GEOMETRIES OF HYPERBOLIC SURFACES WITH AND WITHOUT BOUNDARY

by

Kimberly A. Romanelli

B.S. Mathematics, Slippery Rock University, 2009

M.A. Mathematics, Villanova University, 2011

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Kimberly A. Romanelli

It was defended on

July 23rd, 2018

and approved by

J. DeBlois, PhD, Associate Professor, Dept. of Mathematics, University of Pittsburgh

T. Hales, PhD, Professor, Dept. of Mathematics, University of Pittsburgh

C. Lennard, PhD, Associate Professor, Dept. of Mathematics, University of Pittsburgh

M. Stover, PhD, Assistant Professor, Dept. of Mathematics, Temple University

Dissertation Director: J. DeBlois, PhD, Associate Professor, Dept. of Mathematics, University of Pittsburgh

GEOMETRIES OF HYPERBOLIC SURFACES WITH AND WITHOUT BOUNDARY

Kimberly A. Romanelli, PhD

University of Pittsburgh, 2018

In this dissertation, I will investigate three different points of view of maximizing packings on complete hyperbolic surfaces with finite area, possibly with geodesic boundary. This optimization takes place over the Teichmüller space of each surface.

First I find a sharp upper bound for the packing radius, and consequently the injectivity radius, of a surface with Euler characteristic χ, n cusps, and b geodesic boundary components. In particular, I do not fix these boundary lengths. This is an extension of the results found in DeBlois's papers [3] and [4] to the with-boundary setting.

Second, I find a formula for maximizing the systole of loops on the three-holed sphere with fixed boundary lengths and discuss the more general claim on the general systole of loops formula asserted by Gendulphe in [10].

Finally, I present work towards proving Conjecture 1.2 in Hoffman and Purcell's paper [12] which asserts an upper bound of $10/\sqrt{3}$ on the minimal area of a packing of a hyperbolic surface by horoball cusp neighborhoods, over all such packings. I verify that the geometric decorations they define cover the decorated Teichmüller space of a generic surface. I then find an explicit upper bound on the minimal area over all packings of both the three- and four-punctured spheres using these coordinates as well as the decorations which achieve this maximum.

TABLE OF CONTENTS

1.0 INTRODUCTION 1

1.1 The Injectivity Radius 1

1.2 The Systole of Loops 2

1.3 The Hyperbolic Tammes Problem 4

2.0 BACKGROUND 7

2.1 Models of the hyperbolic plane 8

2.1.1 The Half-Plane Model 8

2.1.2 The Poincaré Disk Model 10

2.1.3 The Hyperboloid Model 10

2.2 Hyperbolic Surfaces 13

2.2.1 The Structure of Non-Compact Hyperbolic Surfaces 15

2.2.2 Surfaces with Boundary 17

2.3 Unique Polygons 18

3.0 THE INJECTIVITY RADIUS 21

3.1 Set-up 21

3.2 The Packing Radius 26

4.0 THE SYSTOLE OF LOOPS 33

4.1 The systole of loops on the three-holed sphere 34

5.0 THE HYPERBOLIC TAMMES PROBLEM 47

5.1 Set-up 50

5.1.1 Spinal Triangulations 53

5.2 The three-punctured sphere 57

5.2.1 Triangulation 1 58
5.2.2 Triangulation 2 59
5.3 The four-punctured sphere 62
5.3.1 Triangulation 1 62
5.3.2 Triangulation 2 68
5.3.3 Triangulations with no valence one ideal points 72
5.3.4 Triangulation 3 74
BIBLIOGRAPHY 80

LIST OF FIGURES

1 The half-plane model 9
2 The Poincaré disk model 10
3 The hyperboloid model (Image source [17, §3.1]) 11
4 F as a quotient space of $\mathbb{H}^2$ 14
5 A horoball cusp neighborhood construction 16
6 Angle measure of the equilateral triangle and horocyclic ideal triangle with compact side length $2r$ 19
7 Angle measure of a Saccheri quadrilateral and one-holed monogon 20
8 Construction of a surface with $\chi = -3,\ n = 2,\ b = 1$ 31
9 A loop homotopic to B_1 36
10 The three-holed sphere as a union of right-angles hexagons 37
11 Surgery on geodesic loop segments 39
12 The two cases of p on the boundary of F 40
13 Conditions on a geometric decoration 51
14 The correspondence between the set of geometric decorations and the set of decorated surfaces 54
15 The two triangulations of the three-punctured sphere 57
16 Three-punctured sphere triangulation 1 58
17 Three-punctured sphere triangulation 2 60
18 Four-punctured sphere triangulation 1 62
19 Four-punctured sphere triangulation 2 68
20 The geometric decoration of triangulation 2 69

21 v_0 has two edges both with the same opposite vertex 72
22 v_0 has two edges with distinct opposite vertices 73
23 Four-punctured sphere triangulation 3 75
24 Flipping an edge of the ideal triangulation 76
25 Flipping the front edge AC to BD 78

1.0 INTRODUCTION

In this dissertation, I investigate three points of view of maximizing packings on complete hyperbolic surfaces, possibly with geodesic boundary. Each of these terms is defined formally in Section 2.2 of this dissertation.

The first two chapters will consider two directions for extending the bound on the maximal injectivity radius of hyperbolic surfaces given in Theorem 5.11 of [4] to the setting of hyperbolic surfaces with geodesic boundary. Here for a point p of a hyperbolic surface F, the *injectivity radius of F at p*, denoted $injrad_p(F)$, is the supremum of all $r > 0$ such that there is a locally isometric embedding of an open metric disk of radius r into F taking the disk's center to p. For any orientable surface S homeomorphic to a complete, hyperbolic surface of finite area, Theorem 5.11 of [4] gave a sharp upper bound on $injrad_p(F)$, taken over all such surfaces F homeomorphic to S and all $p \in F$.

1.1 THE INJECTIVITY RADIUS

We first extend the bound of [4, Theorem 5.11] to a bound on the maximal injectivity radius of hyperbolic surfaces with geodesic boundary. Here for p in the interior of a hyperbolic surface with geodesic boundary, we define $injrad_p(F)$ exactly as above. Noting that $injrad_p(F)$ is bounded above by the distance from p to ∂F, and hence that it approaches 0 uniformly as $p \to \partial F$, we extend it continuously to ∂F by defining $injrad_p(F) := 0$ for any $p \in \partial F$. Our first main result is analogous to [4, Theorem 5.11]: a sharp upper bound on $injrad_p(F)$, taken over *all* hyperbolic surfaces F with geodesic boundary homeomorphic to a fixed S, and all $p \in F$.

Theorem 2. *For a complete, finite-area, n-cusped hyperbolic surface F with b compact geodesic boundary components and Euler characteristic χ, where $\chi < 0$, $n \geq 0$, and $b \geq 0$, the injectivity radius r of F at any $p \in F$ satisfies $r \leq r_{\chi,n,b}$, where $r_{\chi,n,b}$ is the unique solution to:*

$$3\left(2-(2\chi+b+n)\right)\alpha(r_{\chi,n,b})+2n\beta(r_{\chi,n,b})+2b\gamma(r_{\chi,n,b})=2\pi.$$

For a complete, finite-area, n-cusped hyperbolic surface F with b geodesic boundary components and Euler characteristic χ, where $\chi < 0$, $n \geq 0$, and $b \geq 0$, and $p \in F$, $injrad_p(F) = r_{\chi,n,b}$ if and only if p is the unique vertex of a decomposition of F into equilateral triangles with side length $2r_{\chi,n,b}$, n horocyclic ideal triangles with compact side length $2r_{\chi,n,b}$, and b Saccheri quadrilaterals with leg length $r_{\chi,n,b}$ and summit length $2r_{\chi,n,b}$.
Here, $\alpha(r), \beta(r)$, and $\gamma(r)$ each measures the vertex angle of an equilateral triangle with side length $2r$, the angle at the finite vertices of a horocyclic ideal triangle with compact side length $2r$, and the summit angle of a Saccheri quadrilateral with summit length $2r$ and leg lengths r, respectively, as defined in Section 2.3.

We prove Theorem 2 in Section 3.2 using the "centered dual decomposition" defined in [4]. The hyperbolic polygons described above are defined in Section 2.3

1.2 THE SYSTOLE OF LOOPS

The second chapter of this dissertation considers the *systole of loops* function, denoted at $p \in F$ by $sys_p(F)$ and defined as the infimum of the lengths of closed, non-constant geodesic arcs in F based at p, which offers a different direction for extending [4, Theorem 5.11] to the bounded context. If F has no boundary then $sys_p(F) = 2injrad_p(F)$ for each $p \in F$, since a maximal-radius disk based at p determines a minimal-length geodesic arc based at p comprised of two radial arcs that meet at a point of self-tangency. But if F has boundary then

$$injrad_p(F) = \min\left\{\frac{1}{2}sys_p(F), dist(p, \partial F)\right\}$$

is strictly less than $sys_p(F)$ in general, for instance near ∂F.

There is no direct analog of Theorem 2 concerning sys_p. One can however fix both a topology and a collection of boundary lengths, and seek to maximize $sys_p(F)$ over all points p in all hyperbolic surfaces F with geodesic boundary that possess this data. This perspective is taken in the preprint [10]. Indeed, Theorem 1.2 there describes an equation determined by this data and claims that the maximum value is the unique positive solution of this equation.

However for certain collections of boundary lengths, in particular when one is much longer than any other, the equation given in [10, Theorem 1.2] has no solutions. See Lemma 6. We expand on this issue in Section 4.1. We are inclined to believe that the formula given in [10, Theorem 1.2] is correct for "reasonable" boundary length collections and requires adjustment only for collections with exactly one very long component. We confirm this in the simplest case, that of the three-holed sphere, proving the corrected formula:

Theorem 3. *Let F be a hyperbolic three-holed sphere with geodesic boundary components of lengths $b_1 \leq b_2 \leq b_3 \in (0, \infty)$, and let*

$$f_{(b_1,b_2)}(x) := 6\sin^{-1}\left(\frac{1}{2\cosh(x/2)}\right) + 2\sum_{i=1}^{2}\sin^{-1}\left(\frac{\cosh(b_i/2)}{\cosh(x/2)}\right).$$

a) The maximum value of $sys_p(F)$, taken over all $p \in F$, is attained in the interior of F if and only if $f_{(b_1,b_2)}(b_3) > \pi$. In this case it is the unique positive solution x to:

$$f_{(b_1,b_2)}(x) + \sin^{-1}\left(\frac{\cosh(b_3/2)}{\cosh(x/2)}\right) = 2\pi.$$

For a point $p \in int(F)$ at which the maximum value is attained, the systolic loops based at p divide F into an equilateral triangle and three one-holed monogons.

b) If $f_{(b_1,b_2)}(b_3) \leq \pi$, then $sys_p(F)$ attains its maximum on the component B_3 of ∂F of length b_3. Taking $p_2 \in B_3$ to be the endpoint of the shortest arc joining it to B_2, the maximum is attained at $p \in B_3 - \{p_2\}$ if and only if $g(x_2) > \pi$, where

$$g(x) = 2\cos^{-1}\left(\frac{\tanh(b_3/2)}{\tanh(x)}\right) + 2\sin^{-1}\left(\frac{\sinh(b_3/2)}{\sinh(x)}\right) + 2\sum_{i=1}^{2}\sin^{-1}\left(\frac{\cosh(b_i/2)}{\cosh(x/2)}\right)$$

and x_2 satisfies

$$\sinh(x_2/2) = (\cosh(b_2/2)\cosh(b_3/2) + \cosh(b_1/2))/\sinh(b_3/2).$$

If $f_{(b_1,b_2)}(b_3) \leq \pi$ and $g(x_2) > \pi$ then the maximum value of $sys_p(F)$ is the unique $x \in (x_2, b_3]$ such that $g(x) = \pi$. For $p \in F$ where the maximum is attained, the systolic loops based at p divide F into an isosceles triangle and two one-holed monogons.

c) If $f_{(b_1,b_2)}(b_3) \leq \pi$ and $g(x_2) \leq \pi$ then the maximum value x_1 of $sys_p(F)$ is defined by

$$\sinh(x_1/2) = \sqrt{\coth^2(b_3/2)\left[\cosh(b_1/2)\cosh(b_3/2) + \cosh(b_2/2)\right]^2 - \sinh^2(b_1/2)\sinh^2(b_3/2)},$$

attained at p_2, and a systolic loop based at w bounds a one-holed monogon enclosing B_1.

1.3 THE HYPERBOLIC TAMMES PROBLEM

The final chapter of this dissertation was inspired by [12, Conjecture 1.2], which states that any packing by horoball cusp neighborhoods of a hyperbolic surface has at least one horoball cusp neighborhood with area less than $10/\sqrt{3}$. A *horoball cusp neighborhood* is a non-compact but finite-area region of a hyperbolic surface homeomorphic to a punctured disk. See Definition 1.

In geometry, the Tammes problem is a well known packing problem which was posed by the botanist P.L. Tammes in 1930. He sought, for a given number of points on the (positive curvature) sphere $\mathbb{S}^2$, to maximize the minimum distance between them. In other words, what is the maximum diameter (or equivalently, maximum area) of n equal non-overlapping disks on the surface of the sphere? Sharp bounds to this question for $n \leq 14$ and $n = 24$ were found by a variety of authors, and the $n = 14$ case was not solved until 2015 in [16].

At the beginning of Chapter 5, we proffer the max-min problem and the hyperbolic Tammes problem as stated below, motivated by Hoffman and Purcell's conjecture. We prove that these problems are equivalent in Lemma 12.

The Max-Min Problem. *For a given $n \geq 3$, find the supremum area over all packings of all complete, finite-area, n-cusped planar hyperbolic surfaces (i.e. n-punctured spheres) by horoball cusp neighborhoods of the minimum-area such neighborhood. That is, for any $n \geq 3$, find*

$$\sup_{(F,\mathcal{P})} \left\{ \min_{i=1,\ldots,n} \{\text{area}(B_i)\} \right\}$$

where the supremum is taken over all pairs $(F, \mathcal{P})$ with F a complete, finite-area hyperbolic n-punctured sphere and $\mathcal{P} = \{\{B_i\}_{i=1}^n$ a packing of F by horoball cusp neighborhoods.

The Hyperbolic Tammes Problem. *For a given $n \geq 3$, find the supremum area over all packings of all complete, n-cusped planar hyperbolic surfaces (i.e. n-punctured spheres) by horoball cusp neighborhoods of equal area. That is, for any $n \geq 3$, find*

$$\sup_{(F,\mathcal{P})} \left\{ \min_{i=1,\ldots,n} \{\text{area}(B_i)\} \right\}$$

where the supremum is taken over all pairs $(F, \mathcal{P})$ with F a complete, finite-area hyperbolic n-punctured sphere and $\mathcal{P} = \{B_i\}_{i=1}^n$ a packing of F by horoball cusp neighborhoods of equal area.

We also prove the following theorems solving the hyperbolic Tammes problem for both the three- and four-punctured spheres, both of which are found to be strictly less than the conjectured $10/\sqrt{3}$.

Theorem 4. *The maximum area of the minimal-area horoball cusp neighborhood, over all packings of the hyperbolic three-punctured sphere by such neighborhoods, is* 2.

Theorem 5. *The maximum area of the minimal-area horoball cusp neighborhood, over all packings of all hyperbolic four-punctured spheres by such neighborhoods, is* 3.

These results follow from a well known corollary of Böröczky's Theorem (see Proposition 3 and the subsequent remark for an example of how to attain these bounds for $n = 3, 4, 6$, and 12.) I give alternative proofs of the preceding theorems in Sections 5.2 and 5.3 respectively using the idea of a "geometric decoration" defined in [12] and detailed at the beginning of Chapter 5. These coordinates were used to prove Proposition 3.1 of [12] constructing

a surface with minimum horoball area within ϵ of $10/\sqrt{3}$, but does not go so far as to prove the conjecture for *all* surfaces. In Section 5.1, we investigate the relationship between the geometric coordinates of Hoffman and Purcell with the spinal triangulation coordinates defined by Bowditch and Epstein in [2] and Epstein and Penner in [7] to show that the coordinate system provided by geometric decorations will indeed cover the Teichmüller space of any generic surface, i.e.

Proposition 4. *Every complete, finite-area hyperbolic surface equipped with a packing by horoball cusp neighborhoods is the geometric realization of some geometric decoration.*

2.0 BACKGROUND

First let us recall some basic hyperbolic geometric terminology. We will denote by $\mathbb{H}^2$ the hyperbolic plane, that is, the complete and simply connected Riemannian manifold with constant curvature -1 and dimension 2.

Definition. We define the notions of *length* and *distance* in the standard Riemannian way.

- The *length* of a continuously differentiable curve $\gamma : [a, b] \to \mathbb{H}^2$ is defined to be

$$L(\gamma) := \int_a^b \|\gamma'(t)\| dt,$$

  where $\|\gamma'(t)\|$ is the norm induced by the hyperbolic inner product on the tangent space at the point $\gamma(t)$, $T\mathbb{H}^2(\gamma(t))$.
- The *distance* between any two points $p, q \in \mathbb{H}^2$ is defined to be

$$d(p, q) := \inf\{L(\gamma) \mid \gamma : [a, b] \to \mathbb{H}^2 \text{ joins } p \text{ to } q\}.$$

  This gives $\mathbb{H}^2$ the structure of a metric space.

We will largely be considering two important classes of curves, *geodesics* and *horocycles*. We define them here and illustrate what each look like in two different models of $\mathbb{H}^2$ in Sections 2.1.1 and 2.1.2.

Definition. A *geodesic segment* between two points in $\mathbb{H}^2$ is the shortest-length path between them.

Fact. Let us collect some useful facts about geodesics in $\mathbb{H}^2$. (cf. [13, §1.2])

1. Any two points $p, q \in \mathbb{H}^2$ can be joined by a unique geodesic segment.

2. Every geodesic segment extends uniquely to an infinite length *geodesic*, which minimizes the distance between any two of its points.

3. There is a unique geodesic through any point in any direction.

Definition. A *horocycle* is a curve C whose perpendicular geodesics converge asymptotically in the same direction, i.e. there exist continuously varying normal vectors $\vec{v}(p)$ to C such that for any $p, q \in C$, if $\gamma_p, \gamma_q : [0, \infty) \to \mathbb{H}^2$ are geodesic rays based at p and q in the direction of $\vec{v}(p)$ and $\vec{v}(q)$, respectively, then for any $\epsilon > 0$ there exists $T > 0$ such that for every $t \geq T$, $d(\gamma_p(t), \gamma_q(t)) < \epsilon$.

2.1 MODELS OF THE HYPERBOLIC PLANE

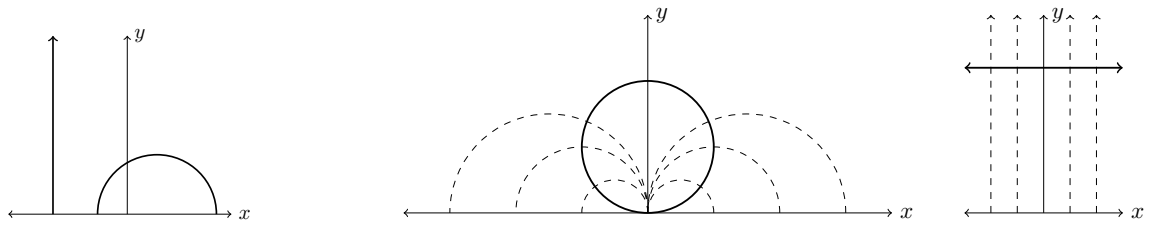
Throughout this dissertation, I utilize three different, though isometric, models of the hyperbolic plane. The goal of this section is to establish the definitions of each model.

2.1.1 The Half-Plane Model

Definition. In the *upper half-plane model* of $\mathbb{H}^2$, we consider the upper half of the Euclidean plane, i.e. $\{(x, y) \,|\, y > 0\}$, and give it the metric $(ds)^2 = \frac{(dx)^2+(dy)^2}{y^2}$. The *boundary* $\partial\mathbb{H}^2$ of this model is the extended x-axis, i.e. the x-axis along with a point at infinity.

Fact. Geodesics in the half-plane model of $\mathbb{H}^2$ are vertical lines and half-circles perpendicular to the x-axis ([13, Theorem 1.2.1]). Horocycles are either Euclidean circles tangent to the x-axis or horizontal lines ([13, §4.2]). See Figure 1.

Here, one can observe that every geodesic has two endpoints on the boundary of $\mathbb{H}^2$ and is determined by these endpoints. The geodesics which determine a horocycle (i.e. the geodesics perpendicular to the curve) all share an endpoint on $\partial\mathbb{H}^2$, which we call the *center* of the horocycle. The center of a horocycle which is a horizontal line is the point at infinity.



(a) Geodesics

(b) Horocycles (thick) and perpendicular geodesics (dashed)

Figure 1: The half-plane model

Fact. ([13, Theorem 1.3.1]) The group of orientation-preserving isometries of the upper half-plane model of $\mathbb{H}^2$ is the projective special linear group

$$PSL_2(\mathbb{R}) = \{M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mid a, b, c, d \in \mathbb{R},\ ad - bc = 1\} / \{\pm I\}.$$

Such a matrix M acts on a point $z \in \mathbb{H}^2$ as a *Möbius transformation*, that is

$$Mz = \frac{az + b}{cz + d}.$$

Definition. An isometry of $\mathbb{H}^2$ is called:

(a) *elliptic* if it has a fixed point in $\mathbb{H}^2$,

(b) *parabolic* if it fixes a unique point on $\partial\mathbb{H}^2$, or

(c) *hyperbolic* if it fixes a pair of points on $\partial\mathbb{H}^2$.

Theorem (Classification of hyperbolic isometries)**.** *Every isometry of* $\mathbb{H}^2$ *is either elliptic, parabolic, or hyperbolic.*

Fact. ([13, §2.1]) A matrix $M \in PSL_2(\mathbb{R})$ is:

(a) elliptic if and only if $|tr(M)| < 2$,

(b) parabolic if and only if $|tr(M)| = 2$, and

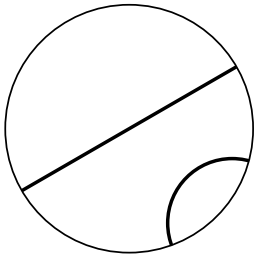
(c) hyperbolic if and only if $|tr(M)| > 2$.

Here $|tr(M)| = |a + d|$ is the *trace* of the matrix M.

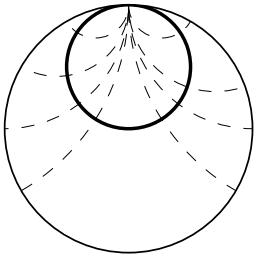
2.1.2 The Poincaré Disk Model

Definition. The *Poincaré disk model* of $\mathbb{H}^2$ consists of the points inside the unit disk, i.e. $\{(x, y) \,|\, x^2 + y^2 < 1\}$, along with the metric given by $(ds)^2 = 4\frac{(dx)^2+(dy)^2}{(1-(x^2+y^2))^2}$. The boundary $\partial\mathbb{H}^2$ of this model is the boundary of the unit disk, i.e. all points with $x^2 + y^2 = 1$.

Here, a geodesic is either a circular arc perpendicular to the boundary of the unit disk or a diameter of the disk. A horocycle is a Euclidean circle tangent to the boundary. See Figure 2.



(a) Geodesics



(b) A horocycle (thick) and perpendicular geodesics (dashed)

Figure 2: The Poincaré disk model

Remark 1. Note that in the two preceding models, the points that make up the boundary $\partial\mathbb{H}^2$ of the hyperbolic plane are not included in the space itself. We call these points *ideal points*. The isometry between these models extends to a homeomorphism between their boundaries (with an appropriate choice of topology on $\mathbb{R} \cup \{\infty\}$).

2.1.3 The Hyperboloid Model

The last of the three models, which is utilized in Chapter 4, is the *hyperboloid model*, also known as the *Lorentz model*, of the hyperbolic plane. In this model, points are represented by vectors on the positive sheet of a *two-sheet hyperboloid* in three-dimensional *Lorentzian space*. Recall that a two-sheet hyperboloid in three-space is defined to be $\{(x_1, x_2, x_3) \in$

$\mathbb{R}^3 \mid -\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = -1\}$. For our purposes, we take $a = b = c = 1$ so that the *positive sheet* is the set of all such points in which $x_1 > 0$. The *Lorentzian inner product* of the vectors $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ is defined to be $x \circ y := -x_1y_1 + x_2y_2 + x_3y_3$. Not surprisingly, two vectors x, y are considered *Lorentz orthogonal* if $x \circ y = 0$. The set of vectors of $\mathbb{R}^3$ with this inner product is Lorentzian space, denoted $\mathbb{R}^{1,2}$.

