $v_1 \in [0, v]$ and v_2. From the definition of the loxodromic elevation map, these points satisfy $v_1 v_2 = u^2 \geq (4.2)^2$. The Lift Principle I, from Section 2.4, gives $2r^2 = v_2 - v_1$. Hence,

$$r^2 = \frac{v_2 - v_1}{2} \geq \frac{1}{2}\left(\frac{(4.2)^2}{v} - v\right). \tag{22.7}$$

Given that the center of mass of C_1^* lies on the segment $(-1/3, 0) \times 0$ and $\pi(C_1^*)$ has radius 1, the Lift Principle II says that $v \in [0, 2/3)$. A bit of calculus shows that the right-hand side of Equation 22.7 exceeds $8.5/v$ for $v \in [0, 2/3]$. □

Remark: We will use Lemma 22.5 explicitly several times in our arguments, but we also use it implicitly. Lemma 22.5 shows that the fibers of Ψ are always large in comparison to the circles C_1^* and E_2. This large size is implicit in our geometric picture of how the fibers of Ψ encounter C_1^* and E_2. If F is a fiber that contains a point of C_1^*, then we will trace F out counterclockwise. Then F rises upwards, encountering C_1^* before rising halfway up. In other words, the first half of F contains the intersection point $F \cap C_1^*$. We check this for a single fiber at a single parameter, and then the large size of all the fibers guarantees that the "first half" condition persists for all fibers. The same condition holds for E_2 for the same reasons.

22.4.3 The Main Argument

Figure 22.3 shows a fairly accurate picture of $\pi(C_1^*)$ and $\pi(E_1)$, with certain arcs thickened. The preimages of the semicircles in the lower half-plane

are the ones with positive vertical coordinate. For $p \in \mathcal{H}$ we write $p = (\pi(p), v(p))$, as in Lemma 20.13.

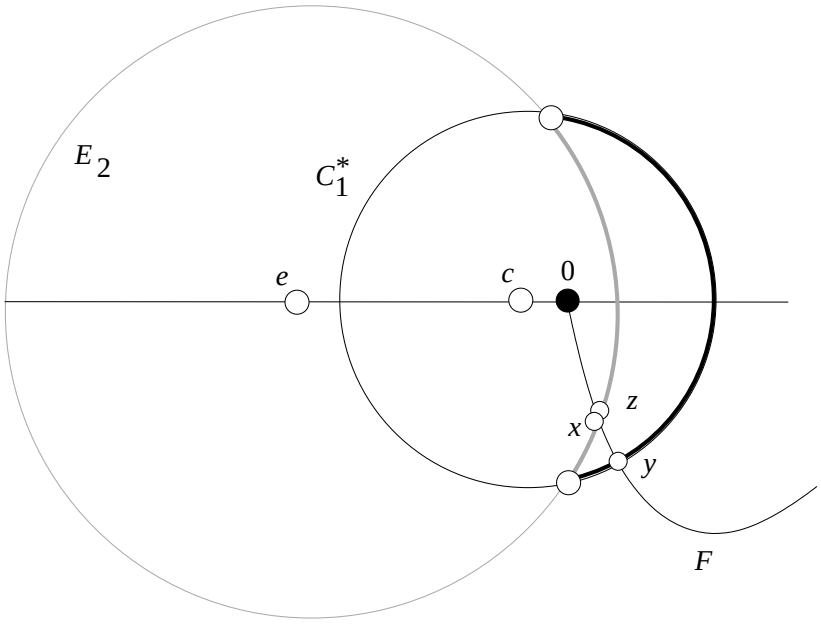

Figure 22.3: Projections to $\boldsymbol{C}$ of C_1^* (black) and E_2 (gray)

Suppose that F is a fiber of Ψ that intersects C_1^* and E_2, respectively, in points x and y, which project to the thickened arcs. By symmetry, it suffices to consider the case when these points project into the lower half-plane. The curve $\pi(F)$ is one lobe of a lemniscate centered at 0. The point $z \in E_2$ is such that $\pi(z)$ is colinear with 0 and $\pi(y)$. Given the rough sizes of our circles and the convexity of a lemniscate lobe, the point $\pi(z)$ lies closer to the real axis than does $\pi(x)$. Hence, $v(x) > v(z)$.

