Pacific Journal of Mathematics

Volume 267 No. 2 February 2014

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

See inside back cover or msp.org/pjm for submission instructions.

The subscription price for 2014 is US $410/year for the electronic version, and $535/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.257

SUMS OF SQUARES IN ALGEBRAIC FUNCTION FIELDS OVER A COMPLETE DISCRETELY VALUED FIELD

KARIM JOHANNES BECHER, DAVID GRIMM AND JAN VAN GEEL

A recently found local-global principle for quadratic forms over function fields of curves over a complete discretely valued field is applied to the study of quadratic forms, sums of squares, and related field invariants.

1. Introduction

Let K be a field of characteristic different from 2 and F/K an algebraic function field (i.e., a finitely generated extension of transcendence degree one). The study of quadratic forms over F is generally difficult, even in such cases where the quadratic form theory over all finite extensions of K is well understood. It can be considered complete in the cases where K is algebraically closed, real closed, or finite, but it is wide open for example when K is a number field.

A breakthrough was obtained recently in the situation where the base field K is a nondyadic local field. Parimala and Suresh [2010] proved that in this case any quadratic form of dimension greater than eight over F is isotropic. Harbater, Hartmann, and Krashen [Harbater et al. 2009] obtained the same result as a consequence of a new local–global principle for isotropy of quadratic forms over F. The local conditions are in geometric terms, relative to an arithmetic model for F. A less geometric version of the local–global principle, in terms of the discrete rank one valuations of F, was obtained by Colliot-Thélène, Parimala, and Suresh [Colliot-Thélène et al. 2012]; see Theorem 6.1 below. Both versions of the local–global principle hold more generally when K is complete with respect to a nondyadic discrete valuation.

In this article we apply the local–global principle to study sums of squares in F and to obtain further results on quadratic forms over F. This is of particular interest in the case where K is the field of Laurent series $k((t))$ over a (formally) real field k.

We outline the structure of this article and the main results. Section 2 provides some necessary basic results on valuations. In Section 3 we discuss discrete

MSC2010: primary 11E04, 11E10, 11E25, 12J10, 14H05; secondary 12D15, 12F20.
Keywords: isotropy, local-global principle, real field, sums of squares, u-invariant, Pythagoras number, valuation, algebraic function fields.

valuations on an algebraic function field over a complete discretely valued field and characterize their residue fields. In Section 4 we move on to the study of sums of squares in fields and corresponding field invariants, in the context of valuations. In Section 5 we do an analogous discussion of the u-invariant in the context of valuations. According to [Elman and Lam 1973], the u-invariant of a field is the supremum on the dimension of anisotropic torsion forms over that field. In Section 6 we finally apply the local-global principle to obtain new results for algebraic function fields and in particular the rational function field. Let us describe some of these results.

In Theorem 6.4 we show that the upper bound on the dimension of anisotropic torsion forms over algebraic function fields over K is the double of the corresponding upper bound for algebraic function fields over k, the residue field of the discrete valuation on K. We thus obtain an upper bound on the u-invariant of an algebraic function field over K. We obtain in Theorem 6.6 a refinement for the case of the rational function field, saying that the u-invariant of $K(X)$ is equal to the supremum of the u-invariant of all $\ell(X)$ where ℓ/k is a finite algebraic extension. In Corollary 6.9 we prove that the Pythagoras number of the rational function field $K(X)$ is equal to the supremum of the Pythagoras numbers of $\ell(X)$ for all finite field extensions ℓ/k. We conjecture in Conjecture 4.16 that this is equal to the Pythagoras number of $k(X)$. This is motivated by the observation — made previously in [Scheiderer 2001] — that both Pythagoras numbers are bounded by the same power of two. In the case where k is real closed we show in Theorem 6.12 that any sum of squares in F can be expressed as a sum of three squares and further prove the finiteness of $\sum F^2/D_F(2)$, the quotient of the group of nonzero sums of squares modulo the subgroup of sums of two squares in F.

Our methods are based on valuation theory and quadratic form theory, for which [Engler and Prestel 2005] and [Lam 2005] are our standard references. We also use some algebraic geometry, namely desingularization of arithmetic surfaces and the properties of blowing-ups in regular points. Our reference on this topic is [Liu 2002].

This article grew out of results obtained in the Ph.D. thesis of D. Grimm under the supervision of K. J. Becher at Universität Konstanz.

2. Valuations

For a ring R we denote by $R^\times$ its group of invertible elements.

Let K be a field. Given a valuation on K, we denote by $\mathcal{O}_v$ the valuation ring of v, by $\mathfrak{m}_v$ its maximal ideal, by κ_v the residue field, by K^v the completion of K with respect to v, and we further call v *dyadic* if κ_v has characteristic 2, *nondyadic* otherwise. Given a local ring R contained in K, we say that a valuation v of K *dominates* R if $\mathfrak{m}_v \cap R$ is the maximal ideal of R. Given a field extension L/K, we say that a valuation v of L is *unramified over* K if $v(L^\times) = v(K^\times)$.

A valuation with value group $\mathbb{Z}$ is called a *$\mathbb{Z}$-valuation*. Any discrete valuation of rank one can be identified (via a unique isomorphism of the value groups) with a $\mathbb{Z}$-valuation. A commutative ring is the valuation ring of a $\mathbb{Z}$-valuation if and only if it is a regular local ring of dimension one (see [Matsumura 1986, Theorem 11.2]); such rings are called *discrete valuation rings*.

Lemma 2.1. *Let w_1 and w_2 be two valuations on K such that $\mathfrak{m}_{w_1} \subseteq \mathcal{O}_{w_2}$. Then $\mathcal{O}_{w_1} \subseteq \mathcal{O}_{w_2}$ or $\mathcal{O}_{w_2} \subseteq \mathcal{O}_{w_1}$.*

Proof. If $\mathfrak{m}_{w_1} \subseteq \mathfrak{m}_{w_2}$, then $\mathcal{O}_{w_1} \supseteq \mathcal{O}_{w_2}$, otherwise for any choice of $t \in \mathfrak{m}_{w_1} \setminus \mathfrak{m}_{w_2}$ we have $t^{-1} \in \mathcal{O}_{w_2}$ and $\mathcal{O}_{w_1} = t^{-1}(t\mathcal{O}_{w_1}) \subseteq t^{-1}\mathfrak{m}_{w_1} \subseteq \mathcal{O}_{w_2}$. □

The property for a valuation to be *henselian* is characterized by a list of equivalent conditions, including the statement of Hensel's Lemma, hence satisfied in particular by complete $\mathbb{Z}$-valuations; see [Engler and Prestel 2005, Section 4.1].

Proposition 2.2. *Let v be a henselian $\mathbb{Z}$-valuation on K. Then v is the unique $\mathbb{Z}$-valuation on K.*

Proof. By [Engler and Prestel 2005, Corollary 2.3.2] for distinct $\mathbb{Z}$-valuations w_1 and w_2 on K one has $\mathcal{O}_{w_1} \not\subseteq \mathcal{O}_{w_2}$ and $\mathcal{O}_{w_2} \not\subseteq \mathcal{O}_{w_1}$. Consider now a $\mathbb{Z}$-valuation w on K. Since v is henselian we have $1 + \mathfrak{m}_v \subseteq K^{\times n}$ for all $n \in \mathbb{N}$ prime to the characteristic of κ_v. As $w(K^\times) = \mathbb{Z}$, this implies that $1 + \mathfrak{m}_v \subseteq \mathcal{O}_w^\times$ and thus $\mathfrak{m}_v \subseteq \mathcal{O}_w$. Now Lemma 2.1 yields that $\mathcal{O}_w = \mathcal{O}_v$. □

Let X always denote a variable over a given ring or field.

Proposition 2.3. *Let R be a local domain with maximal ideal $\mathfrak{m}$ and residue field k. Let $p \in R[X]$ be monic and such that $\bar{p} \in k[X]$, the reduction of p modulo $\mathfrak{m}$, is irreducible. Then $R[X]/(p)$ is a local domain with maximal ideal $(\mathfrak{m}[X]+(p))/(p)$ and residue field $k[X]/(\bar{p})$. The ring $R[X]/(p)$ has the same dimension as R. Moreover, if R is regular, then $R[X]/(p)$ is regular.*

Proof. Note that $\mathfrak{m}[X] + (p)$ is a maximal ideal of $R[X]$. Consider a maximal ideal M of $R[X]$ containing p and set $\mathfrak{p} = M \cap R$. Since $R[X]/(p)$ is an integral extension of R, it follows using [Matsumura 1986, Theorems 9.3 and 9.4] that both rings have the same dimension. Moreover, the field $R[X]/M$ is an integral extension of $R/\mathfrak{p}$, whereby $R/\mathfrak{p}$ is a field. It follows that $\mathfrak{p} = \mathfrak{m}$ and thus $M = \mathfrak{m}[X] + (p)$. This shows that $\mathfrak{m}[X] + (p)$ is the unique maximal ideal of $R[X]$ containing p. Hence, $R[X]/(p)$ is a local domain with maximal ideal $(\mathfrak{m}[X] + (p))/(p)$ and residue field $k[X]/(\bar{p})$. Any set of generators of $\mathfrak{m}$ in R yields a set of generators of $(\mathfrak{m}[X] + (p))/(p)$ in $R[X]/(p)$. In particular, if R is regular, so is $R[X]/(p)$. □

Corollary 2.4. *Let T be a discrete valuation ring of K with residue field k. Let $p \in T[X]$ be monic with $\bar{p} \in k[X]$ irreducible. Then $T[X]/(p)$ is a discrete valuation ring with field of fractions $K[X]/(p)$ and residue field k-isomorphic to $k[X]/(\bar{p})$.*

Proof. Since a discrete valuation ring is the same as a regular local ring of dimension one, the statement follows from Proposition 2.3. □

We want to mention the following partial generalization of Corollary 2.4.

Proposition 2.5. *Let T be a valuation ring of K with residue field k and let ℓ/k be a finite field extension. There exists a finite field extension L/K with $[L:K]=[\ell:k]$ and a valuation v on L dominating T and unramified over K whose residue field is k-isomorphic to ℓ.*

Proof. It suffices to consider the case where $\ell = k[x]$ for some $x \in \ell$. Let $\mathfrak{m}$ denote the maximal ideal of T. Let $p \in T[X]$ be a monic polynomial whose residue $\bar{p}$ in $k[X]$ is the minimal polynomial of x over k. Then p is irreducible in $K[X]$, so $L = K[X]/(p)$ is a field. We obtain from Proposition 2.3 that $R = T[X]/(p)$ is a local domain with maximal ideal $M = (\mathfrak{m}[X] + (p))/(p)$ and residue field $k[X]/(\bar{p})$. Let v be a valuation on L dominating T. Then $T \subseteq R \subseteq \mathcal{O}_v$, and as M is generated by $\mathfrak{m}$, it follows that v dominates R. Hence, $k[X]/(\bar{p})$ embeds naturally into κ_v. In particular $[\kappa_v : k] \geq \deg(\bar{p}) = \deg(p) = [L : K]$. Using the Fundamental Inequality [Engler and Prestel 2005, Theorem 3.3.4] we conclude that v is unramified over K and $[\kappa_v : k] = \deg(\bar{p}) = [L : K]$, whereby κ_v is k-isomorphic to $k[X]/(\bar{p})$ and therefore to ℓ. □

3. Valuations on algebraic function fields

In this section we want to relate algebraic function fields over a valued field to algebraic function fields over the corresponding residue field. In particular we show in Proposition 3.4 that an algebraic function field over the residue field of a valuation on K can be realized as the residue field of an unramified extension to some algebraic function field over K, and we refine this statement in Theorem 3.5 for rational function fields.

In the sequel let T denote a valuation ring, K its field of fractions, and k the residue field of T. (That is, we have $T = \mathcal{O}_v$ for a valuation v on K and $k = \kappa_v$.) We consider the residue fields of valuations dominating T. (The reader may observe that we avoid to speak of extensions of valuations, as this can lead to confusion about the corresponding value groups.) For a field extension F/K and a valuation v on F dominating T, the field k is naturally embedded in the residue field κ_v. We often identify residue fields of valuations dominating T up to k-isomorphism, in order to simplify the language.

A finitely generated field extension F/K of transcendence degree one is called an *algebraic function field*. We say that F/K is *an algebrorational function field* if $F = L(x)$ for a finite extension L/K with $L \subseteq F$ and some element $x \in F$ that is transcendental over L; if this holds already with $L = K$, then F/K is called *a rational function field*.

Proposition 3.1. *Let F/K be an algebraic function field and v a valuation on F dominating T. The extension κ_v/k is either algebraic or an algebraic function field.*

Proof. This is a special case of the Dimension Inequality [Engler and Prestel 2005, Theorem 3.4.3]. □

The following is a refinement of Proposition 3.1 for rational function fields.

Theorem 3.2 (Ohm and Nagata). *Let F/K be a rational function field and v be a valuation on F dominating T. Then κ_v/k is either algebraic or algebrorational.*

Proof. This is shown in [Ohm 1983, Theorem], as a generalization of [Nagata 1967, Theorem 1]. □

We recall a construction to extend a valuation to a rational function field; in [Engler and Prestel 2005, Section 2.2] this is called the Gauss extension.

Proposition 3.3. *Let F/K be a rational function field. Let $x \in F$ be such that $F = K(x)$. Let T' be the localization of $T[x]$ with respect to the prime ideal $\mathfrak{m}[x]$ where $\mathfrak{m}$ is the maximal ideal of T. Then T' is a valuation ring with field of fractions F. The residue field of T' is $k(\bar{x})$ where $\bar{x}$ is the class of x modulo $\mathfrak{m}[x]$, which is transcendental over k. The corresponding valuation v on F with $\mathbb{O}_v = T'$, uniquely determined up to equivalence, is unramified over K.*

Proof. This follows from [Engler and Prestel 2005, Corollary 2.2.2]. □

Proposition 3.4. *Let E/k be an algebraic function field. There exists an algebraic function field F/K and a valuation v on F dominating T and unramified over K whose residue field is E.*

Proof. Let F'/K be a rational function field. Let $x \in F'$ be such that $F' = K(x)$ and let T' denote the valuation ring described in Proposition 3.3. We identify $\bar{x}$ with some element of E transcendental over k. Then $E/k(\bar{x})$ is a finite extension. By Proposition 2.5 there exists a finite field extension F/F' with $[F : F'] = [E : k(\bar{x})]$ and a valuation v on F dominating T' and unramified over F' with residue field E. Using Proposition 3.3 it follows that v is also unramified over K. □

Theorem 3.5. *Assume that $T \neq K$ and let F/K be a rational function field. Let ℓ/k be a finite separable field extension. There exists a valuation v on F dominating T and unramified over K for which κ_v/k is an algebrorational function field with field of constants ℓ.*

Proof. Let $y \in F$ and $\alpha \in \ell$ be such that $F = K(y)$ and $\ell = k(\alpha)$. Let $q \in T[Y]$ be monic and such that the residue $\bar{q}$ in $k[Y]$ is the minimal polynomial of α. Let $\mathfrak{m}$ be the maximal ideal of T. We choose $m \in \mathfrak{m} \setminus \{0\}$ and set $x = m^{-1}q(y) \in F$. Note that x is transcendental over K, and thus $F/K(x)$ is a finite extension. Let T' be the localization of $T[x]$ with respect to $\mathfrak{m}[x]$, the ideal consisting of the polynomials in x with coefficients in $\mathfrak{m}$. Let $\mathfrak{m}'$ be the maximal ideal of T'. By Proposition 3.3 T' is a valuation ring with field of fractions $K(x)$ and residue field $k(\bar{x})$, and $\bar{x}$ is transcendental over k. Note that $\bar{q}$ remains irreducible in $k(\bar{x})[Y]$.

Consider $p = q - q(y) \in T'[Y]$. As $q(y) = mx$, taking residues modulo $\mathfrak{m}'[Y]$ we have $\bar{p} = \bar{q}$ in $k(\bar{x})[Y]$. It follows by Proposition 2.3 that $R = T'[Y]/(p)$ is a local ring with maximal ideal lying over $\mathfrak{m}'$, with field of fractions $K(x)[Y]/(p)$, and residue field $k(\bar{x})[Y]/(\bar{p})$. Note that $K(x)[Y]/(p)$ is $K(x)$-isomorphic to F. Using Chevalley's theorem [Engler and Prestel 2005, Theorem 3.1.1], we obtain a valuation v on F that dominates T'. Then v also dominates T. As $p(y) = 0$, we have that y is integral over T'. Since $\bar{p} = \bar{q}$ is irreducible in $k(\bar{x})[Y]$, we have that $\bar{p}(0) \neq 0$, whereby $p(0) \in T'^{\times}$. As v dominates T' and $p(y) = 0$, we obtain that $v(y) = 0$. Hence, $\bar{x}, \bar{y} \in \kappa_v$, and $\bar{y}$ is algebraic over k, because $\bar{q}(\bar{y}) = \bar{p}(\bar{y}) = 0$. As $\bar{q}$ is irreducible in $k(\bar{x})[Y]$ we obtain that

$$[\kappa_v : k(\bar{x})] \geq [k(\bar{x})[\bar{y}] : k(\bar{x})] = \deg(\bar{p}) = \deg(p) = [F : K(x)].$$

By the Fundamental Inequality [Engler and Prestel 2005, Theorem 3.3.4], it follows that v is unramified over $K(x)$ and $\kappa_v = k(\bar{x})[\bar{y}] = k[\bar{y}](\bar{x})$. Using Proposition 3.3 we obtain that v is unramified over K. Since $q(\bar{y}) = 0 = \bar{q}(\alpha)$ and since we consider residue fields up to k-isomorphism, we can identify $\ell = k[\alpha]$ with $k[\bar{y}]$. □

Remark 3.6. In Theorem 3.5, the hypothesis on the finite extension ℓ/k to be separable is not necessary. Given a finite extension ℓ/k we can obtain a regular model (see below for the definition) for F/T whose special fiber contains a component isomorphic to $\mathbb{P}^1_\ell$ in the following way: We choose $\alpha \in \ell$ and $\ell' = k(\alpha)$. Blowing up $\mathbb{P}^1_T$ in a point on the special fiber $\mathbb{P}^1_k$ with residue field ℓ', we obtain a new regular model whose special fiber has a component given by the exceptional fiber of this blowing-up and thus isomorphic to $\mathbb{P}^1_{\ell'}$. Iterating this process we eventually obtain a regular model for F/T whose special fiber has a component isomorphic to $\mathbb{P}^1_\ell$, and its generic point corresponds to a $\mathbb{Z}$-valuation whose residue field is a rational function field over ℓ.