Definition. We call a vector $x \in \mathbb{R}^{1,2}$ either

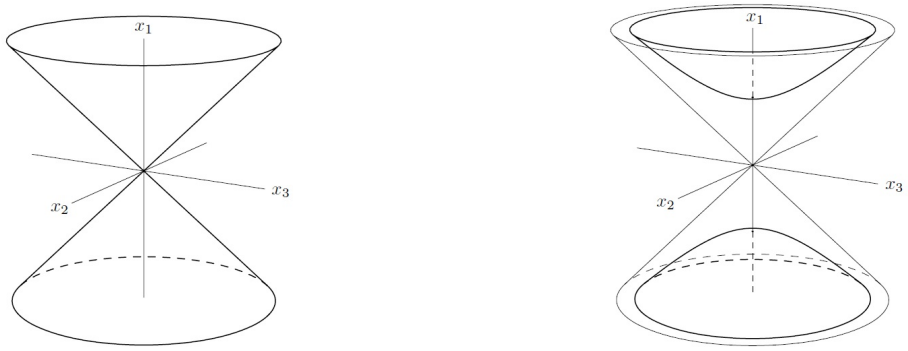
(a) *space-like* if $\|x\| := \sqrt{x \circ x} > 0$,

(b) *light-like* if $\|x\| = 0$, or

(c) *time-like* if $\|x\|$ is imaginary.

It is interesting to note that the light-like vectors of Lorentzian space form a cone, called the *light cone*, in three-space. The exterior of the cone consists of all space-like vectors and the interior of all time-like vectors. See Figure 3a.

Definition. The *hyperboloid model* of hyperbolic space is the set of time-like vectors $H^2 = \{x \in \mathbb{R}^{1,2} \mid \|x\|^2 = x \circ x = -1, x_1 > 0\}$ and is illustrated in Figure 3b. That is, the positive sheet of a two-sheet hyperboloid in Lorentzian 3-space.



(a) The light cone C^2

(b) The hyperboloid model inside C^2

Figure 3: The hyperboloid model (Image source [17, §3.1])

These vectors interact in very interesting geometric ways, which I describe below.

Definition. The *time-like angle* between time-like vectors x, y is defined to be the unique non-negative real number $\eta(x, y)$ such that

$$x \circ y = \|x\| \ \|y\| \cosh \eta(x, y).$$

As shown in [17, §3.2], $\eta(x, y)$ defines a metric on H^2, and so we define the *hyperbolic distance* between x, y to be

$$d_H(x, y) := \eta(x, y).$$

Definition. Let V be a vector subspace of $\mathbb{R}^{1,2}$. Then V is said to be

(a) *time-like* if and only if V has a time-like vector,

(b) *space-like* if and only if every non-zero vector in V is space-like, or

(c) *light-like* otherwise.

In general, we can define *hyperbolic n-space* $H^n := \{x \in \mathbb{R}^{1,n} \,|\, \|x\|^2 = -1, \text{ and } x_1 > 0$ with the same metric as defined above.

Definition. (a) A *hyperbolic m-plane* of H^n is defined to be the intersection of H^n with an $(m + 1)$-dimensional time-like vector subspace of $\mathbb{R}^{1,n}$ and a *hyperplane* of H^n is a hyperbolic $(n - 1)$-plane.

(b) A *hyperbolic line* is defined as the intersection of H^n with a two-dimensional time-like vector subspace of $\mathbb{R}^{1,n}$.

Notice in H^2, a hyperplane P is the intersection of H^2 with a two-dimensional time-like vector subspace of $\mathbb{R}^{1,2}$, and it is Lorentz orthogonal to a vector x if $P = \langle x \rangle^L \cap H^2$, where $\langle x \rangle^L := \{y \in \mathbb{R}^{1,2} \,|\, x \circ y = 0\}$ is the Lorentz orthogonal complement of the vector space spanned by x. This is also the definition of a hyperbolic line. In particular, we have the unique line defined by the distinct vectors x and y in H^2 defined by $L(x, y) = H^2 \cap \langle x, y \rangle$. By [17, Corollary 4], the geodesics of H^2 are precisely its hyperbolic lines.

Definition. The *space-like angle* between space-like vectors x, y which span a space-like vector subspace is defined to be the unique real number $\eta(x, y) \in [0, \pi]$ such that

$$x \circ y = \|x\| \ \|y\| \cos \eta(x, y).$$

Geometrically, if $\lambda, \mu : \mathbb{R} \to H^2$ are geodesics such that $\lambda(0) = \mu(0)$ and $\lambda'(0) = x,\ \mu'(0) = y$, then $\eta(x, y)$ gives the *hyperbolic angle* between λ and μ.

Definition. The *time-like angle* between space-like vectors x, y which span a time-like vector subspace is defined to be the unique positive real number $\eta(x, y)$ such that

$$|x \circ y| = \|x\|\ \|y\| \cosh \eta(x, y).$$

This turns out to be precisely the hyperbolic distance from the hyperplanes P and Q of H^2 which are Lorentz orthogonal to x and y respectively. Moreover, as shown in [17, Theorem 3.2.8], $x \circ y < 0$ in this context if and only if x and y are oppositely oriented tangent vectors of the hyperbolic line N which is orthogonal to both P and Q.

Definition. For x a space-like vector and y positive time-like vector (that is, its first component is positive), we define the *time-like angle* between them as the unique non-negative real number $\eta(x, y)$ such that

$$|x \circ y| = \|x\|\ |||y|||| \sinh \eta(x, y).$$

Geometrically, this gives the hyperbolic distance from $y/|||y|||$ to the hyperplane P of H^2 Lorentz orthogonal to x. Moreover, [17, Theorem 3.2.12] tells us that $x \circ y < 0$ if and only if x and y are on opposite sides of the hyperplane of $\mathbb{R}^{1,2}$ spanned by P.

2.2 HYPERBOLIC SURFACES

Definition. A *hyperbolic surface* F is a Hausdorff, second-countable topological space with an atlas of charts to open subsets of the hyperbolic plane such that the transition functions are restrictions of hyperbolic isometries, i.e. a collection of pairs (U_α, ϕ_α) where each $U_\alpha \subset F$, $\bigcup_\alpha U = F$, and each $\phi_\alpha : U_\alpha \to \mathbb{H}^2$ is a homeomorphism to an open subset of $\mathbb{H}^2$ with the property that $\phi_\alpha \circ \phi_\beta^{-1}$ is the restriction of an isometry whenever $U_\alpha \cap U_\beta \neq \emptyset$.

Note that this definition is equivalent to having a Riemannian metric with constant curvature -1, such as the one defined in Section 2.1.1, since the chart maps can be used to pull back the metric on $\mathbb{H}^2$ locally and the isometric transition functions ensure this is well-defined.

Definition. A hyperbolic surface F is *complete* if and only if every sequence in F which is Cauchy with respect to the Riemannian distance metric converges in F.

Theorem 1 (cf. Theorem 8.5.9 of [17])**.** *Any complete, connected hyperbolic surface F is isometric to the quotient space $\mathbb{H}^2/\Gamma$ for some freely acting discrete group Γ of isometries of $\mathbb{H}^2$.*

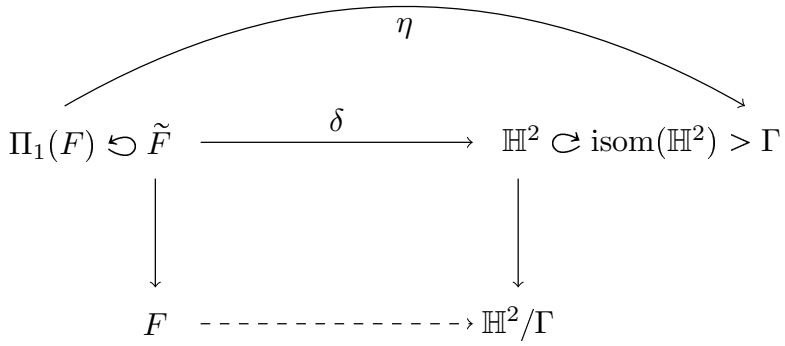


Figure 4: F as a quotient space of $\mathbb{H}^2$

Sketch of proof. Given that F is complete we have that the universal cover $\widetilde{F}$ is complete, when equipped with the pullback metric from F. As the universal cover is simply connected by definition, there exists an isometry (called the developing map) $\delta : \widetilde{F} \to \mathbb{H}^2$. This is due to the uniqueness of complete, simply connected constant curvature spaces (see e.g. [14, Theorem 11.12].)

This isometry δ determines a homomorphism $\eta : \pi_1(F) \to \mathrm{isom}(\mathbb{H}^2)$ given by $\eta(\gamma)(y) := \delta(\gamma.\delta^{-1}(y))$, where γ in the fundamental group $\pi_1(F)$ acts on $\widetilde{F}$ by covering transformations. Finally, the fact that $\Pi_1(F)$ has a covering space action on $\widetilde{F}$ (in the sense of [11, Proposition 1.40]) implies $\Gamma := \eta(\Pi_1(F))$ is discrete and acts freely. Thus by the construction of the holonomy representation, the developing map is equvariant with respect to the actions of

$\Pi_1(F)$ of $\widetilde{F}$ and $\mathbb{H}^2$, so it induces the dashed isometry from F to $\mathbb{H}^2/\Gamma$ shown in the diagram in Figure 4. □

2.2.1 The Structure of Non-Compact Hyperbolic Surfaces

Chapter 5 of this dissertation considers non-compact hyperbolic surfaces with finite area. In this section, we lay out the basic structure theory of such a surface's non-compact regions, which follow from the Margulis Lemma [15, Lemma 4.2.8].

Definition. For a point p of a hyperbolic surface F, the *injectivity radius of F at p*, denoted $injrad_p(F)$, is the supremum of all $r > 0$ such that there is a locally isometric embedding of an open metric disk of radius r into F taking the disk's center to p.

Note that when we consider $F = \mathbb{H}^2/\Gamma$ as a quotient space as in Theorem 1, the structure laid out there identifies the universal cover of F with $\mathbb{H}^2$, so the injectivity radius of any surface F is positive at every point p.

Definition. For a non-compact hyperbolic surface F, the *ϵ-thin part* of F is

$$F_{(0,\epsilon]} := \{p \in F \,|\, injrad_p(F) \leq \epsilon\}$$

for any $\epsilon > 0$. The *thick part* of a non-compact hyperbolic surface F is $F \setminus F_{(0,\epsilon]}$.

Proposition 1. cf. [15, Proposition 4.2.15]). *There exists an $\epsilon_2 > 0$ (called the two-dimensional* Margulis constant*) such that for any complete hyperbolic surface F, the thin part $F_{(0,\epsilon]}$ is a disjoint union of R-tubes and horoball cusp neighborhoods.*

We define these objects below.

Definition. An *R-tube* in a hyperbolic surface F is the closed metric neighborhood $N_R(\gamma)$ of radius R around a closed geodesic γ that is homeomorphic to $S^1 \times [0, 1]$.

Note that Proposition 1 implies in particular that every geodesic of length less than $2\epsilon_2$ has an R-tube neighborhood for some $R > 0$.

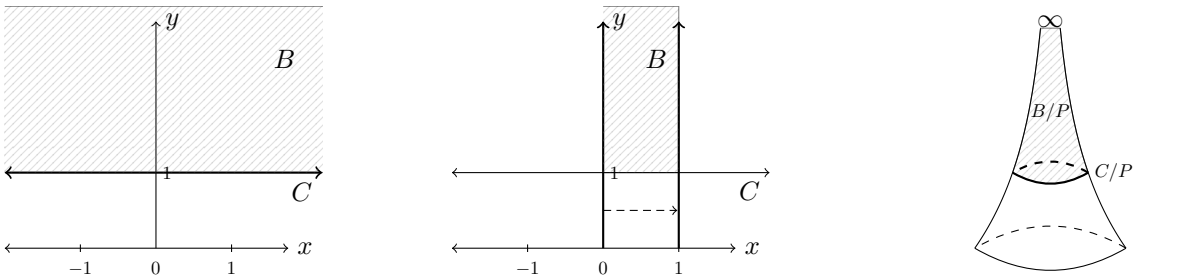
Definition 1. Given a horocycle C with center $p \in \partial\mathbb{H}^2$, let $\{\gamma_q : [0,\infty) \to \mathbb{H}^2\}_{q\in C}$ be the set of geodesic rays perpendicular to C based at $q \in C$ with ideal endpoint p. The *horoball* B with *center* p and boundary C is

$$B := \{\gamma_q(t) \,|\, t \in [0,\infty) \text{ and } q \in C\}.$$

If $F = \mathbb{H}^2/\Gamma$ and $P \leq \Gamma$ is a cyclic subgroup of the stabalizer of p, $stab(\{p\}$, generated by a parabolic isometry of $\mathbb{H}^2$, then the *horoball cusp neighborhood* associated with the horoball B is the projection $int(B)/P \subset F$.

If F has non-compact regions in its thin part, then by Proposition 1, Γ contains parabolic elements which each fix a unique point on the boundary of $\mathbb{H}^2$. For each parabolic element fixing some $p \in \partial\mathbb{H}^2$, the stabilizer P of p in Γ is cyclic and there exists a horoball B with ideal center p such that the projection of $int(B)$ factors through an embedding of $int(B)/P$.

Example 1. Consider the horocycle $C = \{(x,1) \,|\, x \in \mathbb{R}\}$ centered at the ideal point ∞ in the upper half-plane model of $\mathbb{H}^2$ with $P = \langle \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \rangle \subset stab(\{\infty\}) = \{\left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right)\}$ in $PSL_2(\mathbb{R})$. The fundamental domain of $\mathbb{H}^2$ for P is $\{(x,y) \,|\, 0 \leq x \leq 1\}$. Then $\mathbb{H}^2/P$ is the cusp shown in Figure 5. The horoball cusp neighborhood is the projection of the shaded horoball $B = \{(x,y) \,|\, x \in \mathbb{R},\ y > 1\}$.



(a) A horoball with center ∞

(b) A fundamental domain for P

(c) A horoball cusp neighborhood in $\mathbb{H}^2/P$

Figure 5: A horoball cusp neighborhood construction

Computing the area of B/P is quite simple. The *hyperbolic area* of $A \subset \mathbb{H}^2$ in the upper half-plane model is $\text{area}(A) = \int_A \frac{1}{y^2} dxdy$ ([13, §1.4]). Therefore,

$$\text{area}(B/P) = \int_1^\infty \int_0^1 \frac{1}{y^2} dxdy = \int_1^\infty \frac{1}{y^2} dy = 1$$

Note that for any $a > 1$, this computation extends easily to any sub-horoball $B_a := \{(x, y) \,|\, y \geq a\}$ of B and so the area of the horoball cusp neighborhood B_a/P is $\frac{1}{a}$.

Remark 2. A horoball cusp neighborhood of F is homeomorphic to $S^1 \times [a, \infty)$, which is bounded by $S^1 \times \{a\}$.

Proposition. [15, Proposition 4.2.17] *A complete hyperbolic surface has finite volume if and only if its thick part is compact.*

Since R-tubes are compact, every complete, non-compact, finite-area hyperbolic surface is the union of a compact region with a disjoint union of horoball cusp neighborhoods.

2.2.2 Surfaces with Boundary

Definition. A *hyperbolic half-plane* is the closure of a component of $\mathbb{H}^2 \setminus \gamma$, where γ is a geodesic in the hyperbolic plane $\mathbb{H}^2$. Its *boundary* is the geodesic γ.

Definition. A hyperbolic surface F *with geodesic boundary* is a Hausdorff, second countable topological space with an atlas of charts to open subsets of hyperbolic half-planes. The *boundary* ∂F is the set of points mapped into half-plane boundaries by chart maps, and the *interior* is $int\, F = F \setminus \partial F$.

Note that F has the structure of a smooth manifold with boundary, and it obtains a hyperbolic Riemannian metric as in the boundryless case. We define a distance function and completeness analogously as well.

Remark 3. Recall that the injectivity radius at a point p in a hyperbolic surface F is defined to be the supremum of all $r > 0$ such that there is a locally isometric embedding of an open metric disk of radius r into F taking the disk's center to p. If F is a surface with boundary, we must take into account how to define this idea for points on the boundary of F. Noting that $injrad_p(F)$ is bounded above by the distance from p to ∂F, and hence that it approaches

0 uniformly as $p \to \partial F$, we extend it continuously to ∂F by defining $injrad_p(F) := 0$ for any $p \in \partial F$.

Definition. The *systole of loops* function, denoted at $p \in F$ by $sys_p(F)$, is defined as the infimum of the lengths of closed, non-constant geodesic arcs in F based at p.

Remark 4. If F has no boundary then $sys_p(F) = 2injrad_p(F)$ for each $p \in F$, since a maximal-radius disk based at p determines a minimal-length geodesic arc based at p comprised of two radial arcs that meet at a point of self-tangency. But if F has boundary then

$$injrad_p(F) = \min\left\{\frac{1}{2}sys_p(F), dist(p, \partial F)\right\}$$

is strictly less than $sys_p(F)$ in general, for instance near ∂F.

2.3 UNIQUE POLYGONS

Definition. A *triangle* in the hyperbolic plane is the intersection of three half-planes. Its *sides* are the geodesic segments that bound the area and its *vertices* are the intersection points of its sides. These vertices may lie on the boundary $\partial\mathbb{H}^2$ in which case they are not actually on the triangle.

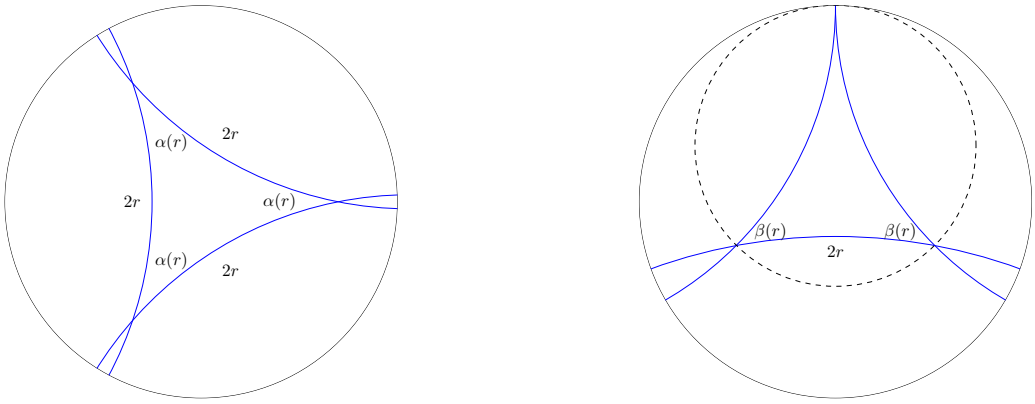
- A triangle is *equilateral* if the length of its sides are equal.
- A triangle is *ideal* if all of its vertices are ideal points, i.e. lie on $\partial\mathbb{H}^2$.
- A *horocyclic ideal triangle* is a triangle with two vertices on a horocycle C and a single ideal vertex at the ideal center of C.

The definitions of being ideal and horocyclic extend to any n-gon in $\mathbb{H}^2$.

Fact. Here we record some trigonometric formulas relating to these polygons. See Figure 6 for illustrations.

- An equilateral triangle of side length $2r$ exists for any $r > 0$ (see [17, Theorem 3.5.9]) and has equal angle measures $\alpha(r) = 2\sin^{-1}\left(\frac{1}{2\cosh r}\right)$. Thus there is a unique such triangle, up to isometry.

- A horocyclic ideal triangle with compact side length $2r$ exists for any $r > 0$ (see [5, Proposition 3.4]) and has equal base angles of measure $\beta(r) = \sin^{-1}\left(\frac{1}{\cosh r}\right)$. The angle at its ideal vertex is 0. Thus there is a unique such triangle, up to isometry.



(a) An equilateral triangle

(b) A horocyclic ideal triangle

Figure 6: Angle measure of the equilateral triangle and horocyclic ideal triangle with compact side length $2r$

The hyperbolic law of sines and the hyperbolic law of cosines

$$\frac{\sin A}{\sinh \alpha} = \frac{\sin B}{\sinh \beta} = \frac{\sin C}{\sinh \gamma}; \qquad \cosh c = \cosh a \cosh b - \sinh a \sinh b \cos \gamma$$

for a hyperbolic triangle with side lengths A, B, and C with opposite angles α, β, and γ respectively allow us to compute the formulas for the angles. See, e.g. [17, Theorem 3.5.2 and Theorem 3.5.3].

Definition. A *Saccheri quadrilateral* is a quadrilateral (i.e. intersection of four half-planes) with two equal length *legs* perpendicular to its *base*. The remaining edge is called the *summit*.

Fact. Some facts about the Saccheri quadrilateral follow (see [9, §VI.3.3]). Figure 7a provides an illustration.

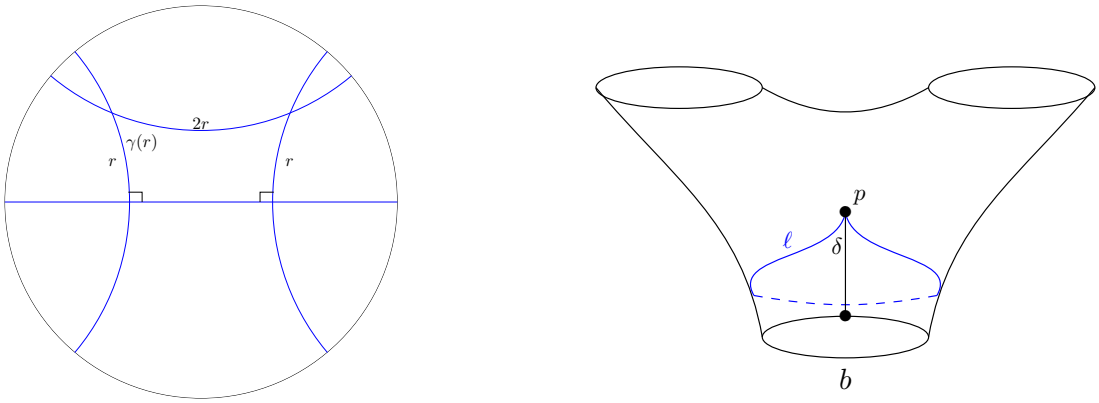
- The base length is determined by the length of the legs and summit. In particular, if the base has length b, the legs length h and the summit length s, we have

$$\cosh b = -\sinh^2 h + \cosh^2 h \cosh s$$

- If a Saccheri quadrilateral has legs of length r and summit length $2r$, the summit angles have equal measure $\gamma(r) = 2\sin^{-1}\left(\frac{1}{\sqrt{2}\cosh r}\right)$.
- It can be shown directly that there is a unique Saccheri quadrilateral of given leg and summit length, up to isometry.

Definition. A *one-holed monogon* is the quotient of a Saccheri quadrilateral by isometrically identifying its legs so that the vertices of the summit are identifies with each other.

Fact. A one-holed monogon has the structure of a hyperbolic surface with geodesic boundary away from the summit vertex quotient, where the total angle is 2δ for $\delta = \frac{\cosh(b/2)}{\cosh(\ell/2)}$. Here b is the length of the base of the Saccheri quadrilateral and ℓ is the length of its summit. See Figure 7b for an illustration of such a construction.



(a) A Saccheri quadrilateral

(b) A one-holed monogon

Figure 7: Angle measure of a Saccheri quadrilateral and one-holed monogon

3.0 THE INJECTIVITY RADIUS

In this chapter we prove Theorem 2. Its upper bound follows from the more general Proposition 2 below, which gives an upper bound $r^k_{\chi,n,b}$ on the radius of packings of hyperbolic surfaces with boundary by k equal-radius disks, for arbitrary $k \in \mathbb{N}$. This result extends Proposition 0.2 of [3], itself a follow-up to [4], to the with-boundary setting.

3.1 SET-UP

When given a hyperbolic surface F with geodesic boundary, we double it across ∂F to produce a boundaryless hyperbolic surface we call DF. Then given a finite set $\mathcal{S} \subset F$, we analyze the centered dual complex of $\mathcal{S} \cup \overline{\mathcal{S}}$, where $\overline{\mathcal{S}} \subset DF$ is the reflection of $\mathcal{S}$ across ∂F. This complex is reflection-invariant, by Lemma 1 below.

Definition. For a hyperbolic surface F with geodesic boundary, the *double* of F is the quotient space $DF = F \cup \overline{F} / \sim$, where $\overline{F}$ is a second copy of F and $x \sim y$ if and only if $y = x$ or $x \in \partial F$ and $y = \bar{x}$ is the point of $\overline{F}$ corresponding to x.

