MCSHANE'S IDENTITY, USING ELLIPTIC ELEMENTS

THOMAS A. SCHMIDT AND MARK SHEINGORN

ABSTRACT. We introduce a new method to establish McShane's Identity. Elliptic elements of order two in the Fuchsian group uniformizing the quotient of a fixed once-punctured hyperbolic torus act so as to exclude points as being highest points of geodesics. The highest points of simple closed geodesics are already given as the appropriate complement of the regions excluded by those elements of order two that factor hyperbolic elements whose axis projects to be simple. The widths of the intersection with an appropriate horocycle of the excluded regions sum to give McShane's value of 1/2. The remaining points on the horocycle are highest points of simple open geodesics, we show that this set has zero Hausdorff dimension.

1. INTRODUCTION

In his 1991 Ph.D. dissertation, G. McShane proved the striking identity

$$\sum_{\gamma} \frac{1}{1 + e^{\ell(\gamma)}} = \frac{1}{2} \; ,$$

where the sum is taken over all simple closed geodesics of any fixed hyperbolic once-punctured torus, and $\ell(\gamma)$ is the length of the geodesic. This has been reproved in various ways: [13], [3], [8]; and generalized variously: [11], [12], [1], [22]. The identity has had deep applications due to Mirzakhani [14] (see also [17]), [15], [16].

We give a proof of the original identity that is, in a sense, intermediate to Mc-Shane's original proof and Bowditch's proof by Markoff triples [3]: it is geometric; but lengths of geodesics do not enter directly. Similar to [8], we take a classical approach; ours involves a uniformizing Fuchsian group. We avoid McShane's invocation of a deep result of Birman and Series; in its place, we show directly that the appropriate complementary set is of Hausdorff dimension zero, and thus certainly of Lebesgue measure zero, as the identity itself requires. This Cantor set is the set of apexes of simple open geodesics that achieve their height.

In the following two paragraphs, we sketch the proof. It is related most directly to the singular punctured sphere that is the quotient of the punctured hyperbolic torus by its elliptic involution. The simple closed geodesics on the torus and this

Date: 18 February 2008.

²⁰⁰⁰ Mathematics Subject Classification. 57M50, 20H10.

sphere are in 1–1 correspondence (and more). On the sphere, each simple closed geodesic bounces between two elliptic fixed points; thus, any hyperbolic element of the uniformizing Fuchsian group whose axis projects to the simple closed geodesic can be factored as a product of elliptics. But, our first lemma shows that any elliptic element of order two increases radii of circles whose apexes lie within its *uplift region*, bounded by Euclidean hyperbolas, in the Poincaré upper half plane, \mathcal{H} .

It is well known (from the work of H. Cohn and others) that there is a lowest horocycle (thus, informally, loop about the cusp) on the hyperbolic torus beyond which no simple geodesic penetrates; this is true as well for the quotient orbifold. The appropriate tree of simple closed geodesics' elliptic factorizations gives a set of uplift regions (suitably trimmed) that fit together so as to raise all apexes (below the lift of the fundamental horocycle) other than those of simple closed geodesics' highest lifts. These regions meet the lift of the fundamental horocycle in disjoint intervals, our *excision intervals*, indexed by the tree of simple closed geodesics. One easily shows that the union of the excision intervals lies in a finite interval; the complement of their union is a Cantor set. We show that along all but countably many branches of the tree, the limit of the ratios of excised interval to ambient interval is 1. The Cantor set thus has Hausdorff dimension zero. It in particular has Lebesgue measure zero; but, the length of each excised interval is a multiple of a corresponding $\frac{1}{1 + e^{\ell(\gamma)}}$, and the full interval has length one-half this multiple, McShane's Identity follows.

1.1. Further Remarks. The Cantor set in our construction is the set of apexes of lifts of those open simple geodesics that have a highest apex lift. The endpoints of our excision intervals correspond to geodesics that spiral about a simple geodesic, the remaining points correspond to 'irrational laminations'. These facts can easily be verified by using [9] (see especially Proposition 18 there) or the more recent [5]. Whereas our excision intervals lie along the fundamental horocycle, [13] finds his gaps along any fixed horocycle closer to the cusp than the fundamental horocycle.

Our approach relies in part on replacing Fricke's equation (in trace coordinates on the Teichmüller space): $a^2+b^2+c^2 = abc$ by an adjusted equation: $x^2+y^2+z^2 = axyz$. In the modular case of a = b = c = 3, it was Cohn's [6] recognition that the adjusted equation is the classical Markoff equation that led him to investigate the geodesy of the corresponding once-punctured hyperbolic torus.

Our approach should recover results of [23] in the setting of a hyperbolic torus with geodesic boundary. Generalizations to higher genus must be carefully pursued: [2] shows that there are non-simple geodesics whose self-intersections are not caused by parabolic elements; such geodesics must then be low in the corresponding height spectrum and thus of highest lifts of apex exterior to all uplift regions. It would be interesting to generalize our techniques to hyperbolic surfaces with more general conical singularities, see [7] and [22].

Finally, we mention that our approach of trees of triples of order two elements is strongly reminiscent of work of L. Yu. Vulakh on the Markoff and Lagrange spectra and their generalizations, see for example [24].

1.2. Outline of Paper. Section 2 provides background material. In §3 we define and give basic results on the basic tool, the uplift regions. In Section 4 we normalize by using work of A. Schmidt so as to combine our earlier work on triples of elliptic elements with standard results on trees of simple closed geodesics on hyperbolic tori. We finish the proof in §5.

1.3. Notation. We use X + iY to denote points in \mathcal{H} . We call a geodesic of \mathcal{H} with its standard hyperbolic metric an *h*-line. To increase legibility, all h-lines mentioned are non-vertical (in the Euclidean sense) except as explicitly stated.

1.4. **Thanks.** We thank Y. Cheung for conversation related to this work. We also thank the referee for suggestions and references.

2. Background

2.1. Tori, Simple Closed Geodesics and Automorphisms. Each hyperbolic once-punctured torus has a Weierstrass involution. The quotient of this torus by its involution gives a singular punctured sphere. Indeed, there is a one-toone correspondence between the sets of uniformizing groups: Fuchsian groups of signature $(0; 2, 2, 2, \infty)$ and Fuchsian groups of signature $(1; \infty)$, see say [18]. Furthermore, the simple closed geodesics of each such pair are also in one-to-one correspondence, see say [21] (here one finds that a simple closed geodesic on a hyperbolic punctured sphere with three elliptic order two singularities actually bounces back and forth between two of the singularities); indeed, the lengths of corresponding simple closed geodesics are the same: in fact, there is a common element (primitive in each of the groups) whose axis projects to both geodesics.