Given that Equation 20.8 holds, the same argument as in Lemma 20.13 shows that $v(z) > v(y)$. Hence, $v(x) > v(y)$. When we trace F counterclockwise, this fiber rises upwards, encountering x first. Hence, $v(x) < v(y)$. (Compare the remark at the end of the last section.) We have a contradiction.

Now suppose $\pi(x)$ and $\pi(y)$ lie in the thin arcs above, as shown in Figure 22.4. From Lemma 22.5 we know that $\pi(F)$ has radius $R > \sqrt{8.5/v(x)}$. Following Section 22.4.1, we interpret the segment from 0 to $\pi(y)$ as a radiant of $\pi(F)$. This radiant divides the region bounded by $\pi(F)$ into two regions, the smaller of which is shaded. (This is one of the places where we are implicitly using the large size of $\pi(F)$ guaranteed by Lemma 22.5.) Let λ be the length of the radiant.

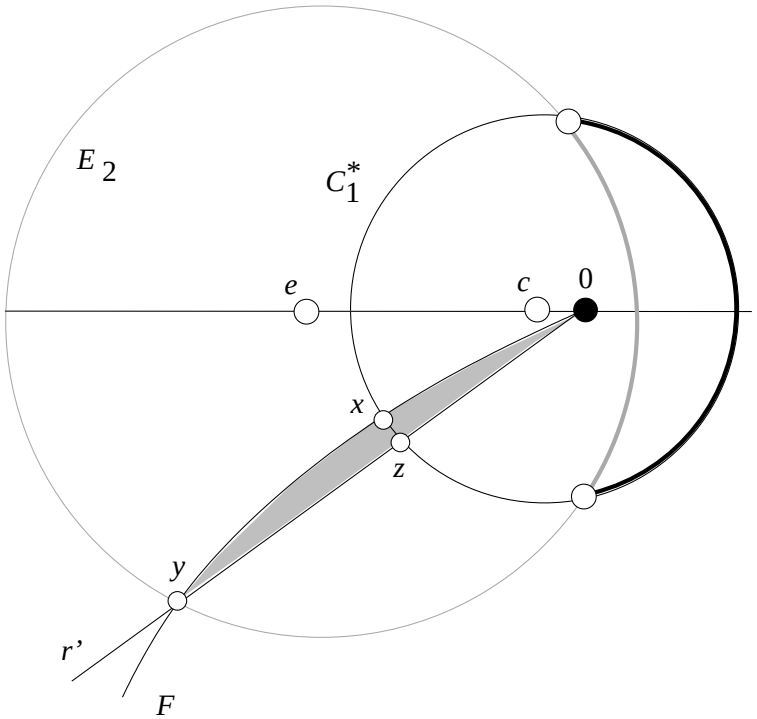

Figure 22.4: Projections to $\boldsymbol{C}$ of C_1^* (black) and E_2 (gray)

Lemma 22.6 $\lambda/R < 4/5$.

Proof: We know that $R > 3.57$ by Lemma 22.5. On the other hand, from our description of $\pi(E_2)$ in Lemma 22.4 we see that $\lambda \leq 9/\sqrt{10} = 2.84506...$. Dividing the one bound by the other gives .797..., which is less than $4/5$. □

Case 1: Suppose $v(z) > v(x)$ and $\lambda \leq 7/3$. Since our points project to the thin arcs, we get a stronger bound for the right-hand side of Equation 20.8:

$$|r' \cap \pi(C_1^*) < |r' \cap \pi(E_2)|,$$

where r' is the ray containing 0, $\pi(y)$, and $\pi(z)$. Hence, $\Delta_e > (7/2)\Delta_c$. Hence, $v(y) > 3.5v(z) > 3.5v(x)$. Hence, $v(y) - v(x) > 2v(x)$. From the Lemniscate Estimate Lemma our shaded region has area at most

$$\zeta = \frac{(7/3)^4}{7 \times (8.5/v(x))} < .5v(x).$$

By the Lift Principle I, the amount that F rises as it travels from its lowest endpoint to $F \cap E_2$ is $4\zeta < 2v(x)$. Hence, $v(y) - v(x) < 2v(x)$, a contradiction.