Assume that the valuation ring T is discrete and consider an algebraic function field F/K. By a *regular model for F/T* we mean a 2-dimensional integral regular projective flat T-scheme $\mathscr{X}$ whose function field is K-isomorphic to F. Given a regular model $\mathscr{X}$ for F/K we denote by $\mathscr{X}_k$ its special fiber; by [Liu 2002, Chapter 8, Lemma 3.3] $\mathscr{X}_k$ is a curve.

Given an integral scheme $\mathscr{X}$, a point $P \in \mathscr{X}$, and a valuation v on the function field of $\mathscr{X}$, we say that v is *centered at P* if v dominates $\mathcal{O}_{\mathscr{X},P}$, the local ring at P.

Proposition 3.7. *Assume that T is a discrete valuation ring. Let F/K be an algebraic function field. Let $\mathscr{X}$ be a regular model for F/T. Let v be a $\mathbb{Z}$-valuation on F dominating T. Then v is centered at a point P of $\mathscr{X}$ lying in $\mathscr{X}_k$. Moreover, if the extension κ_v/k is neither algebraic nor algebrorational, then $\mathbb{O}_v = \mathbb{O}_{\mathscr{X},P}$ where P is the generic point of an irreducible component of $\mathscr{X}_k$.*

Proof. By [Liu 2002, Chapter 8, Definition 3.17] v is centered at a point P of the special fiber $\mathscr{X}_k$. Since $\mathscr{X}_k$ is a curve, P is either a closed point or the generic point of an irreducible component $\mathscr{X}_k$. In either case $\mathbb{O}_{\mathscr{X},P}$ is a regular local ring.

If P is a closed point of $\mathscr{X}_k$, then by [Abhyankar 1956, Proposition 3] the extension κ_v/k is either algebraic or algebrorational. Assume now that P is a generic point of $\mathscr{X}_k$. Then P has codimension one in $\mathscr{X}$, so $\mathbb{O}_{\mathscr{X},P}$ is a regular local ring of dimension one and thus a discrete valuation ring. As $\mathbb{O}_{\mathscr{X},P}$ is dominated by $\mathbb{O}_v$ and both are discrete valuation rings with the same field of fractions, it follows by [Engler and Prestel 2005, Corollary 2.3.2] that $\mathbb{O}_v = \mathbb{O}_{\mathscr{X},P}$. □

Proposition 3.8. *Assume that T is a complete discrete valuation ring. Let F/K be an algebraic function field. Then there exists a regular model for F/T.*

Proof. There exists a regular projective curve C over K whose function field is K-isomorphic to F. If the curve C is smooth, then by [Liu 2002, Chapter 10, Proposition 1.8)] there exists a regular model for F/T. Note that this applies in particular when $\mathrm{char}(K) = 0$. Without assuming that C is smooth, we can follow the first steps in the proof of the proposition cited to obtain a 2-dimensional projective T-scheme $\mathscr{X}$ with function field F. Since the structure morphism $\mathscr{X} \to \mathrm{Spec}(T)$ is surjective, it is flat (see [Liu 2002, Chapter 8, Definition 3.1]). Then T is an excellent ring (see [ibid., Corollary 2.40]), and $\mathscr{X}$, being locally of finite type over T, is excellent (see [ibid., Theorem 2.39]).

Let $\mathscr{X}' \to \mathscr{X}$ be the normalization of $\mathscr{X}$. Since $\mathscr{X}$ is excellent and projective over T, the normalization $\mathscr{X}' \to \mathscr{X}$ is a finite projective birational morphism (see [ibid., Theorem 8.2.39 and Lemma 3.47]). The singular locus of $\mathscr{X}'$ is closed in $\mathscr{X}'$ (see [ibid., Corollary 2.38]). We consider the blowing-up $\mathscr{X}'' \to \mathscr{X}'$ along the singular locus of $\mathscr{X}'$; this is a birational projective morphism (see [ibid., Propositions 1.12 and 1.22]).

We may alternate normalization and blowing-up until we reach a scheme that is regular. At each step we obtain a flat projective 2-dimensional T-scheme whose function field is F. By Lipman's desingularization theorem (see [ibid., Theorem 3.44]), after finitely many steps we come to a situation where the T-scheme is regular. □

Corollary 3.9. *Assume that T is a complete discrete valuation ring. Let F/K be an algebraic function field. Then there exist only finitely many $\mathbb{Z}$-valuations v on F dominating T for which the extension κ_v/k is neither algebraic nor algebrorational.*

Proof. By Proposition 3.8 there exists a regular model for F/T. The statement follows by applying Proposition 3.7 to any such model. □

The result Corollary 3.9 can be extended to the situation where T is an arbitrary discrete valuation ring. Moreover, one may ask to characterize the $\mathbb{Z}$-valuations on an algebraic function field that dominate a given discrete valuation ring of the base field and for which the residue field extension is neither algebraic nor algebrorational. We intend to develop these topics in a forthcoming article.

4. Sums of squares and valuations

From now on let K be a field of characteristic different from 2. We denote by $\sum K^2$ the subgroup of nonzero sums of squares in K and, for $n \in \mathbb{N}$, by $D_K(n)$ the set of nonzero elements that can be written as sums of n squares in K. One calls

$$s(K) = \inf\{n \in \mathbb{N} \mid -1 \in D_K(n)\} \in \mathbb{N} \cup \{\infty\}$$

the *level of* K. Recall that K is *real* if $s(K) = \infty$ and *nonreal* otherwise, and in the latter case $s(K)$ is a power of two (see [Lam 2005, Chapter XI, Section 2]).

The *Pythagoras number of* K is defined as

$$p(K) = \inf\left\{n \in \mathbb{N} \mid D_K(n) = \sum K^2\right\} \in \mathbb{N} \cup \{\infty\}.$$

We further define

$$p'(K) = \begin{cases} p(K) & \text{if } K \text{ is real,} \\ s(K)+1 & \text{if } K \text{ is nonreal.} \end{cases}$$

This field invariant has no independent interest, but it allows us to avoid case distinctions in statements about valuations and Pythagoras numbers, by formulating them for $p'(K)$ rather than for $p(K)$. As for nonreal field K we always have $s(K) \le p(K) \le s(K)+1 = p'(K)$. Hence, $p'(K)$ is always equal to $p(K)$ or to $p(K)+1$.

We now consider valuations in the context of sums of squares. We say that a valuation v on K is *real* or *nonreal*, respectively, if the residue field κ_v has the corresponding property.

Lemma 4.1. *Let v be a valuation on K and $n \in \mathbb{N}$. Then $s(\kappa_v) \ge n$ if and only if $v(a_1^2 + \cdots + a_n^2) = 2\min\{v(a_1), \dots, v(a_n)\}$ holds for all $a_1, \dots, a_n \in K$.*

Proof. Both conditions are easily seen to be equivalent to having that any sum of n squares of elements in $\mathcal{O}_v^\times$ lies in $\mathcal{O}_v^\times$. □

Let $\Omega(K)$ denote the set of nondyadic $\mathbb{Z}$-valuations on K.

Proposition 4.2. *Let $v \in \Omega(K)$. If v is real, then $v\left(\sum K^2\right) = 2\mathbb{Z}$. If v is nonreal, then for $s = s(\kappa_v)$ we have $v(D_K(s)) = 2\mathbb{Z}$ and $v(D_K(s+1)) = v\left(\sum K^2\right) = \mathbb{Z}$.*

Proof. If v is real, then it follows from Lemma 4.1 that $v\big(\sum K^2\big) = 2\mathbb{Z}$. Assume now that v is nonreal and let $s = s(\kappa_v)$. Then it follows from Lemma 4.1 that $v(D_K(s)) = 2\mathbb{Z}$ and that there exist $x_0, \dots, x_s \in K$ such that $v(x_0^2 + \cdots + x_s^2) \neq 2\min\{v(x_0), \dots, v(x_s)\}$. Dividing by one of the elements $x_0, \dots, x_s$ with minimal value, we can assume that $\min\{v(x_0), \dots, v(x_s)\} = 0$. Hence $v(x_0^2 + \cdots + x_s^2) \geq 1$. If $v(x_0^2 + \cdots + x_s^2) > 1$, then we choose $t \in K$ with $v(t) = 1$, and we conclude that $v((x_0 + t)^2 + x_1^2 \cdots + x_s^2) = 1$, as v is nondyadic. We may therefore assume that $v(x_0^2 + \cdots + x_s^2) = 1$. Since $K^{\times 2} \cdot D_K(s+1) = D_K(s+1)$, we conclude that $v(D_K(s+1)) = \mathbb{Z}$, and thus in particular that $v\big(\sum K^2\big) = \mathbb{Z}$. □

Proposition 4.3. *Let $v \in \Omega(K)$. Then $p'(K) \geq p(K) \geq p'(\kappa_v)$. Moreover, if v is henselian, then $p'(K) = p(K) = p'(\kappa_v)$.*

Proof. Note that $p(K) \geq p(\kappa_v)$. If v is real, then κ_v and K are real, and we obtain that $p'(K) = p(K) \geq p(\kappa_v) = p'(\kappa_v)$. If v is nonreal, then for $s = s(\kappa_v)$ we conclude that $D_K(s) \subsetneq D_K(s+1)$ by Proposition 4.2, and therefore $p'(K) \geq p(K) \geq s+1 = p'(\kappa_v)$.

Assume finally that v is henselian. Then $s(K) = s(\kappa_v)$, and further $p(K) = p(\kappa_v)$ in case v is real. This yields that $p'(K) = p'(\kappa_v)$. □

Recall that the completion of K with respect to a valuation v is denoted by K^v.

Corollary 4.4. *For $v \in \Omega(K)$ we have $p(K) \geq p(K^v) = p'(\kappa_v)$.*

Proof. Since v extends to a $\mathbb{Z}$-valuation on K^v with the same residue field κ_v, we obtain using both statements in Proposition 4.3 that $p(K) \geq p'(\kappa_v) = p(K^v)$. □

Corollary 4.5. *We have $p'(K(t)) = p(K(t)) \geq p'(K(\!(t)\!)) = p(K(\!(t)\!)) = p'(K)$.*

Proof. We have $p(K(t)) \geq p(K(\!(t)\!))$ by Corollary 4.4 and $p'(K(\!(t)\!)) = p(K(\!(t)\!)) = p'(K)$ by Proposition 4.3. If K is real, then $K(t)$ is real, thus $p'(K(t)) = p(K(t))$ by the definition. If K is nonreal, then $p(K(t)) = s(K) + 1 = s(K(t)) + 1 = p'(K(t))$. □

Corollary 4.6. *Let F/K be an algebrorational function field. Then $p'(F) = p(F)$.*

Proof. Replacing K by its relative algebraic closure in F, we have $F = K(t)$ for some $t \in F$ transcendental over K. We conclude using Corollary 4.5. □

This does not generalize to arbitrary algebraic function fields:

Example 4.7. Consider the function field F of the curve $Y^2 = -(X^2+1)(X^2+1+t)$ over $\mathbb{R}(\!(t)\!)$. By [Becher and Van Geel 2009, Example 5.13] we have $p(F) = s(F) = 2$, and therefore $p'(F) = 3 > p(F)$. In particular $-1 \notin F^{\times 2}$ whereas -1 is a square in F^v for any $v \in \Omega(F)$ by Corollary 4.4.

We apply Proposition 4.3 to give a short argument for a well-known fact:

Corollary 4.8. *Assume that K is a finitely generated nonalgebraic extension of a subfield. Then $p(K) \geq 2$.*

Proof. It follows from the hypotheses that there exists $v \in \Omega(K)$ such that κ_v is nonreal. From Proposition 4.3 we obtain that $p(K) \geq p'(\kappa_v) = s(\kappa_v) + 1 \geq 2$. □

Remark 4.9. If $K = k(t)$ for a subfield k and $t \in K$ transcendental over k, then $1 + t^2 \notin K^{\times 2}$ and thus $p(K) \geq 2$. An alternative proof of Corollary 4.8 is therefore obtained by reduction to the case of a rational function field via the Diller–Dress Theorem [Lam 2005, Chapter VIII, Theorem 5.7], which says that if $p(K) \geq 2$ then $p(L) \geq 2$ for every finite field extension L/K.

For $S \subseteq \Omega(K)$ we define a homomorphism

$$\Phi_S : K^\times \to \mathbb{Z}^S, \quad x \mapsto (v(x))_{v \in S}.$$

If $S \subseteq \Omega(K)$ is a finite subset, then it follows from the Approximation Theorem (see [Engler and Prestel 2005, Theorem 2.4.1] or [Liu 2002, Chapter 9, Lemma 1.9]) that Φ_S is surjective.

The following statement extends Proposition 4.2 from a single $\mathbb{Z}$-valuation to finitely many $\mathbb{Z}$-valuations on K.

Proposition 4.10. *Let S be a finite subset of $\Omega(K)$ and $n \in \mathbb{N}$. Then*

$$\Phi_S(D_K(n)) = \{(e_v)_{v \in S} \in \mathbb{Z}^S \mid e_v \in 2\mathbb{Z} \textit{ for } v \in S \textit{ with } s(\kappa_v) \geq n\}.$$

Proof. For $v \in \Omega(K)$ with $s(\kappa_v) \geq n$ we have $v(D_K(n)) \subseteq 2\mathbb{Z}$ by Lemma 4.1. This shows that

$$\Phi_S(D_K(n)) \subseteq \{(e_v)_{v \in S} \in \mathbb{Z}^S \mid e_v \in 2\mathbb{Z} \text{ for } v \in S \text{ with } s(\kappa_v) \geq n\}.$$

It remains to show the other inclusion. Consider a tuple $(e_v)_{v \in S} \in \mathbb{Z}^S$ such that $e_v \in 2\mathbb{Z}$ for all $v \in S$ with $s(\kappa_v) \geq n$. The aim is to find an element $x \in D_K(n)$ with $\Phi_S(x) = (e_v)_{v \in S}$. We explain how to obtain such an element, using the Approximation Theorem (see above) several times.

For $v \in S$ with $e_v \notin 2\mathbb{Z}$, as $s(\kappa_v) < n$ we may choose $x_{v,2}, \dots, x_{v,n} \in \mathbb{O}_v$ such that $v(1 + x_{v,2}^2 + \cdots + x_{v,n}^2) > 0$. For $v \in S$ with $e_v \in 2\mathbb{Z}$ we set $x_{v,2} = \cdots = x_{v,n} = 0$. For $i = 2, \dots, n$ we choose $x_i \in K^\times$ such that $v(x_i - x_{v,i}) > 0$ for all $v \in S$. We set $y = x_2^2 + \cdots + x_n^2$. For $v \in S$ we have $v(1 + y) = 0$ if $e_v \in 2\mathbb{Z}$ and $v(1 + y) > 0$ otherwise. We choose $t \in K^\times$ such that, for all $v \in S$, we have $v(t) = 1$ if $v(1 + y) > 1$, and $v(t) > 1$ otherwise. Note that $(1 + t)^2 + y \in D_K(n)$. For any $v \in S$ the value $v((1 + t)^2 + y)$ is either 0 or 1 and such that $v((1 + t)^2 + y) \equiv e_v \mod 2\mathbb{Z}$. Now choose $z \in K^\times$ such that $2v(z) = e_v - v((1 + t)^2 + y)$ for all $v \in S$ and set $x = z^2((1 + t)^2 + y)$. Then $x \in D_K(n)$ and $\Phi_S(x) = (v(x))_{v \in S} = (e_v)_{v \in S}$. □

Corollary 4.11. *Let $n \in \mathbb{N}$ and S a finite subset of $\Omega(K)$ such that $s(\kappa_v) = 2^n$ for all $v \in S$. Then Φ_S induces a surjective homomorphism*

$$D_K(2^{n+1})/D_K(2^n) \to (\mathbb{Z}/2\mathbb{Z})^S.$$

In particular, $|D_K(2^{n+1})/D_K(2^n)| \geq 2^{|S|}$.

Proof. By the hypotheses on S and by Proposition 4.10, we have $\Phi_S(D_K(2^{n+1})) = \mathbb{Z}^S$ and $\Phi_S(D_K(2^n)) = (2\mathbb{Z})^S$. From this the statement follows. □

Theorem 4.12. *Let K be a real field. For $n \in \mathbb{N}$ the following are equivalent*:

(i) $p(K(X)) \leq 2^n$.

(ii) $p(L) < 2^n$ *for all finite real extensions L/K.*

(iii) $s(L) \leq 2^{n-1}$ *for all finite nonreal extensions L/K.*

(iv) $p'(L) < 2^n$ *for all finite extensions L/K with $-1 \notin L^{\times 2}$.*

Proof. See [Becher and Van Geel 2009, Theorem 3.3] for the equivalence of (i)–(iii); the equivalence of these conditions with (iv) is obvious. □

Corollary 4.13. *Let $n \in \mathbb{N}$ be such that $p(K(X)) \leq 2^n$. Then $p(L(X)) \leq 2^n$ for any finite field extension L/K.*

Proof. If K is nonreal, then $p(L(X)) = s(L)+1 \leq s(K)+1 = p(K(X)) \leq 2^n$. If K is real and L is nonreal, then $s(L) \leq 2^{n-1}$ by Theorem 4.12 and thus $p(L(X)) \leq 2^n$. If L is real, then since any finite real extension of L is a finite real extension of K, the equivalence of (i) and (ii) in Theorem 4.12 allows us to conclude that $p(L(X)) \leq 2^n$. □

Theorem 4.14. *Let K be endowed with a $\mathbb{Z}$-valuation with residue field k. Then $p(K(X)) \geq p(k(X))$. Moreover, if the valuation is henselian and $n \in \mathbb{N}$ is such that $p(k(X)) \leq 2^n$, then $p(K(X)) \leq 2^n$.*

Proof. Using Proposition 3.3 the given $\mathbb{Z}$-valuation on K extends to a $\mathbb{Z}$-valuation on $K(X)$ with residue field $k(X)$. Hence, $p(K(X)) \geq p'(k(X)) = p(k(X))$ by Proposition 4.3 and Corollary 4.5.