We refer the reader to Section 2 of [10] for a particularly nice discussion of the structure of the graph of generators of a group of signature $(1; \infty)$, see the left side of Figure 1. Here, vertices represent group elements, up to inverses and conjugation. Edges connect pairs that generate the group (which is a free group on two elements); one calls either element in such a pair a generating element. Any generating element has axis projecting to a simple closed geodesic; its inverse and any conjugate elements give the same curve (up to orientation). Thus, our vertices can be seen as corresponding to the simple closed geodesics of the torus uniformized by the group. As Bowditch (see especially the discussion on p. 49 of [4]) pointed out, the dual graph is particularly helpful when discussing Fricke triples, see the right side of Figure 1. (Each node of the resulting tree corresponds

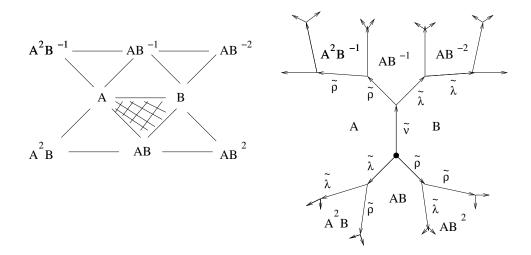


FIGURE 1. Equivalence classes of toral generators and dual graph of triples.

to a triple of simple closed geodesics such that a triple of corresponding open simple "cusped" geodesics is mutually disjoint.)

2.2. Fricke's Equation and Explicit Groups. We are interested in explicit lifts of simple closed geodesics. For this, we use a variation of A. Schmidt's application [18] of work of Fricke. Suppose that positive real a, b, c satisfy the Fricke equation

(1)
$$a^2 + b^2 + c^2 = abc$$

the elements

$$T_0 := \begin{pmatrix} 0 & -a/c \\ c/a & 0 \end{pmatrix}, \ T_1 := \begin{pmatrix} a/c & * \\ b/a & -a/c \end{pmatrix}, \ T_2 := \begin{pmatrix} a-b/c & * \\ 1 & -a+b/c \end{pmatrix}$$

(of determinant one) generate a group of signature $(0; 2, 2, 2; \infty)$. Note that

$$T_2 \cdot T_1 \cdot T_0 = S^a : z \mapsto z + a$$

is the fundamental translation of this group. A full set of orbit representatives under the action of the Teichmüller group is given when one takes $2 < a \leq b \leq c < ab/2$; this can be deduced from [18], see also [25]; we will always assume that our Fricke triples (a, b, c) satisfy this restriction. Note that the modular case of $\Gamma^3 \setminus \mathcal{H}$ corresponds to a = b = c = 3 and in this case T_j is the conjugate of T_0 by the translation $z \mapsto z + j$. (Note that our T_0 is not that of [18].)

2.3. Fixed Point Triples and Fundamental Domains. For ease of presentation, in [19] we restricted to the modular case. However, as we noted, our arguments extend to the full Teichmüller case. **Proposition 1.** [19] Let the signature $(0; 2, 2, 2; \infty)$ -orbifold $\mathcal{U} = \Gamma \setminus \mathcal{H}$ correspond to the Fricke triple (a, b, c). Each simple closed geodesic has a highest lift which is the axis of $S^a E$, where $E \in \Gamma$ is elliptic of order two. There is a factorization of $S^{a}E = GF$ as the product of elliptic elements such that a highest lifting segment of this simple geodesic joins the fixed point f of F to the fixed point q of G. Let e be the fixed point of E. A fundamental domain for Γ is given by the hexagon of vertices: ∞ , e, f, F(e), g, a + e. In particular, $\{E, F, G\}$ generates Γ .

Given a fixed Fricke triple (a, b, c), we have the corresponding *adjusted Fricke* equation

$$(2) x^2 + y^2 + z^2 = a xyz$$

Note that when a = 3 the adjusted Fricke equation is exactly Markoff's equation.

Recall that the imaginary part, Y, of a point X + iY is its *height*. The factorization SE = GF can be used to show that the hyperbolic GF, whose axis projects to a simple closed geodesic on the surface, has trace az, where 1/z is the height of the fixed point of E. One can show that there is such a factorization for every simple closed geodesic, and since the adjusted Fricke equation is satisfied by the traces of appropriate triples of simple hyperbolic elements, one finds the following result.

Corollary 1. [19] Let E, F, G be as above. Then the fixed points of E, F, G have respective heights 1/z, 1/y, 1/x, whose inverses give a triple satisfying the adjusted Fricke Equation, and with $z = \max\{x, y, z\}$. Furthermore, the simple closed geodesic that lifts to the axis of $S^a E$ has height $r_a(z) = \sqrt{a^2/4 - 1/z^2}$. The closed geodesic that lifts to the axis of $ES^a ES^{-a}$ has height $R_a(z) = \sqrt{a^2/4 + 1/z^2}$.

Proof. This follows from Theorems 2 and 3 (and their proofs) of [20], where we give detailed proofs in the modular case. The only aspect of the proof given there that does not hold in general is that by use of the map $w \mapsto -\bar{w}$, in the modular case one can further assume that $y \ge x$.

Corollary 2. [19] Let E, F, G be as above. Then there is a real translation conjugating the triple E, F, G to

$$E_0 = \begin{pmatrix} 0 & * \\ z & 0 \end{pmatrix}, E_1 = \begin{pmatrix} x/z & * \\ y & -x/z \end{pmatrix}, E_2 = \begin{pmatrix} ax - y/z & * \\ x & -ax + y/z \end{pmatrix}.$$

roof. This follows as in the proof of Theorem 2 of [20].

Proof. This follows as in the proof of Theorem 2 of [20].

3. BASIC TOOL: UPLIFT REGIONS

The following elementary result is key to our approach.

Lemma 1. If $A = \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha \end{pmatrix}$ is in $SL(2, \mathbb{R})$, then A increases the height of any h-line with apex (X, Y) satisfying $|(X - \alpha/\gamma)^2 - Y^2| < 1/\gamma^2$.