We give DF the structure of a hyperbolic surface (without boundary) as follows. For $x \in \mathit{int}\, F$ and a chart map $\phi\colon U \to \mathbb{H}^2$, where $U \subset F$ is open, let $U_0 = U \cap \mathit{int}\, F$ and let ϕ_0 be the restriction of ϕ to U_0. For the corresponding point $\bar{x} \in \overline{F}$, let $\overline{U}_0 = \rho(U_0)$ and $\bar{\phi}_0 = \rho \circ \phi_0 \circ \rho$, where $\rho\colon \mathbb{H}^2 \to \mathbb{H}^2$ is the reflection through the boundary geodesic of the half-plane to which ϕ maps. For $x \in \partial F$ we take $U_0 = U \cup \rho(U)$ and define ϕ_0 as ϕ on U and $\rho \circ \phi \circ \rho$ on $\rho(U)$.

It is an exercise to show that these chart maps have isometric transition functions and

that the *reflection across* ∂F, the involution $DF \to DF$ that exchanges each $x \in F$ with $\bar{x} \in \overline{F}$, fixing ∂F, is an isometry of the resulting hyperbolic structure. It is also a fact that DF is complete if and only if F is complete. In the complete case there is a locally isometric universal cover $\pi\colon\ \mathbb{H}^2 \to DF$, and the restriction of this cover to any component of the preimage of F is its universal cover.

Below we will utilize the results of [3, Proposition 5.9] and [4, Proposition 1.2]. In particular, the *Voronoi tessellation* of DF determined by a locally finite subset $\mathcal{S}$ is obtained by projecting the Voronoi tessellation of $\mathbb{H}^2$ determined by $\widetilde{\mathcal{S}} := \pi^{-1}(\mathcal{S})$ to DF under the locally isometric universal cover $\pi : \mathbb{H}^2 \to DF$, as described in [4, §1]. Our first goal is to prove some useful results concerning the Voronoi tessellation and centered dual complex on the double of a hyperbolic surface with geodesic boundary.

We begin by noting that the Voronoi tessellation of $\widetilde{\mathcal{S}} \subset \mathbb{H}^2$ is invariant under all isometries of $\mathbb{H}^2$ that preserve $\widetilde{\mathcal{S}}$, since it is defined in purely geometric terms. In particular, if $\widetilde{\mathcal{S}} = \pi^{-1}(\mathcal{S})$ for a locally finite subset $\mathcal{S}$ of a complete hyperbolic surface F with locally isometric universal cover $\pi\colon\ \mathbb{H}^2 \to F$ then the Voronoi tessellation of $\widetilde{\mathcal{S}}$ is invariant under the lifts of isometries of F that preserve $\mathcal{S}$, and it follows that the Voronoi tessellation of F determined by $\mathcal{S}$ is invariant under these isometries. The following Lemma's first assertion is a special case of this fact.

Lemma 1. *For a complete hyperbolic surface F with geodesic boundary and a finite set $\mathcal{S} \subset F$, the Voronoi tessellation of DF determined by $\mathcal{S} \cup \overline{\mathcal{S}}$ is preserved by the reflection through ∂F, where $\overline{\mathcal{S}} = \{\bar{s} \,|\, s \in \mathcal{S}\}$. This reflection also preserves the non-centered Voronoi subgraph.*

The *non-centered Voronoi subgraph* of $\widetilde{\mathcal{S}} \subset \mathbb{H}^2$, as defined in [4, Dfn. 2.1], is the union of the Voronoi edges that are *non-centered* in the sense that they do not intersect their geometric duals. Here it is helpful to recall (from eg. [4, §1]) that each Voronoi edge is of the form $e = V_s \cap V_t$, where $s, t \in \widetilde{\mathcal{S}}$ determine Voronoi two-cells V_s and V_t, and the *geometric dual* of such an edge e is the geodesic arc γ_{st} joining s to t. As observed in Lemma 5.4 of [4], the non-centered Voronoi subgraph of $\widetilde{\mathcal{S}} = \pi^{-1}(\mathcal{S} \cup \overline{\mathcal{S}})$ is preserved by covering transformations. We will also call its image in DF the non-centered Voronoi subgraph of

the Voronoi tessellation determined by $\mathcal{S} \cup \overline{\mathcal{S}}$.

Proof of Lemma 1. We have already justified the Lemma's first assertion, and the second follows from an analogous observation (also behind [4, L. 5.4]): that *any* isometry g of $\mathbb{H}^2$ that preserves $\widetilde{\mathcal{S}}$ takes the edge $e = V_s \cap V_t$ to $g.e = V_{g(s)} \cap V_{g(t)}$ and its geometric dual γ_{st} to $g.\gamma_{st} = \gamma_{g(s)g(t)}$, so e intersects γ_{st} if and only if $g.e$ intersects $\gamma_{g(s)g(t)}$. This applies in particular to the lift of the reflection across ∂F. □

Lemma 2. *For a complete hyperbolic surface F with geodesic boundary and a locally finite set $\mathcal{S} \subset int\, F$, ∂F lies in the Voronoi graph (i.e. the one-skeleton of the Voronoi tessellation) of DF determined by $\mathcal{S} \cup \overline{\mathcal{S}}$, where $\overline{\mathcal{S}} = \{\bar{s} \,|\, s \in \mathcal{S}\}$.*

Proof. For a set $\widetilde{\mathcal{S}} \subset \mathbb{H}^2$, the Voronoi graph of $\widetilde{\mathcal{S}}$ is characterized as the set of $x \in \mathbb{H}^2$ that have at least two closest points in $\widetilde{\mathcal{S}}$. Let us now take $\widetilde{\mathcal{S}} = \pi^{-1}(\mathcal{S} \cup \overline{\mathcal{S}})$, where $\pi\colon\ \mathbb{H}^2 \to DF$ is a locally isometric universal cover. For any geodesic γ in the preimage of ∂F under π, the reflection $r\colon\ DF \to DF$ across ∂F lifts to a reflection $\tilde{r}$ of $\mathbb{H}^2$ fixing γ. Since $\mathcal{S} \cup \overline{\mathcal{S}}$ is invariant under r, $\widetilde{\mathcal{S}}$ is invariant under $\tilde{r}$, so for any $x \in \gamma$ and closest point $s \in \widetilde{\mathcal{S}}$, $\tilde{r}(s) \in \widetilde{\mathcal{S}}$ is also a closest point to x. And $s \neq \tilde{r}(s)$ since $\mathcal{S} \subset int\, F$, so x lies in the Voronoi graph of $\widetilde{\mathcal{S}}$. □

Lemma 3. *For a complete hyperbolic surface F with compact geodesic boundary and a finite set $\mathcal{S} \subset F$, each component T of the non-centered Voronoi subgraph of DF determined by $\mathcal{S} \cup \overline{\mathcal{S}}$ that intersects the union of the boundary geodesics is taken to itself by reflection across ∂F, it is compact, and its root vertex is in the union of the boundary geodesics.*

Proof. Recall from Lemma 1 that the non-centered Voronoi subgraph is preserved by reflection across ∂F, so T is taken to another component of the non-centered Voronoi subgraph. Since it intersects ∂F, which is fixed by the reflection, it intersects its image and hence is preserved (being a component).

By Proposition 2.9 of [4], T can have at most one non-compact edge, so if it were non-compact then that edge would lie in the fixed locus ∂F of the reflection. This cannot be, since ∂F is compact but each end of a non-compact Voronoi edge exits a cusp. Finally,

Lemma 2.7 of [4] asserts that the root vertex of T is unique, so it must also lie in the fixed locus ∂F of the reflection. □

We have already defined the geometric dual to a Voronoi edge, and below we will discuss the geometric dual to a Voronoi vertex. As described in eg. [4, Prop. 1.1], for a vertex v of the Voronoi tessellation determined by $\widetilde{\mathcal{S}} \subset \mathbb{H}^2$ this is the the convex hull of the set of $s \in \widetilde{\mathcal{S}}$ such that $v \in V_s$. It is a compact, convex polygon with its vertex set in $\widetilde{\mathcal{S}}$, cf. [4, Lemma 1.5]. The geometric dual to a Voronoi two-cell V_s, $s \in \widetilde{\mathcal{S}}$, is s itself. The collection of geometric dual cells is a polyhedral complex (see eg. [4, Theorem 1.2]) that we call the *geometric dual complex* determined by $\widetilde{\mathcal{S}}$.

If $\widetilde{\mathcal{S}} = \pi^{-1}(\mathcal{S})$ for a locally finite universal cover $\pi\colon \mathbb{H}^2 \to F$ to a complete hyperbolic surface F and a locally finite subset $\mathcal{S}$ then π embeds the interior of each geometric dual cell in F (see Remark 5.3 of [4]). Below we will call the *geometric dual complex of F determined by $\mathcal{S}$* the projection of the geometric dual complex of $\widetilde{\mathcal{S}}$, and below we will count the "vertices" of a geometric dual two-cell in F by those of a cell in $\mathbb{H}^2$ projecting to it.

Lemma 4. *For a complete hyperbolic surface F and a finite set $\mathcal{S} \subset \operatorname{int} F$, the geometric dual complex of DF determined by $\mathcal{S} \cup \overline{\mathcal{S}}$ is invariant under the reflection across ∂F, and the geometric dual to each Voronoi vertex in ∂F has an even number of vertices.*

Proof. Let $\widetilde{\mathcal{S}} = \pi^{-1}(\mathcal{S} \cup \overline{\mathcal{S}})$, for a locally finite universal cover $\pi\colon \mathbb{H}^2 \to DF$. Since the geometric dual complex of $\widetilde{\mathcal{S}}$ is defined geometrically (it is indeed the intersections sets defined by distance inequalities,) it is invariant under *all* isometries that preserve $\widetilde{\mathcal{S}}$, and as in previous results this implies that the geometric dual complex of DF determined by $\mathcal{S} \cup \overline{\mathcal{S}}$ is invariant under reflection across ∂F.

Now suppose $C = \pi(\widetilde{C})$ is a two-cell of the geometric dual complex of $\mathcal{S} \cup \overline{\mathcal{S}}$ that is dual to a Voronoi vertex $v \in \partial F$. Then $\widetilde{C}$ is dual to a Voronoi vertex $\tilde{v}$ in a component γ of $\pi^{-1}(\partial F)$. Then as $\tilde{v}$ is invariant under the reflection through γ (which preserves $\widetilde{\mathcal{S}}$ since it is a lift of the reflection across ∂F), $\widetilde{C}$ is also invariant under this reflection. The reflection therefore preserves the vertex set of $\widetilde{C}$, and since $\mathcal{S} \subset \operatorname{int} F$ no vertex is fixed. So the vertices come in pairs exchanged by the reflection, hence there is an even number. □

The geometric dual complex of a complete hyperbolic surface F determined by a locally finite set $\mathcal{S}$ is a subcomplex of the *Delaunay tessellation* of F determined by $\mathcal{S}$. In addition to the geometric dual cells, the Delaunay tessellation has one non-compact two-cell enclosing a horocyclic neighborhood of each cusp of F. Each such "horocyclic" two-cell is the projection of a cell in the universal cover that is the convex hull of $\widetilde{\mathcal{S}} \cap B$, where $\widetilde{\mathcal{S}}$ is the preimage of $\mathcal{S}$ and B is a horoball bounded by a horocycle S with the property that $S \cap \widetilde{\mathcal{S}} = B \cap \widetilde{\mathcal{S}}$. We may divide such cells into horocyclic ideal triangles using a collection of geodesic rays, one from each point of $B \cap \widetilde{\mathcal{S}}$ to the ideal point of B. The Delaunay tessellation and its relation to the geometric dual complex are described in [6].

The *centered dual decomposition* has two-cells of two forms: first, for each component T of the non-centered Voronoi subgraph, the union of geometric dual two-cells dual to vertices of T (together with one of the horocyclic ideal triangles above if T has a non-compact edge); and second, those geometric dual two-cells, necessarily centered, not dual to a vertex of any such component T. This is constructed in [4, §2]. The *centered dual plus* is obtained by throwing in the remaining horocyclic ideal triangles comprising horocyclic Delaunay cells as two-cells, producing a decomposition with underlying space $\mathbb{H}^2$, or, upon projecting to F, underlying space F. See §5 of [4]. We will use this decomposition to prove our bound.

Lemma 5. *Let C be a compact 2-cell of the centered dual complex plus of F which intersects the union of boundary geodesics and that for fixed $d > 0$ each edge of ∂C has length at least d. If C is a quadrilateral, then its area is at least that of a hyperbolic square with side lengths d. If ∂C has $k > 4$ edges, then area$(C) \geq (k-4)A_m(d) + D_0(d,d,d,d)$, where $A_m(d)$ is the area of an isosceles hyperbolic triangle with two sides of length d, inscribed in a hyperbolic circle with its third side a diameter and $D_0(d_1, \ldots, d_m)$ is the area of a hyperbolic m-gon with side lengths $d_1, \ldots, d_m \in (0, \infty)$.*

Proof. By Lemma 4, C has an even number of vertices and therefore $k \geq 4$. So area$(C) \geq (k-2)A_m(D)$ by [4, Theorem 3.31]. For an isosceles hyperbolic triangle T with two sides of length d, inscribed in a hyperbolic circle with its third side a diameter, we observe that the union of T with its reflection across this diameter is a hyperbolic square with side lengths d.

Thus $2A_m(d) = D_0(d,d,d,d)$. It follows that if $k > 4$ then

$$\text{area}(C) \geq (k-2)A_m(d) = (k-4)A_m(d) + 2A_m(d) = (k-4)A_m(d) + D_0(d,d,d,d),$$

and if $k = 4$ then $\text{area}(C) \geq 2A_m(d) = D_0(d,d,d,d)$. □

This bound allows us to extend Proposition 0.2 of [3], which gave the bound of the packing radius of a surface without boundary, to the with-boundary setting.

3.2 THE PACKING RADIUS

Proposition 2. *For a complete, finite-area, n-cusped hyperbolic surface F with b geodesic boundary components and Euler characteristic χ, where $\chi < 0$, $n \geq 0$, and $b \geq 0$, and any packing of F by k disks of radius $r > 0$ ($k \in \mathbb{N}$), we have $r \leq r^k_{\chi,n,b}$, where $r^k_{\chi,n,b}$ is the unique solution to:*

$$3\left(2 - \left(\frac{2\chi + b + n}{k}\right)\right)\alpha(r^k_{\chi,n,b}) + \frac{2n}{k}\beta(r^k_{\chi,n,b}) + \frac{2b}{k}\gamma(r^k_{\chi,n,b}) = 2\pi.$$

A complete, finite-area, n-cusped hyperbolic surface F has an equal-radius packing by disks of radius $r = r^k_{\chi,n,b}$, if and only if the disks are centered at the vertices of a decomposition of F into equilateral triangles, n horocyclic ideal triangles each with compact side length $2r^k_{\chi,n,b}$, and b Saccheri quadrilaterals with legs $r^k_{\chi,n,b}$ and summit $2r^k_{\chi,n,b}$. Here, $\alpha(r), \beta(r)$, and $\gamma(r)$ each measures the vertex angle of an equilateral triangle with side length $2r$, the angle at the finite vertices of a horocyclic ideal triangle with compact side length $2r$, and the summit angle of a Saccheri quadrilateral with summit length $2r$ and leg lengths r, respectively, as defined in Section 2.3.

Proof. Let F be a finite-area, genus g hyperbolic surface of Euler characteristic χ with n cusps and b geodesic boundary components equipped with an equal-radius packing by k disks of radius r. Then DF has no boundary and $\chi(DF) = 2\chi(F) = 2(2-2g-b-n)$. Note that each given disk $B \subset F$ with center p and radius r is taken by reflection to another disk $\overline{B} \subset \overline{F}$ with center $\overline{p}$ and radius r. Thus if the given packing of F is given by $\{B_1, \ldots, B_k\}$,

then $\{B_1, \ldots, B_k, \overline{B_1}, \ldots, \overline{B_k}\}$ is a packing of DF by equal radius disks. Taking $\mathcal{S} \cup \overline{\mathcal{S}}$ to be the set of all disk centers, construct the centered dual complex plus of DF and label the $\pi_1(F)$-orbits of cells $C_1, \ldots, C_m$.

By the Gauss-Bonnet theorem, we have

$$\text{area}(C_1) + \ldots + \text{area}(C_m) = -2\pi\chi(DF) = -4\pi\chi \tag{3.1}$$

For any non-compact cell, we have by [4, Theorem 4.16] that

$$\text{area}(C_i) \geq D_0(\infty, d, \infty) + (n_i - 3)D_0(d, d, d) \tag{3.2}$$

Notice that because the radius-r disks centered at the vertices of each cell are disjoint, each compact edge must have length at least $d := 2r$. Equality holds if and only if $n_i = 3$ and the compact side length is d.

If C_i is a compact cell which intersects the axis of reflection, then by Lemma 5 and [4, Corollary 3.5] we have

$$\text{area}(C_i) \geq (n_i - 4)A_m(d) + D_0(d, d, d, d) \geq (n_i - 4)D_0(d, d, d) + D_0(d, d, d, d) \tag{3.3}$$

In this case, equality holds if and only if C_i is a square with side length d. Notice that we must have at least one such cell per boundary component.

On the other hand, if C_i is a compact cell which does not intersect the axis of reflection of DF, then we apply [4, Theorem 3.31, Corollary 3.5] to obtain the bounds:

$$\text{area}(C_i) \geq (n_i - 2)A_m(d) \geq (n_i - 2)D_0(d, d, d) \qquad \text{if } n_i \geq 4 \tag{3.4}$$

$$\text{area}(C_i) \geq (n_i - 2)D_0(d, d, d) \qquad \text{if } n_i = 3 \tag{3.5}$$

In both cases, equality holds if and only if C_i is an equilateral triangle with side length d.

Order the cells so that C_i is non-compact if and only if $i \leq m_0$ for some fixed $m_0 \leq m$, and for each i let n_i be the number of edges of C_i. Note that $m_0 \geq 2n$, since DF has $2n$ cusps and each centered dual cell has at most one ideal vertex. Additionally, let us list all compact cells which intersect the axis of rotation before we list any compact cells which do not.

Now we can use the bound 3.2 on the non-compact cells (which we have re-ordered to be labeled $1, \dots, m_0$), bound 3.3 on the compact cells which intersect the axis of reflection, and bounds 3.4 and 3.5 on the remaining cells. Note that since the length of the diagonal of a square is longer than that of its sides, $(n_i - 4)D_0(d,d,d) + D_0(d,d,d,d) > (n_i - 2)D_0(d,d,d)$. Thus we use the result of bound 3.3 for precisely b cells which intersect the axis of reflection, and write the bound for the remaining such cells as the bound for the compact cells which do not intersect the axis, i.e. the bound in 3.4/3.5. Applying these bounds to our Gauss-Bonnet formula 3.1 results in

$$\begin{aligned}
-4\pi\chi &\geq \sum_{i=1}^{m_0} (D_0(\infty,d,\infty) + (n_i-3)D_0(d,d,d)) + \sum_{i=m_0+1}^{m_0+b} ((n_i-4)D_0(d,d,d) + D_0(d,d,d,d)) \\
&\qquad + \sum_{i=m_0+b+1}^{m} ((n_i-2)D_0(d,d,d)) \\
&= m_0 \cdot D_0(\infty,d,\infty) + \left(\sum_{i=1}^{m}(n_i-2) - m_0 - 2b\right) \cdot D_0(d,d,d) + b \cdot D_0(d,d,d,d) \\
&\geq 2n \cdot D_0(\infty,d,\infty) + \left(\sum_{i=1}^{m} n_i - 2m - 2b - 2n\right) \cdot D_0(d,d,d) + b \cdot D_0(d,d,d,d) \\
&= 2n \cdot (\pi - 2\beta(r)) + (2e - 2m - 2b - 2n) \cdot (\pi - 3\alpha(r)) + b \cdot (2\pi - 4\gamma(r)) \\
&= (2e-2m)\pi - 4n\beta(r) - 3(2e - 2m - 2b - 2n)\alpha(r) - 4b\gamma(r) \\
&= (4k-4\chi)\pi - 4n\beta(r) - 3(4k - 4\chi - 2b - 2n)\alpha(r) - 4b\gamma(r)
\end{aligned}$$

Note that to move from the second to third line, we think of $m_0 = 2n + m_0 - 2n$ and use the fact that $D_0(\infty,d,\infty) \geq D_0(d,d,d)$. In the last line we used the Euler characteristic of the compactification of our doubled surface by filling in each cusp with a unique point. Thus $2\chi + 2n = v - e + f = (2k + 2n) - e + m \implies 2e - 2m = 4k - 4\chi$.

After some rearrangement, we have that

$$\begin{aligned}
4k\pi &\leq 4n\beta(r) + 3(4k - 4\chi - 2b - 2n)\alpha(r) + 4b\gamma(r) \\
\implies 2\pi &\leq 3\left(2 - \left(\frac{2\chi + b + n}{k}\right)\right)\alpha(r) + \frac{2n}{k}\beta(r) + \frac{2b}{k}\gamma(r)
\end{aligned}$$

Since α, β, and γ are decreasing functions in r (recall Section 2.3) we see that $r \leq r^k_{\chi,n,b}$ and equality holds if and only if the cells away from the axis of reflection are triangles with

compact side length $d = 2r^k_{\chi,n,b}$ and the cells intersecting the axis of reflection are squares also with side length $d = 2r^k_{\chi,n,b}$. Reducing our view back to $F \subset DF$, we have remaining the unaffected triangles away from the boundary and half of each of the squares which crossed the boundary components. These halves are precisely Saccheri quadrilaterals with summit length $2r^k_{\chi,n,b}$ and leg length $r^k_{\chi,n,b}$.

At this point, we have showed that the radius r of a k-disk packing satisfies $r \leq r^k_{\chi,n,b}$ and equality holds if and only if the Delaunay tessellation of DF yields the prescribed decomposition. It remains to show that beginning with such a decomposition, there exists a packing of F by k disks of radius $r^k_{\chi,n,b}$

Let us define a *sector* of a metric disk as its intersection with two half-planes whose boundaries contain the disk's center. To guarantee the disks form a packing, i.e. do not overlap, we must show that they do not overlap within each polygon and that for each disk B with center p, if P is a polygon with vertex p, $B \cap P$ must be a full sector.

Each equilateral triangle T is a centered polygon with side lengths $2r^k_{\chi,n,b}$, thus [4, Lemma 5.12] guarantees that the open disks of radius r centered at each vertex of the triangle intersect with T in full sectors and that these disks do not overlap in T for any $r \leq r^k_{\chi,n,b}$.

For a horocyclic ideal triangle R with compact side length $2r^k_{\chi,n,b}$, notice that the perpendicular bisector γ of the compact side ρ is an axis of reflection for R. Thus, by the hyperbolic law of cosines, the closest point of γ to either vertex of R on ρ is the point at $\gamma \cap \rho$. Since the length of ρ is $2r^k_{\chi,n,b}$, we are guaranteed that γ divides a disk of radius $r \leq r^k_{\chi,n,b}$ centered at one vertex of the side ρ of R from the corresponding disk centered at the other such vertex, so they do not overlap. The result follows from the fact that for any triangle T and a disk B of radius r centered at a vertex v of T, $B \cap T$ is a full sector of B if and only if $d(v, \gamma) \leq r$, where γ is the side of T opposite v.

For a Saccheri quadrilateral Q, again we see that the perpendicular bisector γ of the summit σ is an axis of reflection for Q, as its legs have equal length $r^k_{\chi,n,b}$. Thus the closest point of γ to either vertex of Q on σ is $\gamma \cap \sigma$. Since the length of σ is $2r^k_{\chi,n,b}$, we are guaranteed that γ divides a disk of radius $r \leq r^k_{\chi,n,b}$ centered at one summit vertex of Q from the corresponding disk centered at the other such vertex. Additionally, as the base of Q is perpendicular to the legs of Q by the definition of a Saccheri quadrilateral, the vertices

at the base are also the closest points from the summit vertices on the base. Thus for any disk B of radius $r \leq r^k_{\chi,n,b}$ centered at a vertex on the summit, $B \cap Q$ is a full sector.

Therefore we have proven that for each $r \leq r^k_{\chi,n,b}$, the collection of disks of radius $r \leq r^k_{\chi,n,b}$ centered at the (finite) vertices of the surface composed of equilateral triangles of side length $2r^k_{\chi,n,b}$, ideal horocyclic triangles with compact side length $2r^k_{\chi,n,b}$, and Saccheri quadrilaterals with leg length $r^k_{\chi,n,b}$ and summit length $2r^k_{\chi,n,b}$ is an equal-radius packing of the surface. □

Now to the proof of the main result. Let us recall the statement of the theorem regarding the maximal injectivity radius.