Assume now that the $\mathbb{Z}$-valuation on K is henselian. If K is nonreal, then $p(K(X)) = s(K) + 1 = s(k) + 1 = p(k(X))$. Assume that K is real. Then k and $k(X)$ are real. Let $n \in \mathbb{N}$ be such that $p(k(X)) \leq 2^n$. By Theorem 4.12, to prove that $p(K(X)) \leq 2^n$ it suffices to show that $p'(L) < 2^n$ for all finite extensions L/K with $-1 \notin L^{\times 2}$. Consider such an extension L/K. Then L is endowed with a henselian $\mathbb{Z}$-valuation whose residue field ℓ is a finite extension of k. Then $-1 \notin \ell^{\times 2}$ and thus $p'(L) = p'(\ell) < 2^n$ by Proposition 4.3 and Theorem 4.12. □

The last two statements motivate us to formulate the following two conjectures.

Conjecture 4.15. *For any finite field extension L/K, one has $p(L(X)) \leq p(K(X))$.*

Conjecture 4.16. *If K is complete with respect to a nondyadic $\mathbb{Z}$-valuation with residue field k, then $p(K(X)) = p(k(X))$.*

Note that both conjectures hold trivially if K is a nonreal field. In the case where K is real, Conjecture 4.16 was raised originally by C. Scheiderer [2001, Remark 5.18.2] as a question. We shall prove in Corollary 6.9 that the two conjectures are equivalent.

5. The u-invariant for algebraic function fields

We refer to [Lam 2005] for basic facts and terminology from the theory of quadratic forms over fields of characteristic different from two. The u-invariant of K was defined in [Elman and Lam 1973] as

$$u(K) = \sup\{\dim\varphi \mid \varphi \text{ anisotropic torsion form over } K\} \in \mathbb{N}\cup\{\infty\},$$

where a *torsion form* is a regular quadratic form that corresponds to a torsion element in the Witt ring.

Proposition 5.1. *Let $v \in \Omega(K)$. Let ψ be a torsion form over κ_v. There exist $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{O}_v^\times$, and $t \in K^\times$ with $v(t) = 1$ such that $\langle 1, -t\rangle \otimes \langle a_1, \dots, a_n\rangle$ is a torsion form over K and such that ψ is Witt equivalent to $\langle \bar{a}_1, \dots, \bar{a}_n\rangle$.*

Proof. Assume first that v is nonreal. Then by Proposition 4.2 there exists $t \in \sum K^2$ with $v(t) = 1$. For $n = \dim\psi$ and $a_1, \dots, a_n \in \mathbb{O}_v^\times$ such that ψ is isometric to $\langle \bar{a}_1, \dots, \bar{a}_n\rangle$, we obtain that $\langle 1, -t\rangle \otimes \langle a_1, \dots, a_n\rangle$ is a torsion form over K.

Assume now that v is real. Then ψ is Witt equivalent to a sum of binary torsion forms over κ_v (see [Pfister 1966, Satz 22]). Every binary torsion form over κ_v is of the shape $\langle \bar{a}_1, \bar{a}_2\rangle$ with $a_1, a_2 \in \mathbb{O}_v^\times$ such that $-a_1a_2 \in \sum K^2$. Hence, there exist $r \in \mathbb{N}$ and $a_1, \dots, a_{2r} \in \mathbb{O}_v^\times$ such that ψ is Witt equivalent to $\langle \bar{a}_1, \dots, \bar{a}_{2r}\rangle$ and $-a_{2i-1}a_{2i} \in \sum K^2$ for $i = 1, \dots, r$. Then $\langle a_1, \dots, a_{2r}\rangle$ is torsion form over K. We choose any $t \in K^\times$ with $v(t) = 1$. Then also $\langle 1, -t\rangle \otimes \langle a_1, \dots, a_{2r}\rangle$ is a torsion form over K. □

The following statement was independently obtained in [Scheiderer 2009, Proposition 5] using different arguments, based on the theory of spaces of orderings.

Proposition 5.2. *For $v \in \Omega(K)$ we have $u(K) \geq u(K^v) = 2u(\kappa_v)$.*

Proof. Let $v \in \Omega(K)$. Let ψ be an anisotropic torsion form over κ_v. Using Proposition 5.1 we choose $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{O}_v^\times$, and $t \in K^\times$ with $v(t) = 1$ such that ψ is Witt equivalent to $\langle \bar{a}_1, \dots, \bar{a}_n\rangle$ and such that $\langle 1, -t\rangle \otimes \langle a_1, \dots, a_n\rangle$ is a torsion form over K. Let φ denote its anisotropic part. Then φ is a torsion form and isometric to $\langle b_1, \dots, b_s\rangle \perp -t\langle c_1, \dots, c_r\rangle$ for certain $r, s \in \mathbb{N}$ and $c_1, \dots, c_r, b_1, \dots, b_s \in \mathbb{O}_v^\times$. Applying residue homomorphisms (see [Lam 2005, Chapter VI, §1]), it follows that the forms $\langle \bar{b}_1, \dots, \bar{b}_s\rangle$ and $\langle \bar{c}_1, \dots, \bar{c}_r\rangle$ over κ_v are

Witt equivalent to ψ. As ψ is anisotropic we conclude that $\dim\varphi = r+s \geq 2\dim\psi$. This shows that $u(K) \geq 2u(\kappa_v)$. Using Springer's Theorem for complete discretely valued fields (see [Lam 2005, Chapter VI, §1]), we further obtain that $u(K^v) = 2u(\kappa_v)$. □

A generalization of Proposition 5.2 for arbitrary nondyadic valuations is given in [Becher and Leep 2013, Theorem 5.2].

Corollary 5.3. *Let k be the residue field of a nondyadic $\mathbb{Z}$-valuation on K. For every algebraic function field F/K there exists an algebraic function field E/k such that $u(F) \geq 2u(E)$.*

Proof. Let T denote the discrete valuation ring with field of fractions K and residue field k. Let F/K be an algebraic function field. Choose $x \in F$ transcendental over K. Consider the valuation ring T' in $K(x)$ described in Proposition 3.3. Note that T' is a discrete valuation ring. Since $F/K(x)$ is a finite extension, there exists a $\mathbb{Z}$-valuation v on F dominating T'. The residue field E of v is a finite extension of $k(\bar{x})$, hence an algebraic function field over k. By Proposition 5.2 we obtain that $u(F) \geq 2u(E)$. □

We define

$$\hat{u}(K) = \tfrac{1}{2}\sup\{u(F) \mid F/K \text{ algebraic function field}\}.$$

For nonreal fields $\hat{u}$ coincides with the *strong u-invariant* defined in [Harbater et al. 2009, Definition 1.2], by the following result.

Corollary 5.4. *For any algebraic extension L/K we have*

$$u(L) \leq \tfrac{1}{2}u(K(X)) \leq \hat{u}(K).$$

Proof. If L is a field of odd characteristic p, then the Frobenius homomorphism given by $x \longmapsto x^p$ shows that any quadratic form over L is obtained by scalar extension from a quadratic form defined over L^p. Therefore every torsion form defined over an algebraic extension of K comes from a torsion form defined over a finite separable extension of K. Since any finite separable extension of K is the residue field of a $\mathbb{Z}$-valuation v on $K(X)$, the first inequality now follows from Proposition 5.2. The second inequality is obvious. □

6. Function fields over complete discretely valued fields

In this section we assume that K is the field of fractions of a complete discrete valuation ring T with residue field k of characteristic different from 2. We want to apply the following reformulation of the local-global principle in [Colliot-Thélène et al. 2012, Theorem 3.1] to the study of the u-invariant and the Pythagoras number of algebraic function fields over K.

Theorem 6.1 (Colliot-Thélène, Parimala, Suresh). *Let F be an algebraic function field over K. A regular quadratic form over F of dimension at least 3 is isotropic if and only if it is isotropic over F^v for every $v \in \Omega(F)$.*

Proof. This slightly more general version of the result cited follows from [Harbater et al. 2013, Proposition 9.10]. □

We will apply Theorem 6.1 to obtain upper bounds for the two mentioned field invariants. We have to distinguish two types of $\mathbb{Z}$-valuations on algebraic function fields over K.

Proposition 6.2. *Let F/K be an algebraic function field and $v \in \Omega(F)$. Then either v is trivial on K or it dominates T.*

Proof. This follows from Proposition 2.2. □

The lower bounds that we will obtain are based on more elementary arguments:

Lemma 6.3. *Let F/K be an algebraic function field and v a $\mathbb{Z}$-valuation on F that is trivial on K. Then $p(F^v) = p'(\kappa_v) \le p(k(X))$ and $u(F^v) = 2u(\kappa_v) \le u(k(X))$.*

Proof. By Corollary 4.4 and Proposition 5.2 we have $p(F^v) = p'(\kappa_v)$ and $u(F^v) = 2u(\kappa_v)$. As κ_v is a finite extension of K and T is a complete discrete valuation ring of K, there is a unique $\mathbb{Z}$-valuation w on κ_v with $\mathcal{O}_w \cap K = T$. Then κ_w is a finite extension of k, and κ_v is complete with respect to w, in particular henselian. By Corollary 5.4 and Proposition 4.3 we obtain that $p'(\kappa_v) = p'(\kappa_w)$ and $u(\kappa_v) = 2u(\kappa_w)$. We choose $\alpha \in \kappa_w$ such that $\kappa_w/k(\alpha)$ is purely inseparable. Since k is of characteristic different from 2, it follows that every element of κ_w is a product of a square and an element from $k(\alpha)$. This yields that $p'(\kappa_w) \le p'(k(\alpha))$ and $u(\kappa_w) \le u(k(\alpha))$. Since $k(\alpha)$ is the residue field of a $\mathbb{Z}$-valuation on $k(X)$, we obtain from Proposition 4.3 and Corollary 5.4 that $p'(k(\alpha)) \le p(k(X))$ and $2u(k(\alpha)) \le u(k(X))$. □

We can now extend Theorem 4.10 of [Harbater et al. 2009] to the current setting, thus covering real function fields. C. Scheiderer [2009, Theorem 3] independently gave a more geometric proof.

Theorem 6.4. *We have $\hat{u}(K) = 2\hat{u}(k)$.*

Proof. For any algebraic function field E/k, by Proposition 3.4 there exists an algebraic function field F/K and a $\mathbb{Z}$-valuation on F with residue field E, and using Proposition 5.2 we obtain that $u(E) \le \frac{1}{2}u(F) \le \hat{u}(K)$. This yields that $2\hat{u}(k) \le \hat{u}(K)$.

To prove the converse inequality, we need to show for an arbitrary algebraic function field F/K that $u(F) \le 4\hat{u}(k)$ holds. Fix F/K. By Theorem 6.1, any anisotropic form over F remains anisotropic over F^v for some $v \in \Omega(F)$. It thus suffices to show that $u(F^v) \le 4\hat{u}(k)$ for every $v \in \Omega(F)$. Fix $v \in \Omega(F)$. As

$u(F^v) = 2u(\kappa_v)$, it suffices to show that $u(\kappa_v) \le 2\hat{u}(k)$. If v is trivial on K, we obtain by Lemma 6.3 that $2u(\kappa_v) \le u(k(X)) \le 2\hat{u}(k)$. Assume that v is nontrivial on K. Then $\mathbb{O}_v \cap K = T$ by Proposition 6.2. If κ_v/k is an algebraic function field then $u(\kappa_v) \le 2\hat{u}(k)$ by the definition of $\hat{u}(k)$. Otherwise κ_v/k is an algebraic extension and then $u(\kappa_v) \le \hat{u}(k)$ by Corollary 5.4. □

Corollary 6.5. *Let $m \in \mathbb{N}$. If $u(E) = m$ for every algebraic function field E/k, then $u(F) = 2m$ for every algebraic function field F/K.*

Proof. Let F/K be an algebraic function field over K. Using Theorem 6.4 we obtain that $u(F) \le 2\hat{u}(K) = 4\hat{u}(k)$. By Corollary 5.3 there exists an algebraic function field E/k with $u(F) \ge 2u(E)$. If we assume that $u(E) = m$ holds for every algebraic function field E/k, we obtain that $2\hat{u}(k) = m$ and conclude that $u(F) = 2m$. □

Theorem 6.6. *We have that*

$$u(K(X)) = 2 \cdot \sup\{u(\ell(X)) \mid \ell/k \textit{ finite separable field extension}\}.$$

Proof. Let $F = K(X)$. As $u(F) \ge 2$, it follows from Theorem 6.1 that

$$u(F) \le \sup\{u(F^v) \mid v \in \Omega(F)\}.$$

Consider $v \in \Omega(F)$. If v is trivial on K then $u(F^v) \le 2u(k(X))$ by Lemma 6.3. If v is nontrivial on K, then by Proposition 2.2 and Theorem 3.2 κ_v/k is either an algebraic extension or algebrorational. In any case we obtain that $u(\kappa_v) \le u(\ell(X))$ and thus $u(F^v) = 2u(\kappa_v) \le 2u(\ell(X))$ for a finite extension ℓ/k. Let ℓ'/k be the separable subextension of ℓ/k such that ℓ/ℓ' is purely inseparable. Then $\ell(X)/\ell'(X)$ is purely inseparable and of odd degree, so every element of $\ell(X)$ is a product of a square in $\ell(X)$ with an element of $\ell'(X)$, whereby $u(\ell(X)) \le u(\ell'(X))$. This together shows that

$$u(F) \le 2 \cdot \sup\{u(\ell(X)) \mid \ell/k \text{ finite separable field extension}\}.$$

On the other hand, given a finite separable field extension ℓ/k, it follows from Theorem 3.5 that there exists a $\mathbb{Z}$-valuation on F with residue field $\ell(X)$, which by Proposition 5.2 implies that $u(F) \ge 2u(\ell(X))$. This shows the claimed equality. □

We turn to the study of sums of squares and the Pythagoras number.

Theorem 6.7. *Let F/K be an algebraic function field. For any $m \ge 2$ we have that $D_F(m) = F^\times \cap (\bigcap_{v\in\Omega(F)} D_{F^v}(m))$. Moreover, $p(F) = \sup\{p'(\kappa_v) \mid v \in \Omega(F)\}$.*

Proof. Applying Theorem 6.1 to the quadratic forms $m \times \langle 1\rangle \perp \langle -a\rangle$ for $a \in F^\times$ shows for any $m \ge 2$ the claimed equality of sets. Note that $\Omega(F)$ contains a nonreal valuation v, and we have that $p(F^v) = s(\kappa_v) + 1 \geq 2$. Since $p(F) \geq 2$ by Corollary 4.8, we obtain that

$$\begin{aligned} p(F) &= \inf\{m \geq 2 \mid D_F(m) = D_F(m+1)\} \\ &\leq \inf\{m \geq 2 \mid D_{F^v}(m) = D_{F^v}(m+1) \text{ for all } v \in \Omega(F)\} \\ &= \sup\{p(F^v) \mid v \in \Omega(F)\}. \end{aligned}$$

Moreover, by Proposition 4.3 we have $p(F^v) = p'(\kappa_v)$ for every $v \in \Omega(F)$. □

Theorem 6.8. *Let F/K be an algebraic function field. There exists an algebraic function field E/k such that $p'(E) \geq p'(F)$. Moreover, if F/K is algebrorational, then one may choose E/k to be algebrorational.*

Proof. If $p(F) = \infty$, then as F is a finite extension of a rational function field, we conclude with [Pfister 1995, Chapter 7, Proposition 1.13] that $p(K(X)) = \infty$ and then with Theorem 4.14 we obtain that $p(k(X)) = \infty$, so that for $E = k(X)$ we have $p'(E) = \infty = p'(F)$.

We now suppose that $p(F) < \infty$. By Theorem 6.7 there exists $v \in \Omega(F)$ such that $p(F) = p'(\kappa_v)$.

Assume first that $p'(F) \neq p(F)$. Then F is nonreal with $p(F) = s(F)$, and by Corollary 4.6 F is not algebrorational. It follows that κ_v is nonreal with $s(\kappa_v) = p'(\kappa_v) - 1 = p(F) - 1 = s(F) - 1$, and as $s(\kappa_v)$ and $s(F)$ are both powers of two, we conclude that $s(F) = 2$. Then $s(k) \geq 2$ and for $E = k(X)(\sqrt{-(1+X^2)})$ we have that $s(E) = 2$ and thus $p'(E) = 3 = p'(F)$.

Suppose now that $p'(F) = p(F) = p'(\kappa_v)$. If $v|_K$ is trivial, then we have $p(k(X)) \geq p'(\kappa_v) = p(F)$ by Lemma 6.3 and further $s(k(X)) = s(k) = s(K) \geq s(F)$, so we may choose $E = k(X)$ to have $p'(E) \geq p'(F)$. Suppose that $v|_K$ is nontrivial. By Proposition 6.2 then v dominates T, and the residue extension κ_v/k is either algebraic or it is an algebraic function field. If κ_v/k is an algebraic function field, we may choose $E = \kappa_v$ and have that $p'(E) \geq p'(F)$. Moreover, by Theorem 3.2, if F/K is algebrorational, then so is E/k. If κ_v/k is algebraic, then as $p'(\kappa_v) = p'(F) < \infty$ there exists a finite extension ℓ/k contained in κ_v/k with $p'(\ell) \geq p'(\kappa_v)$, and we may thus choose $E = \ell(X)$ to have $p'(E) \geq p'(\ell) \geq p'(\kappa_v) = p'(F)$. □

Corollary 6.9. *We have $p(K(X)) = \sup\{p(\ell(X)) \mid \ell/k \text{ finite field extension}\}$.*

Proof. The statement is trivial if k is nonreal. Assume that k is real. Given an arbitrary finite extension ℓ/k, by Theorem 3.5 there is a $\mathbb{Z}$-valuation on $K(X)$ with residue field $\ell(X)$, whereby Proposition 4.3 yields that $p'(\ell(X)) \leq p'(K(X))$. On the other hand, by Theorem 6.8, there exists a finite extension ℓ/k with $p'(K(X)) \leq p'(\ell(X))$. Since $p'(K(X)) = p(K(X))$ and $p'(k(X)) = p(k(X))$ the statement follows. □

Note that Corollary 6.9 shows the equivalence of Conjectures 4.15 and 4.16.