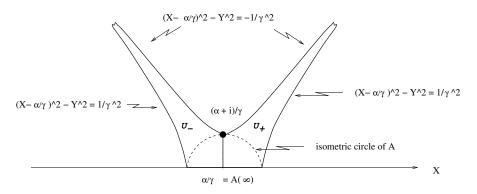


FIGURE 2. Order two A increases heights of h-lines with apex in uplift region $\mathcal{U}(A) = \mathcal{U}_{-}(A) \cup \mathcal{U}_{+}(A)$.

Proof. We first note that $T_0: z \mapsto -1/z$ takes C(c, r), the circle of real center c and radius r, to $C(-c/(c^2-r^2), r/|c^2-r^2|)$. Thus, this element increases radii whenever $|c^2 - r^2| < 1$. Now, an h-line of apex (X_0, Y_0) has center X_0 and radius Y_0 . Thus, the element T_0 increases heights for all h-lines of apex of coordinate (X, Y) with $|X^2 - Y^2| < 1$.

The fixed point of A is $w = (\alpha + i)/\gamma$. Since $w \mapsto \gamma w - \alpha$ on \mathcal{H} respects relative size, A increases radii for h-lines of apex w = X + iY with $|(X - \alpha/\gamma)^2 - Y^2| < i$ $1/\gamma^2$.

3.1. Uplift Regions Defined.

Definition 1. For $A = \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha \end{pmatrix}$ in SL(2, \mathbb{R}), the *uplift region* of A, $\mathcal{U}(A)$, is the subset of $(X,Y) \in \mathcal{H}$ such that $|(X - \alpha/\gamma)^2 - Y^2| < 1/\gamma^2$. We let $\mathcal{U}_{-}(A)$ denote the elements of the uplift region of A with $X < \alpha/\gamma$, and $\mathcal{U}_{+}(A)$ denote the remaining elements. Finally, we call

- $\{ (X,Y) | (X \alpha/\gamma)^2 Y^2 = -1/\gamma^2 \}$ the upper boundary of $\mathcal{U}(A)$, and $\{ (X,Y) | (X \alpha/\gamma)^2 Y^2 = 1/\gamma^2 \}$ the lower boundary of $\mathcal{U}(A)$.

See Figure 2.

Recall that the *isometric circle* of an element of $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R})$ with $\gamma \neq 0$ is the circle of center $-\delta/\gamma$ and radius $1/|\gamma|$. The isometric circle of an order two elliptic element A (as above) is inscribed in the uplift region, with points of intersection at the fixed point of A and at two (ideal) points on the real axis. The elliptic A acts so as to send its isometric circle to itself, by reflection through the vertical line passing through the fixed point.

Lemma 2. Suppose that $A = \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha \end{pmatrix}$ is in $SL(2, \mathbb{R})$ and ℓ is an h-line. Then A preserves the height of ℓ if and only if ℓ either passes through the fixed point of A, or else ℓ meets perpendicularly the isometric circle of A. In the first of these cases, the apex of ℓ lies on the upper boundary of U(A); in the second, this apex lies on the lower boundary.

Proof. From Lemma 1, the height is preserved exactly for ℓ of apex on the boundary of $\mathcal{U}(A)$. To identify the geometry associated to apexes on the components of this boundary, it again suffices to treat the special case of $A = T_0$. This is then a straightforward exercise, easily performed using at most elementary calculus. \Box

3.2. Uplift Regions and Translations. Our main application of uplifting is in the setting of triples of elliptic elements of order two whose product is a translation.

Proposition 2. Suppose that A, B and C are distinct elliptic elements of order two such that the product ABC is a translation. Then the axis of AB meets perpendicularly the isometric circle of C.

Proof. We first show that C fixes the height of the axis of AB. Suppose w lies on this axis, then BAw does as well. Now, $C(BAw) = (ABC)^{-1}w$ and thus we find that the image of the axis of AB under C is simply a translation of itself.

The axis of AB passes through the fixed point of each of A and of B. If it also passes through the fixed point of C, then each of A, B and C send this axis to itself. But, the translation ABC cannot send any h-line to itself. Therefore, in fact the axis of AB cannot pass through the fixed point of C.

Since C fixes the height of the axis of AB, but this axis does not pass through the fixed point of C, by Lemma 2 we conclude that the axis of AB meets perpendicularly the isometric circle of C.

Definition 2. If $A, B \in SL(2, \mathbb{R})$ are elliptic elements of order two, let ap(AB) denote the apex of the h-line passing through their fixed points. (Note that ap(AB) = ap(BA), an ambiguity that causes no harm in what follows.)

For ease of discussion, we will say that uplift regions of two order two elements *bifurcate* at a point p if the upper boundaries of these regions intersect at p. Thus, with A, B as above, their uplift regions bifurcate at ap(AB).

Corollary 3. Suppose that A, B and C are distinct elliptic elements of order two such that the product ABC is a translation. Then ap(BC) lies on the intersection of the lower boundaries of the uplift regions of A and of CBABC.

Proof. That ap(BC) lies on the intersection of the lower boundary of the uplift region of A follows by taking inverses and applying Proposition 2 and Lemma 2. To show that this apex lies on the lower boundary of the uplift region of CBABC,

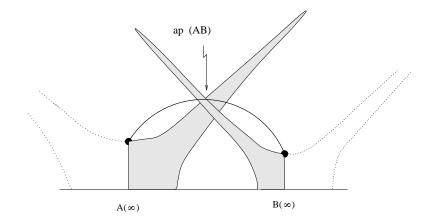


FIGURE 3. Two order two elements: h-line joining fixed points and bicorn region.

we can repeat the above, after replacing A by CBABC and BC by its inverse CB.

3.3. Translates of Uplift Regions.

Definition 3. Fix a real number a > 0. For z > 2/a, let $r_a(z) = \sqrt{a^2/4 - 1/z^2}$ and $R_a(z) = \sqrt{a^2/4 + 1/z^2}$.

The final statement of the following Lemma strengthens Corollary 3 in this setting.

Lemma 3. Suppose $A \in SL(2,\mathbb{R})$, with $A = \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha \end{pmatrix}$ and S^a is the translation by a > 0. Then the lower boundaries of $\mathcal{U}_+(A)$ and $\mathcal{U}_-(S^{\alpha}AS^{-\alpha})$ meet at $(X,Y) = (a/2 + \alpha/\gamma, r_a(\gamma))$, while their upper boundaries meet at $(X,Y) = (a/2 + \alpha/\gamma, R_a(\gamma))$. Furthermore, if $S^a A$ is hyperbolic, then the apex of its axis lies at the point of intersection of the lower boundaries of $\mathcal{U}_+(A)$ and $\mathcal{U}_-(S^aAS^{-a})$; similarly, $ap(S^aAS^{-a}A) = (a/2 + \alpha/\gamma, R_a(\gamma))$.