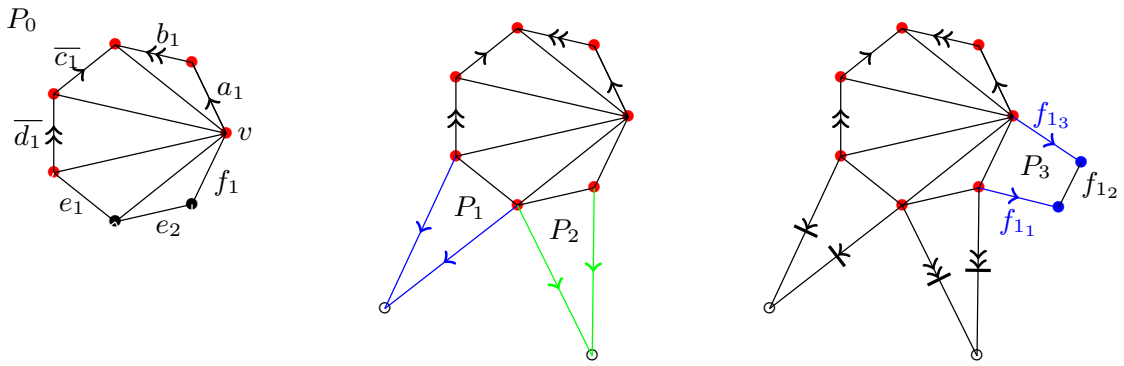
Theorem 2. *For a complete, finite-area, n-cusped hyperbolic surface F with b compact geodesic boundary components and Euler characteristic χ, where $\chi < 0$, $n \geq 0$, and $b \geq 0$, the injectivity radius r of F at any $p \in F$ satisfies $r \leq r_{\chi,n,b}$, where $r_{\chi,n,b}$ is the unique solution to:*

$$3\left(2 - (2\chi + b + n)\right)\alpha(r_{\chi,n,b}) + 2n\beta(r_{\chi,n,b}) + 2b\gamma(r_{\chi,n,b}) = 2\pi.$$

For a complete, finite-area, n-cusped hyperbolic surface F with b geodesic boundary components and Euler characteristic χ, where $\chi < 0$, $n \geq 0$, and $b \geq 0$, and $p \in F$, $\mathrm{injrad}_p(F) = r_{\chi,n,b}$ if and only if p is the unique vertex of a decomposition of F into equilateral triangles with side length $2r_{\chi,n,b}$, n horocyclic ideal triangles with compact side length $2r_{\chi,n,b}$, and b Saccheri quadrilaterals with leg length $r_{\chi,n,b}$ and summit length $2r_{\chi,n,b}$.
Here, $\alpha(r), \beta(r)$, and $\gamma(r)$ each measures the vertex angle of an equilateral triangle with side length $2r$, the angle at the finite vertices of a horocyclic ideal triangle with compact side length $2r$, and the summit angle of a Saccheri quadrilateral with summit length $2r$ and leg lengths r, respectively, as defined in Section 2.3.

Proof. The $k = 1$ case of Proposition 2 gives the upper bound of Theorem 2. To prove the Theorem in full, it remains to give examples that attain this bound, and this is what we do.

Fix $\chi < 0$, $n \geq 0$, $b \geq 0$ such that $g = \frac{1}{2}(2-\chi-n-b)$ is a non-negative integer. Let $r_{\chi,n,b}$ be as defined in Theorem 2. Take $4g + n + b - 2$ equilateral triangles each with side length $2r_{\chi,n,b}$ and arrange them fan-like with common vertex v to form a triangulated $(4g+n+b)$-gon,



(a) Step 1: genus (b) Step 2: cusps (c) Step 3: boundary

Figure 8: Construction of a surface with $\chi = -3,\ n = 2,\ b = 1$

P_0 in $\mathbb{H}^2$. Label the edges cyclically $a_1,\ b_1,\ c_1,\ d_1, \ldots, a_g,\ b_g,\ c_g,\ d_g, e_1, \ldots, e_n,\ f_1, \ldots, f_b$ and give each edge the counter-clockwise orientation with v the initial vertex of a_1. See Figure 8a.

For each $1 \leq i \leq g$, let A_i be the orientation-preserving isometry pairing the edge a_i with $\overline{c_i}$ (the segment c_i with clock-wise orientation) and similarly define B_i as the orientation-preserving isometry paring b_i with $\overline{d_i}$. See Figure 8b. Notice that the result has made the first $4g + 1$ vertices equivalent to v under the quotient.

Next, for $1 \leq j \leq n$ we attach to each e_j a horocyclic ideal triangle P_j with base length $2r_{\chi,n,b}$ and let E_j be the parabolic isometry fixing the respective ideal vertex and taking one of the infinite sides of the triangle to the other. See Figure 8c. Thus we have added the next $2n$ vertices as equivalent to v under the quotient.

Finally for $1 \leq k \leq b$ we attach to each f_k a Saccheri quadrilateral P_{n+k} with summit length $2r_{\chi,n,b}$ and leg length $r_{\chi,n,b}$, assigning counter-clockwise orientation to the additional three sides, f_{k_1}, f_{k_2}, and f_{k_3}. Denote by F_k the orientation-preserving isometry which exchanges opposite sides f_{k_1} and $\overline{f_{k_3}}$, thus making the final $2b-2$ vertices of P_0 also equivalent to v under the quotient.

By the definition of r taken from Theorem 2, the sum of the angles about vertex v is

$$(4g+n+b-2)\cdot 3\alpha(r_{\chi,n,b})+n\cdot 2\beta(r_{\chi,n,b})+b\cdot 2\gamma(r_{\chi,n,b})=2\pi.$$

The group $G=\langle A_1,\ B_1,\ldots,A_g,\ B_g,\ E_1,\ldots,E_n,F_1,\ldots,F_b\rangle$ is a Fuchsian group with the polygon $P=\bigcup_{i=0}^{n+b}P_i$ its fundamental domain by Poincare's Polygon Theorem (see [13, Theorem 4.3.2] and the following remarks.) Hence the quotient $F=\mathbb{H}^2/G$ is a complete hyperbolic surface. Inspecting the edge pairing, we see that F has n cusps and b boundary components. Additionally, we see that the number of vertices, edges, and faces of the quotient space is:

$$\begin{aligned}V&=1+b\\E&=(2g+2n+3b)+(4g+n+b-3)=6g+3n+4b-3\\V&=(4g+n+b-2)+n+b=4+2n+2b-2\end{aligned}$$

The original polygon P had just one vertex under the quotient. The addition of the ideal triangles added no new vertices, and the addition of the Saccheri quadrilaterals contributed one each. Counting the edges, we see that the first $4g$ edges were paired off and so only $2g$ remain in the final triangulation. Each ideal triangle contributes two new edges, and each quadrilateral contributes 3. The number of interior edges (which are not identified with any other edge by G) is the number of triangles less one, i.e. $4g+n+b-3$. Thus by the Gauss-Bonnet theorem it also has genus g, as desired.

Notice that by the trigonometric relations between the sides of Saccheri quadrilaterals (see [9, §VI.3.3]) and the use of the hyperbolic double- and half-angle identities, this forces the length of each boundary component to be $2\sinh^{-1}(\tanh(r_{\chi,n,b}))$.

To show that the injectivity radius of P is in fact $r_{\chi,n,b}$, recall that in the proof of Proposition 2 we proved that each of these polygons will intersect a disk of radius $r\leq r^1_{\chi,n,b}(=r_{\chi,n,b})$ as a full sector. Therefore after passing to the quotient $F=\mathbb{H}^2/G$, any such open disk will not overlap itself and $r_{\chi,n,b}$ is indeed the injectivity radius of F at v. □

4.0 THE SYSTOLE OF LOOPS

Now we turn our attention to the systole of loops function. Recall from the introduction that this is defined at a point p of a hyperbolic surface F (possibly with geodesic boundary) as the infimum of the lengths of closed, non-constant geodesic arcs in F based at p and denoted by $sys_p(F)$. The main goal of this chapter is to prove Theorem 3, characterizing the maximum value of $sys_p(F)$ for a three-holed sphere F in terms of its boundary lengths. First we recall some motivation.

Theorem. [10, **Theorem 1.2**] *For a fixed $k \geq 0$ and $(b_1, \ldots, b_k) \in [0, \infty)^k$, for any complete, finite-area hyperbolic surface F with k geodesic boundary components of respective lengths b_i (where if $b_i = 0$ the boundary component is replaced by a cusp), the value of $sys_p(F)$ is bounded above for all $p \in F$ by the unique positive solution x to:*

$$6(-2\chi(F) + 2 - k)\sin^{-1}\left(\frac{1}{2\cosh(x/2)}\right) + 2\sum_{i=1}^{k}\sin^{-1}\left(\frac{\cosh(b_i/2)}{\cosh(x/2)}\right) = 2\pi. \tag{4.1}$$

It further asserts that this maximum is attained at $p \in F$ if and only if the systolic loops based at p divide F into equilateral triangles and one-holed monogons. However, we have:

Lemma 6. *For $(b_1, \ldots, b_k) \in [0, \infty)^k$ such that $b_k = \max\{b_1, \ldots, b_k\}$, the equation (4.1) has a positive solution if and only if $f_{(b_1,\ldots,b_{k-1})}(b_k) \geq \pi$, where $f_{(b_1,\ldots,b_{k-1})}$ is defined by*

$$f_{(b_1,\ldots,b_{k-1})}(x) = 6(-2\chi(F) + 2 - k)\sin^{-1}\left(\frac{1}{2\cosh(x/2)}\right) + 2\sum_{i=1}^{k-1}\sin^{-1}\left(\frac{\cosh(b_i/2)}{\cosh(x/2)}\right).$$

This is a continuous and decreasing function of x that limits to 0 as $x \to \infty$.

Proof. The function $f := f_{(b_1,\dots,b_{k-1})}$ records the left-hand side of the formula (4.1) but excludes its final term: the summand $\sin^{-1}\left(\frac{\cosh(b_k/2)}{\cosh(x,2)}\right)$. We begin by observing that formula (4.1) only makes sense for (positive) $x \in [b_k, \infty)$ (recalling that $b_k = \max_i\{b_i\}$). This is because the domain of the arcsine is the range of the sine function, $[-1, 1]$, and if $0 \leq x < b_k$ then $\cosh(b_k/2)/\cosh(x/2) > 1$.

We next note that both f and the left side of formula (4.1), regarded as a function of x, are decreasing on $[b_k, \infty)$. This follows from the facts that $\cosh(x/2)$ and the inverse sine are increasing functions, and that the only x-dependence in the definitions of f and (4.1) is the appearance of $\cosh(x/2)$ in the denominator of each summand. Therefore if equation (4.1) has a solution then its left-hand side's value at $x = b_k$ is at least 2π. This is equivalent to $f(b_k) = \pi$ because the final term of the left side of (4.1) is $\sin^{-1}\left(\frac{\cosh(b_k/2)}{\cosh(b_k/2)}\right) = \pi/2$ for $x = b_k$.

On the other hand, $f(x)$ limits to 0 as $x \to \infty$, since $\sin^{-1}(0) = 0$ and the inverse sine is continuous. By the intermediate value theorem, this implies that if $f(b_k) \geq \pi$ then there exists $x \in [b_k, \infty)$ (unique, since f is decreasing) such that $f(b_k) = \pi$ and hence (4.1) holds. □

For any given $(b_1, \dots, b_{k-1}) \in [0, \infty)^{k-1}$, since $f_{(b_1,\dots,b_{k-1})}(x) \to 0$ as $x \to \infty$ there is an unbounded interval consisting of possible choices of b_k for which equation (4.1) has no solutions. In the remainder of this chapter, we will analyze the case of three-holed spheres using an elementary direct method, providing a corrected version of the theorem in this case.

4.1 THE SYSTOLE OF LOOPS ON THE THREE-HOLED SPHERE

For every three-holed sphere F with geodesic boundary (see Figure 10a for an example of such a surface) it is well known that there is a unique shortest arc between any two components of ∂F, meeting both of them at right angles, and that the collection of all three such arcs cuts F into a pair of isometric right-angled hexagons exchanged by a reflective involution of F. (Establishing this is a key step in the description of "Fenchel–Nielsen coordinates" for the Teichmüller spaces of arbitrary hyperbolic surfaces, see eg. [8, §10.6].) Below we will denote

the geodesic boundary components of F as B_1, B_2, and B_3 and their lengths as b_1, b_2, and b_3, respectively. Then the resulting right-angled hexagons have alternate side lengths $\frac{b_1}{2}$, $\frac{b_2}{2}$, and $\frac{b_3}{2}$, which determine them up to isometry (see eg. [17, Theorem 3.5.14]), and it follows that F itself is determined up to isometry by the triple (b_1, b_2, b_3).

We now begin the process of understanding $sys_p(F)$ for fixed (b_1, b_2, b_3).

Lemma 7. *For any $p \in F$, $sys_p(F) = \min\{\ell_1, \ell_2, \ell_3\}$, where ℓ_i is the length of a simple loop freely homotopic to B_i for each $i \in \{1, 2, 3\}$.*

Proof. Fix $p \in F$, and let H be a right-angled hexagon containing p such that $F = H \cup \bar{H}$, where $\bar{H}$ is the mirror image of H. Without loss of generality, let us say $p \in H$.

By standard results there is a unique geodesic (parametrized proportional to arclength) in the based homotopy class of every loop based at p, which minimizes length in this homotopy class. Let $\gamma\colon [0,1] \to F$ be such a geodesic, and assume further that this γ minimizes the length of a loop over all non-trivial homotopy classes, i.e. assume the length of γ is $\ell = sys_p(F)$. Partition $[0,1]$ using γ's intersections with H and $\bar{H}$. That is, let

$$0 = t_0 < t_1 < \ldots < t_n = 1,$$

such that for each $i > 0$, $\gamma|_{[t_{i-1},t_i]}$ maps into either H or $\bar{H}$, and if $\gamma|_{[t_{i-1},t_i]}$ maps into H then $\gamma|_{[t_i,t_{i+1}]}$ maps into $\bar{H}$ and vice-versa. We observe that $n > 1$ since neither H nor $\bar{H}$ contains a closed geodesic loop, each being isometric to a convex subset of $\mathbb{H}^2$.

We now assume that p does not lie in one of the three arcs of $H \cap \bar{H}$. In this case $n \geq 3$ since both $\gamma|_{[t_0,t_1]}$ and $\gamma|_{[t_{n-1},t_n]}$ map into H. We claim that in fact $n = 3$; i.e., only the single segment $\gamma|_{[t_1,t_2]}$ of γ lies in $\bar{H}$.

If this is not so, then for any $k \geq 2$ such that $\gamma|_{[t_k,t_{k+1}]}$ lies in $\overline{H}$ replace the segment with it's mirror image $\overline{\gamma}|_{[t_k,t_{k+1}]}$ in H. Since the reflection exchanging H with $\overline{H}$ leaves $H \cap \overline{H}$ fixed, we thus obtain a continuous broken geodesic

$$\gamma|_{[t_0,t_1]}.\gamma|_{[t_1,t_2]}.\left(\gamma|_{[t_2,t_3]}.\overline{\gamma}|_{[t_3,t_4]}.\cdots.\gamma|_{[t_{n-1},t_n]}\right)$$

based at p, with the same length as γ. Now the entire broken geodesic within the parentheses above maps into H, which as we previously mentioned is isometric to a convex subset of $\mathbb{H}^2$,

so it can be replaced in the concatenation by the strictly shorter geodesic arc in H that joins its endpoints. Noting that each geodesic segment of γ must have its endpoints on different edges of $H \cap \overline{H}$ (otherwise it wouldn't be a geodesic), this yields a loop at p which is strictly shorter than γ and not null-homotopic, contradicting our hypothesis.

Therefore $n = 3$ and only $\gamma|_{[t_1,t_2]}$ maps into $\bar{H}$. The two edges of $\bar{H}$ containing the endpoints of $\gamma|_{[t_1,t_2]}$ have endpoints on a common component B_i of ∂F, and it is now easy to see directly that γ is simple and freely homotopic to B_i. The case that $p \in H \cap \bar{H}$ is similar, with only bookkeeping changes: in that case we must have $n = 2$, and $\gamma = \gamma|_{[t_0,t_1]}.\gamma|_{[t_1,t_2]}$ is freely homotopic to the boundary component containing the common endpoints of the edges of H containing p and $\gamma(t_1)$. We leave details to the reader. □

Lemma 8. *For any $p \in F$ and any boundary component B_i, the simple loop based at p and freely homotopic to B_i has length ℓ_i determined by*

$$\sinh\left(\frac{\ell_i}{2}\right) = \cosh X_i \sinh\left(\frac{b_i}{2}\right),$$

where X_i is the length of the shortest geodesic arc in F from p to B_i. (See Figure 9.) The angle δ_i between the loop and the arc satisfies $\sin(\delta_i) = \frac{\cosh(b_i/2)}{\cosh(\ell_i/2)}$.

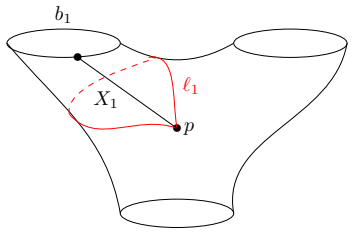
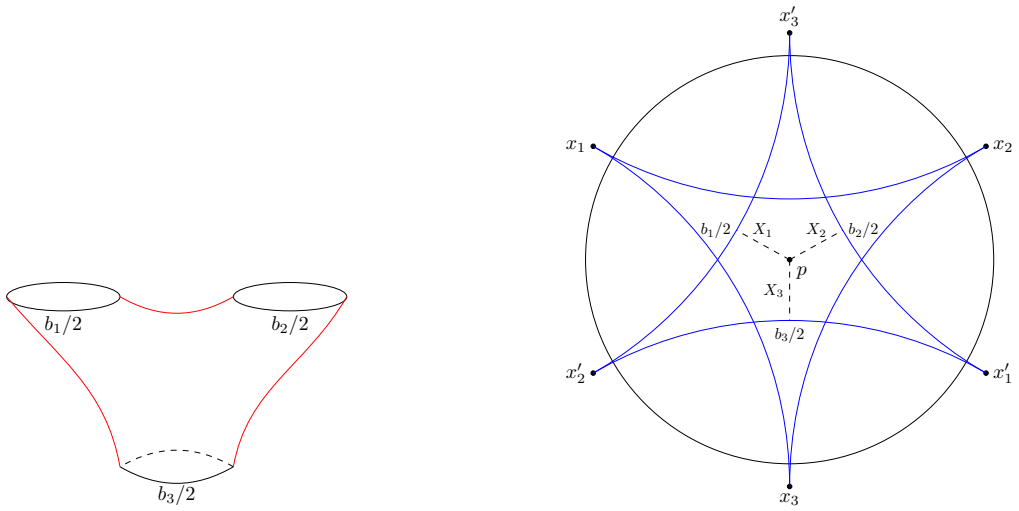


Figure 9: A loop homotopic to B_1

Proof. Consider the one-holed monogon bounded by a simple geodesic loop based at p and freely homotopic to B_i. Cutting it along the geodesic arc connecting p to B_i results in a Saccheri quadrilateral with legs of length X_i, base of length b_i, and summit of length ℓ_i. Since this quadrilateral's legs have equal length, the arc joining the midpoint of its base to that of its summit divides it into isometric sub-quadrilaterals exchanged by a reflection, each

with three right angles. Standard hyperbolic trigonometric identities for such quadrilaterals now give the formulas claimed, compare eg. [9, VI.3.3]. □

Lemma 9. *Write $F = H \cup \bar{H}$ for isometric right-angled hexagons H and $\bar{H}$ such that $H \cap \bar{H}$ is a disjoint union of three edges. If $sys_p(F)$ attains a maximum at $p \in int(H)$ or $int(\bar{H})$, then the simple loops based at p and freely homotopic to each boundary component all have equal lengths.*



(a) The three-holed sphere

(b) Hexagon in Lorentzian space

Figure 10: The three-holed sphere as a union of right-angles hexagons

Proof. By Lemma 8, the distance of p from a boundary component and the length of the loop based at p around that component increase or decrease together. Supposing that the loop about B_3 gives the systolic value with the loop about B_1 strictly longer, i.e. $\ell_3 \leq \ell_2$, and $\ell_3 < \ell_1$, the idea is to move p further from both B_2 and B_3 simultaneously, increasing the systolic loop length and yielding a contradiction.

It is easy to see how to do this in the hyperboloid model for $\mathbb{H}^2$, contained in the Lorentzian space $\mathbb{R}^{1,2}$ as described in Chapter 3 of [17]. To this end, assume that $p \in int(H)$ and regard H as a subset of $\mathbb{H}^2$. Since H is right-angled, the shortest arcs from p to each of the B_i all lie in H as well, so we may take the whole picture to lie in $\mathbb{H}^2$. In particular,

for each $i \in \{1,2,3\}$ the edge of H that was contained in B_i — which we will simply call B_i here, see Figure 10 — now lies in a hyperbolic geodesic that is itself the intersection of a two-dimensional time-like subspace of $\mathbb{R}^{1,2}$ with $\mathbb{H}^2$. (See the Definition above Theorem 3.2.6 in [17].)

For each $i \in \{1,2,3\}$, let x_i be the unit space-like normal in $\mathbb{R}^{1,2}$ to the subspace containing B_i that has the property that the Lorentzian inner product $p \circ x_i$ is less than 0. (Every time-like subspace of codimension one in $\mathbb{R}^{1,n}$ has a space-like normal vector, cf. [6, Lemma 1.1].) By [17, Theorem 3.2.8], the distance X_i from p to B_i satisfies $\sinh X_i = -p \circ x_i$.

The tangent space to $\mathbb{H}^2$ at p is $p^\perp = \{v \in \mathbb{R}^{1,2} \,|\, v \circ p = 0\}$, a space-like subspace by [17, Theorem 3.1.5]. A unit vector $v \in p^\perp$ determines a geodesic $\lambda\colon \mathbb{R} \to \mathbb{H}^2$ parametrized by arclength, given by $\lambda(t) = (\cosh t)p + (\sinh t)v$, with $\lambda(0) = p$ and $\lambda'(0) = v$. Thus by the paragraph above, in order to ensure that the distance between $\lambda(t)$ and B_i increases for $i = 2, 3$, we must make $\lambda(t) \circ x_i$ *decrease*, for instance by choosing v so that $v \circ x_2 < 0$ and $v \circ x_3 < 0$. In fact, since $\lambda''(0) = p$ pairs negatively with x_2 and x_3, it is enough to have $v \circ x_i \leq 0$ for $i = 2, 3$.

The set of all vectors v with $v \circ x_2 \leq 0$ is a closed half-space of $\mathbb{R}^{1,2}$ that intersects $p^\perp$ in a closed half-plane (since $p^\perp \neq x_2^\perp$). Similarly, the set $v \circ x_3 \leq 0$ is a closed half-plane in $p^\perp$, and the result now follows from the standard exercise that in a two-dimensional vector space, the intersection of any two closed half-planes contains a line. □

The hypotheses of Lemma 9 require p to lie in the *interior* of H or $\bar{H}$ because its proof requires p to move in an arbitrary direction and remain in H or $\bar{H}$, so that we can continue to interpret $sys_p(F)$ only in terms of the hexagon. We now consider cases when p lies in an edge.

Lemma 10. *Writing $F = H \cup \bar{H}$ for isometric right-angled hexagons H and $\bar{H}$ such that $H \cap \bar{H}$ is a disjoint union of three edges, the maximum of $sys_p(F)$ is not attained at any point p in the interior of an edge of $H \cap \bar{H}$.*

Proof. Suppose p lies on a shortest geodesic arc between two boundary components, say the arc B_1' joining B_2 to B_3. We first claim that every simple loop based at p that encircles the third component has length strictly larger than $sys_p(F)$. This can be seen by a surgery

(a) A loop and its mirror image

(b) A loop about B_3

Figure 11: Surgery on geodesic loop segments

argument akin to that of Lemma 7. By the proof of that lemma, a shortest simple geodesic loop $\gamma\colon I \to F$ based at p and freely homotopic to B_1 must have both its initial and terminal endpoints in either H or $\bar{H}$; and assuming the former, as in the proof of Lemma 7 we may write

$$\gamma = \gamma|_{[0,t_1]}.\gamma|_{[t_1,t_2]}.\gamma|_{[t_2,1]},$$

where $\gamma|_{[0,t_1]}$ and $\gamma|_{[t_2,1]}$ map entirely into H and $\gamma|_{[t_1,t_2]}$ maps into $\bar{H}$. Note in this case that the mirror image $\bar{\gamma}$ of γ is also a simple geodesic loop based at p that is freely homotopic to B_1, this time with the opposite orientation, with the same length as γ but its initial and terminal segments in $\bar{H}$. See the red and blue loops of Figure 11(a).