Theorem 6.10. *Let $n \in \mathbb{N}$. Assume that $p(k(X)) \le 2^n$ and that $\sum E^2/D_E(2^n)$ is finite for every algebraic function field E/k. Then $p(K(X)) \le 2^n$ and $\sum F^2/D_F(2^n)$ is finite for every algebraic function field F/K.*

Proof. By Theorem 4.14 we have $p(K(X)) \le 2^n$. Consider an algebraic function field F/K. By Theorem 6.7 the natural homomorphism

$$\sum F^2/D_F(2^n) \longrightarrow \prod_{v \in \Omega(F)} \sum (F^v)^2/D_{F^v}(2^n)$$

is injective. To prove that $\sum F^2/D_F(2^n)$ is finite, it thus suffices to show that the set

$$S = \{v \in \Omega(F) \mid p(F^v) > 2^n\}$$

is finite and that $\sum (F^v)^2/D_{F^v}(2^n)$ is finite for each $v \in S$. Consider $v \in \Omega(F)$. If v is trivial on K, then $p(F^v) \le p(k(X)) \le 2^n$ by Lemma 6.3. Otherwise $\mathcal{O}_v \cap K = T$ by Proposition 6.2 and κ_v is an extension of k. If the extension κ_v/k is algebraic, then $p(F^v) = p'(\kappa_v) \le p(k(X)) \le 2^n$. If κ_v/k is an algebraic function field which is algebrorational, then using Corollary 6.9 we obtain that $p(F^v) = p'(\kappa_v) \le p(K(X)) \le 2^n$. This proves that, for any $v \in S$, we have $\mathcal{O}_v \cap K = T$ and κ_v/k is an algebraic function field that is not algebrorational. The finiteness of S thus follows from Corollary 3.9, and for any $v \in S$ we have $\left|\sum (F^v)^2/D_{F^v}(2^n)\right| \le 2 \cdot \left|\sum (\kappa_v)^2/D_{\kappa_v}(2^n)\right|$, which is finite by the hypothesis. □

Theorem 6.11. *Assume that $n \in \mathbb{N}$ is such that $p(E) \le 2^n$ for any algebraic function field E/k. Let F/K be an algebraic function field. Then $p(F) \le 2^n + 1$ and the set $S = \{v \in \Omega(F) \mid s(\kappa_v) = 2^n\}$ is finite with $\left|\sum F^2/D_F(2^n)\right| = 2^{|S|}$. Moreover, $\Phi_S : F^\times \to \mathbb{Z}^S$ induces an isomorphism $\sum F^2/D_F(2^n) \to (\mathbb{Z}/2\mathbb{Z})^S$.*

Proof. Consider $v \in \Omega(F)$. If $v|_K$ is trivial, then $p'(\kappa_v) \le p(k(X)) \le 2^n$ by Lemma 6.3 and the hypothesis. Suppose now that $v|_K$ is nontrivial. By Proposition 6.2 then $\mathcal{O}_v \cap K = T$ and the residue field extension κ_v/k is either algebraic or it is an algebraic function field. If κ_v/k is algebraic, then $p'(\kappa_v) \le 2^n$. Suppose that κ_v/k is an algebraic function field. Then $p(\kappa_v) \le 2^n$ by the hypothesis. Moreover, if κ_v/k is algebrorational, then Corollary 4.6 yields that $p'(\kappa_v) = p(\kappa_v) \le 2^n$.

Hence, in any case we have that $p(\kappa_v) \le 2^n$, and thus $p(F^v) = p'(\kappa_v) \le 2^n + 1$ by Corollary 4.4. Furthermore, we conclude that $p(F^v) = 2^n + 1$ if and only if $v \in S$, and in this case the residue field extension κ_v/k is an algebraic function field but not algebrorational.

By Theorem 6.7 we conclude that $p(F) \le p'(F) \le 2^n + 1$ and furthermore

$$\sum F^2 = \Big(\bigcap_{v \in S} D_{F^v}(2^n{+}1)\Big) \cap \Big(\bigcap_{v \in S^c} D_{F^v}(2^n)\Big) \cap F^\times,$$

where $S^c = \Omega(F) \setminus S$. Moreover, using Corollary 3.9 we obtain that S is finite. By Corollary 4.11 then $\Phi_S : F^\times \to \mathbb{Z}^S$ induces a surjective homomorphism $\sum F^2/D_F(2^n) \to (\mathbb{Z}/2\mathbb{Z})^S$. It remains to show that this homomorphism is also injective. In view of Theorem 6.7 and the above equality for $\sum F^2$, it suffices to verify that $\Phi_S^{-1}((2\mathbb{Z})^S) \subseteq \bigcap_{v\in S} D_{F^v}(2^n)$. Consider $x \in \sum F^2$ and $v \in S$ with $v(x) \in 2\mathbb{Z}$. Then $x = t^2 y$ with $t \in F^\times$ and $y \in \mathbb{O}_v^\times \cap \big(\sum F^2\big)$, whereby $y + \mathfrak{m}_v \in \sum \kappa_v{}^2$. Since F^v is complete and $p(\kappa_v) \le 2^n$, it follows that $x = t^2 y \in D_{F^v}(2^n)$. This shows the claim. □

Recall that the field K is said to be *hereditarily quadratically closed* if $L^\times = L^{\times 2}$ for every finite field extension L/K. The following result applies in particular to the situation where R is a real closed field.

Theorem 6.12. *Let $n \in \mathbb{N}$ and $K = R(\!(t_1)\!)\dots(\!(t_n)\!)$ for a field R such that $R(\sqrt{-1})$ is hereditarily quadratically closed. Let F/K be an algebraic function field. Then $u(F) = 2^{n+1}$, $2 \le p(F) \le 3$, and the group $\sum F^2/D_F(2)$ is finite.*

Proof. We prove this by induction on n. For $n = 0$ we obtain from [Elman and Wadsworth 1987, Theorem] that $u(F) = 2$, and we conclude by [Lam 2005, Chapter XI, Corollary 6.26] and Corollary 4.8 that $p(F) = 2$, whereby $\sum F^2 = D_F(2)$ and $2 \le p(F) \le p'(F) \le 3$. Assume that $n > 0$. Applying the induction hypothesis to all algebraic function fields over $k = R(\!(t_1)\!)\dots(\!(t_{n-1})\!)$, we obtain by Corollary 6.5 that $u(F) = 2^{n+1}$, by Corollary 4.8 and Theorem 6.8 that $2 \le p(F) \le p'(F) \le 3$, and by Theorem 6.10 that $\sum F^2/D_F(2)$ is finite. □

For certain real function fields over $\mathbb{R}(\!(t)\!)$, it was asked in [Becher and Van Geel 2009, Question 5.15] whether their Pythagoras number is three or four. We can now answer this question:

Corollary 6.13. *Let $h \in \mathbb{R}[X]$ be a nonconstant square-free polynomial with no roots in $\mathbb{R}$. Let F be the function field of the curve $Y^2 = (tX - 1)h$ over $\mathbb{R}(\!(t)\!)$. Then $p(F) = 3$.*

Proof. We have $p(F) \ge 3$ by [Becher and Van Geel 2009, Theorem 5.3 and Corollary 4.2] and $p(F) \le 3$ by Theorem 6.12. □

Acknowledgments

We wish to thank Jean-Louis Colliot-Thélène and Yong Hu for helpful discussions in the context of Proposition 3.7, and furthermore Arno Fehm and Claus Scheiderer for comments on the text. We further wish to acknowledge the scrutiny of the referee, whose comments helped us to improve the presentation. This work was supported by the Deutsche Forschungsgemeinschaft (project *Quadratic Forms and Invariants*, BE 2614/3), by the Swiss National Science Foundation (Grant 200020-124785/1), and by the Zukunftskolleg, Universität Konstanz.

References

[Abhyankar 1956] S. Abhyankar, "On the valuations centered in a local domain", *Amer. J. Math.* **78** (1956), 321–348. MR 18,556b Zbl 0074.26301

[Becher and Leep 2013] K. J. Becher and D. B. Leep, "Real fields, valuations, and quadratic forms", *Manuscripta Math.* **141**:3-4 (2013), 737–754. MR 3062606 Zbl 06179653

[Becher and Van Geel 2009] K. J. Becher and J. Van Geel, "Sums of squares in function fields of hyperelliptic curves", *Math. Z.* **261**:4 (2009), 829–844. MR 2009m:12001 Zbl 1253.11050

[Colliot-Thélène et al. 2012] J.-L. Colliot-Thélène, R. Parimala, and V. Suresh, "Patching and local-global principles for homogeneous spaces over function fields of p-adic curves", *Comment. Math. Helv.* **87**:4 (2012), 1011–1033. MR 2984579 Zbl 06104833

[Elman and Lam 1973] R. Elman and T. Y. Lam, "Quadratic forms and the u-invariant, I", *Math. Z.* **131** (1973), 283–304. MR 48 #2072 Zbl 0244.10020

[Elman and Wadsworth 1987] R. Elman and A. R. Wadsworth, "Hereditarily quadratically closed fields", *J. Algebra* **111**:2 (1987), 475–482. MR 89g:12004 Zbl 0669.12010

[Engler and Prestel 2005] A. J. Engler and A. Prestel, *Valued fields*, Springer, Berlin, 2005. MR 2007a: 12005 Zbl 1128.12009

[Harbater et al. 2009] D. Harbater, J. Hartmann, and D. Krashen, "Applications of patching to quadratic forms and central simple algebras", *Invent. Math.* **178**:2 (2009), 231–263. MR 2010j:11058 Zbl 1259.12003

[Harbater et al. 2013] D. Harbater, J. Hartmann, and D. Krashen, "Local-global principles for torsors over arithmetic curves", preprint, 2013. In versions 1 and 2 (dated 2011) the result cited appears as 9.9 rather than 9.10. arXiv 1108.3323v4

[Lam 2005] T. Y. Lam, *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics **67**, American Mathematical Society, Providence, RI, 2005. MR 2005h:11075 Zbl 1068.11023

[Liu 2002] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Graduate Texts in Mathematics **6**, Oxford University Press, 2002. MR 2003g:14001 Zbl 0996.14005

[Matsumura 1986] H. Matsumura, *Commutative ring theory*, Cambridge Studies in Advanced Mathematics **8**, Cambridge University Press, 1986. MR 88h:13001 Zbl 0603.13001

[Nagata 1967] M. Nagata, "A theorem on valuation rings and its applications", *Nagoya Math. J.* **29** (1967), 85–91. MR 34 #7503 Zbl 0146.26302

[Ohm 1983] J. Ohm, "The ruled residue theorem for simple transcendental extensions of valued fields", *Proc. Amer. Math. Soc.* **89**:1 (1983), 16–18. MR 85b:12010 Zbl 0523.12021

[Parimala and Suresh 2010] R. Parimala and V. Suresh, "The u-invariant of the function fields of p-adic curves", *Ann. of Math.* (2) **172**:2 (2010), 1391–1405. MR 2011g:11074 Zbl 1208.11053

[Pfister 1966] A. Pfister, "Quadratische Formen in beliebigen Körpern", *Invent. Math.* **1** (1966), 116–132. MR 34 #169 Zbl 0142.27203

[Pfister 1995] A. Pfister, *Quadratic forms with applications to algebraic geometry and topology*, London Mathematical Society Lecture Note Series **217**, Cambridge University Press, 1995. MR 97c:11046 Zbl 0847.11014

[Scheiderer 2001] C. Scheiderer, "On sums of squares in local rings", *J. Reine Angew. Math.* **540** (2001), 205–227. MR 2002j:13031 Zbl 0991.13014

[Scheiderer 2009] C. Scheiderer, "The u-invariant of one-dimensional function fields over real power series fields", *Arch. Math.* (*Basel*) **93**:3 (2009), 245–251. MR 2011d:11090 Zbl 1241.11047

Received June 22, 2012. Revised December 2, 2013.

KARIM JOHANNES BECHER
DEPARTEMENT WISKUNDE EN INFORMATICA
UNIVERSITEIT ANTWERPEN
MIDDELHEIMLAAN 1
2020 ANTWERP
BELGIUM

and

ZUKUNFTSKOLLEG
UNIVERSITÄT KONSTANZ
BOX 216
78457 KONSTANZ
GERMANY

becher@maths.ucd.ie

DAVID GRIMM
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
EPFL-SB-MATHGEOM-CSAG
STATION 8
CH-1015 LAUSANNE
SWITZERLAND

david.grimm@epfl.ch

JAN VAN GEEL
UNIVERSITEIT GENT
VAKGROEP WISKUNDE
KRIJGSLAAN 281
9000 GHENT
BELGIUM

jvg@cage.ugent.be

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.277

ON THE EQUIVALENCE PROBLEM FOR TORIC CONTACT STRUCTURES ON S^3-BUNDLES OVER S^2

CHARLES P. BOYER AND JUSTIN PATI

We study the contact equivalence problem for toric contact structures on S^3-bundles over S^2. That is, given two toric contact structures, one can ask the question: when are they equivalent as contact structures while inequivalent as toric contact structures? In general this appears to be a difficult problem. To show that two toric contact structures with the same first Chern class are contact inequivalent, we use Morse–Bott contact homology. To find inequivalent toric contact structures that are contact equivalent, we show that the corresponding 3-tori belong to distinct conjugacy classes in the contactomorphism group. We treat a subclass of contact structures which includes the Sasaki–Einstein contact structures $Y^{p,q}$ studied by physicists with the anti-de Sitter/conformal field theory conjecture. In this case we give a complete solution to the contact equivalence problem by showing that $Y^{p,q}$ and $Y^{p',q'}$ are inequivalent as contact structures if and only if $p \neq p'$.

Introduction

It is well known that contact structures have only discrete invariants; that is, Gray's theorem says that the deformation theory is trivial. Apparently, the crudest such

Boyer was partially supported by grant #245002 from the Simons Foundation. Pati was partially supported by the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine.
MSC2010: 53D10, 53D20, 53D42.
Keywords: toric contact geometry, equivalent contact structures, orbifold Hirzebruch surface, contact homology, extremal Sasakian structures.

invariant is the first Chern class of the contact bundle $\mathcal{D}$. Indeed, the mod 2 reduction of $\mathcal{D}$ is a topological invariant, namely the second Stiefel–Whitney class. A much more subtle and powerful invariant is contact homology, a small part of the more general symplectic field theory (SFT) of Eliashberg, Givental, and Hofer [Eliashberg et al. 2000] — which can be used to distinguish contact structures belonging to the same isomorphism class of oriented $2n$-plane bundle.

On the other hand, given two contact structures with the same invariants, when can one show that they are equivalent? In full generality this appears to be a very difficult problem. However, if we restrict ourselves to toric contact structures in dimension five, we can begin to get a handle on things. The problem is of particular interest when applied to toric contact manifolds since they have been classified [Lerman 2003a]. Thus, one is interested in when two inequivalent toric contact structures are equivalent as contact structures. Specializing further we consider all toric contact structures on S^3-bundles over S^2. It is well known that such manifolds are classified by $\pi_1(\mathrm{SO}(4)) = \mathbb{Z}_2$, so there are exactly two such bundles, the trivial bundle $S^2 \times S^3$ and one nontrivial bundle X_∞ (in the notation of [Barden 1965]). They are distinguished by their second Stiefel–Whitney class $w_2 \in H^2(M, \mathbb{Z}_2)$. The problem of determining when two such toric contact structures belong to equivalent contact structures is now somewhat tractable owing to the work of Karshon [2003] and Lerman [2003b].

The general toric contact structures on $S^2 \times S^3$ or X_∞ depend on four integers (p_1, p_2, p_3, p_4) that satisfy $\gcd(p_i, p_j) = 1$ for $i = 1, 2$ and $j = 3, 4$. We write the contact structures as $\mathcal{D}_{\boldsymbol{p}}$, using vector notation $\boldsymbol{p}$ for the quadruple. However, this general situation appears somewhat intractable, so we consider the special case when $p_1 = p_2$ or $p_3 = p_4$, which is more tractable because then a certain quotient is a Hirzebruch surface with branch divisors. We may as well assume that $p_3 = p_4$. It is often convenient to further divide this case into two subcases, as follows. We set $\boldsymbol{p} = (j, 2k-j, l, l)$ for $S^2 \times S^3$ and $\boldsymbol{p} = (j, 2k-j+1, l, l)$ for X_∞ with $1 \le j \le k$. We denote either one of these contact structures by $\mathcal{D}_{p_1,p_2,l,l}$, $\mathcal{D}_{j,2k-j,l,l}$, $\mathcal{D}_{j,2k-j+1,l,l}$ or simply as $\mathcal{D}_{j,k,l}$, depending on which notation is more convenient. The first Chern class $c_1(\mathcal{D})$ of the contact bundle is a classical algebraic invariant of the contact structure. For the contact structure $\mathcal{D}_{\boldsymbol{p}}$ we will give an explicit formula for $c_1(\mathcal{D}_{\boldsymbol{p}})$ which for $\mathcal{D}_{j,2k-j,l,l}$ and $\mathcal{D}_{j,2k-j+1,l,l}$ equals $2(k-l)$ and $2(k-l)+1$, respectively. Using contact homology we show that even if $\mathcal{D}_{j,k,l}$ and $\mathcal{D}_{j',k',l'}$ are not distinguished by the first Chern class, they are inequivalent if $k \ne k'$.

Our main result about inequivalence is this:

Theorem 1. *The two toric contact structures $\mathcal{D}_{p_1,p_2,l,l}$ and $\mathcal{D}_{p'_1,p'_2,l',l'}$ on $S^2 \times S^3$ or X_∞ are inequivalent contact structures if $p'_1 + p'_2 \ne (p_1 + p_2)$.*

For our main result about equivalence, we need to specialize a bit further. In this case we require that $\gcd(p_2 - p_1, l)$ be constant. We have

Theorem 2. *The two contact structures $\mathcal{D}_{p_1,p_2,l,l}$ and $\mathcal{D}_{p'_1,p'_2,l,l}$ satisfying $p'_1+p'_2 = p_1+p_2$ are equivalent if $\gcd(l, p_2 - p_1) = \gcd(l, p'_2 - p'_1)$.*

Recently there has been a great deal of focus on certain toric contact structures $Y^{p,q}$ with vanishing first Chern class on $S^2 \times S^3$ discovered by Gauntlett, Martelli, Sparks, and Waldram [2004a], and used in their study of the anti-de Sitter/conformal field theory conjecture [Gauntlett et al. 2004b; 2005] (see Chapter 11 of [Boyer and Galicki 2008] and [Sparks 2011] and references therein). In our notation the contact structures $Y^{p,q}$ correspond to $\mathcal{D}_{p-q,p+q,p,p}$. Remarkably our results give a complete answer to the contact equivalence problem for these structures.