Proof. This is a trivial computation. (Note that the closures of these uplift regions also meet at two points of Y-coordinate $\sqrt{a^2/4 + 1/\gamma^4}$.) See Figure 4.

Definition 4. With notation as above, we call the intersection of Y = a/2 with the union of $\mathcal{U}_+(A)$ and $\mathcal{U}_-(S^aAS^{-a})$ the excision interval of A.

Lemma 4. With notation and hypotheses as above, the excision interval of A has width $w_a(A) = a - 2r_a(\gamma)$.

Proof. This is also a trivial computation. See Figure 4.

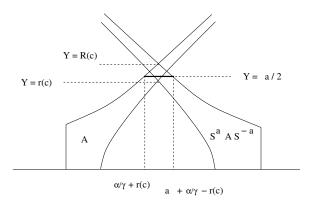


FIGURE 4. Bicorn region for A and $S^a A S^{-a}$ meets Y = a/2 in Excision Interval of A.

4. FRICKE-INDEXED FUNDAMENTAL DOMAINS

Convention For the remainder of the paper, unless otherwise stated, we fix a Fricke triple (a, b, c). Note that the fundamental translation length is thus a.

In [19] and [20], we showed that $\Gamma^3 \setminus \mathcal{H}$ admits particularly nice fundamental domains indexed by solutions to the original Markoff equation. Here we summarize this and its direct generalization to the general hyperbolic orbifold of signature $(0; 2, 2, 2, \infty)$.

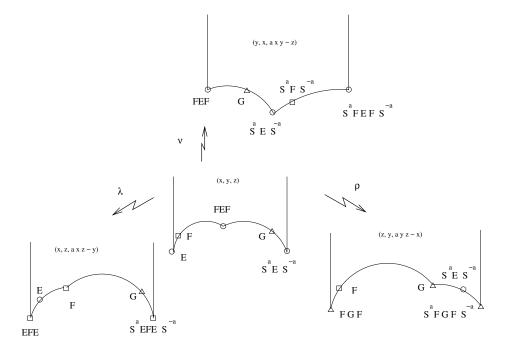


FIGURE 5. Moving through the adjusted Fricke tree.

4.1. Fundamental Domains, Relating Uplift Regions.

Definition 5. For any (E, F, G) as in Corollary 1, we define the following maps to triples of elliptic elements of order two.

$$\nu : (E, F, G) \mapsto (FEF, G, S^a F S^{-a})$$
$$\rho : (E, F, G) \mapsto (FGF, F, S^a E S^{-a})$$
$$\lambda : (E, F, G) \mapsto (EFE, E, G).$$

The following is a straightforward computation, compare with Figure 1.

Lemma 5. Fix some triple E, F, G as above; let A = EF and B = FEFG. For a homomorphism $\phi : \Gamma \to \Gamma$, let $\tilde{\phi}$ denote the induced homomorphism on the unique index two subgroup of Γ that is of signature $(1; \infty)$ applied to ordered triplets of elements of this subgroup. Then

$$\begin{split} \tilde{\nu}(A,B,AB) &= (B,B^{-1}A^{-1}B,A^{-1}B)\\ \tilde{\rho}(A,B,AB) &= (AB,B,AB^2)\\ \tilde{\lambda}(A,B,AB) &= (A,AB,A^2B) \,. \end{split}$$

With the above identifications, the triple of simple cusped geodesics paired to the triple (A, B, AB) (mentioned in the final sentence of subsection 2.1) as seen

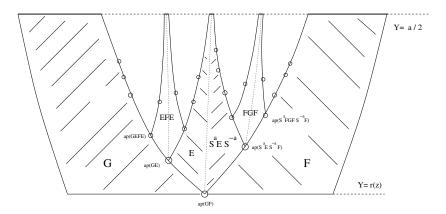


FIGURE 6. Uplift regions and apexes: Subtree generated by λ and ρ ; (cross-hatched) trident of Lemma 6; some excision intervals.

on $\Gamma \setminus \mathcal{H}$ is nothing other than the projection of the rays emanating vertically up from the fixed points of E, F and G.

Proposition 3. Let (E, F, G) be as above. Then each of $\nu(E, F, G)$, $\lambda(E, F, G)$, and $\rho(E, F, G)$ is a generating triple of Γ . For each of these triples, the corresponding triple of fixed points gives rise to a solution of the adjusted Fricke equation, by taking inverses of heights, as indicated in Figure 5. Furthermore, if $z \ge \max\{x, y\}$ then the analogous inequality holds upon applying either of λ or ρ .

Proof. That each of these triples also generate Γ is easily checked.

By use of the translated version of the matrices given in Lemma 2, one verifies that the triples of multiplicative inverses of the heights of the fixed points of each of the elements involved in $\nu(E, F, G)$, $\lambda(E, F, G)$, and $\rho(E, F, G)$ are as indicated in Figure 5. Our hypotheses on *a* imply that *FGF* has the lowest fixed point of the triple $\rho(E, F, G)$ and similarly for *EFE* and $\lambda(E, F, G)$.

Definition 6. For (E, F, G) as above, we call the union of $\mathcal{U}_{-}(G)$ and $\mathcal{U}_{+}(F)$ the associated uplift *bicorn region*. See Figure 3.

Figure 6 indicates regions discussed in the following two results. In particular, the cross-hatched region of the figure shows the once-punctured trident formed by the union of the bicorn regions of (E, F, G), $\rho(E, F, G)$ and $\lambda(E, F, G)$ for one triple (E, F, G). Each non-horizontal dotted curve indicates the splitting of a region into a union of some $\mathcal{U}_+(A)$ and $\mathcal{U}_-(S^aAS^{-a})$.

Lemma 6. Let (E, F, G) be as above, with $z \ge \max\{x, y\}$. The intersection of $r_a(z) \le Y < a/2$ with the union of the bicorn regions of (E, F, G), $\lambda(E, F, G)$ and $\rho(E, F, G)$ has the form of a once-punctured trident. The single puncture occurs at ap(GF); the bifurcations of the trident are at ap(GE) and $ap(FS^aES^{-a})$.