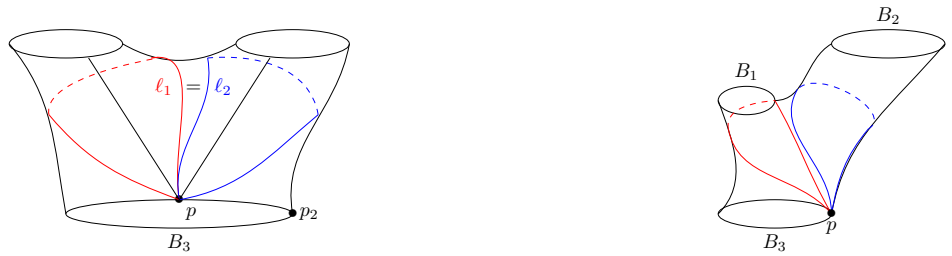
We can form broken geodesic loops encircling B_2 and B_3 by concatenating each of the initial and terminal segments of γ with their mirror images: $\gamma|_{[0,t_1]}.\bar{\gamma}|_{[0,t_1]}$ and $\bar{\gamma}|_{[t_2,1]}.\gamma|_{[t_2,1]}$. See Figure 11(b). But at least one of these broken geodesics already has length strictly shorter than that of γ and $\bar{\gamma}$, proving the claim.

We finish the proof by showing that the lengths ℓ_2 and ℓ_3 of the simple loops based at p and freely homotopic to B_2 and B_3, respectively, are simultaneously increased by moving p off of the geodesic B_1' perpendicularly into H. We lift H to $\mathbb{H}^2$ and use notation from the proof of Lemma 9, in particular taking x_2 and x_3 to be space-like unit normals to the geodesics containing the edges of H that lie in B_2 and B_3, respectively. Since by hypothesis p lies in the interior of B_1' — that is, not in B_2 or B_3 — we may again choose x_2 and x_3 so

that $p \circ x_2 < 0$ and $p \circ x_3 < 0$.

The proof of Theorem 3.2.7 of [17] implies that the hyperbolic geodesic N containing B_1' is the intersection between $\mathbb{H}^2$ and the span of x_2 and x_3 in $\mathbb{R}^{1,2}$. Therefore a vector $v \in p^\perp$ and normal to N is normal to each of x_2 and x_3. Taking v to point into H and arguing as in the proof of Lemma 9 shows that moving p along the geodesic λ with $\lambda(0) = p$ and $\lambda'(0) = v$ thus increases its distances to B_2 and B_3, hence also ℓ_2 and ℓ_3 by Lemma 8. □

Lemma 11. *If $sys_p(F)$ attains a maximum at $p \in \partial F$, say $p \in B_3$ without loss of generality, and p is not the endpoint of a shortest geodesic arc joining B_3 to B_1 or B_2, then $\ell_1 = sys_p(F) = \ell_2$ (where ℓ_i is the length of the simple loop based at p and freely homotopic to B_i). If $p \in B_3$ is the endpoint of the shortest geodesic arc joining B_3 to B_i, for $i = 1, 2$, then $sys_p(F) = \ell_{3-i}$.*



(a) p lies on the interior of B_3

(b) p lies on the intersection

Figure 12: The two cases of p on the boundary of F

Proof. Continuing to assume $p \in B_3$, we first observe that this implies that $\ell_3 = b_3 \geq \min\{\ell_1, \ell_2\}$. For if ℓ_3 is less than both ℓ_1 and ℓ_2 then $sys_p(F) = \ell_3$ has this property as well. But ℓ_3, and hence in this case also $sys_p(F)$, can be increased by just moving p off of B_3 into F (compare Lemma 8). So no such point maximizes $sys_p(F)$.

Now let p_i be the endpoint on B_3 of the shortest arc joining it to B_i, for $i = 1, 2$ (see Figure 12). These arcs comprise two of the three edges of the intersection $H \cap \bar{H}$, where (as before) H and $\bar{H}$ are right-angled hexagons such that $F = H \cup \bar{H}$ and $H \cap \bar{H}$ is a union of edges. It is a fundamental hyperbolic trigonometric fact that the distance to B_i increases in

both $H \cap B_3$ and $\bar{H} \cap B_3$ as one moves away from p_i. (It can be showed by an exercise using the notation from the proof of Lemma 9, for instance.)

It follows from this fact that if $sys_p(F)$ attains a maximum at any $p \in B_3 - \{p_2\}$ then $\ell_1 \geq \ell_2$, and if $sys_p(F)$ attains a maximum at $p \in B_3 - \{p_1\}$ then $\ell_2 \geq \ell_1$. These follow from the same line of reasoning, so let us consider the first case, $p \neq p_2$, and by way of contradiction suppose $\ell_1 < \ell_2$. By the first paragraph $\ell_3 \geq \ell_1$ at p so there are two possibilities: either $sys_p(F) = \ell_1$ is less than both ℓ_2 and ℓ_3, whence it is not a maximum since it can be increased by moving p away from p_1 along B_3, or $sys_p(F) = \ell_1 = \ell_3 < \ell_2$. In this case moving p away from p_1 along B_3 again increases ℓ_1, yielding a new point p' with $sys_p(F) = sys_{p'}(F)$, but with B_3 the only systolic loop based at p'. But this never occurs at a maximum of $sys_p(F)$, by the first paragraph, contradicting our hypothesis on p.

The combination of the observations above directly implies the Lemma's conclusion. □

Lemma 11 notably does *not* assert that the simple loops freely homotopic to the different boundary components all have the same length at a maximizer for $sys_p(F)$ on ∂F. As we will show, Lemma 6 in fact implies that this *cannot* occur for certain triples (b_1, b_2, b_3) of boundary component lengths. This leads to the corrected statement of [10, Theorem 1.2] in this case.

Theorem 3. *Let F be a hyperbolic three-holed sphere with geodesic boundary components of lengths $b_1 \leq b_2 \leq b_3 \in (0, \infty)$, and let*

$$f_{(b_1,b_2)}(x) := 6\sin^{-1}\left(\frac{1}{2\cosh(x/2)}\right) + 2\sum_{i=1}^{2}\sin^{-1}\left(\frac{\cosh(b_i/2)}{\cosh(x/2)}\right).$$

a) The maximum value of $sys_p(F)$, taken over all $p \in F$, is attained in the interior of F if and only if $f_{(b_1,b_2)}(b_3) > \pi$. In this case it is the unique positive solution x to:

$$f_{(b_1,b_2)}(x) + \sin^{-1}\left(\frac{\cosh(b_3/2)}{\cosh(x/2)}\right) = 2\pi.$$

For a point $p \in int(F)$ at which the maximum value is attained, the systolic loops based at p divide F into an equilateral triangle and three one-holed monogons.

b) If $f_{(b_1,b_2)}(b_3) \leq \pi$, then $sys_p(F)$ attains its maximum on the component B_3 of ∂F of length b_3. Taking $p_2 \in B_3$ to be the endpoint of the shortest arc joining it to B_2, the maximum is attained at $p \in B_3 - \{p_2\}$ if and only if $g(x_2) > \pi$, where

$$g(x) = 2\cos^{-1}\left(\frac{\tanh(b_3/2)}{\tanh(x)}\right) + 2\sin^{-1}\left(\frac{\sinh(b_3/2)}{\sinh(x)}\right) + 2\sum_{i=1}^{2}\sin^{-1}\left(\frac{\cosh(b_i/2)}{\cosh(x/2)}\right)$$

and x_2 satisfies

$$\sinh(x_2/2) = (\cosh(b_2/2)\cosh(b_3/2) + \cosh(b_1/2))/\sinh(b_3/2).$$

If $f_{(b_1,b_2)}(b_3) \leq \pi$ and $g(x_2) > \pi$ then the maximum value of $sys_p(F)$ is the unique $x \in (x_2, b_3]$ such that $g(x) = \pi$. For $p \in F$ where the maximum is attained, the systolic loops based at p divide F into an isosceles triangle and two one-holed monogons.

c) If $f_{(b_1,b_2)}(b_3) \leq \pi$ and $g(x_2) \leq \pi$ then the maximum value x_1 of $sys_p(F)$ is defined by

$$\sinh(x_1/2) = \sqrt{\coth^2(b_3/2)\left[\cosh(b_1/2)\cosh(b_3/2) + \cosh(b_2/2)\right]^2 - \sinh^2(b_1/2)\sinh^2(b_3/2)},$$

attained at p_2, and a systolic loop based at w bounds a one-holed monogon enclosing B_1.

Proof. By Lemma 7 the function $p \mapsto sys_p(F)$ is the minimum of functions ℓ_1, ℓ_2, ℓ_3 each of which depends continuously on p by the formula given in Lemma 8. Therefore since F is compact, $p \mapsto sys_p(F)$ attains a maximum at some $p \in F$.

If the maximum is attained at a point $p \in int(F)$ then by Lemma 9, for each i the simple geodesic loop based at p and freely homotopic to B_i has length equal to $sys_p(F)$, i.e. it is a systolic loop. Each systolic loop bounds an annulus (i.e. a one-holed monogon) with the boundary component it encircles, and the loops do not intersect except at p: if they did then this would give rise to a bigon in F bounded by geodesic arcs, which is impossible. Therefore the monogons bounded by the three systolic loops meet only at p. By inspection, the closure of their complement is the image of an equilateral triangle in $\mathbb{H}^2$, embedded in F except that its vertices are all identified to p, with its edges sent to the systolic loops. The vertex angles of the monogons and the equilateral triangle therefore sum to 2π, since these all meet at p.

As an equilateral triangle with side length $2r$ has vertex angle $\alpha(r)$ given in Theorem 2, and a one-holed monogon with hole radius b_i has vertex angle δ_i given in Lemma 8, we find

that $f_{(b_1,b_2)}(x)$ described above records the vertex angle sum of an equilateral triangle with one-holed monogons of hole radii b_1 and b_2, all with side length x. Therefore $x = sys_p(F)$ satisfies

$$f_{(b_1,b_2)}(x) + 2\sin^{-1}\left(\frac{\cosh(b_3/2)}{\cosh(x/2)}\right) = 2\pi \tag{4.2}$$

as claimed. Since $f_{(b_1,b_2)}$ is the three-holed sphere case of the function of Lemma 6, that Lemma's conclusion implies that $f_{(b_1,b_2)}(b_3) \geq \pi$. But we claim that in fact strict inequality holds in this case. To this end, note that $p \in int(F)$ has positive distance to each boundary component B_i and hence $x = \ell_i > b_i$ for each i by Lemma 8. So since $x > b_3$ the inverse sine term above has value less than π and hence $f_{(b_1,b_2)}(x) > \pi$. Since $f_{(b_1,b_2)}(x)$ decreases with x, the claim follows.

On the other hand, if $f_{(b_1,b_2)}(b_3) > \pi$ then by the other direction of Lemma 6's conclusion there exists a unique $x > b_3$ for which equation (4.2) holds. For this x, we may identify the edges of an equilateral triangle to those of monogons with hole radii b_1, b_2, and b_3, all with side length x to produce an identification space F homeomorphic to the three-holed sphere. This identification space further inherits a hyperbolic structure from its constituent pieces. In particular, because equation (4.2) is satisfied a standard argument shows that the vertex quotient $p \in F$ has a neighborhood isometric to an open disk in $\mathbb{H}^2$ that is a union of wedges: three of angle $\alpha(x/2)$ and one each of angle δ_i for $i = 1, 2, 3$. By construction, F has boundary components B_1, B_2 and B_3 of lengths b_1, b_2 and b_3, respectively, and the simple loops based at p and freely homotopic to each B_i all have length x. Therefore by Lemma 8, $sys_p(F) = x$.

As noted above for this x we have $x > b_3 \geq b_2 \geq b_1$. On the other hand, for any point p' in a boundary component B_i, $sys_{p'}(F) \leq \ell_i = b_i$. Therefore for the three-holed sphere F constructed above, the function $p \mapsto sys_p(F)$ attains its maximum in the interior of F — in fact at p, by what was already showed — and we have established the Theorem's assertion (a).

Henceforth suppose that $f_{(b_1,b_2)}(b_3) \leq \pi$, and hence for a hyperbolic three-holed sphere F with boundary lengths b_1, b_2 and b_3, that $sys_p(F)$ attains its maximum at some $p \in \partial F$. We first claim that by our hypothesis that $b_1 \leq b_2 \leq b_3$, p lies on the component B_3 with

length b_3. This follows from the fact that for every $p \in B_1$, $sys_p(F) \leq \ell_1 = b_1$, and similarly $sys_p(F) \leq b_2$ for each $p \in B_2$. So to prove the claim we just have to exhibit a point $p \in B_3$ such that $sys_p(F) > \max\{b_1, b_2\} = b_2$. For this we take $p = p_2$, the endpoint on B_3 of the shortest geodesic arc B_1' joining B_2 to B_3. At $p = p_2$ we have $\ell_3 = b_3$, and we observe that this is strictly larger than b_2 since $f_{(b_1,b_2)}(x)$ decreases with x and $f_{(b_1,b_2)}(b_2) > \pi$ by direct computation. The formula of Lemma 8 shows also that $\ell_2 > b_2$, since $X_2 = d(B_2, B_3) > 0$.

We now show that $\ell_1 > b_2$ as well, by a somewhat more involved computation. Lemma 8 gives $\sinh(\ell_1/2) = \cosh X_1 \sinh(b_1/2)$, where X_1 is the distance from p_2 to B_1. This can in turn be computed using the same hyperbolic trigonometric identity underlying Lemma 8, since a shortest geodesic arc from w to B_1 (which lies entirely in one of the right-angled hexagons that comprise F) bounds a Saccheri quadrilateral together with three other arcs: half of B_3 joining p_2 to the endpoint of the shortest geodesic arc B_2' joining B_1 to B_3, B_2' itself, and a segment of B_1. From this we obtain $\sinh X_1 = \cosh(b_3/2) \sinh d_2'$, where d_2' is the length of B_2'. The hyperbolic law of cosines for right-angled hexagons supplies the final piece of the puzzle, giving

$$\begin{aligned}\cosh d_2' &= \frac{\cosh(b_1/2)\cosh(b_3/2) + \cosh(b_2/2)}{\sinh(b_1/2)\sinh(b_3/2)} \\ &= \coth(b_1/2)\coth(b_3/3) + \frac{\cosh(b_2/2)}{\sinh(b_1/2)\sinh(b_3/2)}.\end{aligned}$$

The hyperbolic cotangent $\coth x$ decreases to 1 as $x \to \infty$, so fixing b_1 and b_2 above and taking $b_3 \to \infty$ yields the inequality $\cosh d_2' > \coth(b_1/2)$. Putting these together yields

$$\begin{aligned}\sinh(\ell_1/2) &= \sinh(b_1/2)\sqrt{\cosh^2(b_3/2)\sinh^2 d_2' + 1} \\ &> \sinh(b_1/2)\sqrt{\cosh^2(b_3/2)(\coth^2(b_1/2) - 1) + 1} = \sqrt{\cosh^2(b_3/2) + \sinh^2(b_1/2)}\end{aligned}$$

In the last equality above we have used the fact that $\coth^2 x - 1 = 1/\sinh^2 x$, and distributed the $\sinh(b_1/2)$ under the square root. This inequality shows in fact that $\ell_1 > b_3$, so it is certainly larger than b_2 as desired, proving the claim.

Note that in the course of proving the last claim, we explicitly computed the value of ℓ_1 at w. Lemma 11 asserts that this ℓ_1 is the maximum value of $p \mapsto sys_p(F)$, if the maximum is

attained at p_2. The value for x_1 given in part (c) of the Theorem equals this ℓ_1; its equation is obtained from the one above by a sequence of algebraic manipulations.

It remains to show the Theorem's part (b): that $p \mapsto sys_p(F)$ attains its maximum at some $p \in B_3 - \{p_2\}$ if and only if $g(x_2) > \pi$, for g and x_2 as described in part (b) of the Theorem, and to show that the maximum value x satisfies $g(x) = \pi$ in this case. If the maximum of $p \mapsto sys_p(F)$ occurs at $p \in B_3 - \{p_2\}$ then by Lemma 11, $\ell_1 = \ell_2 = sys_p(F)$. By an argument similar to the case that $p \in int(F)$, we find in this case that the closure of the complement of the union of the monogons enclosing B_1 and B_2 is an isosceles triangle with two sides of length $sys_p(F)$ and one of length b_3. All three vertices of this triangle are identified to p, as are the monogons' vertices, so since p lies in ∂F the sum of their vertex angles is π. For $x = sys_p(F)$ we thus have the equation

$$2\cos^{-1}\left(\frac{\tanh(b_3/2)}{\tanh(x)}\right) + 2\sin^{-1}\left(\frac{\sinh(b_3/2)}{\sinh(x)}\right) + 2\sum_{i=1}^{2}\sin^{-1}\left(\frac{\cosh(b_i/2)}{\cosh(x/2)}\right) = \pi,$$

as $\cos^{-1}\left(\frac{\tanh(b_3/2)}{\tanh x}\right)$ is the measure of the equal angles of an isosceles triangle with edge lengths x, x, and b_3 while $2\sin^{-1}\left(\frac{\sinh(b_3/2)}{\sinh x}\right)$ gives the third angle of the same triangle. That is, $g(x) = \pi$.

For this x we have $x = sys_p(F) \leq \ell_3(p) = b_3$ and $x = \ell_2(p) > \ell_2(p_2)$, since $\ell_2(p_2)$ is the minimum value of ℓ_2 over all points of B_3. This follows from the equation of Lemma 8, which shows that ℓ_2 increases with the distance X_2 to B_2, and the fact that p_2 is the endpoint on B_3 of the shortest arc B_1' joining it to B_2. The value of X_2 at p_2 is the length d_1' of this arc, given by the law of cosines for right-angled hexagons analogously to the equation for d_2' above, and plugging this into the formula for $x_2 = \ell_2(p_2)$ given in part (b) of the Theorem. Since $x > x_2$, $g(x) = \pi$, and g is a decreasing function of x (as can easily be seen), we find that $g(x_2) > \pi$.

Now suppose that $g(x_2) > \pi$. We note that $g(b_3) = f(b_3)$ — this can be shown by a series of trigonometric manipulations, or by recalling that $2\cos^{-1}\left(\frac{\tanh(b_3/2)}{\tanh(x)}\right) + 2\sin^{-1}\left(\frac{\sinh(b_3/2)}{\sinh(x)}\right)$ measures the vertex angle sum of an isosceles triangle T with two sides of length x and one of length b_3 — so by our standing assumption that $f(b_3) \leq \pi$ there is a unique $x \in (x_2, b_3]$ such that $g(x) = \pi$. (Recall that $g(x)$ decreases with x.) Arguing as in the case $f(b_3) > \pi$,

we assemble identify the edges of an isosceles triangle T as above to those of monogons with edge length x and hole radii b_1 and b_2, producing an identification space F isometric to a three-holed sphere with boundary components B_i of lengths b_i, for $i = 1, 2, 3$ with the vertex quotient $p \in B_3$. By construction we have $\ell_1(p) = \ell_2(p) = x$, and since $p \in B_3$, $\ell_3(p) = b_3 \geq x$. Therefore $x = sys_p(F)$, and this is a maximum value of $p \mapsto sys_p(F)$ by what we have already showed. We note also that p is not the endpoint w of the shortest arc joining B_3 to B_2, since by construction $\ell_2(p) = x > x_2 = \ell_2(p_2)$. $\square$

5.0 THE HYPERBOLIC TAMMES PROBLEM

In this final chapter, we look to attack the question of maximizing the minimum area horoball for a packing of n horoballs on a hyperbolic surface F with n punctures and genus g. This is motivated by Conjecture 1.2 of [12], which states:

Conjecture. [12, Conjecture 3.2] *Let F be a planar surface admitting a hyperbolic structure. For any horoball packing of F, there is at least one horoball of area less than $10/\sqrt{3}$.*

This leads us to the problem we call the max-min problem, stated below:

The Max-Min Problem. *For a given $n \geq 3$, find the supremum area over all packings of all complete, finite-area, n-cusped planar hyperbolic surfaces (i.e. n-punctured spheres) by horoball cusp neighborhoods of the minimum-area such neighborhood. That is, for any $n \geq 3$, find*

$$\sup_{(F,\mathcal{P})} \left\{ \min_{i=1,\ldots,n} \{\text{area}(B_i)\} \right\}$$

where the supremum is taken over all pairs $(F, \mathcal{P})$ with F a complete, finite-area hyperbolic n-punctured sphere and $\mathcal{P} = \{\{B_i\}_{i=1}^n$ a packing of F by horoball cusp neighborhoods.

This problem is strikingly similar to the Tammes Problem for the hyperbolic plane. The Tammes problem is a well-known question in geometry which looks to find a packing of a given number of disks on the surface of a sphere such that the minimum distance between the disk centers is maximized. The equivalent hyperbolic version is as follows:

The Hyperbolic Tammes Problem. *For a given $n \geq 3$, find the supremum area over all packings of all complete, n-cusped planar hyperbolic surfaces (i.e. n-punctured spheres) by horoball cusp neighborhoods of equal area. That is, for any $n \geq 3$, find*

$$\sup_{(F,\mathcal{P})} \left\{ \min_{i=1,\ldots,n} \{\text{area}(B_i)\} \right\}$$

where the supremum is taken over all pairs $(F, \mathcal{P})$ with F a complete, finite-area hyperbolic n-punctured sphere and $\mathcal{P} = \{B_i\}_{i=1}^n$ a packing of F by horoball cusp neighborhoods of equal area.

It is interesting to note that Theorems 4 and 5 proven in this chapter lack the requirement that the collection of horoball cusp areas have equal area, and thus are the solution the the max-min problem. One might think that the results there will be larger than that of the solution to the hyperbolic Tammes problem. In fact, the solution will always be the same, as proven in the following Lemma.

Lemma 12. *For a given $n \geq 3$, let r_1 be the supremal area of the minimal area horoball cusp neighborhood over all such packings of all complete, n-cusped, planar hyperbolic surfaces. Let r_2 be the supremal area over all packings of the same surface F by horoball cusp neighborhoods of equal area. Then $r_1 = r_2$.*

Proof. Choose any $n \geq 3$ and let F be an n-punctured sphere. Since the collection of all equal-area horoball cusp neighborhood packings is a subset of the collection of all horoball cusp neighborhood packings, $r_2 \leq r_1$.

To see the other inequality, recall that given a horoball cusp neighborhood of given area A, our calculations in Example 1 from Section 2.2.1 allow us to find a horoball cusp neighborhood of any area less than A that lies entirely within the original neighborhood. With this in mind, we see that every packing $\mathcal{C} := \{B_1, \ldots, B_n\}$ by horoball cusp neighborhoods can be replaced by an equal-area packing $\widetilde{\mathcal{C}} := \{\widetilde{B_1}, \ldots, \widetilde{B_n}\}$ of horoball cusp neighborhoods with $\text{area}(\widetilde{B_k}) = \min\{\text{area}(B_i) \,|\, 1 \leq i \leq n\}$ for every $1 \leq k \leq n$ as these sub-neighborhoods will remain embedded and non-overlapping.

Since we have exhibited for any given packing an equal-area packing with radius equal to the minimal radius of the original, we have $r_1 \leq r_2$.

We have shown both inequalities and so $r_1 = r_2$, as desired. □

A final observation is that Böröczky's bound from [1] gives an upper bound to the minimal area horoball cusp neighborhood for any such packing. For all $n \leq 52$, the bound indeed provides a solution less than the $10/\sqrt{3}$ that we are seeking for Hoffman and Purcell's conjecture. However it does not complete the conjecture for *all* n. The following proposition

implies the results we spend much more effort to prove in the remainder of the chapter. We expect the strategy employed in these sections can ultimately give sharper bounds than this in many cases.

Proposition 3. *For a complete, finite area hyperbolic n-punctured sphere F with a packing by horoball cusp neighborhoods of minimum area A, we have*

$$\frac{nA}{2\pi\,(n-2)} \leq \frac{3}{\pi} \quad i.e. \quad A \leq \frac{6\,(n-2)}{n}.$$

Remark 5. This bound is sharp for $n = 3, 4, 6,$ and 12. Note that in these cases we have $A = 2, 3, 4,$ and 5. The triangulations given by the boundaries of the tetrahedron, octahedron, and icosahedron have 4, 6, and 12 vertices respectively, each of valence 3, 4, and 5 respectively. The geometric decoration of $[1, 1, 1]$ for each triangle (see Section 5.1 below for an introduction to this concept) thus attains Böröczky's bound for $n = 4, 6,$ and 12. The $n = 3$ case is similarly attained by doubling the triangle with decoration $[1, 1, 1]$ along its boundary. We will prove in Sections 5.2 and 5.3 that these decorations which provide the bound are not the *unique* such decorations of these surfaces.