Theorem 3. *Let ϕ denote the Euler ϕ-function. The toric contact structures $Y^{p,q}$ and $Y^{p',q'}$ on $S^2 \times S^3$ belong to equivalent contact structures if and only if $p' = p$, and for each fixed integer $p > 1$ there are exactly $\phi(p)$ toric contact structures $Y^{p,q}$ on $S^2 \times S^3$ that are equivalent as contact structures, denoted by $\mathcal{D}_p$. Moreover, the contactomorphism group of $\mathcal{D}_p$ has at least $\phi(p)$ conjugacy classes of maximal tori of dimension three.*

A partial result, namely that $Y^{p',1}$ and $Y^{p,1}$ are inequivalent contact structures if $p' \neq p$, was recently given by Abreu and Macarini [2012], and an outline of the proof of Theorem 3 was recently given by one of us [Boyer 2011a].

As a bonus we also obtain the following results concerning extremal Sasakian structures:

Corollary 4. *For each such contact structure $\mathcal{D}_p$ there are $\phi(p)$ compatible Sasaki–Einstein metrics that are inequivalent as Riemannian metrics.*

Corollary 5. *For both $S^2 \times S^3$ and X_∞ the moduli space of extremal Sasakian structures has a countably infinite number of components. Moreover, each component has extremal Sasakian metrics of positive Ricci curvature whose isometry group contains T^3.*

This corollary follows already from the results of [Pati 2009; 2010; Boyer 2011b], but Theorem 1 actually gives a much larger class in the sense that there are countably many new components. As shown in [Boyer 2011b], many of these components are themselves non-Hausdorff.

Corollary 6. *The moduli space of Sasaki–Einstein metrics on $S^2 \times S^3$ has a countably infinite number of components. Moreover, each such component has Sasaki–Einstein metrics whose isometry group contains T^3.*

This corollary also follows from [Abreu and Macarini 2012].

1. Contact structures and cones

It is well known that contact geometry is equivalent to the geometry of certain symplectic cones. However, for certain contact structures there are several cones that become important, and as we shall see they are all related.

A warning about notation. In contact topology the contact bundle is usually denoted by ξ, whereas in Sasakian geometry ξ is almost always a Reeb vector field. To avoid confusion we eschew the use of ξ completely, and use $\mathcal{D}$ for the contact bundle and R for a Reeb vector field.

Contact structures. Recall that a *contact structure*[1] on a connected oriented manifold M is an equivalence class of 1-forms η satisfying $\eta \wedge (d\eta)^n \neq 0$ everywhere on M where two 1-forms η and η' are equivalent if there exists a nowhere-vanishing function f such that $\eta' = f\eta$. We shall also assume that our contact structure has an orientation, or equivalently, the function f is everywhere positive. More conveniently the contact structure can be thought of as the oriented $2n$-plane bundle defined by $\mathcal{D} = \ker \eta$, and we denote by $\mathfrak{C}^+(\mathcal{D})$ the set of all contact 1-forms representing the oriented bundle $\mathcal{D}$.

Recall that an almost-complex structure J on $\mathcal{D}$ is *compatible* with the contact structure if these two conditions hold for any smooth sections X and Y of $\mathcal{D}$:

$$d\eta(JX, JY) = d\eta(X, Y), \quad d\eta(JX, Y) > 0.$$

It is easy to see that these conditions are independent of the choice of 1-form η representing $\mathcal{D}$. The space of almost-complex structures that are compatible with $\mathcal{D}$ is contractible which implies that the Chern classes are invariants of the contact bundle $\mathcal{D}$. In particular, the first Chern class $c_1(\mathcal{D})$ will play an important role for us. Notice also that the pair $(\mathcal{D}, J)$ defines a strictly pseudoconvex almost-CR structure on M, and a choice of contact form η gives a choice of Levi form — essentially $d\eta$.

Also for every choice of contact 1-form η there exists a unique vector field R, called the *Reeb vector field*, that satisfies $\eta(R) = 1$ and $R \lrcorner d\eta = 0$. Such vector fields and the orbits of their flows will play a crucial role for us. We can now extend J to an endomorphism Φ of TM by defining $\Phi|_{\mathcal{D}} = J$ and $\Phi R = 0$. The triple (R, η, Φ) canonically defines a Riemannian metric on M by setting $g = d\eta \circ (\Phi \otimes \mathbb{1}) + \eta \otimes \eta$, and the quadruple (R, η, Φ, g) is known as a *contact metric structure* on M.

Notice that R defines a one-dimensional foliation $\mathcal{F}_R$ on M, often called the *characteristic foliation*. We say that the foliation $\mathcal{F}_R$ is *quasiregular* if there is a positive integer k such that each point has a foliated coordinate chart (U, x) such that each leaf of $\mathcal{F}_R$ passes through U at most k times. If $k = 1$ then the foliation is

[1] This is not the most general definition of a contact structure, but it suffices in most situations (compare [Boyer and Galicki 2008, Chapter 6]), and certainly for us here.

called *regular*. We also say that the corresponding contact 1-form η is *quasiregular* (*regular*), and more generally that a contact structure $\mathcal{D}$ is *quasiregular* (*regular*) if it has a quasiregular (regular) contact 1-form. A contact 1-form (or characteristic foliation) that is not quasiregular is called *irregular*. On a compact manifold any quasiregular contact form is necessarily K-contact, and then the foliation $\mathcal{F}_R$ is equivalent to a locally free circle action (compare [Boyer and Galicki 2008, §7.1]) preserving the quadruple (R, η, Φ, g). This is the case that we are interested in. The quotient space $\mathcal{Z} = M/\mathcal{F}_R$ is a compact orbifold with a naturally defined symplectic structure ω and compatible almost-complex structure $\hat{J}$ satisfying $\pi^*\omega = d\eta$ and J is the horizontal lift of $\hat{J}$, that is, $(\omega, \hat{J})$ defines an almost-Kähler structure on the orbifold $\mathcal{Z}$. Moreover, η can be interpreted as a connection 1-form in the principal S^1 orbibundle $\pi : M \to \mathcal{Z}$ with curvature 2-form $\pi^*\omega$.

In this paper we are interested in the case when both J and $\hat{J}$ are integrable. Then the quadruple (R, η, Φ, g) is a *Sasakian structure* on M, and $(\omega, \hat{J})$ defines a is projective algebraic orbifold structure on $\mathcal{Z}$ with an orbifold Kähler metric. This construction has a converse, that is, beginning with a compact almost-Kähler orbifold one can construct a K-contact structure on the total space of a certain S^1 orbibundle over $\mathcal{Z}$. This is often referred to as the *orbifold Boothby–Wang construction*.[2] It lies at the heart of the proof of Theorem 2. Indeed, we shall show the equivalence of certain contact structures by exhibiting a symplectomorphism between their corresponding quotient orbifolds.

Orbifolds. As just described, orbifolds will play an important role for us in this paper. We refer to [Boyer and Galicki 2008, Chapter 4] for the basic definitions and results. Here we want to emphasize several aspects. First, many cohomology classes that are integral classes on manifolds are only rational classes on the underlying topological space of an orbifold, in particular, the orbifold first Chern class of a complex line orbibundle or circle orbibundle is generally a rational class. However, not all rational classes occur as such. To determine which rational classes can be used to classify line orbibundles, it is convenient to pass to Haefliger's classifying space $B\mathcal{X}$ (see [Haefliger 1984] and/or [Boyer and Galicki 2008, Chapter 4]) of an orbifold $\mathcal{X}$ where, as with smooth manifolds, all complex line orbibundles correspond to integral cohomology classes. Let $\mathcal{X}$ be a complex orbifold with underlying topological space X. Then Haefliger's orbifold cohomology $H^*_{\mathrm{orb}}(\mathcal{X}, \mathbb{Z})$ equals $H^*(B\mathcal{X}, \mathbb{Z})$, which is generally different than $H^*(X, \mathbb{Z})$, but satisfies $H^*_{\mathrm{orb}}(\mathcal{X}, \mathbb{Z}) \otimes \mathbb{Q} = H^*(X, \mathbb{Z}) \otimes \mathbb{Q}$. So, for example, we obtain an integral cohomology class $p^*c_1^{\mathrm{orb}}(\mathcal{X}) \in H^2_{\mathrm{orb}}(\mathcal{X}, \mathbb{Z})$ for complex line orbibundles from the rational class $c_1^{\mathrm{orb}}(\mathcal{X}) \in H^2(X, \mathbb{Q})$. This amounts to clearing the order of the orbifold in the denominator. Here $p : B\mathcal{X} \to \mathcal{X}$ is the

[2]In [Eliashberg et al. 2000] a contact manifold constructed in this way is called a *prequantization space*.

natural projection. We warn the reader that the orbifold cohomology $H^*_{\mathrm{orb}}(\mathcal{X}, \mathbb{Z})$ is not the Chen–Ruan cohomology.

The orbifolds that occur in this paper are of a special type. They are all complex orbifolds whose underlying space is a smooth projective algebraic variety with an added orbifold structure. In such cases it is convenient to view an orbifold $\mathcal{X}$ as a pair (X, Δ) where X is a smooth algebraic variety and Δ is a certain $\mathbb{Q}$-divisor, called a *branch divisor* [Boyer et al. 2005; Ghigi and Kollár 2007; Boyer and Galicki 2008, Chapter 4]. We write $(X, \varnothing)$ to denote the algebraic variety X with the trivial orbifold structure, that is, the charts are just the standard manifold charts. In this situation, as emphasized in [Ghigi and Kollár 2007], we consider the map $\mathbb{1}_X : (X, \Delta) \to (X, \varnothing)$, which is the identity as a set map, and a Galois cover with trivial Galois group.

Symplectic cones. Given a contact structure $\mathcal{D}$ on M we recall the symplectic cone $C(M) = M \times \mathbb{R}^+$ with its natural symplectic structure (the *symplectization* of $(M, \mathcal{D})$) $\Omega = d(r^2\eta)$, where r is a coordinate on $\mathbb{R}^+$. Recall the Liouville vector field $\Psi = r\partial/\partial r$ on the cone $C(M)$.

Now for each choice of contact form $\eta \in \mathfrak{C}^+(\mathcal{D})$ there is a natural extension of the almost-complex structure J on $\mathcal{D}$ to an almost-complex structure I on the cone $C(M)$ defined uniquely by

$$I = \Phi + \Psi \otimes \eta, \quad I\Psi = -R, \tag{1}$$

where Φ is the extension of J to TM defined by $\Phi R = 0$. We can also check that there is a one-to-one correspondence between the compatible almost-complex structures I on $C(M)$ and elements of $\mathfrak{C}^+(\mathcal{D})$, and that (1) hold, so we recover the full-contact metric structure for each $\eta \in \mathfrak{C}^+(\mathcal{D})$. Given an almost-complex manifold, W, with complex structure j, a C^∞ map, u, from W into the almost-complex manifold (N, J) is called J-holomorphic if $du + J(u)d(u \circ j) = 0$. We are specifically interested in pseudoholomorphic maps into the cone, that is, maps which are pseudoholomorphic with respect to the almost-complex structure given by (1). Such maps from a Riemann surface into $C(M)$ are of particular interest and are known as pseudoholomorphic curves.

Summarizing, we have these correspondences:

(1) symplectic cone $(C(M), \Omega) \leftrightarrow$ contact structure $(M, \mathcal{D})$,

(2) almost-Kähler cone $(C(M), \Omega, I) \leftrightarrow$ contact metric structure (M, R, η, Φ, g),

(3) almost-Kähler cone $(C(M), \Omega, I)$ with $\Psi - iR$ pseudoholomorphic $\leftrightarrow$ K-contact structure (M, R, η, Φ, g), and

(4) Kähler cone $(C(M), \Omega, I)$ with $\Psi - iR$ holomorphic $\leftrightarrow$ Sasakian structure (M, R, η, Φ, g).

Remark 1.1. In the sequel when we study pseudoholomorphic curves, it is customary to parametrize the cone by $\mathbb{R}$ so that the singularity, which appears at 0 in the parametrization above, appears at $-\infty$ instead. This amounts to choosing $\Omega = d(e^r\eta)$ for the symplectic form on $C(M)$.

Sasakian structures. The contact structures considered in this paper are all of *Sasaki type*, that is, there is a contact form η and compatible metric g such that $\mathcal{S} = (R, \eta, \Phi, g)$ is a Sasakian structure on M. In this case not only is the cone $C(M)$ discussed above Kähler, but the geometry transverse to the characteristic foliation $\mathcal{F}_R$ is also Kähler. This gives rise to a basic cohomology ring $H^*_B(\mathcal{F}_R)$ (see [Boyer and Galicki 2008, §7.2]), and a transverse Hodge theory. This gives basic Chern classes $c_k(\mathcal{F}_R)$ which, if (R, η, Φ, g) is quasiregular, are the pullbacks of the orbifold Chern classes $c_k^{\mathrm{orb}}(\mathcal{Z})$ on the base orbifold $\mathcal{Z}$. In particular we are interested in the basic first Chern class $c_1(\mathcal{F}_R) \in H_B^{1,1}(\mathcal{F}_R)$. A Sasakian structure $\mathcal{S} = (R, \eta, \Phi, g)$ is said to be *positive* (*negative*) if its basic first Chern class $c_1(\mathcal{F}_R)$ can be represented by a positive (negative)-definite $(1, 1)$-form. It is *null* if $c_1(\mathcal{F}_R) = 0$, and *indefinite* otherwise. It follows from [Boyer 2011b, Lemma 5.1] that all Sasakian structures occurring in toric contact structures of Reeb type are either positive or indefinite. We mention that these types occur in rays, that is, performing a transverse homothety (see [Boyer and Galicki 2008, p. 228]), preserves the type.

The Sasaki cone. Let $\mathfrak{CR}(\mathcal{D}, J)$ denote the group of almost-CR transformations of $(\mathcal{D}, J)$ on M. If M is compact, it is a Lie group which is compact except when $(\mathcal{D}, J)$ is the standard CR structure on the sphere S^{2n+1} by, in various stages, a theorem of Frances, Lee, and Schoen (see [Boyer 2013]). We let $\mathfrak{cr}(\mathcal{D}, J)$ denote the Lie algebra of $\mathfrak{CR}(\mathcal{D}, J)$. Recall [Boyer et al. 2008] that the subset

$$\mathfrak{cr}^+(\mathcal{D}, J) = \{X \in \mathfrak{cr}(\mathcal{D}, J) \mid \eta(X) > 0\}$$

is independent of the choice of $\eta \in \mathfrak{C}^+(\mathcal{D})$ and is an open convex cone (without the cone point) in $\mathfrak{cr}(\mathcal{D}, J)$. Now the adjoint action of the group $\mathfrak{CR}(\mathcal{D}, J)$ on its Lie algebra leaves $\mathfrak{cr}^+(\mathcal{D}, J)$ invariant, and the quotient space

$$\kappa(\mathcal{D}, J) = \mathfrak{cr}^+(\mathcal{D}, J)/\mathfrak{CR}(\mathcal{D}, J)$$

is known as the (*reduced*) *Sasaki cone* of $(\mathcal{D}, J)$. One should think of $\kappa(\mathcal{D}, J)$ as the *moduli space* of K-contact structures associated to the strictly pseudoconvex almost-CR structure $(\mathcal{D}, J)$. In the case that the almost-CR structure is integrable, $\kappa(\mathcal{D}, J)$ is the moduli space of Sasakian structures associated to $(\mathcal{D}, J)$. It is often convenient to work with the *unreduced Sasaki cone* given by choosing a maximal torus T of $\mathfrak{CR}(\mathcal{D}, J)$. Then the unreduced Sasaki cone is $\mathfrak{t}^+(\mathcal{D}, J) = \mathfrak{t} \cap \mathfrak{cr}^+(\mathcal{D}, J)$ where $\mathfrak{t}$ is the Lie algebra of T.

Of course, many contact structures do not have a Sasaki cone. In fact, a contact structure has a nonempty Sasaki cone if and only if it is of K-contact type. It is important to realize that the Sasaki cone depends on the choice of transverse almost-complex structure J. Indeed by changing J in a given K-contact structure, we can have more than one Sasaki cone. These occur in bouquets related to the conjugacy classes of maximal tori in the contactomorphism $\mathfrak{Con}(M, \mathcal{D})$ of $(M, \mathcal{D})$ [Boyer 2011b; 2013].

The moment cone. Now let T be a torus subgroup of $\mathfrak{Con}(M, \mathcal{D})$, and let $\mathfrak{t}$ be its Lie algebra. Consider the annihilator $\mathcal{D}^o$ of $\mathcal{D}$ which is a trivial real line bundle over M. The orientation on $\mathcal{D}$ allows us to write $\mathcal{D}^o \setminus \{0\} = \mathcal{D}^o_+ \cup \mathcal{D}^o_-$, and we can identify $\mathcal{D}^o_+ \approx M \times \mathbb{R}^+ = C(M)$. Then the contact moment map $\Upsilon : \mathcal{D}^o_+ \to \mathfrak{t}^*$ is defined by

$$\langle \Upsilon(x, p), \tau \rangle = \langle p, \tau_x \rangle, \tag{2}$$

where $\tau \in \mathfrak{t}$ and τ_x denotes the fundamental vector field associated to τ at the point x. The *moment cone* $C(\Upsilon)$ is defined [Lerman 2003a] as the union of the image set with the cone point, that is,

$$C(\Upsilon) = \Upsilon(\mathcal{D}^o_+) \cup \{0\}. \tag{3}$$

By averaging over T we can choose a T-invariant contact form η which gives an equivariant moment map $\mu_\eta : M \to \mathfrak{t}^*$ satisfying

$$\mu_\eta = \Upsilon \circ \eta. \tag{4}$$

Again by averaging we can choose an almost-complex structure J that is T-invariant, so $\mathfrak{t}$ is an Abelian subalgebra of $\mathfrak{cr}(\mathcal{D}, J)$. Furthermore, the contact form η is K-contact (with respect to J) if and only if its Reeb vector field R_η lies in the Lie algebra $\mathfrak{t}$. In this case we also say that the torus action is of *Reeb type* [Boyer and Galicki 2000a]. It is easy to see that this is equivalent to the existence of an element $\tau \in \mathfrak{t}$ such that $\eta(\tau)$ is strictly positive on M. When the contact structure $\mathcal{D}$ is of Reeb type $C(\Upsilon)$ is a convex rational polyhedral cone, and we have the following result of Lerman [2003a].

Lemma 1.2. *A T-invariant contact form η is K-contact if and only if the image $\mu_\eta(M)$ lies in the intersection of a hyperplane H_η with the moment cone $C(\Upsilon)$. Moreover, in the K-contact case the intersection $P_\eta = H_\eta \cap C(\Upsilon)$ is a simple convex polytope which is rational if and only if η is quasiregular.*

The hyperplane H_η is called the *characteristic hyperplane*.