Proof. From the original triple we have that GF has axis projecting to a simple closed geodesic; applying ρ , the same is true for FS^aES^{-a} ; applying λ , also for GE. Lemma 2 shows that $\operatorname{ap}(GF)$ lies on the upper boundary of the uplift regions of G and F. Furthermore, since by construction, the h-line segment joining the fixed points of F and G is a highest lifting segment of the simple geodesic, and these respective fixed points satisfy $\Re(f) < \Re(g)$, we conclude that the union of $\mathcal{U}_{-}(G)$ and $\mathcal{U}_{+}(F)$ bifurcates at $\operatorname{ap}(GF)$. Similar roles are played by $\operatorname{ap}(GE)$ and $\operatorname{ap}(FS^aES^{-a})$.

By Lemma 3 we have that ap(GF) lies on the lower boundary of both $\mathcal{U}_+(E)$ and $\mathcal{U}_-(S^a E S^{-a})$; furthermore, these regions meet for Y between the height of ap(GF) and a value greater than a/2.

The result follows by now considering the union.

Definition 7. For (E, F, G) as above, let $\mathcal{T}_{\lambda,\rho}(E, F, G)$ denote the tree formed by applying to the triple all finite compositions (including the identity) of λ and ρ to (E, F, G), and let $\mathcal{U}_{\lambda,\rho}(E, F, G)$ denote the union of all of the corresponding bicorn regions.

Proposition 4. Let (E, F, G) be a triple as above, with $z \ge \max\{x, y\}$. Then $\mathcal{U}_{\lambda,\rho}(E, F, G)$ meets the strip $r_a(z) < Y < a/2$ in an infinitely punctured domain bounded by the lower boundary of $\mathcal{U}_{-}(G)$ and the lower boundary of $\mathcal{U}_{+}(F)$.

Proof. Since λ and ρ preserve the property $z \geq \max\{x, y\}$, we can repeatedly invoke the previous lemma. We thus need only show that successive "generations" of uplift triples overlap appropriately. But, as each of λ and ρ retains one of For G in its original position and promotes either E or $S^a E S^{-a}$ to the other, this also easily follows. \Box

4.2. Tree of Triples and Simple Closed Geodesics. For each Fricke equation, a unique minimum (with respect to the sum of the x, y and z) solution exists as [18] p. 352 deduces from his Theorem 3.1 (see also [3]); this thus also holds true for the adjusted Fricke equations and [18] implies that this minimum solution is given by the multiplicative inverses of the heights of the T_i . All solutions (up to cyclic ordering of x, y, z) are derived from this minimal solution by sequences of ν, λ and ρ . We form a tree completely analogous to that of [18], see his p. 351.

Definition 8. Let $\mathcal{T}_{\lambda,\rho}^{\nu}$ denote the tree formed by joining $\mathcal{T}_{\lambda,\rho}(T_0, T_1, T_2)$ to $\mathcal{T}_{\lambda,\rho}(\nu(T_0, T_1, T_2))$ with an edge (labeled by ν). The *uplift configuration* is the union of the bicorn regions of the nodes of this tree:

$$\mathcal{U}_{\lambda,\rho}^{\nu} := \mathcal{U}_{\lambda,\rho}(T_0, T_1, T_2) \cup \mathcal{U}_{\lambda,\rho}(\nu(T_0, T_1, T_2)).$$

The normalized uplift configuration is given by replacing $\mathcal{U}_+(T_2)$ in $\mathcal{U}_{\lambda,\rho}^{\nu}$ with its horizontal translation by -a.

Definition 9. Let C denote the set of all simple closed geodesics on $\Gamma \setminus \mathcal{H}$.

Figure 6 indicates some of the geometry of the following result.

Theorem 1. Fix an adjusted Fricke equation. The normalized uplift configuration meets the strip $r_a(1) \leq Y < a/2$ in an infinitely punctured (half-open) domain. Let \mathcal{P} be the set of these punctures. Then \mathcal{P} is in one-to-one correspondence with \mathcal{C} : each $p \in \mathcal{P}$ is the apex of a highest lift of some element of \mathcal{C} and each element of \mathcal{C} has a highest lift with apex in \mathcal{P} .

Proof. For ease of notation, let (x, y, z) denote the solution to the adjusted Fricke equation associated to (T_0, T_1, T_2) . Since this is a minimal solution, ν sends (T_0, T_1, T_2) to a triple whose lowest fixed point is given by its *E*-entry, $T_1T_0T_1$. Let w be the corresponding entry in the resulting solution to the adjusted Fricke equation. Now, Proposition 4 shows that $\mathcal{U}_{\lambda,\rho}(\nu(T_0, T_1, T_2))$ meets the horizontal open strip $r_a(w) < Y < a/2$ in an infinitely punctured domain whose right hand boundary is the right hand boundary of $\mathcal{U}_+(T_2)$ and whose left hand boundary is that of $\mathcal{U}_-(S^aT_1S^{-a})$. But, by Lemma 3, this left hand boundary is contained in $\mathcal{U}_+(T_1)$ for $r_a(y) < Y < a/2$. Thus, $\mathcal{U}_{\lambda,\rho}(\nu(T_0, T_1, T_2))$ and $\mathcal{U}_{\lambda,\rho}(T_0, T_1, T_2)$ have non-trivial intersection. The union, $\mathcal{U}^{\nu}_{\lambda,\rho}$ thus meets the strip $r_a(z) < Y < a/2$ in an infinitely punctured domain.

The normalization simply replaces $\mathcal{U}_+(T_2)$ by $\mathcal{U}_+(S^{-a}T_2S^a)$; due to Lemma 3, the intersection with the strip remains an infinitely punctured domain.

From Lemma 6, each of $\mathcal{U}_{\lambda,\rho}(T_0, T_1, T_2)$ and $\mathcal{U}_{\lambda,\rho}(\nu(T_0, T_1, T_2))$ contributes elements to \mathcal{P} that are appears. That the lifts in question are highest lifts of simple closed geodesics follows by observing the geometry of the fundamental domains (each of which has a single ideal vertex).

There are exactly two remaining elements of \mathcal{P} : one introduced by taking the union of $\mathcal{U}_{\lambda,\rho}(T_0, T_1, T_2)$ and $\mathcal{U}_{\lambda,\rho}(\nu(T_0, T_1, T_2))$; the second an artifact of our normalization. The first is the puncture lying on the intersection of the lower boundaries of $\mathcal{U}_+(T_1)$ and $\mathcal{U}_-(S^aT_1S^{-a})$. By Lemma 3 this is the apex of the axis of S^aT_1 . Similarly, our normalization introduces the puncture given by apex of the axis of T_2S^{-a} . Now, $\operatorname{ap}(S^aT_1)$ lies on $Y = r_a(y)$ which is lies above or is the line $Y = r_a(x)$, the horizontal line upon which lies $\operatorname{ap}(T_2S^{-a})$; this as x = 1is the minimum of the triple of inverse of heights of the fixed points of T_0, T_1 and T_2 . These two apexes lie on highest lifts of the simple closed geodesics (seen by using conjugation and taking inverses) that are the projections of the axes of $B = T_1T_0T_1T_2$ and $A = T_0T_1$, respectively.