Case 2: Suppose $v(z) > v(x)$ and $\lambda > 7/3$. Since $\pi(C_1^*)$ has radius 1

and is centered on a point within 1/3 of the origin, we have $|\pi(z)| \leq 4/3$. But this means

$$|r' \cap \pi(C_1^*)| < \frac{4}{7}|r' \cap \pi(E_2)|.$$

Hence, $\Delta_e/\Delta_c \geq (7/2) \times (7/4) > 6$. Hence, $v(y) - v(x) > 5v(x)$. Given the description of $\pi(E_2)$ in Lemma 22.4, we see that $\lambda < 9/\sqrt{10}$. From the Lemniscate Estimate Lemma we now have

$$\zeta = \frac{(9/\sqrt{10})^4}{7 \times (8.5/v(x))} < 1.15v(x).$$

Hence, $v(y) - v(x) < 4\zeta < 5v(x)$, a contradiction.

Case 3: Suppose $v(z) \leq v(x)$. This happens only if $\pi(z)$ is closer to $\boldsymbol{R}$ than is $\pi(x)$. By symmetry $\pi(C_1^*)$ and $\pi(E_2)$ intersect in the imaginary axis. Hence, $\mathrm{Re}(\pi(x)) \in (2c, 0] \subset (-2/3, 0]$. The line segment connecting c to $\pi(x)$ has length 1 and travels horizontally at most 1/3. Hence, both $\pi(x)$ and $\pi(y)$ are at least $\sqrt{8}/3$ from $\boldsymbol{R}$. Hence, Δ_e has area at least $1/2 \times 7/6 \times \sqrt{8}/3 > 1/2$. The Lifting Principle II Corollary gives $v(y) > 2$. On the other hand, $v(x) < 2/3$ for any choice of $x \in C_1^*$. (See Lemma 22.5.) Hence, $v(y) - v(x) > 2v(x)$.

Given that $\pi(x)$ is farther from the real axis than from the imaginary axis, the ray r' has slope at least 1. The radius of $\pi_1(E_2)$ is at most $\sqrt{10}/2$. Hence, λ has length at most $\sqrt{5} < 7/3$. But now the argument in Case 1 shows that $v(y) - v(x) < 2v(x)$, a contradiction.

Chapter Twenty Three

The Groups far from the Critical Value

23.1 DISCUSSION OF PARAMETERS

In [**GP**] Goldman and Parker proved that the complex hyperbolic ideal triangle group $\Gamma = \rho_s(G)$ is discrete, provided that $s \in [0, \underline{s}]$. A certain transition occuring at the parameter $\underline{s}$ made it impossible for their method to work in $(\underline{s}, \overline{s}]$. The transition at $\underline{s}$ is invisible to our proof given in the previous chapter. However, our proof in the previous chapter becomes much more difficult for parameters much less than $\underline{s}$ for other technical reasons. For this reason, we proved Theorem 1.3 for parameters in $(\underline{s} - \epsilon, \overline{s})$, avoiding very small parameters but still covering the somewhat tricky parameter $\underline{s}$. Here we prove Theorem 1.3 for $s \in [0, \underline{s})$. Our proof uses a picture we developed in [**S1**] but not any of the computer-aided parts of [**S1**]. One could also base the proof on the work in [**GP**].

23.2 THE CLIFFORD TORUS PICTURE

The three $\boldsymbol{C}$-circles C_0, C_1, C_2 lie on the *Clifford torus*

$$T = \{(z, w) |\ |z| = |w| = \sqrt{2}/2\} \subset S^3. \tag{23.1}$$

Let $\mathcal{C}$ denote the orbit of T under Γ. We say that two distinct $T_1, T_2 \in \mathcal{C}$ are *almost disjoint* if the closure of one component of $S^3 - T_1$ contains T_2 (and vice versa). Here we will sketch the argument, given in [**S1**], that every two elements of $\mathcal{C}$ are either disjoint or almost disjoint for $s \in [0, \underline{s}]$.