2. Toric contact structures of Reeb type

Toric contact structures on manifolds of dimension greater than three come in two types, those where the action of the torus is free, and those where it is not [Banyaga and Molino 1993; Lerman 2003a]. The latter contain an important special subclass known as toric contact structures of Reeb type [Boyer and Galicki 2000a]. These are precisely the toric contact where the torus action is not free and the moment cone contains no nonzero linear subspace. When the moment cone contains a nontrivial linear subspace, the toric contact manifold will have infinite fundamental group. Thus, any toric contact structure on an S^3-bundle over S^2 must be of Reeb type, and these correspond precisely to convex polyhedral cones in the dual of the Lie algebra of the torus that are cones over a polytope [Boyer and Galicki 2000a; Lerman 2003a].

Definition 2.1. A *toric contact manifold* $(M, \mathcal{D}, \mathcal{A})$ is a contact manifold of dimension $2n+1$ together with an effective action of a torus T of dimension $n+1$ that leaves the contact structure invariant, that is, if $\mathcal{A} : T \times M \to M$ denotes the action map then $\mathcal{A}_* \mathcal{D} = \mathcal{D}$.

By averaging over T we can always find a contact 1-form η representing $\mathcal{D}$ such that $\mathcal{A}^*\eta = \eta$. In this case we also have $\mathcal{A}_* R = R$ for the Reeb vector field. A toric contact manifold is said to be of *Reeb type* if there is a contact form $\eta \in \mathfrak{C}^+(\mathcal{D})$ whose Reeb vector field lies in the Lie algebra $\mathfrak{t}$ of T.

Two toric contact manifolds $(M, \mathcal{D}, \mathcal{A})$ and $(M', \mathcal{D}', \mathcal{A}')$ are said to be *equivariantly equivalent* (or *equivalent toric contact manifolds*) if there exists a contactomorphism between them that conjugates the torus actions $\mathcal{A}$ and $\mathcal{A}'$. Toric contact manifolds were classified in [Lerman 2003a]. In this paper we are interested in inequivalent toric contact manifolds that are equivalent as contact manifolds. In this case the tori generated by the actions $\mathcal{A}$ and $\mathcal{A}'$ belong to distinct conjugacy classes in the contactomorphism group $\mathfrak{Con}(M, \mathcal{D})$. Furthermore, to each such conjugacy class there is an associated toric CR structure $(\mathcal{D}, J)$ which by [Boyer 2013, Theorem 7.6] is unique up to biholomorphism.

Contact reduction. It is well known (see [Boyer and Galicki 2000a; Lerman 2003a]) that every contact toric structure of Reeb type can be obtained by symmetry reduction of the standard sphere by a compact Abelian group A, and that this is equivalent to the symplectic reduction of the standard symplectic structure on $\mathbb{C}^N \setminus \{0\}$ by a compact Abelian group which commutes with the action of dilations of the cone. For this one must choose the zero level set of the toral moment map. This equivalence can be described by the commutative diagram

$$\begin{array}{ccc} S^{2N-1}_{\boldsymbol{w}} & \longleftrightarrow & \mathbb{C}^N \setminus \{0\} \\ \Downarrow & & \Downarrow \\ M^{2n-1} & \longleftrightarrow & C(M^{2n-1}), \end{array} \tag{5}$$

with $\dim A = N - n$. See [Boyer and Galicki 2008, p. 293].

Lemma 2.2. *Let M be an S^3-bundle over S^2. Every toric contact structure on M can be obtained by contact circle reduction of the standard contact structure on S^7.*

Proof. As stated above, every compact toric contact manifold of Reeb type can be obtained by symmetry reduction of the standard sphere S^{2N-1} by a compact Abelian group A [Boyer and Galicki 2000a; Lerman 2003a]. Now by the homotopy exact sequence M is simply connected and $\pi_2(M) = \mathbb{Z}$. By a result of Lerman [2004] $\pi_1(M) = \pi_0(A)$ and $\pi_2(M) = \pi_1(A) = \mathbb{Z}^{\dim T}$. Thus, A is a torus of dimension one, that is, a circle. Since $n = 3$, $N = n + \dim A = 4$, so M is obtained by contact reduction from S^7. □

Remark 2.3. It is well known that there are exactly two S^3-bundles over S^2, distinguished by their second Stiefel–Whitney class, the trivial bundle $S^3 \times S^2$ and the nontrivial bundle denoted by X_∞ in the Barden–Smale classification [Barden 1965] of simply connected 5-manifolds. We will show their relation with the reduction parameters in Theorem 2.6.

We now describe this reduction. First, the standard T^4 action on $\mathbb{C}^4$ is $z_j \mapsto e^{i\theta_j} z_j$, and its moment map $\Upsilon_4 : \mathbb{C}^4 \setminus \{0\} \to \mathfrak{t}_4^* = \mathbb{R}^4$ is given by

$$\Upsilon_4(z) = (|z_1|^2, |z_2|^2, |z_3|^2, |z_4|^2). \tag{6}$$

Now we consider the circle group $T(\boldsymbol{p})$ acting on $\mathbb{C}^4 \setminus \{0\}$ by

$$(z_1, z_2, z_3, z_4) \mapsto (e^{ip_1\theta} z_1, e^{ip_2\theta} z_2, e^{-ip_3\theta} z_3, e^{-ip_4\theta} z_4), \tag{7}$$

where $\boldsymbol{p}$ denotes the quadruple (p_1, p_2, p_3, p_4) with $p_i \in \mathbb{Z}^+$ and we assume $\gcd(p_1, p_2, p_3, p_4) = 1$. We have an exact sequence of commutative Lie algebras

$$0 \to \mathfrak{t}_1(\boldsymbol{p}) \to \mathbb{R}^4 \xrightarrow{\tilde{\varpi}} \mathfrak{t}_3(\boldsymbol{p}) \to 0, \tag{8}$$

where $\mathfrak{t}_1(\boldsymbol{p})$ is the Lie algebra of $T(\boldsymbol{p})$ generated by the vector field $L_{\boldsymbol{p}} = p_1 H_1 + p_2 H_2 - p_3 H_3 - p_4 H_4$.

Dualizing (8) gives

$$0 \to \mathfrak{t}_3^*(\boldsymbol{p}) \xrightarrow{\tilde{\varpi}^*} (\mathbb{R}^4)^* \to \mathfrak{t}_1^*(\boldsymbol{p}) \to 0. \tag{9}$$

The moment map $\Upsilon_1 : \mathbb{C}^4 \setminus \{0\} \to \mathfrak{t}_1^* = \mathbb{R}$ for this action is given by

$$\Upsilon_1(z) = p_1|z_1|^2 + p_2|z_2|^2 - p_3|z_3|^2 - p_4|z_4|^2. \tag{10}$$

Now consider the 1-form

$$\eta_0 = -\frac{i}{2} \sum_{j=0}^{n} (z_j d\bar{z}_j - \bar{z}_j dz_j) \tag{11}$$

on $\mathbb{C}^4 \setminus \{0\}$ together with the vector field

$$R_{\boldsymbol{p}} = \sum_j p_j H_j \tag{12}$$

where $H_j = -i(z_j\,\partial/\partial z_j - \bar{z}_j\,\partial/\partial \bar{z}_j)$. Imposing the constraint $\eta_0(R_{\boldsymbol{p}}) = 1$ gives S^7 represented as $\sum_j p_j|z_j|^2 = 1$. Then η_0 pulls back to a contact form on S^7, also denoted by η_0, with Reeb vector field $R_{\boldsymbol{p}} = p_1H_1 + p_2H_2 + p_3H_3 + p_4H_4$. By a change of coordinates one easily sees that this represents the standard contact structure on S^7.

So the zero level set $\Upsilon_1^{-1}(0)$ is diffeomorphic to a cone over $S^3 \times S^3$, or equivalently restricting to S^7, the zero level set of μ_{η_0} is $S^3 \times S^3$, represented by

$$p_1|z_1|^2 + p_2|z_2|^2 = \tfrac{1}{2}, \quad p_3|z_3|^2 + p_4|z_4|^2 = \tfrac{1}{2}. \tag{13}$$

The action of $T(\boldsymbol{p})$ is free on this zero set if and only if $\gcd(p_i, p_j) = 1$ for $i = 1, 2$ and $j = 3, 4$. So assuming these gcd conditions our reduced contact manifold is the $M_{\boldsymbol{p}} = (S^3 \times S^3)/T(\boldsymbol{p})$ whose contact form is the unique 1-form $\eta_{\boldsymbol{p}}$ satisfying $\iota^*\eta_0 = \rho^*\eta_{\boldsymbol{p}}$, where $\iota : \mu_{\eta_0}^{-1}(0) \to S^7$ and $\rho : \mu_{\eta_0}^{-1}(0) \to M_{\boldsymbol{p}}$ are the natural inclusion and projection, respectively. In order to identify $M_{\boldsymbol{p}}$ we consider the $T^2(\boldsymbol{p})$ action on $\mu_{\eta_0}^{-1}(0) \approx S^3 \times S^3$ generated by the S^1 action (7) together with the S^1 action generated by the Reeb vector field $R_{\boldsymbol{p}}$. To guarantee a smooth quotient we have:

Definition 2.4. We say that the quadruple $\boldsymbol{p} = (p_1, p_2, p_3, p_4)$ of positive integers is *admissible* if $\gcd(p_i, p_j) = 1$ for $i = 1, 2$ and $j = 3, 4$. We denote the set of admissible quadruples by $\mathcal{A}$.

Let us describe some obvious equivalences. We can interchange the coordinates $z_1 \leftrightarrow z_2$, likewise $z_3 \leftrightarrow z_4$. Thus, without loss of generality we can assume that $p_1 \le p_2$ and $p_3 \le p_4$. We can also interchange the pairs (z_1, z_2) and (z_3, z_4). It is also convenient to set $k = \gcd(p_1, p_2)$ and $l = \gcd(p_3, p_4)$ and define $(p_1, p_2) = k(\bar{p}_1, \bar{p}_2)$ and $(p_3, p_4) = l(\bar{p}_3, \bar{p}_4)$ with $\gcd(\bar{p}_1, \bar{p}_2) = \gcd(\bar{p}_3, \bar{p}_4) = 1$. Note that $\boldsymbol{p} \in \mathcal{A}$ implies $\gcd(k, l) = 1$. We will need the "standard" Kähler forms on the weighted projective spaces for which we take the Bochner-flat Kähler forms of area 2 described in [Bryant 2001; David and Gauduchon 2006; Gauduchon 2009]. We denote the corresponding Kähler forms by $\omega_{\bar{p}_1,\bar{p}_2}$ and $\omega_{\bar{p}_3,\bar{p}_4}$. We note that these Kähler forms are just those obtained by quotienting from the weighted Sasakian 3-sphere.

Lemma 2.5. *Let $\boldsymbol{p}$ be admissible. Then quotient space of $\mu_{\eta_0}^{-1}(0) \approx S^3 \times S^3$ by the $T^2(\boldsymbol{p})$ action is the orbifold $\mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4)$. Moreover, the cohomology class in $H^2_{\mathrm{orb}}(\mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4), \mathbb{Z})$ of this orbibundle is the class of the Kähler form $\omega_{\boldsymbol{p}} = l\omega_{\bar{p}_1,\bar{p}_2} + k\omega_{\bar{p}_3,\bar{p}_4}$.*

Proof. The $T^2(\boldsymbol{p})$ action on $S^3 \times S^3$ splits as a weighted S^1 action on each factor. Setting $k = \gcd(p_1, p_2)$ and $l = \gcd(p_3, p_4)$, we see after reparametrizing that the quotient of the first factor is $\mathbb{CP}(\bar{p}_1, \bar{p}_2)$, and similarly for the second factor.

We have an exact sequence of groups

$$0 \to T(\boldsymbol{p}) \to T^2(\boldsymbol{p}) \to S^1(R_{\boldsymbol{p}}) \to 0,$$

where $S^1(R_{\boldsymbol{p}})$ is the circle generated by the Reeb vector field $R_{\boldsymbol{p}}$. Thus, we have the commutative diagram

$$\begin{tikzcd} S^3 \times S^3 \arrow[d, "\pi_1 \times \pi_2"'] \arrow[dr, "\rho"] & \\ \mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4) & M_{\boldsymbol{p}}. \arrow[l, "\pi"] \end{tikzcd} \tag{14}$$

We want to determine the integral orbifold first Chern class (Euler class) of the S^1 orbibundle given by π. That is, we look for the class $a\alpha + b\beta \in H^2_{\mathrm{orb}}(\mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4), \mathbb{Z})$ which transcends to the zero class on $M_{\boldsymbol{p}}$, where α and β are primitive classes in each factor. (See [Boyer and Galicki 2008, Chapter 4] for a discussion of these orbifold classes.) For this we take $\alpha = [\omega_{\bar{p}_1, \bar{p}_2}]$, and $\beta = [\omega_{\bar{p}_3, \bar{p}_4}]$. Now according to the action (7) the circle wraps around k times on the first factor and l times with the reverse orientation on the second. So if we take the Kähler form to be

$$\omega_{\boldsymbol{p}} = l\omega_{\bar{p}_1, \bar{p}_2} + k\omega_{\bar{p}_3, \bar{p}_4}, \tag{15}$$

its class pulls back to zero under π, since $\pi^*[\omega_{\bar{p}_1, \bar{p}_2}] = k\gamma$ and $\pi^*[\omega_{\bar{p}_3, \bar{p}_4}] = -l\gamma$, where γ is a generator of $H_2(M_{\boldsymbol{p}}, \mathbb{Z}) \approx \mathbb{Z}$. □

The first Chern class and diffeomorphism types. In this subsection we relate the diffeomorphism type of our manifolds M to the reduction parameters $\boldsymbol{p}$. We do this by giving a formula for the first Chern class of the contact bundle in terms of $\boldsymbol{p}$.

Theorem 2.6. *$M_{\boldsymbol{p}}$ is diffeomorphic to $S^2 \times S^3$ if $p_1 + p_2 - p_3 - p_4$ is even, and diffeomorphic to X_∞, the nontrivial S^3-bundle over S^2, if $p_1 + p_2 - p_3 - p_4$ is odd.*

Proof. We know from the reduction procedure and Lemma 2.2 that $M_{\boldsymbol{p}}$ is simply connected and $\pi_2(M_{\boldsymbol{p}}) = \mathbb{Z}$. So by the Barden–Smale classification of simply connected 5-manifolds $M_{\boldsymbol{p}}$ is determined by its second Stiefel–Whitney class $w_2(M)$. Moreover, since $TM_{\boldsymbol{p}}$ splits as $\mathcal{D}_{\boldsymbol{p}}$ plus a trivial line bundle, $w_2(M)$ is the mod 2 reduction of $c_1(\mathcal{D})$. So the theorem will follow immediately from the following lemma. □

Lemma 2.7. *The first Chern class of the contact bundle $\mathcal{D}_{\boldsymbol{p}} = \ker \eta_{\boldsymbol{p}}$ on $M_{\boldsymbol{p}}$ is given by*

$$c_1(\mathcal{D}_{\boldsymbol{p}}) = (p_1 + p_2 - p_3 - p_4)\gamma,$$

where γ is the positive generator of $H^2(M_{\boldsymbol{p}}, \mathbb{Z}) \approx \mathbb{Z}$.

Proof. We begin by computing the orbifold first Chern class of

$$\mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4).$$

From [Boyer and Galicki 2008, Chapter 4] we see that $p^* c_1^{\mathrm{orb}}$ is given by

$$(\bar{p}_1+\bar{p}_2)[\omega_{\bar{p}_1,\bar{p}_2}]+(\bar{p}_3+\bar{p}_4)[\omega_{\bar{p}_3,\bar{p}_4}] \in H^2_{\mathrm{orb}}(\mathbb{CP}(\bar{p}_1, \bar{p}_2)\times\mathbb{CP}(\bar{p}_3, \bar{p}_4), \mathbb{Z}), \tag{16}$$

which pulls back to the basic first Chern class $c_1(\mathcal{F}_{R_{\boldsymbol{p}}})$ in the basic cohomology group $H^2_B(\mathcal{F}_{R_{\boldsymbol{p}}})$ under the natural projection $\pi : M_{\boldsymbol{p}} \to \mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4)$ by the circle action of $R_{\boldsymbol{p}}$. Now we have an exact sequence

$$\begin{tikzcd}
 & & & 0 \arrow[d] & \\
 & & & H^2(M_{\boldsymbol{p}}, \mathbb{Z}) \arrow[d] & \\
0 \arrow[r] & H^0_B(\mathcal{F}_{R_{\boldsymbol{p}}}) \arrow[r, "\delta"] & H^2_B(\mathcal{F}_{R_{\boldsymbol{p}}}) \arrow[r, "\iota_*"] & H^2(M_{\boldsymbol{p}}, \mathbb{R}) \arrow[r] & \cdots
\end{tikzcd}$$

(see [Boyer and Galicki 2008, p. 245]), with $\iota_* c_1(\mathcal{F}_{R_{\boldsymbol{p}}}) = c_1(\mathcal{D}_{\boldsymbol{p}})_{\mathbb{R}}$ and $\delta a = a[d\eta_{\boldsymbol{p}}]_B$. So $c_1(\mathcal{D}_{\boldsymbol{p}})_{\mathbb{R}}$ is $c_1(\mathcal{F}_{R_{\boldsymbol{p}}}) \bmod [d\eta_{\boldsymbol{p}}]_B$, where $\eta_{\boldsymbol{p}}$ is the contact form on $M_{\boldsymbol{p}}$. Now since $\pi^*\omega_{\boldsymbol{p}} = d\eta_{\boldsymbol{p}}$, we know from the proof of Lemma 2.5 that $\pi^*[\omega_{\bar{p}_1,\bar{p}_2}] = k\gamma$ and $\pi^*[\omega_{\bar{p}_3,\bar{p}_4}] = -l\gamma$ holds over $\mathbb{Z}$. Thus, since $\pi_1(M_{\boldsymbol{p}}) = \{\mathbb{1}\}$ we have over $\mathbb{Z}$

$$\begin{aligned} c_1(\mathcal{D}_{\boldsymbol{p}}) &= (\bar{p}_1 + \bar{p}_2)\pi^*[\omega_{\bar{p}_1,\bar{p}_2}] + (\bar{p}_3 + \bar{p}_4)\pi^*[\omega_{\bar{p}_3,\bar{p}_4}] \\ &= k(\bar{p}_1 + \bar{p}_2)\gamma - l(\bar{p}_3 + \bar{p}_4)\gamma = (p_1 + p_2 - p_3 - p_4)\gamma. \qquad \square \end{aligned}$$

It is easy to see that the argument in [Lerman 2003b] for $S^2 \times S^3$ can be generalized to the nontrivial bundle X_∞ to give the following.