Proof. (of Proposition 3) Given a packing of F by horoball cusp neighborhoods, the pre-image in the universal cover $\mathbb{H}^2$ is a horoball packing. To each horoball B in this packing we associate its *Voronoi cell*

$$V(B) := \{x \in \mathbb{H}^2 : d(x, B) \leq d(x, B') \text{ for all other horoballs } B' \text{ in the packing}\}.$$

By standard results, each Voronoi cell $V(B)$ is an infinite area polygon and the collection of Voronoi cells is invariant under the action of $\pi_1(F)$ on $\mathbb{H}^2$ by isometric covering transformations. Recall that for each horoball B, the stabilizer in the fundamental group $\pi_1(F)$ of its center is a parabolic subgroup P, and $int(B)/P$ embeds in F. The same holds for $V(B)$.

Böröczky's theorem implies that

$$\frac{\text{area}(B/P)}{\text{area}(V(B)/P)} \leq \frac{3}{\pi} \tag{5.1}$$

for each such horoball B.

The left hand side of the inequality is the density of the horoball cusp neighborhood B/P in the projected Voronoi cell $V(B)/P$, which is also the "density" of B in $V(B)$. (Here, as both B and $V(B)$ are infinite-area, we take the limit of $\frac{B\cap R_i}{V(B)\cap R_i}$ for regions $R_i \subset \mathbb{H}^2$ of finite area which are increasing to infinity.) The right hand side of the inequality is the ratio of the area of the intersection of the three mutually tangent horoballs in $\mathbb{H}^2$ with the ideal simplex spanned by their centers to the area of the simplex itself (which is π).

Since the union of all Voronoi cells is $\mathbb{H}^2$, F is the union of the projection of these cells, i.e. $F = \bigcup_{i=1}^{n} V(B_i)/P_i$, where $\{B_1/P_1, \ldots, B_n/P_n\}$ is the collection of all horoball cusp neighborhoods of F. If the minimal-area horoball cusp neighborhood is given by A, then

$$\begin{aligned} nA &\leq \sum_{i=1}^{n} \text{area}(B_i/P_i) \\ &\leq \sum_{i=1}^{n} \frac{3}{\pi} \cdot \text{area}\left(V(B_i)/P_i\right) \\ &= \frac{3}{\pi} \sum_{i=1}^{n} \text{area}(V(B_i)/P_i) = \frac{3}{\pi} \cdot \text{area}(F) \\ &= \frac{3}{\pi} \cdot 2\pi(n-2) \end{aligned}$$

The inequality 5.1 was used to move from the first to second line and the Gauss-Bonnet theorem gives the last line, concluding the proof. □

5.1 SET-UP

In the proofs in Sections 5.2 and 5.3, I will be working in the *geometric decoration* coordinate system defined by Hoffman and Purcell in [12, §3] as they are nice to work in. Before we can define the system, we need to understand what it is to have a *decorated triangulation* of a surface.

Definition. A *triangulation* of a surface is a homeomorphism to the quotient space of a finite collection of triangles in $\mathbb{H}^2$ by homeomorphically pairing all edges.

Definition. A *decoration* of a triangle T is a triplet of positive real numbers $[a, b, c]$, each corresponding to a vertex of T.

We associate each number in the triplet to the area of a horoball cusp neighborhood cross-section contained within the triangle and centered at the vertex it decorates.

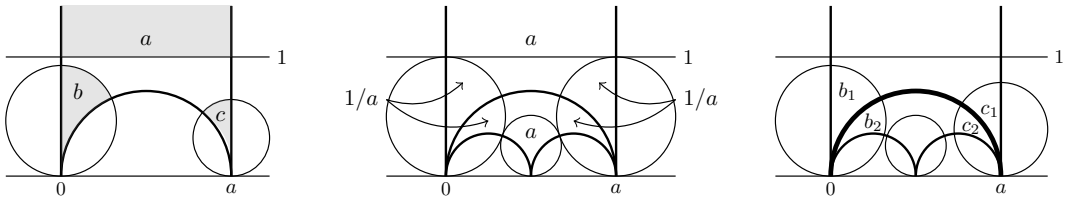
Definition. A *decorated triangulation* is a triangulation of a surface along with a decoration for each triangle in the collection.

Definition. [12, **Definition 3.2**] A decorated triangulation of a surface is said to be *geometric* if the following conditions hold.

(A) No vertex of a triangle is decorated with area more than two.

(B) If one vertex is decorated with A, the remaining vertices are decorated with areas no more than $1/A$.

(C) If two triangles share an edge in a triangulation, with ideal vertices of that edge labeled a_1 and b_1 in one triangle, and a_2 and b_2 in the other, then $a_1 b_1 = a_2 b_2$.



(a) A typical decoration (b) Condition (B) (c) Condition (C)

Figure 13: Conditions on a geometric decoration

They further explain that the conditions on being a geometric decoration are precisely those needed to guarantee the existence of a surface with these markings. More precisely:

Lemma. [12, **Lemma 3.4**] *A decoration of a triangulated surface S that is geometric, in the sense of [12, Definition 3.2], determines a complete hyperbolic structure on the non-compact surface $F = S \setminus V$ with a choice of horoball cusp neighborhood at each cusp, such*

that horoball cusp neighborhoods are embedded and non-overlapping. Here, V is the set of vertices of the triangulation.

Definition. We call the complete hyperbolic structure that results from a geometric decoration as dictated by the Lemma above the *geometric realization* of the decoration.

The idea of the proof of [12, Lemma 3.4] is that F is constructed by gluing together ideal hyperbolic triangles corresponding to the triangles of S. Each entry of the decoration of a triangle T represents the area of a horoball sector of the ideal hyperbolic triangle corresponding to T centered at the respective ideal vertex. Here, we define a *horoball sector* as the intersection of a horoball with two half-planes whose boundaries have the horoball's center as an ideal endpoint.

Figure 13a gives an example of a typical decoration $[a, b, c]$ for an ideal triangle with vertices 0, a, and ∞. In particular, condition (A) ensures that the horoball sectors are embedded in the triangle; condition (B) ensures that the horoball sectors are disjoint in each triangle (see Figure 13b); and condition (C) ensures that triangles can be glued together across edges of the triangulation. For example, in Figure 13c, we must have $b_1c_1 = b_2c_2$ as these abutting triangles share the thick edge made up of the semi-circle of radius a centered at $(a/2, 0)$ and $-\log(b_ic_i)$ is the distance between the horoball sectors in each triangle (cf. [12, Lemma 3.3]).

Fact ([12])**.** The area of the horoball cusp neighborhood with ideal center p comes from summing the decoration corresponding to p for each triangle of which it is an ideal vertex.

While this is not explicitly stated in [12], they do indeed use it.

Remark 6. The labeling of $[1, 1, 1]$ described in Remark 5 attaining the maximal area of the minimal area horoball cusp neighborhood in such a packing of an n-punctured sphere for $n = 3, 4, 6$ and 12 provided by Böröczky's theorem is indeed geometric as it satisfies conditions (A)-(C) in Definition [12, Definition 3.2].

To guarantee the use of these coordinates will attain the maximum value the hyperbolic Tammes problem is seeking, we must ensure that they do in fact cover the entire Teichmüller space of a generic surface. Theorem 3.6 of [7] does this for us.

Proposition 4. *Every complete, finite-area hyperbolic surface equipped with a packing by horoball cusp neighborhoods is the geometric realization of some geometric decoration.*

Proof. Let F be a complete, finite-area hyperbolic surface with a packing by horoball cusp neighborhoods. By [7, Theorem 3.6], for Γ a finitely generated discrete subgroup of isometries of $\mathbb{H}^2$ so that the hyperbolic surface $F := \mathbb{H}^2/\Gamma$ has finite volume, there is a canonical decomposition of the surface by disjointly embedded geodesics into a finite number of cells. Additionally, in the remarks preceding the statement of the theorem, they also state that the structure provides a locally finite tessellation of totally geodesic faces with no zero-dimensional cells (they all lie in the positive light cone). That is, each cell is indeed an ideal polygon.

After possibly sub-dividing these cells into triangles by adding the necessary geodesics, we can simply measure the area of each horoball cusp neighborhood with center p intersected with each triangle which has p as a vertex to obtain a decoration. As we began with a valid packing on a hyperbolic structure, we are guaranteed that this decoration will be geometric. □

Below we relate the geometric decorations described above on an ideal triangulation of the surface F with the coordinate system described by Bowditch and Epstein in [2] via their *spinal triangulation*. While this point of view is not essential, it gives an alternative and more concrete approach to showing the existence of geometric decorations toward the proof of Proposition 3. I will describe the system now. Figure 14 describes the relationship between the two systems as a preview to the construction.

5.1.1 Spinal Triangulations

While Hoffman and Purcell begin with a geometric decoration which can be realized as a surface with a packing by horoball cusp neighborhoods, Bowditch and Epstein created a triangulation by beginning with a surface and packing.

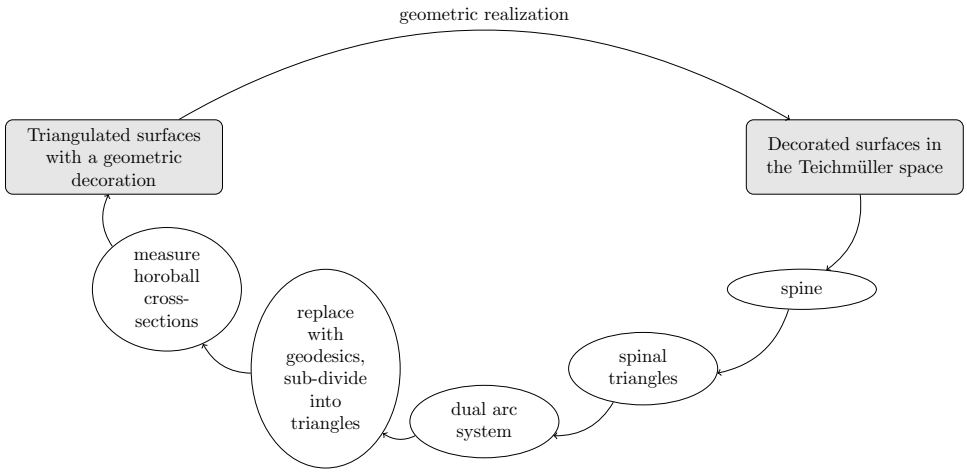


Figure 14: The correspondence between the set of geometric decorations and the set of decorated surfaces

Definition. Given a hyperbolic surface F with a packing by horoball cusp neighborhoods of areas $\{x_y > 0 : y \text{ is an cusp of } F\}$, there exist *spinal coordinates* $\theta(e^*)$ such that for each cusp y, we have

$$x(y) = \frac{1}{2}\sum\{\theta(e^*) : e^* \text{ is incident at } y\}.$$

Here, $x(y)$ is the ratio of the cusp neighborhood area compared to the total sum of all x_y (i.e. $x(y_0) := x_{y_0}/\sum x_y$) and e^* is an edge of Harer's complex, the construction of which is found below.

To determine these coordinates, we must first define the *spinal triangulation* of F, given a set of cusps P and a function $x\colon\ P \to [0,1]$ such that $\sum_P x(y) = 1$. We denote by $B(x,y)$ the horoball cusp neighborhood with center y and horocyclic boundary of length $x(y)$ and define $B(x) := \bigcup_{y\in P} B(x,y)$.

Definition. The *spine* (or *cut locus*) Σ of (S,P,x) is defined to be the set of all points in $F \setminus P$ such that the number of distinct shortest geodesics to $B(x)$ is at least two. That is, all points which have at least two "closest" horoball cusp neighborhoods. We define the *vertices* of Σ by V to be all points in $F \setminus P$ with at least three distinct shortest geodesics. For $v \in V$, we define the set of *ribs* incident on v to be the collection of retraction lines ending at v, i.e.

the set of all geodesics $\gamma(v)$ from y to v for each y such that $B(x, y)$ is one of the horoball cusp neighborhoods closest to v. The collection of all ribs over all $v \in V$ is denoted by R.

The previous constructions give the spinal triangulation of F, as proven in [2, Theorem 2.2.4].

Theorem. [2, Theorem 2.2.4](Existence of spinal triangulations) *The edges and vertices of Σ, together with the ribs R and the set P give a finite triangulation of F, which depends only on the hyperbolic structure on $F \setminus P$ and the function $x : P \to [0, 1]$. If e is an edge of Σ, the two triangles abutting on e are congruent by reflection in e. Each triangle has exactly one vertex in P, and exactly one side in Σ, namely, the opposite side to that vertex. Reflection in the edge e of Σ, restricted to the two adjacent triangles, interchanges the piece of $B(x)$ in one triangle with the piece of $B(x)$ in the other triangle.*

By [2, Lemma 2.2.1], V is a finite set of points; $\Sigma \setminus V$ consists of a finite union of open geodesic arcs which each endpoint a point of V; and each point of V is at the head of at least three directed arcs of Σ. Thus, each triangle in the spinal triangulation has precisely one ideal vertex and two non-ideal vertices. To find an *ideal* triangulation such as we found from the geometric decoration, we must continue on to Harer's complex, the construction of which is described in [2, §3].

Definition. Denote by E the set of edges of Σ. As described in [2, Theorem 2.2.4], for each $e \in E$, let Δ_1 and Δ_2 be the two triangles abutting e, with the ideal vertex of each of these triangles y_1, y_2, respectively. Define $\alpha(e)$ to be the length of the horocycle boundary of $B(x, y_i)$ inside Δ_i. By the symmetry described in the theorem, this value is independent of $i = 1, 2$. We then define

$$\theta(e) := 2\alpha(e).$$

We are now ready to define Harer's complex. Define e^* to be the arc from y_1 to y_2 which crossed e once and Σ nowhere else. As e^* is not defined to be a *geodesic* arc, we also assume that e^* crosses no rib. By [2, Lemmas 3.1, 3.2], the collection of all e^* form a non-trivial, proper arc system which is dual to Σ. An *arc system* is a set of arcs in F that are disjoint except for their endpoints in P. It is considered *non-trivial* if no arc bounds a disk in $F \setminus P$ and no two arcs bound a disk in $F \setminus P$. It is called *proper* if each complementary region is

simply connected an contains at most one point of P.

Lemma. [2, **Lemmas 3.1, 3.2**] *The edges e^* ($e \in E$) form a non-trivial, proper arc system.*

Finally, we define the coordinates of the ideal triangulation. We denote by σ the collection of all e^* dual to Σ.

Theorem. [2, **Theorem 4.1**] *There is a hyperbolic structure on $F \setminus P$, unique up to isotopy rel P, such that σ is the dual of the associated spine Σ, and such that $\theta(F, P, x)(e) = \theta(e^*)$ for each edge e of Σ.*

Here, $\theta(e^*)$ is defined to be such that $\theta : \sigma \to (0, 1]$, $\sum_{e^* \in \sigma} \theta(e^*) = 1$, and for each $y \in P$, $x(y) = \frac{1}{2}\{\theta(e^*) : e^* \text{ is incident at } y\}$. Note that if both ends of e^* are incident at y, we add in $2\theta(e^*)$ instead of $\theta(e^*)$.

So for the ideal triangulation resulting in replacing e^* by a geodesic with the same ideal endpoints (if necessary), the coordinates $\theta(e) = \theta(e^*)$ describe what we call the *spinal coordinates*.

Beginning with a geometric decoration, we can produce the geometric realization by [12, Lemma 3.4] and go through the spinal triangulation process to measure the spinal coordinates. It turns out that we can also go in the other direction, though it is a longer process. See Figure 14.

Starting with a decorated surface in the Teichmüller space of F, i.e. a collection of horoball cusp neighborhood areas, we can find the spinal triangulation associated to the packing by horoball cusp neighborhoods, followed by the dual arc system. This may result in non-geodesic ideal arcs. After replacing these with ideal geodesics, and possibly subdividing any non-triangular regions, we have an ideal triangulation. As we began with a decorated hyperbolic surface, measuring the length of the intersections of the horoballs with each ideal triangle results in a geometric decoration for the surface.

Given a set of horoball cusp neighborhood areas, there is a unique spinal triangulation and therefore a unique set of spinal coordinates. However any given surface may have more than one ideal triangulation, and therefore there is not a unique geometric decoration. Certainly beginning with spinal coordinates and following the process to a geometric decoration then back again will bring us back to the original coordinates given. However it may not be

the case that starting with a geometric decoration, measuring the spinal coordinates, and then going through the process of finding the dual ideal triangulation to this surface will result in the original geometric decoration. That is to say, not every geometric decoration triangulation is dual to the spine of its geometric realization. The three-punctured sphere discussed in Section 5.2 below exemplifies this circumstance.

5.2 THE THREE-PUNCTURED SPHERE

The main goal of this section is to prove Theorem 4:

Theorem 4. *The maximum area of the minimal-area horoball cusp neighborhood, over all packings of the hyperbolic three-punctured sphere by such neighborhoods, is* 2.

Lemma 13. *There are only two triangulations of the three-punctured sphere.*

Proof. The three-punctured sphere is a sphere with three marked points. Call them v_0, v_1, v_2. Beginning with v_0, since each vertex must have valence at least one and an edge cannot bound a disk, let us without loss of generality connect it to v_1. Similarly v_2 must connect to either v_0 or v_1. Let's call the endpoint v_0. Since we also cannot allow any two edges to bound a disk, if v_1 has valence greater than 1, it must be connected to v_2. This is our first triangulation. Otherwise both v_1 and v_2 both have valence 1 and thus v_0 has valence 4, as shown below in Figure 15. □

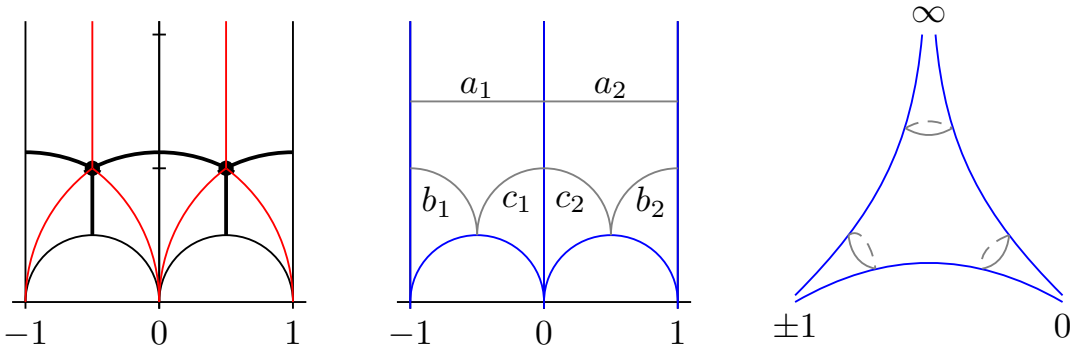
(a) First triangulation (b) Second triangulation

Figure 15: The two triangulations of the three-punctured sphere

5.2.1 Triangulation 1

Lemma 14. *For the triangulation shown in Figure 16, the maximum of the minimal-area horoball cusp neighborhood, over all such packings, is* 2. *This maximum is attained when each horoball has equal area and the geometric decoration of each triangle is* $[1, 1, 1]$

The spinal triangulation attained from the spine in Figure 16a results in the dual spinal triangulation shown in Figures 16b and 16c. In Figures 16a and 16b we use a fundamental domain in the half-plane model of $\mathbb{H}^2$.



(a) Spinal triangulation (b) Geometric decoration (c) Triangulated surface

Figure 16: Three-punctured sphere triangulation 1

This results in horoball cusp neighborhood areas $x_\infty = a_1 + a_2,\ x_{\pm 1} = b_1 + b_2,\ x_0 = c_1 + c_2$. Thus we are looking to maximize $\min\{a_1 + a_2, b_1 + b_2, c_1 + c_2\}$ such that

(1) $a_1 b_1 = a_2 b_2 \leq 1$

(2) $a_1 c_1 = a_2 c_2 \leq 1$

(3) $b_1 c_1 = b_2 c_2 \leq 1$

(4) $0 < a_i,\ b_i,\ c_i \leq 2$,

where these conditions come from Hoffman and Purcell's definition of being a geometric decoration. Notice that the symmetry group of this triangulation is isomorphic to the symmetric group S_3, and so without loss of generality we may assume that $a_1 + a_2 \leq b_1 + b_2 \leq c_1 + c_2$.

Claim 5.1. *At a local maximum,* $a_1 + a_2 = b_1 + b_2$.

Proof. Suppose instead that $a_1 + a_2 < b_1 + b_2$ and consider the continuous transformation which sends $a_i \mapsto ta_i,\ b_i \mapsto \frac{1}{t}b_i,\ c_i \mapsto \frac{1}{t}c_i$, for $t > 1$. This deformation maintains conditions (1), (2), and (3) from above. Condition (4) holds for small enough t as long as $a_i \neq 2$. Notice that if $a_1 = 2$, then $b_1 = \frac{a_2 b_2}{2}$ and $c_1 = \frac{a_2 c_2}{2}$ by conditions (1) and (2), respectively. Thus (3) becomes $\frac{a_2^2 b_2 c_2}{4} = b_2 c_2$, forcing $a_2 = 2$. Hence $4 = a_1 + a_2 < b_1 + b_2 \leq 4$, a contradiction. Similarly for $a_2 = 2$. Thus we are guaranteed that $a_i \neq 2$ for $i = 1, 2$.

Now, notice that this deformation increases $a_1 + a_2 \mapsto t(a_1 + a_2)$ and decreases $b_1 + b_2 \mapsto \frac{1}{t}(b_1 + b_2)$. Thus at a local maximum it must be that $a_1 + a_2 = b_1 + b_2 \leq c_1 + c_2$. □

Claim 5.2. *At a local maximum, $a_1 = a_2 = b_1 = b_2$ and $c_1 = c_2$.*

Proof. By conditions (2) and (3), we get $c_1(a_1 - b_1) = c_2(a_2 - b_2)$. From Claim 5.1 above, $a_1 - b_1 = b_2 - a_2$ and so $-c_1(a_2 - b_2) = c_2(a_2 - b_2)$. Since $c_i > 0$ by condition (4), it must be that $a_2 = b_2$. This implies also that $a_1 = b_1$ by the result in Claim 5.1. Now condition (1) requires $a_1 b_1 = a_2 b_2 \implies a_1^2 = a_2^2$ and since $a_i > 0$ we have $b_1 = a_1 = a_2 = b_2 =: a$. Finally condition (3) allows us to conclude $c_1 = c_2 =: c$ □

Updating the problem statement with this knowledge, we are to maximize $2a$ subject to

(1) $a^2 \leq 1$

(2) $ac \leq 1$

(3) $0 < a \leq c \leq 2$,

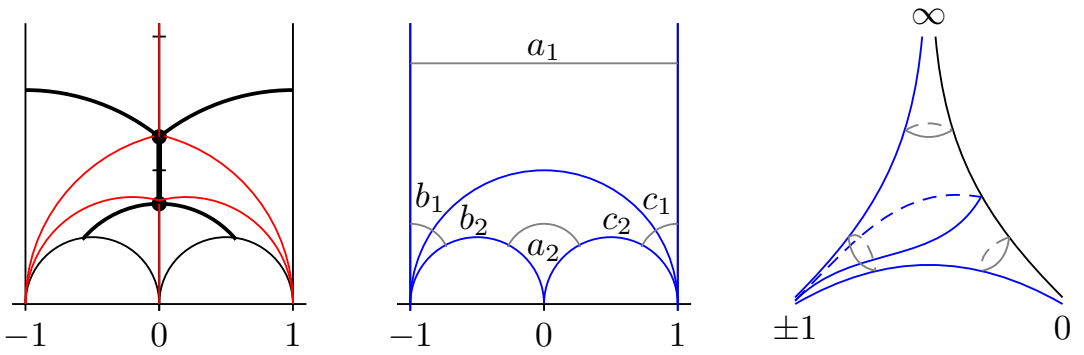
which gives a maximal solution of $a = c = 1$ and the horoball areas all being equal to 2. Thus Lemma 14 is proved.

5.2.2 Triangulation 2

Lemma 15. *For the triangulation shown in Figure 17, the maximum of the minimal-area horoball cusp neighborhood, over all such packings, is 2. This maximum is attained when each horoball has equal area and the geometric decoration of each triangle is $[2, \frac{1}{2}, \frac{1}{2}]$.*

This triangulation of the three-punctured sphere can be formed by flipping the diagonal of the quadrilateral formed by $0, \pm 1,$ and ∞ from the geodesic with ideal endpoints 0 and

∞ of the triangulation from Lemma 14 to the geodesic with ideal endpoints 1 and -1. This triangulation, shown in Figures 17b and 17c, is dual to the spine shown in Figure 17a.