Let

$$h_1 = I_1, \qquad I = I_0, \qquad h_2 = I_0 I_2 I_0. \tag{23.2}$$

Let H be the subgroup generated by h_1 and h_2. Also, let $\pi : S^3 \to \{0\} \times \boldsymbol{C}$ be given by the map $\pi(z, w) = (0, w)$. The image $\pi(S^3)$ is the closed unit disk, which we interpret as the union of the hyperbolic disk and its ideal boundary. The set $\pi(T)$ is the circle of radius $\sqrt{2}/2$ centered at the origin. The subgroup H commutes with the action of π and acts as the infinite dihedral group, generated by 2 rotations. It follows from these facts that the projection $\pi(h(T))$ is a hyperbolic circle for all $h \in H$. In Figure 23.1 we color in these circles to make them more visible.

The two tori $h_1(T)$ and $h_2(T)$ each intersect T along a $\boldsymbol{C}$-circle. The projections $\pi(h_1(T))$ and $\pi(h_2(T)$ each intersect $\pi(T)$ along points. Figure

21.1 shows the pattern of intersection in $\{0\} \times \boldsymbol{C}$. To see the picture in S^3 we note that the fibers of each point in the open unit disk are $\boldsymbol{C}$-circles. Thus, all the circles in Figure 23.1 become tori, and all the points of intersection become $\boldsymbol{C}$-circles. All the points of intersection lie on the axis A, which is the axis of the loxodromic element $\pi \circ h_1 h_2 \circ \pi^{-1}$. We have just drawn the first 4 circles because they quickly get too small to draw.

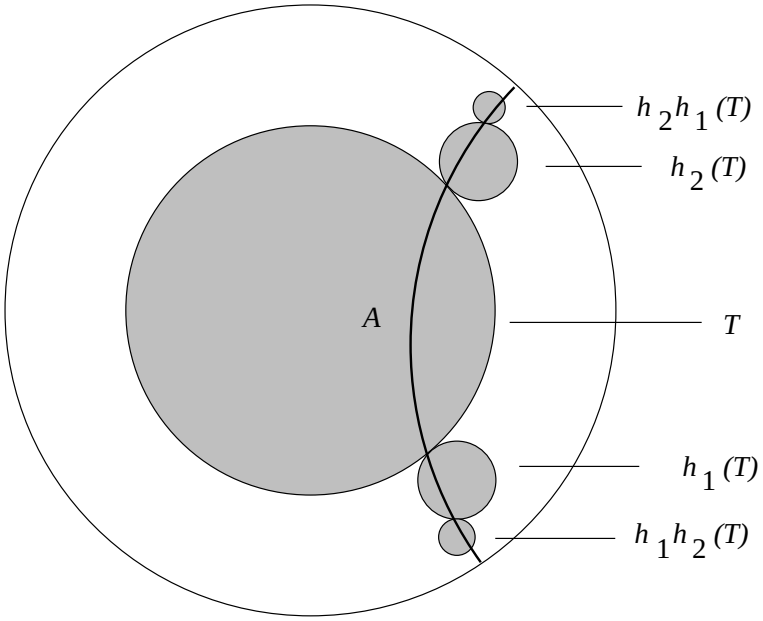

Figure 23.1: Projection to $\boldsymbol{C}$ of the H-orbit of T.

Once we know the intersection patterns of all the tori of the form $h(T)$ for $h \in H$, we use a ping-pong argument to finish the discreteness proof. The point is that Γ' is generated by H and I. Thus, if the tori in the orbit $H(T)$ are pairwise almost disjoint, then basic separation properties imply that the tori in the full orbit $\mathcal{C}$ are pairwise disjoint. For instance, $Ih_1h_2(T)$ is almost disjoint from $h_1h_2h_1(T)$ because these are both almost disjoint from T and lie in different components of $S^3 - T$.