Proposition 2.8. *As a complex vector bundle, $\mathcal{D}_{\boldsymbol{p}}$ is determined uniquely by $p_1 + p_2 - p_3 - p_4$.*

3. Contact homology

Here we apply pseudoholomorphic curve theory as briefly described on page 282 to the Morse–Bott formulation of contact homology. The study of pseudoholomorphic curves in symplectic manifolds was initiated in the seminal paper by Gromov [1985]. Since then, these objects have become a basic tool in understanding symplectic geometry and topology, perhaps most notably in the work of Floer, which is the main motivation behind symplectic field theory (SFT) and contact homology. The latter is a small part of the larger SFT of Eliashberg, Givental, and Hofer [Eliashberg et al. 2000]. The original idea, inspired by Floer homology, was to create a homology theory from the chain complex generated by closed orbits of the Reeb vector field.

Given a contact manifold $(M, \mathcal{D})$, we choose a contact form η for $\mathcal{D}$, and an almost-complex structure J on the symplectization of M which extends the almost-complex structure on $\mathcal{D}$ such that the Reeb vector field is the purely imaginary direction. For the moment, assume that periodic orbits of the Reeb vector field are isolated. This is a generic property of contact forms, and one can always find such a Reeb field for any contact structure. We consider the set of all closed Reeb orbits. We consider two orbits as different if they have different periods, even if they geometrically trace out the same set. Orbits with period one are often called *simple* orbits. We will consider the chain complex whose generators are periodic orbits of the Reeb vector field. The grading is given by the Robbin–Salamon index, which in the case of isolated orbits is the same as the well-known Conley–Zehnder index [Robbin and Salamon 1993]. The differential is given by an algebraic count of rigid J-holomorphic curves from a twice-punctured two-sphere into the symplectization which are *asymptotically cylindrical* over closed Reeb orbits, that is, they are curves for which there exist polar coordinates about each puncture, such that for sufficiently small radius the curve behaves like a cylinder over a closed Reeb orbit. If we look at such a curve in standard coordinates in the symplectization, we call punctures which correspond to limits as the real coordinate approaches positive infinity positive punctures; the others are called negative punctures [Hofer et al. 1996; Eliashberg et al. 2000].

Both the Robbin–Salamon indices arise from the Maslov index for a path of symplectic matrices. We compute the index of a closed Reeb orbit as follows: first, let us assume that $H_1(M, \mathbb{Z}) = 0$ and consider a closed Reeb orbit γ together with an embedded Riemann surface $\Sigma \subset M$ such that $\partial\Sigma = \gamma$. To find the relevant path of symplectic matrices with which to compute the Maslov index, one then pulls back the contact bundle $\mathcal{D}$ to Σ, which then admits a trivialization, since it is a symplectic vector bundle over a Riemann surface with boundary. Then one considers the linearized Reeb flow about a Reeb orbit. This linearized flow gives the desired path of symplectic matrices. It is important to understand that in a contact manifold, these indices depend on the choice of capping disk used to trivialize $\mathcal{D}$. In particular, if the closed Reeb orbit γ is contractible (which is always the case in this article), one trivializes $\mathcal{D}$ by choosing a capping disk Σ of γ. If we consider another capping surface of the form $\Sigma' = \Sigma \# S_A$ where S_A represents a two-dimensional homology class A in M, then the Conley–Zehnder (and Robbin–Salamon) index of the orbit computed with Σ' will differ from that computed using Σ by twice the first Chern class of $\mathcal{D}$ evaluated on A, namely

$$\mu_{\mathrm{CZ}}(\gamma;\, \Sigma_\gamma \# S_A) = \mu_{\mathrm{CZ}}(\gamma;\, \Sigma_\gamma) + 2\langle c_1(\mathcal{D}), A\rangle. \tag{17}$$

Thus, the grading depends on the choice of capping surface.

In order to address this dependence one considers the coefficients for contact homology to be elements in a Novikov ring as follows. Give $H_2(M,\mathbb{Z})$ a grading $|\cdot|$ by setting $|A| = -2\langle c_1(\mathcal{D}), A\rangle$ for any $A \in H_2(M,\mathbb{Z})$. Let $\mathcal{R}$ be a submodule of $H_2(M,\mathbb{Z})$ with zero grading. Then the Novikov ring is the graded group ring $\mathbb{Q}[H_2(M,\mathbb{Z})/\mathcal{R}]$ whose elements are formal power series of the form $\sum_i q_i e^{A_i}$, where $q_i \in \mathbb{Q}$ and $A_i \in H_2(M,\mathbb{Z})/\mathcal{R}$. Here as usual the notation e^A is used to encode the multiplicative structure of a commutative ring with unit (see [McDuff and Salamon 2004, Chapter 11]).

There are some Reeb orbits for which the moduli space of holomorphic curves in $C(M)$ cannot be given a coherent orientation [Bourgeois and Mohnke 2004] — these "bad" Reeb orbits must be discarded. Let γ be a Reeb orbit with minimal period T, and γ_m be a Reeb orbit that covers γ with multiplicity m, so the period of γ_m is mT. The *bad* orbits are those for which the parity of the even multiples $|\gamma_{2m}|$ disagrees with the parity for the odd multiples $|\gamma_{2m-1}|$. A Reeb orbit that is not bad is said to be *good*.

Now that we have a grading, under favorable circumstances we can define a graded chain complex C_* generated by certain closed Reeb orbits with coefficients in the ring $\mathbb{Q}[H_2(M,\mathbb{Z})/\mathcal{R}]$.

Definition 3.1. We define C_* to be the graded chain complex freely generated by all good closed Reeb orbits with coefficients in the Novikov ring $\mathbb{Q}[H_2(M,\mathbb{Z})/\mathcal{R}]$. By convention, we shift all degrees by $n-2$, where $2n+1$ is the dimension of the contact manifold. The *contact homology*, denoted $HC(\mathcal{D})$, is the homology of the differential graded algebra C_* with differential given by (18).

The differential ∂ of this chain complex is given by an algebraic count of pseudoholomorphic curves in the symplectization $C(M)$ of M which come in one-dimensional families. Explicitly, for γ a good closed orbit of the Reeb vector field, M simply connected, and A a two-dimensional homology class, the differential is given by the formula

$$\partial\gamma = \sum_{A \in H_2(M,\mathbb{Z})} \sum_{\gamma'} \frac{1}{\kappa_\gamma} n_{\gamma,\gamma',A} e^{\bar{A}} \gamma', \tag{18}$$

where $\bar{A}$ denotes the image in $H_2(M,\mathbb{Z})/\mathcal{R}$ of the homology class A, κ_γ is the multiplicity of the Reeb orbit γ, and $n_{\gamma,\gamma'}$ is the algebraic count of elements in the moduli space $\mathcal{M}^A(\gamma,\gamma')$ of J-holomorphic curves into the symplectization of M

which are asymptotically cylindrical over the closed Reeb orbits γ, γ' representing the homology class A.

Note that $n_{\gamma,\gamma',A}$ is nonzero only if the dimension of this moduli space is 1. This indeed gives a reasonable homology. The proofs that $\partial^2 = 0$ and that the homology does not depend on choices of a contact form or an almost-complex structure come from analysis of the boundary of moduli spaces of rigid curves and are discussed in [Eliashberg et al. 2000]. These results, in general, depend on abstract transversality results for the $\bar{\partial}_J$ operator. We will make the standing assumption that such transversality can be achieved, either by abstract perturbations or by the amenable geometry of the situation at hand. The signs which appear in the algebraic count depend on coherent orientations of the moduli space are explained in [Bourgeois and Mohnke 2004].

Remark 3.2. Due to the lack of compactness of moduli spaces of pseudoholomorphic cylinders, ∂^2 is not always zero. If it is, then the homology is often called *cylindrical* contact homology. Indeed, the boundary of the compactification of this space can, in general, contain curves with more than two punctures. However, we can instead consider the supercommutative algebra generated by periodic orbits. This means that instead of counting only cylinders, we now count curves with an arbitrary number of negative punctures. In this paper, it suffices to count cylinders.

Morse–Bott contact homology. In the above constructions we needed to make an assumption that the closed Reeb orbits are isolated in order to get a good index, that is, we have to assume that the Poincaré return map constructed about any periodic Reeb orbit has no eigenvalue equal to 1. This condition is generic; however, many natural contact forms, especially those which arise from circle orbibundles, are as far from generic as possible. In order to calculate contact homology for such manifolds one must make some sort of perturbation. It is only in very nice situations that this is not extremely difficult. The Morse–Bott version [Eliashberg et al. 2000; Bourgeois 2002; 2003] allows us to use the symmetries of nice contact structures and symmetric almost-complex structures, by exploiting rather than excluding nonisolated orbits. This is accomplished by considering Morse theory on the quotient space, and relating critical points and gradient trajectories of a Morse function to pseudoholomorphic curves in the symplectization of the contact manifold. Since toric contact manifolds of Reeb type are always total spaces of circle orbibundles admitting Hamiltonian actions of tori and they admit nice Morse functions, the Morse–Bott formalism works quite well for us. We follow a combination of [Eliashberg et al. 2000] and [Bourgeois 2002] in what follows, applying the Morse–Bott setup to our special case.

Let $(M, \mathcal{D})$ be a contact manifold with contact form η, and consider the action functional $\mathfrak{A} : C^\infty(S^1; M) \to \mathbb{R}$, defined by

$$\mathfrak{A}(\gamma) = \int_\gamma \eta. \tag{19}$$

The critical points of $\mathfrak{A}$ are the closed orbits of the Reeb vector field of η. The *action spectrum* is defined to be

$$\sigma(\eta) = \{r \in \mathbb{R} \mid r = \mathfrak{A}(\gamma)\}$$

for γ a periodic orbit of the Reeb vector field. Given $T \in \sigma(\eta)$, define

$$N_T = \{p \in M \mid \phi_p^T = p\}, \quad S_T = N_T/S^1,$$

where S^1 acts on M via the Reeb flow. Then S_T is called the *orbit space* for period T.

When M is the total space of an S^1-orbibundle the orbit spaces are precisely the orbifold strata. This is a special case of a contact form being of Morse–Bott type:

Definition 3.3. A contact form, η is said to be of *Morse–Bott type* if

(i) The action spectrum $\sigma(\eta)$ is discrete.

(ii) The sets N_T are closed submanifolds of M, such that the rank of $d\eta|_{N_T}$ is locally constant and

$$T_p(N_T) = \ker(d\phi_T - I).$$

Remark 3.4. These conditions are the Morse–Bott analogues for the functional on the loop space of M.

Rather than set up Morse–Bott contact homology in full generality, let us do this for the special case of an S^1-orbibundle. In this case the contact form is of Morse–Bott type [Bourgeois 2002]. Let $T_1, \dots, T_m$ be all possible simple periods for closed Reeb orbits. Let ϕ_x^t denote the flow of the Reeb vector field. Let

$$N_{T_j} = \{x \in M \mid \phi_x^{T_j} = x\}, \quad S_{T_j} = N_{T_j}/S^1.$$

For each j, we choose a Morse function f_j on S_{T_j} and, using appropriate bump functions, build a Morse function f on all of M which descends under the quotient by the Reeb action to each orbit space. Now, we perturb η by

$$\eta_f = (1 + \epsilon f)\eta. \tag{20}$$

For almost all ϵ, the closed Reeb orbits of η_f are isolated, and, for bounded action, they correspond to critical points of f. Note that the Reeb orbits of η within each stratum all have the same Robbin–Salamon index. The following formula [Cieliebak et al. 1996; Bourgeois 2002] computes the Conley–Zehnder index of η_f in terms of the Robbin–Salamon index of any Reeb orbit in a particular orbit space:

$$|\gamma| = \mu_{\mathrm{CZ}}(\gamma) = \mu_{\mathrm{RS}}(\gamma_T) - \tfrac{1}{2}\dim(S_{T_j}) + \mathrm{ind}_p(f_j). \tag{21}$$

We now have a contact form with isolated closed Reeb orbits. Since these orbits correspond to critical points of a Morse function, we may think of the generators of contact homology either as isolated orbits, or as critical points of a Morse function on each orbit space. This gives the generators and their gradings for the chain complex (or differential graded algebra) of Morse–Bott contact homology when we add, by convention, the degree shift $n-2$, which is 0 in the five-dimensional case. We now describe the differential ∂; however, we shall be very brief, as the differential vanishes identically in our case, as proven in Theorem 3.9.

Consider orbit spaces S_{T_1} and S_{T_2}, for T_1, $T_2 \in \sigma(\eta)$. Now define the moduli space $\mathcal{M}_J(S_{T_1}, S_{T_2})$ to be the space of pseudoholomorphic curves (with respect to J) into $C(M)$ with one positive and one negative puncture which are asymptotically cylindrical over a closed Reeb orbit in S_{T_1} near the positive puncture, and to S_{T_2} near the negative puncture. We require, moreover, that these curves have finite, nonzero area.

In the case at hand the differential splits into two pieces:

$$\partial p = \partial_{\mathrm{MSW}}\, p + d_{\mathrm{CH}} S, \tag{22}$$

where ∂_{MSW} is the differential on the Morse–Smale–Witten complex determined by our choice of Morse function on the orbit space S containing p, and roughly speaking $d_{\mathrm{CH}} S$ gives a count of rigid pseudoholomorphic curves in $\mathcal{M}_J(S, S')$. Here S' is some orbit space with action less than that of S. The count is over all S' such that the dimension of $\mathcal{M}_J(S, S')$ is equal to 1. We refer to [Bourgeois 2002; Pati 2009] for more details. The following proposition shows the vanishing of d_{CH} for orbibundles.

Proposition 3.5. *When M is the total space of an orbibundle of a symplectic orbifold, then there are no rigid holomorphic curves into the symplectization of M.*

Proof. There is an effective $\mathbb{R}$-action, as well as that of a circle on the moduli spaces; hence, these spaces have dimension at least two. So they can never be rigid. □

Remark 3.6 (on transversality). Though in some special cases we can use the nice properties of toric manifolds to determine regularity of the moduli spaces of curves defined above, the proof of invariance of contact homology as well as the proof that the Morse-Bott complex actually computes the homology of the perturbed complex requires the use of abstract perturbations of the $\bar{\partial}_J$ operator. We believe that the results of Hofer, Wysocki, and Zehnder's polyfold theory will provide a good framework for this problem; however, we make it a standing assumption that there exists an abstract perturbation of the $\bar{\partial}_J$ operator which makes its linearization surjective. Proposition 3.5 requires no transversality result for the $\bar{\partial}_J$ operator since we can get at least these two dimensions without any appeal to abstract Fredholm theory. This does not make the transversality problem go away, however, since it is still needed in proofs of invariance, and independence of choices. Moreover, when one wishes to analyze higher-dimensional moduli spaces by adding marked points, one needs the relevant dimension formulae to hold, although this can be handled in many cases using the fact that J can be chosen to be integrable in these toric situations. We should also mention that, even without the transversality assumption mentioned here, we can obtain a weaker version of invariance, as we shall see later.

Contact homology for toric contact* 5*-manifolds. Let us consider the differential graded algebra discussed above. We start with the set of critical points of a Morse function as picked earlier. Since we are working with toric manifolds of Reeb type in dimension 5 we actually know that the fixed points of the T^3-action are isolated, hence the norm squared of the symplectic moment map on $\mathcal{Z}$ is a perfect Morse function.

We are interested in the orbit structure of the $T^3(\boldsymbol{p})$ action on $M_{\boldsymbol{p}}$.

Lemma 3.7. *Consider the toric contact structure $\mathcal{D}_{\boldsymbol{p}}$ on $M_{\boldsymbol{p}}$, an S^3-bundle over S^2. There are exactly four one-dimensional simple closed orbits under the action of $T^3(\boldsymbol{p})$. Moreover, these four orbits are Reeb orbits for all Reeb fields in the Sasaki cone $\mathfrak{t}_3^+(\boldsymbol{p})$, as well as for a Reeb vector field in $\mathfrak{con}(M_{\boldsymbol{p}}, \eta_{\boldsymbol{p}})$ that is arbitrarily close to one in the Sasaki cone. Moreover, for a generic such Reeb vector field these are the only closed orbits.*

Proof. We have an exact sequence of groups,

$$\{0\} \to T(\boldsymbol{p}) \to T^4 \to T^3(\boldsymbol{p}) \to \{0\},$$

and we consider the action of $T^3(\boldsymbol{p})$ on the level set given by (13) thought of as $T^4/T(\boldsymbol{p})$. If $z_1 \neq 0$, then we can choose $\theta = \theta_1$ of the standard T^4 angles. The remaining T^3 orbit will be one-dimensional only if $z_2 = 0$ and one of z_3 or z_4 is zero. This gives two closed S^1-orbits. On the other hand if $z_1 = 0$, then we must have $z_2 \neq 0$, so we choose $\theta = \theta_2$, and as above this gives exactly the two close orbits with either z_3 or z_4 vanishing. Clearly, any Reeb vector field in $\mathfrak{t}_3^+$ leaves these Reeb orbits invariant, and since every Reeb vector field in the Sasaki cone is arbitrarily close to a quasiregular one, the last statement follows from a result of Bourgeois [2002]. □

Let $\mathfrak{t}_2(\boldsymbol{p})$ denote the Lie algebra of $T^2(\boldsymbol{p})$. It is generated by the two vector fields $L_{\boldsymbol{p}}$ and $R_{\boldsymbol{p}}$. We have an exact sequence of Lie algebras

$$\{0\} \to \mathfrak{t}_2(\boldsymbol{p}) \to \mathfrak{t}_4 \xrightarrow{\rho} \mathfrak{g}_2(\boldsymbol{p}) \to \{0\}, \tag{23}$$

where $\mathfrak{g}_2(\boldsymbol{p})$ is generated by the vector fields $\overline{H}_1 = \rho(H_1)$ and $\overline{H}_3 = \rho(H_3)$. We have a toric symplectic orbifold

$$(\mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4), \omega_{\boldsymbol{p}}), \tag{24}$$

where the symplectic form is given by (15), and the torus $\mathfrak{G}_2(\boldsymbol{p})$ is generated by the Lie algebra $\mathfrak{g}_2(\boldsymbol{p})$. The moment map

$$\mu_2 : \mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4) \to \mathfrak{g}_2(\boldsymbol{p})^*$$

is given by $\mu_2(\boldsymbol{z}) = (|z_1|^2, |z_3|^2)$.