(a) Spinal triangulation (b) Geometric decoration (c) Triangulated surface

Figure 17: Three-punctured sphere triangulation 2

Here $x_\infty = a_1,\ x_{\pm 1} = b_1 + b_2 + c_1 + c_2$ and $x_0 = a_2$ and our constraints are:

(1) $a_1 b_1 = a_1 c_1 \leq 1$

(2) $a_2 b_2 = a_2 c_2 \leq 1$

(3) $b_1 c_1 = b_2 c_2 \leq 1$

(4) $0 < a_i,\ b_i,\ c_i \leq 2$

Claim 5.3. *For any geometric decoration,* $b_1 = c_1 = b_2 = c_2$.

Proof. The first equality comes from condition (1), the last from condition (2). The middle equality then follows by using condition (3). From here on we can define $b_1 = c_1 = b_2 = c_2 =: b$. □

Updating the problem, we are to maximize the minimum of $\{a_1, a_2, 4b\}$ such that

(1) $a_1 b \leq 1$

(2) $a_2 b \leq 1$

(3) $b^2 \leq 1$

(4) $0 < a_i, b \leq 2$

Observe that $\min\{a_1, a_2, 4b\} \leq 2$, since $a_1, a_2 \leq 2$. Then the maximum minimal horoball cusp neighborhood area would be realized only when $a_1 = a_2 = 4b = 2$, with $b = \frac{1}{2}$. This proves Lemma 15.

We prove that this is the unique local maximum with the following two claims. Without loss of generality, let us assume that $a_1 \leq a_2$.

Claim 5.4. *At a local maximum, $4b \geq a_1$, i.e. $4b$ is not the sole minimal horoball cusp neighborhood area.*

Proof. The proof is similar to that of Claim 5.1 from Triangulation 1 in Section 5.2.1. Suppose instead that $4b < a_1$. Since $a_1 \leq 2$, this means that $b < \frac{1}{2}$. Then the transformation that sends $b \mapsto tb$ for $t > 1$ will increase the minimal area horoball and keep all conditions for small enough t. This contradicts the assumption we are at a local maximum. So $min\{a_1, a_2, 4b\} = a_1$. □

Claim 5.5. *At a local maximum, $a_1 = a_2$.*

Proof. Assume instead that $a_1 < a_2$. Consider the continuous transformation which sends $a_1 \mapsto ta_1$ while a_2 and b are sent $x \mapsto \frac{1}{t}x$ for $t > 1$. Then all conditions hold for small enough t as long as $a_1 \neq 2$. Notice that $2 = a_1 < a_2 \leq 2$ cannot be, so we have proven that if $a_1 < a_2$, we can increase the smallest horoball area. Thus at a local maximum we must have $a_1 = a_2 =: a$. □

Thus we are to maximize a subject to $0 < a \leq 2$. By condition (1) we know then that $b \leq \frac{1}{2}$ and $x_0 = x_\infty = 2 \leq x_{\pm 1} \leq 4 \cdot \frac{1}{2} \implies b = \frac{1}{2}$. Thus each horoball has area 2.

This completes the proof of Theorem 4 and exemplifies how distinct geometric decorations can give the same spinal coordinates, though only the first triangulation is the dual to the spinal coordinates $(2, 2, 2)$.

5.3 THE FOUR-PUNCTURED SPHERE

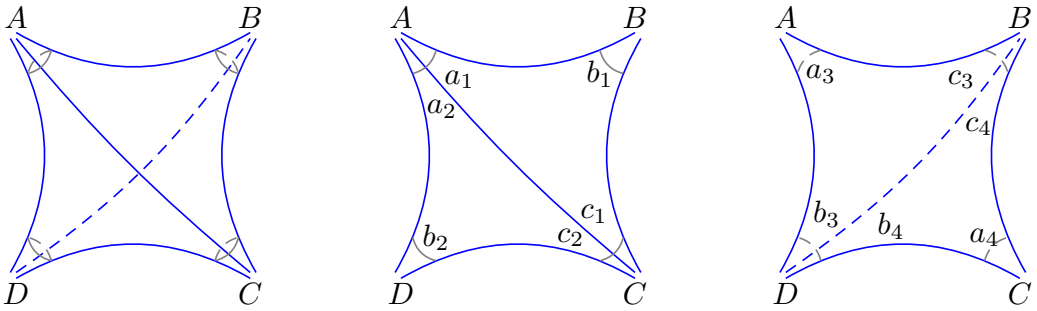
Next we turn our attention to the four-punctured sphere. The Teichmüller space of this surface is much larger than that of the three-punctured sphere, but we will be able to eliminate ta large swath of them from being able to give us the maximal value of our function.

Theorem 5. *The maximum area of the minimal-area horoball cusp neighborhood, over all packings of all hyperbolic four-punctured spheres by such neighborhoods, is* 3.

5.3.1 Triangulation 1

Let us start with the most symmetric triangulation below. As there is no longer a single choice of fundamental domain for the four-punctured sphere, I will not illustrate them in any of the models of $\mathbb{H}^2$, but instead focus on the surface itself.

Lemma 16. *For the triangulation shown in Figure 18, the maximum of the minimal-area horoball cusp neighborhood, over all such packings, is* 3. *This maximum is attained when each horoball has equal area and the geometric decoration is either* $[2, \frac{1}{2}, \frac{1}{2}]$ *for each triangle or* $[1, 1, 1]$ *for each triangle.*



(a) Triangulated surface with horoballs

(b) Geometric decoration: front triangles

(c) Geometric decoration: back triangles

Figure 18: Four-punctured sphere triangulation 1

Notice that the horoball centered at A has area $x_A = a_1 + a_2 + a_3$, at B we have $x_B = b_1+c_3+c_4$, at C is $x_C = a_4+c_1+c_2$, and at D the horoball is made up of $x_D = b_2+b_3+b_4$. Thus we look to maximize $\min\{a_1 + a_2 + a_3, b_2 + b_3 + b_4, a_4 + c_1 + c_2, b_1 + c_3 + c_4\}$ subject to the constraints:

(1) $a_1b_1 = a_3c_3 \leq 1$

(2) $a_1c_1 = a_2c_2 \leq 1$

(3) $b_1c_1 = a_4c_4 \leq 1$

(4) $a_2b_2 = a_3b_3 \leq 1$

(5) $b_2c_2 = a_4b_4 \leq 1$

(6) $b_3c_3 = b_4c_4 \leq 1$

(7) $0 < a_i,\ b_i,\ c_i \leq 2$.

Looking at Figure 18, one sees that this triangulation is the boundary of the tetrahedron, and so its symmetry group is the tetrahedral group S_4 acting by permutations of the cusps. Thus, up to renaming, let us assume that they are listed above in order smallest to largest.

Claim 5.6. *At a local maximum,* $a_1 + a_2 + a_3 = b_2 + b_3 + b_4$.

Proof. Suppose instead that $a_1 + a_2 + a_3 < b_2 + b_3 + b_4$ ($\leq a_4 + c_1 + c_2 \leq b_1 + c_3 + c_4$.) First we show that none of the components of the minimal horoball can be 2, i.e. $a_1, a_2, a_3 < 2$. Notice that:

$$b_1 = \frac{a_3c_3}{a_1},\ c_1 = \frac{a_2c_2}{a_1} \qquad \text{conditions (1), (2)}$$

$$\frac{a_2a_3c_2c_3}{a_1^2} = a_4c_4 \qquad \text{condition (3)}$$

$$b_4^2\frac{a_2a_3c_2c_3}{a_1^2} = a_4b_4c_4b_4 = b_2c_2b_3c_3 \qquad \text{conditions (5), (6)}$$

$$\frac{b_4^2a_2a_3}{a_1^2} = b_2b_3$$

From condition (4), we can solve for $a_3 = \frac{a_2b_2}{b_3}$ or we can solve for $a_2 = \frac{a_3b_3}{b_2}$. Replacing a_3 as such in the line above yields $b_4^2a_2^2 = a_1^2b_3^2$. If we instead replace a_2 with the result from condition (4) we would get $b_4^2a_3^2 = a_1^2b_2^2$. Solving each of these for a_2, a_3 respectively leaves

us with $a_2 = a_1\frac{b_3}{b_4}$ and $a_3 = a_1\frac{b_2}{b_4}$.
Plugging these into our assumption that $a_1 + a_2 + a_3 < b_2 + b_3 + b_4$ results in

$$(b_2 + b_3 + b_4)(b_4 - a_1) > 0,$$

i.e. $b_4 > a_1$. Thus $a_1 \neq 2$.

We could instead solve for $a_1,\ a_3$ in terms of a_2 by following the same outline above but using the following conditions: conditions (2) and (4) in the first line, multiplying each side of condition (5) by c_4^2 instead of b_4^2 as in line 3 above, followed by replacements from conditions (3) and (6) and simplifying as in line 4 to see $\frac{a_1a_3c_4^2}{a_2^2} = b_1c_3$. Then we solve condition (1) once for a_3 and separately for a_1, to solve for $a_1,\ a_3$ respectively in terms of a_2. Then making these replacements in the inequality $a_1 + a_2 + a_3 < b_1 + c_3 + c_4$ yields $a_2 < 2$.

To see that $a_3 \neq 2$, we will not surprisingly follow the same pattern but this time use: conditions (1) and (4) in the first line, multiplying each side of condition (6) by a_4^2, followed by replacements from conditions (3) and (5) and simplifying as in line 4 to see $\frac{a_1a_2a_4^2}{a_3^2} = c_1c_2$. Then we solve condition (2) once for a_2 and separately for a_1, to solve for $a_1,\ a_2$ respectively in terms of a_3. Then making these replacements in the inequality $a_1 + a_2 + a_3 < a_4 + c_1 + c_2$ yields $a_3 < 2$.

Now consider the continuous transformation which sends $a_i \mapsto ta_i$ for $i = 1,\ 2,\ 3$ and sends all other coordinates $x \mapsto \frac{1}{t}x$. This maintains the order of the horoball areas as well as conditions (1) - (7) for small enough t and increases the minimum area horoball. Thus if it is strictly smaller, we are not at a local maximum. □

We may from now on take $a_1 + a_2 + a_3 = b_2 + b_3 + b_4$. In fact, the proof above shows that when $a_1 + a_2 + a_3 = b_2 + b_3 + b_4$, we have $a_1 = b_4,\ a_2 = b_3$, and $a_3 = b_2$. I record the updated conditions below:

(1) $a_1b_1 = a_3c_3 \leq 1$
(2) $a_1c_1 = a_2c_2 \leq 1$
(3) $b_1c_1 = a_4c_4 \leq 1$
(4) $a_2a_3 = a_3a_2 \leq 1$
(5) $a_3c_2 = a_4a_1 \leq 1$

(6) $a_2c_3 = a_1c_4 \leq 1$

(7) $0 < a_i,\ b_1,\ c_i \leq 2$.

Claim 5.7. *For any geometric decoration,*

(i) if $a_1 + a_2 + a_3 < a_4 + c_1 + c_2$, *then* $a_1 < c_2$ *and if* $a_1 + a_2 + a_3 = a_4 + c_1 + c_2$, *then* $a_1 = c_2$.

(ii) if $a_1 + a_2 + a_3 < b_1 + c_3 + c_4$, *then* $a_3 < b_1$ *and if* $a_1 + a_2 + a_3 = b_1 + c_3 + c_4$, *then* $a_3 = b_1$

Proof of (i). By condition (2), $a_2 = \frac{a_1c_1}{c_2}$. By condition (5), $a_3 = \frac{a_1a_4}{c_3}$. Thus

$$a_1 + a_2 + a_3 = a_1 + \frac{a_1c_1}{c_2} + \frac{a_1a_4}{c_3} = \frac{a_1}{c_2}(c_2 + c_1 + a_4).$$

If $a_1 + a_2 + a_3 < a_4 + c_1 + c_2$, then

$$\frac{a_1}{c_2}(c_2 + c_1 + a_4) < a_4 + c_1 + c_2 \implies a_1 < c_2.$$

Replacing the $''<''$ by $''=''$ gives the second result. □

Proof of (ii). By condition (1), $a_1 = \frac{a_3c_3}{b_1}$. By condition (6), $a_2 = \frac{a_1c_4}{c_3} = \frac{a_3c_4}{b_1}$ after substitution. Thus

$$a_1 + a_2 + a_3 = \frac{a_3c_3}{b_1} + \frac{a_3c_4}{b_1} + a_3 = \frac{a_3}{b_1}(c_3 + c_4 + b_1).$$

If $a_1 + a_2 + a_3 < b_1 + c_3 + c_4$, then

$$\frac{a_3}{b_1}(c_3 + c_4 + b_1) < b_1 + c_3 + c_4 \implies a_3 < b_1.$$

Replacing the " < " by " = " gives the second result. □

Claim 5.8. *At a local maximum, either* $a_1+a_2+a_3 = b_1+c_3+c_4$ *or* $a_1+a_2+a_3 = a_4+c_1+c_2$.

Proof. Suppose instead that both $a_1 + a_2 + a_3 < b_1 + c_3 + c_4$ and $a_1 + a_2 + a_3 < a_4 + c_1 + c_2$. Then by Claim 5.7, $a_1 < c_2$ and $a_3 < b_1$. Thus we can utilize the continuous transformation which increases a_1 by a factor of t and decreases b_1, c_1, a_4, c_4 by a factor of $\frac{1}{t}$. This maintains $a_1 + a_2 + a_3$ as the minimal-area horoball cusp neighborhood as well as conditions (1) - (7) for small enough t. Since this transformation increases the minimal area, we could not have been at a local maximum. □

Without loss of generality, let us say it is $b_1 + c_3 + c_4 = a_1 + a_2 + a_3$. Thus by Claim 5.7, $a_1 = c_2$. It follows from (2) that $a_2 = c_1$ and thus from (5) $a_3 = a_4$. The updated conditions are as follows:

(1) $a_1 b_1 = a_3 c_3 \leq 1$

(2) $a_1 a_2 = a_2 a_1 \leq 1$

(3) $b_1 a_2 = a_3 c_4 \leq 1$

(4) $a_2 a_3 = a_3 a_2 \leq 1$

(5) $a_3 a_1 = a_3 a_1 \leq 1$

(6) $a_2 c_3 = a_1 c_4 \leq 1$

(7) $0 < a_i,\ b_1,\ c_i \leq 2$.

Claim 5.9. *At a local maximum, all four horoball cusp neighborhoods are of equal area.*

Proof. Suppose that $a_1 + a_2 + a_3 < a_4 + c_1 + c_2$. Then by Claim 5.7, $a_3 < b_1$. Then the continuous transformation which increases a_3 by a factor of t and decreases c_3, c_4 by a factor of $\frac{1}{t}$ will maintain the order of the horoball cusp neighborhood areas as well as conditions (1) - (7) for small enough t. Since this transformation increases the minimal area horoball cusp neighborhood, we could not have been at a local maximum. □

This again gives $a_3 = b_1$ by Claim 5.7, leaving $a_1 = c_3$ by (1) and $a_2 = c_4$ by (3). One last time, let us update the entire problem statement. We have found that at a local maximum, all horoball cusp areas must be equal and so we are trying to maximize $a_1 + a_2 + a_3$ subject to:

(1) $a_1 a_3 \leq 1$

(2) $a_1 a_2 \leq 1$

(3) $a_3 a_2 \leq 1$

(4) $a_2 a_3 \leq 1$

(5) $a_3 a_1 \leq 1$

(6) $a_1 a_2 \leq 1$

(7) $0 < a_i \leq 2$.

Notice the equivalence of conditions (1) and (5), conditions (2) and (6), and conditions (3) and (4), so we condense these to:

(1) $a_1 a_3 \leq 1$

(2) $a_1 a_2 \leq 1$

(3) $a_3 a_2 \leq 1$

(4) $0 < a_i \leq 2$.

Claim 5.10. *If one of the $a_1 = 2$ for some $i = 1, 2, 3$ then the maximal area is 3 with the geometric decoration of each triangle being $[2, \frac{1}{2}, \frac{1}{2}]$.*

Proof. Suppose without loss of generality that $a_1 = 2$. The we have $2a_3 \leq 1 \implies a_3 \leq \frac{1}{2}$ and $2a_2 \leq 1 \implies a_2 \leq \frac{1}{2}$. This means $a_2 a_3 \leq \frac{1}{4} < 1$ and so condition (3) holds. We want to maximize $2 + a_2 + a_3$, which will occur when $a_2 = a_3 = \frac{1}{2}$. □

Claim 5.11. *If $a_i \neq 2$ for all $i = 1, 2, 3$ then the maximal area is 3 with the geometric decoration of each triangle either $[1, 1, 1]$ or $[2, \frac{1}{2}, \frac{1}{2}]$.*

Proof. If none of the $a_i = 2$, then we can increase each of them by a factor of $t \geq 1$ until either some $a_i = 2$ or some $a_i a_j = 1$. In the first case, we no longer satisfy our hypothesis and instead are in the case of Claim 5.10. In the second, let us without loss of generality assume it is $a_1 a_3 = 1$. Then $a_3 = \frac{1}{a_1}$ and substituting this into condition (3) gives $a_2 \leq a_1$. Then $a_2^2 \leq a_1 a_2 \leq 1$ by condition (2) and $a_2 \leq 1$.

We can increase a_2 by a factor of t until either $a_2 = 1$ or $a_1 a_2 = 1$. In the first case, the horoball cusp area is $a_1 + a_2 + a_3 = a_1 + 1 + \frac{1}{a_1}$. This is maximal when $a_1 = 1$ by simple calculus and so $a_1 = a_2 = a_3 = 1$ and our area is 3, as desired.

In the second case, $a_1 + a_2 + a_3 = a_1 + \frac{2}{a_1}$. Notice that if $a_2 = a_3 = \frac{1}{a_1}$, then condition (3) becomes $\frac{1}{a_1^2} \leq 1$ and so $1 \leq a_1 \leq 2$. Thus $a_1 + \frac{2}{a_1}$ is maximized when $a_1 = 1$ and when $a_1 = 2$. We then have either $a_2 = a_3 = 1$ or $a_2 = a_3 = \frac{1}{2}$. In either case $a_1 + a_2 + a_3 = 3$. □

This concludes the proof of Lemma 16.

5.3.2 Triangulation 2

The first triangulation was as symmetric as we could get. Our next thought may be to consider a triangulation with one vertex seeing many triangles and the others seeing as few as possible. This is achievable with loops based at an ideal endpoint, such as the following:

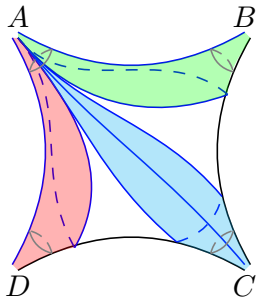


Figure 19: Four-punctured sphere triangulation 2

Cutting along the edge AD opens the area around D into a triangle with edges DA, the loop based at A bounding the "bottom" of this area, and AD. Notice that the edge $DA = AD$ is used twice for this triangle. See the first triangle in Figure 20 for an illustration including the geometric decoration. Similarly we get the second triangle below by cutting along the edge AB and the third triangle by cutting along the edge AC. The final triangle is the "back" of the surface, and is bounded on each side by each of the three loops based at A.

Lemma 17. *For the triangulation shown in Figure 19, the maximum of the minimal-area horoball cusp neighborhood, over all such packings, is* 2*. This maximum is attained when each of the horoballs centered at ideal points B, C, D (as labeled in Figure 19) has area* 2 *but there is not a unique geometric decoration. The decoration for each triangle is $[a_i, a_i, 2]$ for $i = 1, 2, 3$ for any $a_i \leq \frac{1}{2}$ and $[a_4, b_4, c_4]$ for any $0 < a_4, b_4, c_4 \leq 2$ that satisfy the conditions to be a geometric decoration. In this case,*

(1) $a_1^2 = a_4 c_4 \leq \frac{1}{4}$
(2) $a_2^2 = b_4 c_4 \leq \frac{1}{4}$
(3) $a_3^2 = a_4 b_4 \leq \frac{1}{2}$

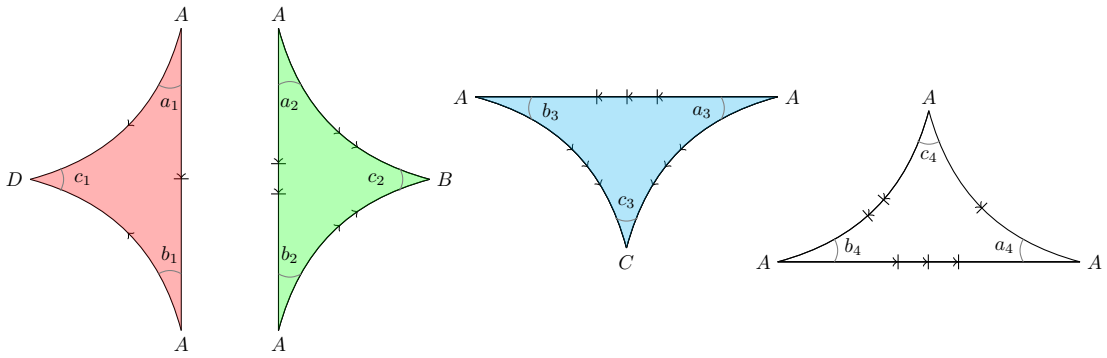


Figure 20: The geometric decoration of triangulation 2

According to the decoration described in Figure 20, we have horoball cusp neighborhood areas $x_A = a_1+a_2+a_3+a_4+b_1+b_2+b_3+b_4+c_4,\ x_B = c_2,\ x_C = c_3,\ x_D = c_1$. Again, we are looking to prove Lemma 17 by maximizing $\min\{a_1+a_2+a_3+a_4+b_1+b_2+b_3+b_4+c_4, c_1, c_2, c_3\}$ subject to the constraints:

(1) $a_1b_1 = a_4c_4 \leq 1$

(2) $a_2b_2 = b_4c_4 \leq 1$

(3) $a_3b_3 = a_4b_4 \leq 1$

(4) $a_1c_1 = b_1c_1 \leq 1$

(5) $a_2c_2 = b_2c_2 \leq 1$

(6) $a_3c_3 = b_3c_3 \leq 1$

(7) $0 < a_i,\ b_i,\ c_i \leq 2$.

Constraints (4), (5), and (6) imply $a_i = b_i \leq 1$ for $i = 1, 2, 3$. Without loss of generality, let us take $c_1 \leq c_2 \leq c_3$. Updating our problem then becomes to maximize $\min\{2a_1 + 2a_2 + 2a_3 + a_4 + b_4 + c_4, c_1\}$ subject to the constraints:

(1) $a_1^2 = a_4c_4 \leq 1$

(2) $a_2^2 = b_4c_4 \leq 1$

(3) $a_3^2 = a_4b_4 \leq 1$

(4) $a_1c_1 \leq 1$

(5) $a_2c_2 \leq 1$

(6) $a_3c_3 \leq 1$

(7) $0 < a_4,\ b_4,\ c_i \leq 2,\ 0 < a_1, a_2, a_3 \leq 1$.

Observe that $\min\{2a_1 + 2a_2 + 2a_3 + a_4 + b_4 + c_4, c_1\} \leq 2$ since $c_1 \leq 2$ and this value is attained when $c_1 = c_2 = c_3 = 2,\ a_1 = a_2 = a_3 = a_4 = b_3 = c_4 = \frac{1}{2}$. This finishes the proof of Lemma 17.

The claims below prove that this is a local maximum.

Claim 5.12. *At a local maximum, x_A cannot be the only horoball of minimum area.*

Proof. Suppose instead that $x_A < x_D = c_1 \leq c_2 \leq c_3$. Then the transformation which sends $x \mapsto tx$ for $x \in \{a_i, b_4, c_4\}$ and $x \mapsto \frac{1}{t}x$ for $x \in \{c_1, c_2, c_3\}$ will increase x_A while decreasing all others. In order to do this, we need to assure three things:

1. $a_1, a_2, a_3 \neq 1$: If $a_1 = 1$, then $c_1 \leq 1$ by (4). But then $x_A > 2 > 1 \geq c_1$, a contradiction.
2. $a_4c_4, b_4c_4, a_4b_4 \neq 1$: If $a_4c_4 = 1$, then by condition (1) $a_1 = 1$, which leads to a contradiction as shown above.
3. $a_4, b_4, c_4 \neq 2$: If $a_4 = 2$, then again $x_A > 2 \geq c_1$, a contradiction.