When $s \to \underline{s}$ the two circles $\pi \circ h_1(T)$ and $\pi \circ h_2(T)$ approach each other, becoming tangent at $\underline{s}$. For larger parameters the circles overlap. In particular, the orbit $\mathcal{C}$ does not consist of almost disjoint tori for $s > \underline{s}$.

23.3 THE HOROTUBE GROUP STRUCTURE

For $s \in [0, \underline{s})$ the discreteness of Γ follows immediately from the fact that $\mathcal{C}$ consists of almost disjoint tori. Moreover, looking at the pattern of tori we see that every element $g \in \Gamma$ is either loxodromic or parabolic and that g is parabolic iff it is conjugate to a power of $I_i I_j$.

Looking at Figure 23.1 we see that there is one component of $S^3 - \bigcup h(T)$ whose boundary contains all the tori of the form $h(T)$. This component is the preimage, under π, of the complement of the unshaded region. We let F denote the closure of this component, minus all the fixed points of elements of H. We claim that every point in Ω is equivalent to a point in F. The technical difficulty of this claim is that, for example, one might worry that some nested sequence of tori does not shrink to a point. The following result takes care of this detail.

Lemma 23.1 *For any $\epsilon > 0$ there are only finitely many tori in $\mathcal{C}$ having spherical diameter greater than ϵ.*

Proof: Let $\{\tau_n\}$ be a sequence of counterexamples. Passing to a subsequence we have one of two situations. Either we can find nested solid tori $\{U_n\}$ such that $\tau_n = \partial U_n$, or else we can find solid tori $\{V_n\}$ with disjoint interiors such that $\tau_n = \partial V_n$.

Consider the latter case first. In this case, we can use the fact that Γ is generated by I and H to reduce to the case that $\tau_n = h_n(T)$ for some $h_n \in H$. In this case we can look at Figure 23.1 and see that actually τ_n shrinks to one of the fixed points of $h_1 h_2$.

In the former case we can pass to a subsequence and use the symmetry of the group to reduce to one of two cases.

Case 1: $\tau_n = (I_1 I_2)^n(T)$. In this case the same argument as in Case 3 of the Shrinking Lemma (Chapter 20) shows that $\bigcap \tau_n$ is a single point.

Case 2: There are infinitely many indices k_n such that $(\tau_{k_n}, \tau_{k_n+1})$ consists of disjoint tori. In this case there might be infinitely many such pairs mod Γ, but every pair is equivalent to a pair of the form $(T, h(T))$ for some $h \in H$. For all such pairs there is a uniform upper bound to the cross ratio $[a, b, c, d]$, where $a, d \in T$ and $c, b \in h(T)$. Thus, an argument just like the one in Lemma 2.6 works here. □

Now we know that Ω is tiled by translates of F, and we use this fact to analyze Ω/Γ. For two parameters $s_1, s_2 \in [0, \underline{s})$ there is a clear homeomorphism from F_1 to F_2. Here F_j stands for the set associated to s_j. The point is that the two pictures like Figure 23.1 are homeomorphic. The homeomorphism from F_1 to F_2 can be chosen to respect the two actions of H. In this way, we can extend the homeomorphism from F_1 to F_2 to a homeomorphism from Ω_1 to Ω_2 which intertwines the two group actions. Hence, Ω_1/Γ_1 and Ω_2/Γ_2 are homeomorphic. In Chapter 4 we we analyzed the quotient Ω_1/Γ_1 for the parameter $s_1 = 0$. Hence, Ω_2/Γ_2 is homeomorphic to the union of a compact set and a finite number of horocusps. The only fine point is that the noncompact ends of Ω_2/Γ_2 are really horocusps and not just homeomorphic copies of horocusps. To deal with this fine point, we note that the noncompact ends of Ω_2/Γ_2 are actually CR-isomorphic to the noncompact ends of

Ω_1/Γ_1 because both ends are based on conjugate parabolic elements.