Finally, by the discussion in the previous subsection, replacing each node (E, F, G)of $\mathcal{T}_{\lambda,\rho}^{\nu}$ by the triple (EF, FEFG, FG) gives the tree of all triples of associated simple closed geodesics on the canonical hyperbolic once punctured torus double (ramified) covering $\Gamma \setminus \mathcal{H}$. But, the bicorn region at each node of $\mathcal{T}_{\lambda,\rho}^{\nu}$ gives the element $\operatorname{ap}(FG) \in \mathcal{P}$. That is, associated to each node of this tree, is the apex of a highest lift of the simple closed geodesic whose face in the dual graph has (directed) edges labeled by $\tilde{\lambda}$ and $\tilde{\rho}$ emanating from the given node (EF, FEFG, FG). Since our initial node is (T_0, T_1, T_2) , we conclude that these elements of \mathcal{P} are the apexes of highest lifts for all elements of \mathcal{C} other than the projection of the axes of A and B (as defined above). But, we have already seen that the remaining points of \mathcal{P} account exactly for these two simple closed geodesics. We thus conclude that \mathcal{P} is exactly in one-to-one equivalence with \mathcal{C} , by associating apexes to projections of corresponding h-lines.

4.3. The Line Y = a/2.

Proposition 5. The normalized uplift configuration meets the line Y = a/2 in the union of disjoint intervals:

$$\bigsqcup_{T} \mathcal{U}_{+}(T) \cup \mathcal{U}_{-}(S^{a}TS^{-a}),$$

where the union is over all order two elements T appearing in the triple for any node of the tree $\mathcal{T}_{\lambda,\rho}^{\nu}$.

Proof. By Lemma 3, each $\mathcal{U}_+(E) \cup \mathcal{U}_-(S^a E S^{-a})$ meets Y = a/2 in an interval. Observing the action of ν , λ and ρ , one easily sees that the normalized uplift configuration meets the line Y = a/2 in the union of the intervals indexed by the various E.

It thus suffices to show that the various $\mathcal{U}_+(E) \cup \mathcal{U}_-(S^aTS^{-a})$ meet the line disjointly. But, we already know that \mathcal{P} lies below Y = a/2; by Lemma 4, \mathcal{P} contains the set of bifurcation points of the (normalized) uplift configuration. Disjointness follows.

Corollary 4. The sum of the $w_a(T) = a - \sqrt{a^2 - 4/z^2}$, indexed over the order elements T appearing in the triple for any node of the tree $\mathcal{T}^{\nu}_{\lambda,\rho}$, is at most a.

Proof. The uplift configuration fills in from $\mathcal{U}_{-}(T_2)$ to $\mathcal{U}_{+}(T_2)$; the closure of the normalized uplift region thus meets the line in a region contained in the interval from the left endpoint of the intersection with $\mathcal{U}_{+}(S^{-a}T_2S^a)$ to the left endpoint of the intersection with $\mathcal{U}_{+}(T_2)$. This ambient interval is of length a.

5. FINAL ARGUMENTS

5.1. Upper Bound: Lengths of Excision Intervals, Lengths of Geodesics. Recall that the length of a closed geodesic on a hyperbolic surface is $\ell(\gamma) = 2 \ln \epsilon_{\gamma}$, where ϵ_{γ} is the larger solution of $\epsilon_{\gamma} + 1/\epsilon_{\gamma} = t$ for t = |t(M)| the absolute value of the trace of a primitive element whose axis projects to γ . From Theorem 1, each simple closed geodesic γ is the projection of the h-line of apex some element of \mathcal{P} . In general, this gives γ as the projection of the axis of $S^{a}E$ with (E, F, G)a uniquely corresponding node of $\mathcal{T}^{\nu}_{\lambda,\rho}$; the corresponding triple (x, y, z) is such that t = az. (As in the proof of Theorem 1, the two simple closed geodesics distinguished as artifice of our indexing are the projections of the axes of S^aT_1 and S^aT_2 , of trace t = ab and t = ac, respectively.)

One easily calculates that

(3)
$$\frac{1}{1+e^{\ell(\gamma)}} = \frac{w_a(E)}{2a}$$

Corollary 4 thus implies that the sum of all $\frac{1}{1 + e^{\ell(\gamma)}}$ is at most 1/2.

5.2. Lower Bound: Hausdorff Dimension Zero. We must show that the upper bound given in Corollary 4 is also a lower bound. To do this, it suffices to show that the Cantor set formed by deleting the union of the excision intervals indexed by the nodes of $\mathcal{T}^{\nu}_{\lambda,\rho}$ has Lebesgue measure zero. In fact, an analysis not unlike that in [3] of limits along branches of our tree, reveals that much more is true. Using Proposition 6 below, we show the following.

Theorem 2. The complement to the union of the excision intervals is a set of zero Hausdorff dimension.

Proof. It suffices to show that s = 0 is an upper bound for this Hausdorff dimension.

Consider first a Cantor set constructed iteratively by removing a centered subinterval with fixed ratio of $k \in (0,1)$ from each interval remaining at the *n*-th step. If the original interval has finite length *L*, then at the *n*-th iteration there are 2^n intervals each of length $((1-k)/2)^n L$. An upper bound for the Hausdorff dimension is then obtained by finding the unique value of *s* such that $\lim_{n\to\infty} 2^n ((1-k)/2)^{ns} L^s$ is finite and non-zero — this is $s = \log 2/\log(2/(1-k))$.

In the case where a Cantor set is formed by removing possibly non-centered subintervals, but with a constant ratio of k, a naive upper bound for the lengths of intervals at the k-th iteration is simply $(1 - k)^n L$. The corresponding upper bound on the Hausdorff dimension is $s(k) := \log 2/\log(1/(1-k))$. Note that s tends to zero as k increases to 1.