Thus the transformation holds all of our conditions and we cannot have $x_A < x_D$. □

From here we break into two cases, if one of $a_ic_i = 1$ or none of them do for $i = 1, 2, 3$.

Claim 5.13. *If none of $a_ic_i = 1$ for $i = 1, 2, 3$, then the maximum of the minimum area horoball is 2.*

Proof. Since we have assumed that $c_1 < c_2 \leq c_3$, this means $c_1 < 2$. Along with the hypothesis that $a_1c_1 \neq 1$, we can simply send $c_1 \mapsto tc_1$ and this will maintain all conditions while increasing only the minimum area horoball. Thus at a local maximum, $c_1 = c_2$. Applying the argument again will give $c_1 = c_3$. A third application may not result in equality of all areas as we are not guaranteed that $x_A \leq 2$. However since $min\{x_A, x_D\} = x_D = c_1$ by the proof of Claim 5.12 above, the maximum value is when $c_1 = c_2 = c_3 = 2$. □

Claim 5.14. *If one of the $a_ic_i = 1$ for $i = 1, 2, 3$, then the maximum of the minimum area horoball is 2.*

Proof. Notice we can we write all variables in the inequality in terms of a_1, a_2, a_3.

$$\begin{aligned} c_4 &= \frac{a_1^2}{a_4} && \text{condition (1)} \\ &= a_1^2 \cdot \frac{b_4}{a_3^2} && \text{condition (3)} \\ &= \frac{a_1^2}{a_3^2} \cdot \frac{a_2^2}{c_4} && \text{condition (2)} \\ \implies c_4 &= \frac{a_1 a_2}{a_3} \end{aligned}$$

Plugging this into condition (2) gives $b_4 = \frac{a_2 a_3}{a_1}$ and using that in (3) makes $a_4 = \frac{a_1 a_3}{a_2}$.

Preceding to the proof, we again use that $c_1 \leq x_A$ by Claim 5.12. First let us show that if for each $i = 1, 2, 3$ we have $a_i c_i = 1$, then $c_1 < x_A$. Suppose instead that $x_A = c_i = \frac{1}{a_i}$. Then using our above calculations, we have

$$2a_1 + 2a_2 + 2a_3 + \frac{a_1 a_3}{a_2} + \frac{a_1 a_2}{a_3} + \frac{a_2 a_3}{a_1} = \frac{1}{a_i}.$$

With the condition that all variables be positive, all solutions to this equation require $a_i < \frac{1}{2}$. But since $c_i \leq 2$ and we assumed $a_i c_i = 1$, that would require $a_i \geq \frac{1}{2}$, a contradiction. Therefore $c_1 \neq x_\infty$ and must be strictly smaller.

Now we only need to maximize c_1 subject to our constraints. Let us simply increase our minimum by sending $c_1 \mapsto t c_1$ and $a_1 \mapsto \frac{1}{t} a_1$. Since c_1 must be the minimum area at a local maximum we are in no danger of switching the order of areas and thus can increase it until it reaches 2. Thus we again have that $c_1 = c_2 = c_3 = 2$. □

It is interesting to note that this solution does not force values on our remaining variables. We instead get additional conditions, namely (4), (5), and (6) dictate that $a_1, a_2, a_3 \leq \frac{1}{2}$.

The natural next question is why this triangulation resulted in a smaller minimum area that the first. The condition that limited growth was the fact that $c_1 \leq 2$, which is the condition that assures the horoball remains completely inside its respective triangle.

Fact. For any triangulation T of an n-punctured sphere, if p is an ideal endpoint of precisely one edge of T then the horoball cusp neighborhood centered at p has area less than or equal to two.

Since our first triangulation in Section 16 found the maximum to be three, we can eliminate all triangulations with a vertex one valence as they cannot exceed the current "best" result.

5.3.3 Triangulations with no valence one ideal points

All ideal triangulations in which an ideal point is a vertex of exactly one triangle have been eliminated as producing the maximal minimal area horoball. So the question remains: how many triangulations are there in which every vertex has valence at least two?

Lemma 18. *There are precisely two distinct triangulations of the four-punctured sphere in which every vertex has valence at least two.*

Proof. The four-punctured sphere is a sphere with four marked points. Call them v_0, v_1, v_2, v_3. Beginning with v_0, since it must have valence at least two, it is an endpoint to at least two edges. Either these two edges have the same other endpoint, or distinct other endpoints. Notice that since ours is an ideal triangulation, the lines making up our triangulation cannot cross. Also we cannot have any monogons or bigons in our final product, meaning no areas bounded by only one or two geodesics. Figure 21 shows that if both edges from v_0 both end at, say, v_2 then since both v_1 and v_3 must also have valence at least two they must both have opposite endpoints v_0 and v_2. This leaves us with a triangulation with valence list (2,2,4,4).

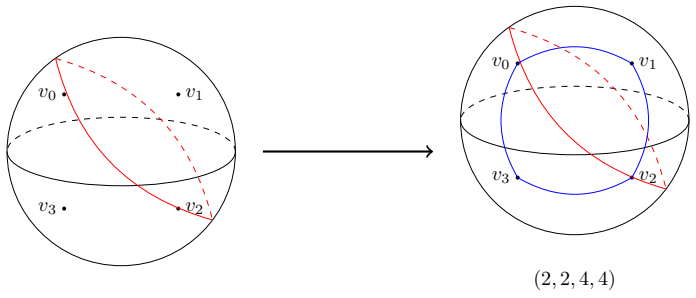


Figure 21: v_0 has two edges both with the same opposite vertex

The more complicated case is when two edges emanating from v_0 have two distinct endpoints, say v_1 and v_3. Figure 22 below follows the options for this case.

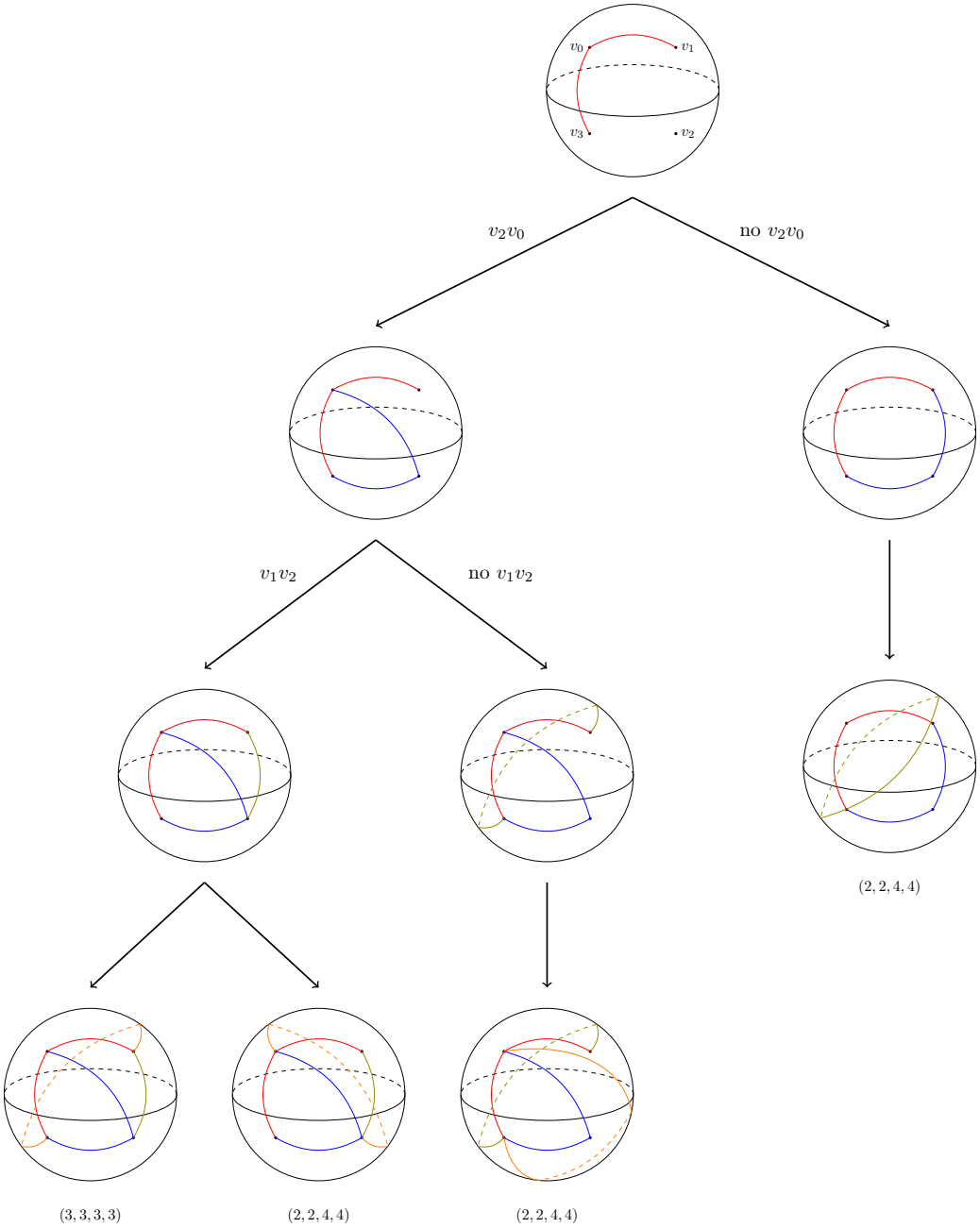


Figure 22: v_0 has two edges with distinct opposite vertices

Consider two edges from v_2. Either the edge from v_2 to v_0 is in the triangulation or it is not. In the latter case, we are forced to have the edges v_2v_1 and v_2v_3. Thus since v_2v_0 is not

in our triangulation, the only way to finish is with v_1v_3 cutting each of the quadrilaterals. Thus we again have valence count (2,2,4,4).

If our triangulation does include edge v_2v_0, then let us without loss say the other edge is v_2v_3. As v_1 must have valence at least two, let us draw its next edge. Either v_1v_2 is in our triangulation or it is not. In the latter case, as we cannot have monogons or bigons, it must be that v_1 is connected to v_3. Hence the only way to finish this case is to connect v_0v_3 as shown below, leaving a valence count of (2,2,4,4).

If the triangulation does include edge v_1v_2, then we are left with two triangles and a single quadrilateral which we can cut one of two ways. If we use v_0v_2 we have the same exact picture as in Figure 21. The opposite cut uses v_1v_3 and gives us Triangulation 1 from Section 5.3.1, with valence count (3,3,3,3).

Notice that all of the cases in which we have valence count (2,2,4,4) are topologically equivalent. □

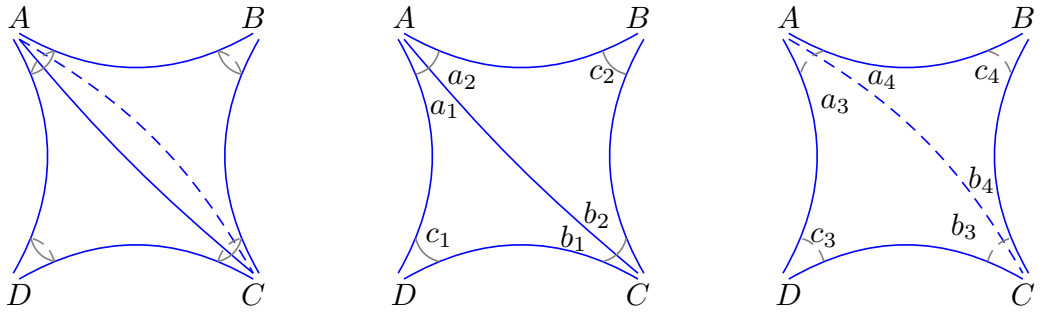
Since we have solved the maximum value of the minimum area horoball for the (3,3,3,3) case to be three, we now turn our attention to the (2,2,4,4) case.

5.3.4 Triangulation 3

The triangulation for a four-punctured sphere with valence count (2,2,4,4) is shown in Figure 23 below. With the decoration described there, we see that the horoball areas are $x_A = a_1 + a_2 + a_3 + a_4,\ x_B = c_2 + c_4,\ x_C = b_1 + b_2 + b_3 + b_4,\ x_D = c_1 + c_3$.

Notice that this triangulation is nearly identical to the triangulation in Section 5.3.1. The only difference is a "flip" of the back geodesic, AC in Figure 23c versus BD in Figure 18c. It is interesting to note that for certain geometric decorations, this flip will result in another valid geometric decoration. We record when it is geometrically possible to flip an edge in the following lemma.

Lemma 19. *Suppose* $[a_1, b_1, c_1]$ *and* $[a_2, b_2, c_2]$ *are the decorations of two abutting triangles on a hyperbolic surface which are part of a geometric decoration. Then there is a geometric decoration obtained by flipping the edge of intersection of these triangles, where* $[a_1, b_1, c_1]$ *and* $[a_2, b_2, c_2]$ *are replaced by* $[\alpha, \gamma_1, \gamma_2]$ *and* $[\beta, \delta_1, \delta_2]$*, as shown in Figure 24, with* $\alpha = a_1 + a_2$



(a) Triangulated surface with horoballs

(b) Geometric decoration: front triangles

(c) Geometric decoration: back triangles

Figure 23: Four-punctured sphere triangulation 3

and $\beta = b_1 + b_2$ *if and only if both* $\alpha = a_1 + a_2 \leq 2$ *and* $\beta = b_1 + b_2 \leq 2$. *If this is the case, the following conditions determine the new decoration.*

(1) $\gamma_1 + \delta_1 = c_1$

(2) $\gamma_2 + \delta_2 = c_2$

(3) $\gamma_1\gamma_2 = \delta_1\delta_2 \leq 1$

(4) $\alpha\gamma_i = a_i c_i$

(5) $\beta\delta_i = b_i c_i$

Moreover, the resulting geometric decoration determines the same surface and horoball packing as the first decoration.

Proof. Suppose two abutting triangles have the geometric decoration illustrated in the left side of Figure 24. As we wish for the decoration on the right side to be of the same point in the decorated Teichmüller space, i.e. have the same collection of horoball cusp neighborhood areas, we must have both $\alpha = a_1 + a_2$ and $\beta = b_1 + b_2$, as well as conditions (1) and (2). For the second decoration to be geometric, condition (A) in Definition 3.2 of [12], recorded in Section 5.1 of this dissertation then forces the inequalities $a + 1 + a_2, b_1 + b_1 \leq 2$. Notice that

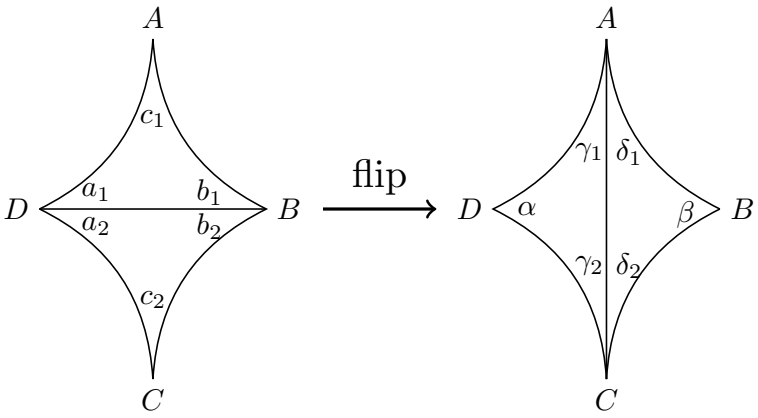


Figure 24: Flipping an edge of the ideal triangulation

for positive solutions, conditions (1) and (2) force each γ_i and δ_i to be less than or equal to 2 as well, so condition (A) is satisfied for any solution.

Condition (B) of the definition [12, Definition 3.2] is satisfied by the inequality in condition (3) above and the equality of (3) is necessary for condition (C). Recall that condition (C) ensured that triangles can be glued together across edges. This is a measure of how far apart the horoball cusp neighborhoods are from each other. Since the outer ideal geodesics of the quadrilaterals in Figure 24 remain unchanged, we must have the same distance between horoball cusp neighborhoods on each side of the flip. This necessitates conditions (4) and (5).

Next, we verify that this set of equations and inequality has a solution. Notice that condition (4) and (5) above give $\gamma_i = \frac{a_i c_i}{a_1+a_2}$ and $\delta_i = \frac{b_i c_i}{b_1+b_2}$. Replacing these in condition (1) and (2) give

$$\begin{aligned}\gamma_i + \delta_i &= \frac{a_i c_i}{a_1 + a_2} + \frac{b_i c_i}{b_1 + b_2} \\ &= c_i \left(\frac{a_i(b_1 + b_2) + b_i(a_1 + a_2)}{(a_1 + a_2)(b_1 + b_2)} \right) \\ &= c_i \left(\frac{a_i b_1 + a_i b_2 + a_1 b_i + a_2 b_i}{a_1 b_1 + a_1 b_2 + a_2 b_1 + a_2 b_2} \right)\end{aligned}$$

Recall that since the original decoration $\{[a_i, b_i, c_i]\}_{i=1,2}$ was assumed to be geometric, we have $a_1 b_1 = a_2 b_2$ by condition (C) of [12, Definition 3.2]. Thus the denominator will be

$2a_1b_1 + a_2b_1 + a_2b_1$. When $i = 1$ the numerator looks like $a_1b_1 + a_1b_2 + a_1b_1 + a_2b_1 = 2a_1b_1 + a_2b_1 + a_2b_1$. When $i = 2$ the numerator becomes $a_2b_1 + a_2b_2 + a_1b_2 + a_2b_2 = 2a_1b_1 + a_2b_1 + a_2b_1$. Hence in both cases, the fraction is 1 and $\gamma_i + \delta_i = c_i$, as desired.

We can also substitute these solutions of equations (4) and (5) into

$$\begin{aligned}\gamma_1\gamma_2 - \delta_1\delta_2 &= \frac{a_1c_1a_2c_2}{(a_1+c_2)^2} - \frac{b_1c_1b_2c_2}{(b_1+b_2)^2} \\ &= c_1c_2\left(\frac{a_1a_2(b_1+b_2)^2 - b_1b_2(a_1+a_2)^2}{(a_1+a_2)^2(b_1+b_2)^2}\right)\end{aligned}$$

Now the numerator is

$$\begin{aligned}&a_1a_2b_1^2 + 2a_1a_2b_1b_2 + a_1a_2b_2^2 - b_1b_2a_1^2 - 2b_1b_2a_1a_2 - b_1b_2a_2^2 \\ =&a_1b_1(b_1a_2 - b_2a_1) - a_2b_2(-a_1b_2 + a_2b_1) \\ =&(a_1b_1 - a_2b_2)(b_1a_1 - b_2a_1) = 0,\end{aligned}$$

again using the fact that $a_1b_1 = a_2b_2$. Thus $\gamma_1\gamma_2 - \delta_1\delta_2 = 0$ and the equality in condition (3) holds.

To attain the inequality of (3), notice that

$$\gamma_1\gamma_2 = \frac{a_1a_2c_1c_2}{(a_1+c_2)^2} \leq \frac{a_1a_2c_1c_2}{4a_1a_2} = \frac{c_1c_2}{4} \leq 1,$$

where the first inequality comes from the arithmetic-geometric mean inequality and the last inequality comes from the fact that by assuming the original decoration was geometric, condition (A) of the definition [12, Definition 3.2] requires all variables be at most 2.

Thus a solution exists and, assuming both $\alpha = a_1 + a_2 \leq 2$ and $\beta = b_1 + b_2 \leq 2$, the remaining solution will leave us with a new decoration that satisfies all of the conditions for being geometric. □

Lemma 20. *For the triangulation shown in Figure 23, the maximum of the minimal-area horoball cusp neighborhood, over all such packings, is less than or equal to* 3. *If* $a_1 + a_2 > 2$, *then the maximum is strictly less than 3. If* $a_1 + a_2 \leq 3$, *then the maximum is 3 and is attained when each horoball has equal area and the geometric decoration is either* $[2, \frac{1}{2}, \frac{1}{2}]$ *for each triangle or* $[1, 1, 1]$ *for each triangle.*

Proof. (of Lemma 20) The max-min problem asks us to maximize $\min\{a_1+a_2+a_3+a_4, c_1+c_3, b_1+b_2+b_3+b_4+c_4, c_2+c_4\}$ subject to the constraints:

(1) $a_1b_1 = a_2b_2 \leq 1$

(2) $a_3b_3 = a_4b_4 \leq 1$

(3) $a_1c_1 = a_3c_3 \leq 1$

(4) $b_1c_1 = b_3c_3 \leq 1$

(5) $a_2c_2 = a_4c_4 \leq 1$

(6) $b_2c_2 = b_4c_4 \leq 1$

(7) $0 < a_i,\ b_i,\ c_i \leq 2.$

First, suppose that $a_1 + a_2 > 2$. Notice that if both $a_1, a_2 \leq 1$ then $a_1 + a_2 \leq 2$, therefore at least one of them is strictly greater than 1. Without loss of generality, let us say $a_1 > 1$. Then condition (3) gives us that $c_1 = \frac{a_3c_3}{a_1} \leq \frac{1}{a_1} < 1$. By condition (7) $c_3 \leq 2$.

Therefore $c_1 + c_3 < 1 + 2 = 3$ and

$$\min\{a_1 + a_2 + a_3 + a_4, c_1 + c_3, b_1 + b_2 + b_3 + b_4 + c_4, c_2 + c_4\} \leq c_1 + c_3 < 3.$$

If on the other hand $a_1 + a_2 \leq 2$, then by Lemma 19, we can flip the front geodesic to obtain Figure 25, as shown below.

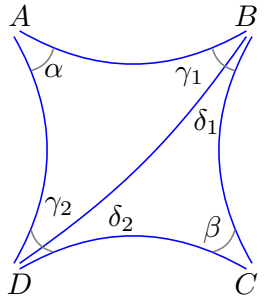


Figure 25: Flipping the front edge AC to BD

while maintaining a geometric decoration. This is precisely the triangulation from Section 5.3.1! (See Figure 18.) The bound follows from Lemma 16. □

Since the minimum area horoball cusp neighborhood area was at most 2 for any triangulation containing a vertex of valence one and the only other triangulation with all vertices at least valence two had its smallest horoball area less than or equal to three, the packing found in Triangulation 1 with all horoball cusp neighborhoods having equal area 3 is our maximal solution and Theorem 5 is thus proven.

BIBLIOGRAPHY

[1] K. Böröczky. Packing of spheres in spaces of constant curvature. *Acta Math. Acad. Sci. Hungar.*, 32(3-4):243–261, 1978.

[2] B. H. Bowditch and D. B. A. Epstein. Natural triangulations associated to a surface. *Topology*, 27(1):91–117, 1988.

[3] Jason DeBlois. Bounds for several-disk packings of hyperbolic surfaces. Preprint. arXiv:1701.07770.

[4] Jason DeBlois. The centered dual and the maximal injectivity radius of hyperbolic surfaces. *Geom. Topol.*, 19(2):953–1014, 2015.

[5] Jason DeBlois. The geometry of cyclic hyperbolic polygons. *Rocky Mountain Journal of Mathematics*, 46(3):801–862, jun 2016.

[6] Jason DeBlois. The Delaunay tessellation in hyperbolic space. *Math. Proc. Cambridge Philos. Soc.*, 164(1):15–46, 2018.

[7] D. B. A. Epstein and R. C. Penner. Euclidean decompositions of noncompact hyperbolic manifolds. *Journal of Differential Geometry*, 27(1):67–80, 1988.

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

[9] Werner Fenchel. *Elementary geometry in hyperbolic space*, volume 11 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1989. With an editorial by Heinz Bauer.

[10] Mathieu Gendulphe. The injectivity radius of hyperbolic surfaces and some Morse functions over moduli spaces. Preprint. arXiv:1510.02581, October 2015.

[11] A. Hatcher. *Algebraic Topology*. Algebraic Topology. Cambridge University Press, 2002.

[12] Neil R. Hoffman and Jessica S. Purcell. Geometry of planar surfaces and exceptional fillings. *Bulletin of the London Mathematical Society*, 49(2):185–201, 2017.

[13] S. Katok. *Fuchsian Groups*. Chicago Lectures in Mathematics. University of Chicago Press, 1992.

[14] J.M. Lee. *Riemannian Manifolds: An Introduction to Curvature*. Graduate Texts in Mathematics. Springer New York, 1997.

[15] B. Martelli. *An Introduction to Geometric Topology*. Bruno Martelli, 2016.

[16] Oleg R. Musin and Alexey S. Tarasov. The Tammes problem for $N = 14$. *Exp. Math.*, 24(4):460–468, 2015.

[17] John G. Ratcliffe. *Foundations of hyperbolic manifolds*, volume 149 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2006.