In the Fuchsian case there is a beautiful horotube assignment for the group. The orthogonal projection $\boldsymbol{CH}^2 \to \boldsymbol{RH}^2$ carries the Clifford torus to the boundary of the square bounded by the sides $x = \pm\sqrt{2}/2$ and $y = \pm\sqrt{2}/2$. This is the boundary of a regular geodesic quadrilateral. The other tori in $\mathcal{C}$ project to translates of this quadrilateral, and the whole pattern gives the usual tiling of $\boldsymbol{RH}^2$ by ideal quadrilaterals. We can take the tree in $\boldsymbol{RH}^2$ dual to this tiling and lift to S^3. This gives the complex Ψ^∞, whose complementary pieces are horotubes. Note that each torus of $\mathcal{T}$ intersects 4 horotubes, each of which is based at some point of this torus. We can use the homeomorphism above to transfer this horotube assignment to the groups associated to the other parameters in $[0, \underline{s})$. The only detail left in our proof is the demonstration that Λ is porous. In light of our horotube assignment and the porosity argument we gave in Section 20.2, the following result implies that Λ is porous.

Lemma 23.2 *Suppose that $g \in PU(2,1)$ is such that $g(\Omega)$ contains no ball of radius $1/100$. Then there is some Clifford torus $\tau \in \mathcal{C}$ such that $g(\tau)$ has spherical diameter at least $1/100$.*

Proof: Let $\mathcal{Q}$ denote the regular ideal quadrilateral tiling of $\boldsymbol{H}^2$. As mentioned above, there is a canonical bijection between $\mathcal{Q}$ and $\mathcal{C}$.

Let $\tau \in \mathcal{C}$. Then $g(\Lambda) \cap g(\tau)$ consists of 4 points, and $g(\Lambda) - g(\tau)$ consists of 4 arcs. We call these the arcs *associated* to τ. Two consecutive associated arcs lie in different components of $S^3 - g(\tau)$. Two opposite associated arcs lie in the same component.

Since $g(\Lambda)$ does not lie in a hemisphere of S^3 and $g(\Lambda)$ is the union of (4 points and) the 4 arcs associated to τ, we see that at least one arc associated to τ has diameter at least $1/8$. Let $|\tau|$ denote the second largest diameter of an arc associated to τ. We claim that there exists some τ for which $|\tau| > 1/100$.

Suppose that our claim is false. Then we can find an infinite sequence $\{\tau_n\}$ such that the quadrilaterals in $\mathcal{Q}$ corresponding to τ_n and τ_{n+1} share an edge, and the long arc associated to τ_{n+1} is a proper subset of the long arc associated to τ_n. The picture here is that all the vertices of $g(\tau_n)$ are crowded together in a small part of S^3, with one of the arcs associated to τ_n going most of the way around S^3 before connecting up two of the vertices of $g(\tau_n)$. However, the intersection of all our long arcs is a point, by Lemma 2.6. This is impossible for long arcs.

Now we know that there is some torus τ such that at least two arcs associated to τ have diameter at least $1/100$. Suppose first that two consecutive arcs associated to τ have big diameter. These arcs lie in different components of $S^3 - g(\tau)$. We conclude that both components of $S^3 - g(\tau)$ have big diameter. Hence, $g(\tau)$ has big diameter as well.

On the other hand, suppose that the two opposite arcs L_1, L_3 associated to τ have big diameter, but nonetheless $g(\tau)$ has vanishingly small diameter.

Let a_j be an endpoint of L_j. For each $j = 1, 3$, we can find points on L_j that are very far from $a_1 \cup a_3$. Hence,

$$\sup_{(b_1,b_3)\in L_1\times L_3} \frac{\langle a_1, b_1\rangle\langle a_3, b_3\rangle}{\langle a_1, a_3\rangle\langle b_1, b_3\rangle} \tag{23.3}$$

is unboundedly large. Note that the quantity in Equation 23.3 is finite because b_1 and b_3 lie on disjoint compact sets. There are only finitely many pairs (L_1, L_3) mod $PU(2, 1)$, and so the quantity in Equation 23.3 is uniformly bounded. This is a contradiction. □

Bibliography

[**ACT**] D. Allcock, J. A. Carlson, and D. Toledo, *The complex hyperbolic geometry of the moduli space of cubic surfaces*, J. Algebraic Geom. **11** (2002), 659–724.