We now turn to our Cantor set. By Proposition 6 (below), for any $\epsilon > 0$ there are at most finitely many nodes of $T^{\nu}_{\lambda,\rho}$ such that the ratio of the corresponding excision interval to its ambient interval is less than $1-\epsilon$. Excision of these intervals leads to finitely many subintervals; restricting the excision process to each gives a Cantor set, of Hausdorff dimension at most $s(1-\epsilon)$. The Hausdorff dimension of their union, our Cantor set, thus has this same upper bound. Letting ϵ tend to zero, we find that s = 0 is indeed an upper bound.

Proposition 6. Fix a directed branch beginning at (T_0, T_1, T_2) in $\mathcal{T}^{\nu}_{\lambda,\rho}$. The limit of the ratio of lengths of the excision interval of E to the ambient interval (on Y = a/2) bounded by $\mathcal{U}_{-}(G)$ and $\mathcal{U}_{+}(F)$ equals 1 unless the branch eventually

ends in an infinite sequence of exactly one of λ or ρ . In this purely periodic case, there is an x as above such that the limit is $\frac{2\sqrt{a^2 - 4/x^2}}{a + \sqrt{a^2 - 4/x^2}}$.

Proof. We delete the excision interval of E from the ambient interval lying between the excision interval of $S^{-a}GS^a$ and that of F. Thus, by Lemma 4, we excise an interval of length $a - 2r_a(z)$ from one of length $F(\infty) + r_a(y) - (G(\infty) - r_a(x))$. But, using the translated versions of our matrices, given in Corollary 2 on page 5, we find that this latter interval has length $x/yz + y/xz - a + r_a(x) + r_a(y)$. Now $x/yz + y/xz - a = -z^2/xyz$; solving Equation (2), the adjusted Fricke equation, for z, (with z sufficiently large) allows us to write $z/(xy) = a/2 + \sqrt{a^2/4 - 1/x^2 - 1/y^2}$. We are thus to find the limit of

(4)
$$\frac{a - 2r_a(z)}{r_a(x) + r_a(y) - (a/2 + \sqrt{a^2/4 - 1/x^2 - 1/y^2})}$$

Throughout our proof we use Taylor series approximation of $f(\delta) = \sqrt{s^2 - \delta}$ around $\delta = 0$:

$$\sqrt{s^2 - \delta} = s - \delta/(2s) - \sum_{j=2}^N c_j \frac{\delta^j}{s^{2j-1}} + O(\delta^{N+1}) ,$$

with $c_j = \frac{1 \cdot 3 \cdots (2j-3)}{2^j j!}$. In particular, the numerator of our ratio is

(5)
$$a - 2r_a(z) = \frac{2}{az^2} + O(z^{-4}).$$

Our denominator is symmetric in x and y; we can and do relabel each pair such that $x \leq y$ (we thus no longer demand that F fixes the point whose height is 1/y). We now treat three cases: our branch eventually ends in repeating exactly one of ρ or λ ; it has unbounded blocks of either ρ or λ ; and finally, it has bounded blocks of either.

Eventually repeating ρ or λ . We first treat the case of the branch eventually repeating in one of ρ or λ . Here, the smallest value of each triple, x (eventually) remains constant, whereas y and z both go to infinity. Since $z/xy = a/2 + \sqrt{(r_a(x))^2 - 1/y^2}$, two term approximation gives $a/2 + r_a(x) - \frac{1}{2r_a(x)y^2} + O(y^{-4})$. Using two term approximation on $r_a(y)$ as well, we find that the denominator is

$$-\frac{1}{ay^2} + \frac{1}{2r_a(x)y^2} + O(y^{-4}) = \frac{a - 2r_a(x)}{2ar_a(x)y^2} + O(y^{-4}).$$

Finally, $z/y = \frac{x}{2}(a + 2r_a(x)) + O(y^{-2})$, and $(a + 2r_a(x))(a - 2r_a(x)) = 4/x^2$. Hence we find

$$\frac{a - 2r_a(z)}{r_a(x) + r_a(y) - z/xy} = \frac{2/a + O(z^{-2})}{(z/y)^2 (\frac{a - 2r_a(x)}{2ar_a(x)} + O(y^{-2}))}$$
$$= \frac{2/a + O(z^{-2})}{\frac{a + 2r_a(x)}{2ar_a(x)} + O(y^{-2})}$$
$$= \frac{4r_a(x) + O(z^{-2})}{a + 2r_a(x) + O(y^{-2})}$$

Taking the limit with x fixed and y, z tending to infinity gives

$$\frac{4r_a(x)}{a+2r_a(x)} = \frac{2\sqrt{a^2-4/x^2}}{a+\sqrt{a^2-4/x^2}} \; ,$$

as claimed. Note that these values tend to 1 as x itself tends to infinity.

Unbounded blocks of ρ , λ . Consider any branch where the number of consecutive nodes of ρ or of λ is unbounded. For each positive integer N there is an infinite set of disjoint blocks of N consecutive applications of λ or ρ on the branch. But, as N increases, we thus find that the ratios of lengths of excised to ambient interval give ever better approximations to the limit ratios along (eventually) constant branches. Moreover, on our branch we must have that these corresponding values of x are also (eventually) increasing. Thus, the limit of ratios along this branch equals $\lim_{x\to\infty} \frac{2\sqrt{a^2-4/x^2}}{a+\sqrt{a^2-4/x^2}}$. But, this limit equals 1.

Bounded blocks of ρ , λ . On any branch not given by eventually repeating λ or ρ , each of x, y and z goes to infinity. We again first concentrate on the denominator, using Taylor series with our assumption that $x \leq y$,

$$\sqrt{a^2/4 - 1/x^2} + \sqrt{a^2/4 - 1/y^2} - (a/2 + \sqrt{a^2/4 - 1/x^2 - 1/y^2})$$
$$= \sum_{j=2}^{N} (\frac{2}{a})^{2j-1} c_j \left[(x^{-2} + y^{-2})^j - (x^{-2j} + y^{-2j}) \right] + O(x^{-2(N+1)})$$

$$= \frac{2}{a^3 x^2 y^2} \cdot \left(1 + \frac{3}{a^2} (x^{-2} + y^{-2}) + \dots + \frac{2^{N-2} c_N}{a^{2N-4}} \sum_{k=1}^{N-1} \binom{N}{k} x^{-2(N-k)+2} y^{-2k+2}\right) \\ + O(x^{-2(N+1)}) .$$

We thus find that our ratio is

$$\frac{1+O(z^{-2})}{\frac{z^2}{a^2x^2y^2}\cdot\left(1+\cdots+\frac{2^{N-2}}{N!}\frac{1\cdot3\cdots(2N-3)}{a^{2N-4}}\sum_{k=1}^{N-1}\binom{N}{k}x^{-2(N-k)+2}y^{-2k+2}\right)+z^2O(x^{-2(N+1)})}$$

Since $\frac{z^2}{x^2y^2} = a^2 + O(x^{-2} + y^{-2})$, we find that the limit equals 1 if there is some N such that z^2/x^{2N} has a finite limit (on our given branch). We apply the next Lemma (replacing N here by at worst 2N + 4, since we can assume that a < x).