[**AGG**] S. Anan'in, H. Grossi, and N. Gusevskii, *complex hyperbolic structures on disc bundles over surfaces, I* (2003), preprint.

[**AG**] S. Anan'in and N. Gusevskii, *complex hyperbolic structures on disc bundles over surfaces, II: Example of a trivial bunble* (2005), preprint.

[**A**] J. W. Anderson, *Incommensurability criteria for kleinain groups*, Proc. Amer. Math. Soc. **130** (2002), 253–258.

[**B**] A. F. Beardon, *The Geometry of Discrete Groups*, Grad. Texts in Math. **91** Springer, New York, 1995.

[**Bo**] B. H. Bowditch, *Geometrical finiteness for hyperbolic groups*, J. Funct. Anal. **113** (1993), 245–317.

[**Br**] U. Brehm, *The shape invariant of triangles and trigonometry in two-point homogeneous spaces*, Geom. Dedicata **33** (1990), 59–76.

[**BH**] M. R. Bridson and A. Haefliger, *Metric Spaces of Non-positive Curvature*, Grund. Math. Wiss. **319**, Springer, Berlin, 1999.

[**CEG**] R. D. Canary, D.B.A. Epstein, and P. Green, "Notes on notes of Thurston", in *Analytical and Geometric Aspects of Hyperbolic Space (Coventry/Durham, 1984)*, London Math. Soc. Lecture Note Ser. **111**, Cambridge Univ. Press, Cambridge, 1987, 3–92.

[**CG**] S. S. Chen and L. Greenberg, "Hyperbolic spaces", in *Contributions to Analysis (A Collection of Papers Dedicated to Lipman Bers)*, Academic Press, New York, 1974, 49–87.

[**DM**] P. Deligne and G. D. Mostow, *Commensurabilities among Lattices in* PU(1,n), Annals of Mathematics Studies **132**, Princeton Univ.

Press, Princeton, NJ, 1993.

[**El**] Y. Eliashberg, *Topological characterization of Stein manifolds of dimension >2*, Internat. J. Math. **1** (1990), 29–46.

[**E**] D.B.A. Epstein, "Complex Hyperbolic Geometry", in *Analytical and Geometric Aspects of Hyperbolic Space (Coventry/Durham, 1984)*, London Math. Soc. Lecture Note Ser. **111**, Cambridge Univ. Press, Cambridge, 1987, 93–111.

[**F**] E. Falbel, *A Branched Spherical CR Structure on the complement of the figure eight know with discrete holonomy* (2005), preprint.

[**FP**] E. Falbel and J. R. Parker, *The moduli space of the modular group in complex hyperbolic Geometry*, Invent. Math. **152** (2003), 57–88.

[**FZ**] E. Falbel and V. Zocca, *A Poincaré's polyhedron theorem for complex hyperbolic geometry*, J. Reine Angew Math. **516** (1999), 133–158.

[**G**] W. M. Goldman, *Complex Hyperbolic Geometry*, Oxford Mathematical Monographs, Clarendon Press/Oxford University Press, New York, 1999.

[**GKL**] W. M. Goldman, M. Kapovich, and B. Leeb, *Complex Hyperbolic manifolds homotopy equivalent to a Riemann surface*, Comm. Anal. Geom. **9** (2001), 61–95.

[**GP**] W. M. Goldman and J. R. Parker, *Complex hyperbolic ideal triangle groups*, J. Reine Agnew Math. **425** (1992), 71–86.

[**Go**] R. E. Gompf, *Handlebody construction of Stein surfaces*, Ann. of Math. **148** (1998), 619–693.

[**GLT**] M. Gromov, H. B. Lawson Jr., W. Thurston, *Hyperbolic* 4*-manifolds and conformally flat* 3*-Manifolds*, Inst. Hautes Études Sci. Publ. Math. **68** (1988), 24–45.

[**KT**] Y. Kamishima and T. Tsuboi, *CR-structures on Seifert manifolds*, Invent. Math. **104** (1991), 149–163.

[**La**] S. Lawton, SL_3 *Character Varieties and* $\boldsymbol{RP}^2$ *Structures on a Trinion*, Ph.D. Thesis, University of Maryland, College Park, MD, 2006.

[**L**] P. Lisca, *Symplectic fillings and positive scalar curvature*, Geom. Topol. **2** (1998), 103–116.

[**M**] B. Maskit, *Kleinian Groups* Grund. Math. Wiss. **287**, Springer, Berlin, 1988.

[**Mat**] S. Matveev, *Algorithmic Topology and Classification of* 3-*Manifolds*, Algorithms and Computation in Mathematics **9**, Springer, Berlin, 2003.

[**McM**] C. McMullen, *Self-similarity of Siegel disks and Hausdorff dimension of Julia sets*, Acta Math. **180** (1998), 247–292.

[**M0**] G. D. Mostow, *Strong Rigidity of Locally Symmetric Spaces*, Annals of Math. Studies **78**, Princeton Univ. Press, Princeton, NJ, 1973.

[**M1**] —, *On a remarkable class of polyhedra in complex hyperbolic space*, Pacific. J. Math. **86** (1980), 171–276.

[**NZ**] W. D. Neumann and D. Zagier, *Volumes of hyperbolic three-manifolds*, Topology **24** (1985), 307–332.

[**Pa**] J. R. Parker, *Unfaithful Complex Hyperbolic Triangle Groups* (2005), preprint.

[**PP**] C. Petronio and J. Porti, *Negatively oriented ideal triangulations and a proof of Thurston's hyperbolic Dehn filling theorem*, Expo. Math. **18** (2000), 1–35. arXiv:math.GT/9901045

[**P**] A. Pratoussevitch, *Traces in complex hyperbolic triangle groups*, Geom. Dedicata **111** (2003), 159–185.

[**R**] J.G. Ratcliffe, *Foundations of Hyperbolic Manifolds*, Graduate Texts in Math. **149**, Springer, New York, 1994.

[**Sa**] H. Sandler, *Traces in* SU(2,1) *and complex hyperbolic ideal triangle groups*, Algebras Groups Geom. **12** (1995), 139–156.

[**S0**] R. E. Schwartz, *Degenerating the complex hyperbolic ideal triangle groups*, Acta Math. **186** (2001), 105–154.

[**S1**] —, *Ideal triangle groups, dented tori, and numerical analysis*, Ann. of Math. **153** (2001), 533–598.

[**S2**] —, *Real hyperbolic on the outside, complex hyperbolic on the inside*, Invent. Math. **151** (2003), 221–295.

[**S3**] —, "Complex hyperbolic triangle groups", in *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*,

Higher Ed. Press, Beijing, 2002, 339–349.

[**S4**] —-, *Modular circle quotients and PL limit sets*, Geom. Topol. **8** (2004), 1–34.

[**S5**] —, *A better proof of the Goldman-Parker conjecture*, Geom. Topol. **9** (2005), 1539–1601.

[**S6**] —, *An interactive proof of the G.-P. conjecture*, Java applet, 2004, **www.math.umd.edu/∼res/Java/App45/test1.html**

[**Se**] C. Series, "Geometrical methods of symbolic coding", in *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces (Trieste, 1989)*, Oxford Sci. Publ., Oxford Univ. Press, New York, 1991.

[**T0**] W. Thurston, *The Geometry and Topology of 3-Manifolds*, Princeton University notes, 1978.

[**T1**] —, "Shapes of polyhedra and triangulations of the sphere" in *The Epstein Birthday Schrift*, Geom. Topol. Monographs **1** (1998), 511–549.

[**W-G**] Justin Wyss-Gallifent, *Discreteness and Indiscreteness Results for Complex Hyperbolic Triangle Groups*, PhD thesis, University of Maryland, College Park, MD, 2000.

[**W**] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer*, 4th ed., Wolfram Media, Champaign, IL, 2000.

Index