For clarity's sake, we use (x, y, z) as in previous sections, and at each node let $l = \min(x, y)$.

Lemma 7. Fix a directed branch beginning at (T_0, T_1, T_2) in $\mathcal{T}^{\nu}_{\lambda,\rho}$. If the length of blocks of consecutive ρ or λ along the branch is bounded by N, then at each node (beyond the first change between λ and ρ), one has $z < (al)^{N+2}$.

Proof. Recall that $\rho : (x, y, z) \mapsto (z, y, ayz - x) =: (x', y', z')$ and $\lambda : (x, y, z) \mapsto (x, z, axz - y) =: (x'', y'', z'')$. Beginning at any node of corresponding triple (x, y, z), induction shows that y' and x'' are the minimum of their respective triples. (The main base case relies on the minimality of the solutions from (T_0, T_1, T_2) ; the secondary base case arising from $\nu(T_0, T_1, T_2)$ is easily verified.)

We always have z < axy. Thus, if l = x, then $z' < a^2 y'^3$ and if l = y then $z'' < a^2 x''^3$. Therefore, under these respective assumptions, we find $z < a^2 l^3$ holds for this new generation.

Now suppose that $z \leq a^k l^n$. Then with l = y we have $z' < a^{k+1}y^{n+1}$. With l = x, we have $z'' < a^{k+1}x^{n+1}$. That is, under these assumptions, we find $z \leq a^{k+1}l^{n+1}$ holds for this new generation.

In summary, after each change to either ρ to λ , we have $z < a^2 l^3$; this followed by n-1 more (consecutive) applications of the current ρ or λ then gives $z < a^{1+n}l^{2+n}$. The result follows.

References

- H. Akiyoshi, H. Miyachi, M. Sakuma, A refinement of McShane's identity for quasifuchsian punctured torus groups, in: In the tradition of Ahlfors and Bers, III, 21–40, Contemp. Math., 355, Amer. Math. Soc., Providence, RI, 2004.
- [2] A. F. Beardon, J. Lehner, M. Sheingorn, Closed geodesics on a Riemann surface with application to the Markov spectrum, Trans. Amer. Math. Soc. 295 (1986), no. 2, 635–647.
- B. Bowditch, A proof of McShane's identity via Markoff triples, Bull. London Math. Soc. 28 (1996), no. 1, 73–78.
- B. Bowditch, C. Maclachlan, A. Reid, Arithmetic hyperbolic surface bundles, Math. Ann. 302 (1995), 31–60.
- [5] F. Bonahon and X. Zhu, The metric space of geodesic laminations on a surface. II. Small surfaces in Proceedings of the Casson Fest, 509–547, Geom. Topol. Monogr., 7, Geom. Topol. Publ., Coventry, 2004.
- [6] H. Cohn, Approach to Markoff's minimal forms through modular functions, Ann. of Math.
 (2) 61 (1955), 1–12.
- [7] N. Do, P. Norbury, Weil-Petersson volumes and cone surfaces, preprint: Arxiv: math.GT/06033406, 2006
- [8] C. Goodman-Strauss, Y. Rieck, Simple geodesics on a punctured surface, Top. Appl. 154 (2007), no. 1, 155–165.
- [9] A. Haas, Diophantine approximation on hyperbolic Riemann surfaces, Acta Math. 156 (1986), 33–82
- [10] T. Jørgensen, A. Marden, Two doubly degenerate groups, Quart. J. Math. 30 (1979), 143– 156
- G. McShane, Simple geodesics and a series constant over Teichmuller space, Invent. Math. 132 (1998), no. 3, 607–632
- [12] _____, Weierstrass points and simple geodesics, Bull. London Math. Soc. 36 (2004), 181– 187
- [13] _____, Length series on Teichmüller space, preprint: Arxiv: math.GT/0403041, 2004
- [14] M. Mirzakhani, Growth of the number of simple closed geodesics on hyperbolic surfaces, Ann. Math. to appear.
- [15] <u>Weil-Petersson volumes and intersection theory on the moduli space of curves</u>, J. Amer. Math. Soc. 20 (2007), no. 1, 1–23
- [16] _____ Simple geodesics and Weil-Petersson volumes of the moduli spaces of bordered Riemann surfaces, Invent. Math. 167 (2007), no. 1, 179–222
- [17] I. Rivin, A simpler proof of Mirzakhani's simple curve asymptotics, Geom. Dedicata 114 (2005), 229–235
- [18] A. L. Schmidt, Minimum of quadratic forms with respect to Fuchsian groups. I, J. Reine Angew. Math. 286/287 (1976), 341–368
- [19] T. A. Schmidt, M. Sheingorn, Parametrizing simple closed geodesy on $\Gamma^3 \backslash \mathcal{H}$, J. Aust. Math. Soc. 74 (2003), no. 1, 43–60
- [20] _____, Low height geodesics on $\Gamma^3 \setminus \mathcal{H}$: height formulas and examples, Int. J. Number Theory 3 (2007), no. 3, 475–501.
- [21] M. Sheingorn, Characterization of simple closed geodesics on Fricke surfaces, Duke Math. J. 52 (1985), 535–545
- [22] S. P. Tan, Y. L. Wong, Y. Zhang, Generalizations of McShane's identity to hyperbolic cone-surfaces, J. Differential Geom. 72 (2006), no. 1, 73–112.
- [23] _____, Generalized Markoff maps and McShane's identity, Advances Math. 217 (2008), 761–813.

- [24] L. Yu. Vulkh, The Markov spectra for triangle groups, J. Number Theory 67 (1997), no. 1, 11–28.
- [25] S. Wolpert, On the Kähler form of the moduli space of once punctured tori, Comment. Math. Helv. 58 (1983), no. 2, 246–256

OREGON STATE UNIVERSITY, CORVALLIS, OR 97331 *E-mail address*: toms@math.orst.edu

CUNY - BARUCH COLLEGE, NEW YORK, NY 10010 E-mail address: marksh@alum.dartmouth.org