Proposition 3.8. *The function $f = |\mu_2|^2$ is a perfect Morse function on the quotient $M_{\boldsymbol{p}}/S^1 \approx \mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4)$ whose critical points are precisely the four Reeb orbits of Lemma 3.7.*

Proof. Since the critical points are isolated f is a Morse function, and Morse–Bott functions that are the norm squared of a moment map are perfect [Lerman and Tolman 1997]. It is easy to check directly (see also [Kirwan 1984, Lemma 3.1]), using the relations

$$H_1 \equiv a H_2 \mod \mathfrak{g}_2(\boldsymbol{p}), \quad H_3 \equiv b H_4 \mod \mathfrak{g}_2(\boldsymbol{p}),$$

for some $a, b \in \mathbb{R}$, that f has precisely the four critical points

$$[1,0] \times [1,0], \quad [1,0] \times [0,1], \quad [0,1] \times [1,0], \quad [0,1] \times [0,1],$$

and these correspond to the four Reeb orbits of Lemma 3.7. □

Theorem 3.9. *In the case of circle reductions in dimension five, which have four Reeb orbits fixed by the T^3-action, the differential in Morse–Bott contact homology vanishes. Moreover, the elements of contact homology $HC(\mathcal{D})$ are given by the good Reeb orbits including multiplicity. More precisely, it is given by the homology groups of each stratum of its orbit space. The degree of each generator is given by* (21).

Proof. By Proposition 3.5 there are no rigid holomorphic curves. So d_{CH} vanishes. But also by Proposition 3.8 $|\mu|_2^2$ is a perfect S^1-invariant Morse function, and the Morse–Smale–Witten differential ∂_{MSW} vanishes as well. Thus, the full differential (22) vanishes. It then follows that the elements of $HC(\mathcal{D})$ are simply the chains of the complex C_*, that is, good closed Reeb orbits including multiplicity. For each period of the Reeb flow we get a different Reeb orbit, corresponding to some critical point of f. Since f is perfect, these critical points correspond not just to chains but to actual homology classes. The statement about the grading follows from (21). □

The next proposition, though not a general proof of invariance of contact homology, does tell us that we do get an invariant in the world of S^1-orbibundles, whose bases admit a perfect Morse function.

Proposition 3.10. *Let M be a quasiregular contact manifold, such that its quotient by the Reeb vector field is a symplectic orbifold which admits a perfect Morse function. Then if M' is contactomorphic to M, is quasiregular, and the quotient by its Reeb vector field is also a symplectic orbifold which admits a perfect Morse function, then the two contact homology algebras are isomorphic.*

Proof. The conditions on M, M', and their bases ensure that all parts of the differential vanish. Therefore we may construct a map between these two algebras as in [Eliashberg et al. 2000] counting rigid curves in a symplectic cobordism between M and M'. The main difficulty is in seeing that this map is a chain map. However, since the differentials vanish on both ends, the map is trivially a chain map, hence the two contact homology algebras are isomorphic. □

The next proposition gives an nonequivalence statement about toric contact manifolds of type (p_1, p_2, l, l).

Proposition 3.11. *Let (p_1, p_2, l, l) and (p_1', p_2', l', l') be two admissible 4-tuples. If $p_1 + p_2 \neq p_1' + p_2'$, then the corresponding contact manifolds cannot be contactomorphic.*

The proof of Proposition 3.11 is essentially an index calculation in light of Theorem 3.9. Let us first collect some information about the contact structures in question in convenient coordinates. To compute the grading on contact homology it is useful to consider a special case of the *join* construction [Boyer et al. 2007]. Since we can view our toric sphere bundles as quotients of $S^3 \times S^3$ we have a convenient way to compute indices. This is of particular interest for strata of positive codimension, since the orbits in the codimension-zero stratum behave exactly as in the regular case. To define the join construction we start with two quasiregular contact manifolds, M_1 and M_2, with contact forms η_1 and η_2, and bases $\mathcal{Z}_1$ and $\mathcal{Z}_2$ with symplectic forms ω_1 and ω_2. Then the product $M_1 \times M_2$ is a T^2-bundle over $\mathcal{Z}_1 \times \mathcal{Z}_2$. We take the quotient of $M_1 \times M_2$ by the action of the circle obtained by gluing together Reeb orbits on each piece, that is,

$$(z, w) \mapsto (e^{ik_1\theta} z, e^{-ik_2\theta} w). \tag{25}$$

The admissibility conditions of Definition 2.4 are precisely the conditions that guarantee that the quotient by this action is smooth in which case it yields a new quasiregular contact manifold with base $\mathcal{Z}_1 \times \mathcal{Z}_2$, contact form $\eta_1 + \eta_2$, contact distribution given by $\mathcal{D}_1 \oplus \mathcal{D}_2$, and Reeb vector field $R_{\eta_1} + R_{\eta_2}$. This contact structure is exactly the one coming from the principal circle bundle obtained by requiring that its curvature form is the pullback of the sum of the two symplectic forms on each base space. We obtain new Reeb orbits as equivalence classes of pairs of Reeb orbits, one from each of M_1 and M_2. When k_1 and k_2 are different from 1, we have a similar contact manifold, except the curvature is given by pulling back k_1 and k_2 multiples of the symplectic forms, namely $d\alpha = \pi^*(k_2\omega_1 + k_1\omega_2)$. In this case Reeb orbits in the new total space will correspond to pairs, one wrapping k_1 and the other wrapping k_2 times (in addition to the multiplicity of the orbit as a Reeb orbit in one of the three-spheres). In the following, M_1 and M_2 are both standard three-spheres. Index calculations on three-dimensional spheres are standard; however, we present the details in Lemma 3.12 for completeness and also to illustrate the inherent role of the orbifold structure.

Let us consider the contact structure on the quotient of the product of two standard weighted three-spheres with weights p_1, p_2, p_3, and p_4. As before we take $k_1 = \gcd(p_1, p_2)$, $k_2 = \gcd(p_3, p_4)$, and $\bar{p}_i = p_i/k_1$ for $i = 1, 2$, and $\bar{p}_j = p_j/k_2$ for $j = 3, 4$. We view this as a product of hypersurfaces in $\mathbb{C}^4$ with coordinates (z_1, z_2, z_3, z_4), $z_j = x_j + iy_j$, subject to the action (25). This manifold is the total space of an orbibundle over an orbifold $S^2 \times S^2$ with orbifold singularities at the products of the north and south poles, and for the products of the north and south poles with copies of S^2. These singularities correspond to setting one or two of the z_j to 0. The Reeb vector field is given by

$$p_1y_1\partial_{x_1} - x_1p_1\partial_{y_1} + p_2y_2\partial_{x_2} - p_2x_2\partial_{y_4} + p_3y_3\partial_{x_3} - x_3p_3\partial_{y_3} + p_4y_4\partial_{x_4} - p_4x_4\partial_{y_4}$$

and the contact distribution is given by the span of the vectors

$$-\frac{1}{p_1}x_2\partial_{x_1} + \frac{1}{p_1}y_2\partial_{y_1} + \frac{1}{p_2}x_1\partial_{x_2} - \frac{1}{p_2}y_1\partial_{y_2}, \tag{26}$$

$$-\frac{1}{p_1}y_2\partial_{x_1} - \frac{1}{p_1}x_2\partial_{y_1} + \frac{1}{p_2}y_1\partial_{x_2} + \frac{1}{p_2}x_1\partial_{y_2}, \tag{27}$$

$$-\frac{1}{p_3}x_4\partial_{x_3} + \frac{1}{p_3}y_4\partial_{y_3} + \frac{1}{p_4}x_3\partial_{x_4} - \frac{1}{p_4}y_3\partial_{y_4}, \tag{28}$$

$$-\frac{1}{p_3}y_4\partial_{x_3} - \frac{1}{p_3}x_4\partial_{y_3} + \frac{1}{p_4}y_3\partial_{x_4} + \frac{1}{p_4}x_3\partial_{y_4}. \tag{29}$$

In the following we restrict ourselves to the case where $p_3 = p_4 = k_2$. To get our hands on an orbit in the quotient, we must, for each time around the fiber, pick an appropriate circle out of the fiber of the torus bundle. It is easy to see that the equivalence relation gives us a circle obtained by wrapping around the first circle k_2 times and around the second circle k_1 times. Let us now parametrize the fiber. We may choose a coordinate for a Reeb orbit by

$$\gamma(t) = (0, \cos(k_2\bar{p}_2t) + i\sin(k_2\bar{p}_2t), 0, \cos(p_2t) + i\sin(p_2t)).$$

Now when $t = 1/\bar{p}_2$ we have wrapped around the first orbit k_2-times and the second k_1 times. Here the action is $1/p_2$. This is the smallest action since we have assumed that $p_2 > p_1$. What about when the first orbit wraps around more than once in S^3? Let us see how to look at such an orbit. This corresponds to taking

$$\gamma(t) = \left(\cos(k_2\bar{p}_1 t) + i\sin(k_2\bar{p}_1 t), 0, 0, \cos\left(\frac{p_1}{m}t\right) + i\sin\left(\frac{p_1}{m}t\right)\right), \tag{30}$$

where m is the multiplicity. Now when $t = m/p_1$, we wrap around the first orbit mk_2 times and the second k_1 times. As long as $m < \min\{k_2, p_1\}$ we do not enter a higher-dimensional orbit space. Similar considerations remain true for $z_1 = 0$. Let us compute the Robbin–Salamon index of the orbits (30).

Lemma 3.12. *The Robbin–Salamon indices of the orbits in* (30) *of multiplicity m are given by*

$$2k_1 m + 2\lfloor mp_2/p_1\rfloor - 1, \quad 2k_1 m + 2\lfloor mp_1/p_2\rfloor - 1. \tag{31}$$

Proof. To do this we must choose a disk D with boundary γ. Such a disk can be written explicitly. We begin by producing a disk in $S^3 \times S^3$:

$$(\cos(\theta), \sin(\theta)e^{2\pi i k_2\bar{p}_2 t}, \cos(\theta), \sin(\theta)e^{2\pi i p_2 t/m}). \tag{32}$$

The above disk clearly has boundary γ (the boundary occurs when $\theta = \pi/2$) and we have $\theta \in [0, \pi/2]$. To pull back the contact distribution we plug the coordinates into (26)–(29):

$$-\frac{1}{p_1}\sin(\theta)\cos(2\pi i k_2\bar{p}_2 t)\partial_{x_1} + \frac{1}{p_1}\sin(\theta)\sin(2\pi i k_2\bar{p}_2 t)\partial_{y_1} + \frac{1}{p_2}\cos(\theta)\partial_{x_2}, \tag{33}$$

$$-\frac{1}{p_1}\sin(\theta)\sin(2\pi i k_2\bar{p}_2 t)\partial_{x_1} - \frac{1}{p_1}\sin(\theta)\cos(2\pi i k_2\bar{p}_2 t)\partial_{y_1} + \frac{1}{p_2}\cos(\theta)\partial_{y_2}, \tag{34}$$

$$-\frac{1}{p_3}\sin(\theta)\cos\left(2\pi i\frac{\bar{p}_2 t}{m}\right)\partial_{x_3} + \frac{1}{p_3}\sin(\theta)\sin\left(2\pi i\frac{\bar{p}_2 t}{m}\right)\partial_{y_3} + \frac{1}{p_4}\cos(\theta)\partial_{x_4}, \tag{35}$$

$$-\frac{1}{p_3}\sin(\theta)\sin\left(2\pi i\frac{\bar{p}_2 t}{m}\right)\partial_{x_3} - \frac{1}{p_3}\sin(\theta)\cos\left(2\pi i\frac{\bar{p}_2 t}{m}\right)\partial_{y_3} + \frac{1}{p_4}\cos(\theta)\partial_{y_4}. \tag{36}$$

When $\theta = \pi/2$, these four vectors become

$$\frac{1}{p_1}(-\cos(2\pi i k_2\bar{p}_2 t)\partial_{x_1} + \sin(2\pi i k_2\bar{p}_2 t)\partial_{y_1}),$$
$$\frac{1}{p_1}(-\sin(2\pi i k_2\bar{p}_2 t)\partial_{x_1} - \cos(2\pi i k_2\bar{p}_2 t)\partial_{y_1}),$$
$$\frac{1}{p_3}\left(-\cos\left(2\pi i\frac{p_2 t}{m}\right)\partial_{x_3} + \sin\left(2\pi i\frac{p_2 t}{m}\right)\partial_{y_3}\right),$$
$$\frac{1}{p_3}\left(-\sin\left(2\pi i\frac{p_2 t}{m}\right)\partial_{x_3} - \cos\left(2\pi i\frac{p_2 t}{m}\right)\partial_{y_3}\right).$$

Disks for the other orbits mapping into branch divisors have a similar expression. The key point is that we only see vectors corresponding to the coordinates which have been set to zero. Now we can easily compute the Robbin–Salamon index of these orbits. Recall that given a path of symplectic matrices, $\Phi(t)$, a number t is called a crossing if $\Phi(t)$ has an eigenvalue equal to 1. To compute the Robbin–Salamon index of a path of symplectic matrices on $[0, T]$ one computes

$$\tfrac{1}{2}\,\mathrm{signature}(\Gamma(0)) + \sum_{\substack{\text{crossings } t\\ t\neq 0,T}} \mathrm{signature}(\Gamma(t)) + \tfrac{1}{2}\,\mathrm{signature}(\Gamma(T)).$$

Here the crossing form is

$$\emptyset(\dot{\Phi}(t)v, v)$$

restricted to the subspace on which Φ has eigenvalues equal to 1. In this case at each crossing the crossing form is just $\emptyset(v, J_0 v)$, so this gives signature 2 on each two-dimensional subspace consisting of eigenvectors with eigenvalue 1. At crossings the vectors above spanning $\mathcal{D}$ above become $-(1/p_1)\partial x_1$, $-(1/p_1)\partial y_1$, $-(1/l)\partial x_3$, and $-(1/l)\partial y_3$.

Recall that the linearized Reeb flow is of the form

$$\begin{bmatrix} e^{2\pi i p_1 t} & 0 & 0 & 0 \\ 0 & e^{2\pi i p_2 t} & 0 & 0 \\ 0 & 0 & e^{2\pi i k_2 t} & 0 \\ 0 & 0 & 0 & e^{2\pi i k_2 t} \end{bmatrix},$$

and each complex block of the matrix looks like

$$\begin{bmatrix} \cos(2\pi p_j t) & -\sin(2\pi p_j t) \\ \sin(2\pi p_j t) & \cos(2\pi p_j t) \end{bmatrix}.$$

The time derivative of each block looks like

$$\begin{bmatrix} -2\pi p_j \sin(2\pi p_j t) & -2\pi p_j \cos(2\pi p_j t) \\ 2\pi p_j \cos(2\pi p_j t) & -2\pi p_j \sin(2\pi p_j t) \end{bmatrix}.$$

At crossings these blocks become

$$\begin{bmatrix} 0 & -2\pi p_j \\ 2\pi p_j & 0 \end{bmatrix}.$$

The crossings which have the first two vectors as 1-eigenvectors occur at integers multiples of $1/(k_2 p_1)$, and those for the second two occur at integer multiples of $k_1 m/k_2$. As we saw above the flow splits into two parts, that corresponding to the first two coordinates and that corresponding to the second two. This means the second part, for multiplicity m, is $2mk_1$. Now we add the normal part. For orbits of multiplicity m we get contribution $1+2\lfloor m/p_1\rfloor$. Therefore for multiplicity m these orbits have Robbin–Salamon index

$$2k_1m+2\lfloor mp_2/p_1\rfloor-1,$$

and similarly, setting p_2 to zero, we obtain

$$2k_1m+2\lfloor mp_1/p_2\rfloor-1. \qquad \square$$

We shall label these orbits $\gamma_{m,i}$. With this information we can now prove Proposition 3.11.

Proof of Proposition 3.11. The goal here is to distinguish contact structures. We will show that given two 4-tuples as in Proposition 3.11 that the contact homology algebras cannot be isomorphic. We know that $(p_2-1)/p_2 > (p_1-1)/p_1$; hence the index of γ is $2(p_1-1)$. This tells us that we have p_1-1+p_2-1 orbits of index less than $2(p_2-1)$.

Theorem 3.9 gives us a complete picture of the contact homology of the manifolds given by admissible 4-tuples up to knowing the Robbin–Salamon indices. Let us spell this out in the case (p_1, p_2, k_2, k_2). In this case there are essentially two different kinds of orbit spaces. We have two-dimensional orbit spaces which project to two-spheres in the base, and we have copies of the whole manifold. The two-dimensional orbits spaces consist of orbits having action k_2m/p_i for $p_i \nmid m$. The four-dimensional orbit spaces consist of orbits of integer action. For each two-dimensional orbit space, S_{k_2m/p_i}, we obtain exactly two orbits contributing to contact homology with grading difference two. We denote these orbits $\hat{\gamma}_{m,i}$, and $\check{\gamma}_{m_i}$ corresponding to the maximum and minimum of the Morse function on S_{k_2m/p_i}. For such orbits with action less than 1 we have grading

$$\begin{aligned} |\hat{\gamma}_{m,i}| &= \mu_{\mathrm{RS}}(\gamma_{m,i})+1 \\ |\check{\gamma}_{m,i}| &= \mu_{\mathrm{RS}}(\gamma_{m,i})-1. \end{aligned} \tag{37}$$

For each four-dimensional orbit space we have four generators for contact homology, again corresponding to critical points. We label these

$$\hat{\gamma}_m, \check{\gamma}_m, \gamma_m^{s_1}, \gamma_m^{s_2}$$

for the maximum, minimum, and two saddle points, respectively. With a choice of disk D projecting to the spherical homology class $\Sigma \in H_2(\mathcal{Z}, \mathbb{Q})$ we have

$$\begin{aligned} |\hat{\gamma}_m| &= \mu_{\mathrm{RS}}(\gamma_m, D)+2, & |\check{\gamma}_m| &= \mu_{\mathrm{RS}}(\gamma_m, D)-2, \\ |\gamma_m^{s_1}| &= \mu_{\mathrm{RS}}(\gamma_m, D), & |\gamma_m^{s_2}| &= \mu_{\mathrm{RS}}(\gamma_m, D). \end{aligned} \tag{38}$$

In (38) $\mu_{\mathrm{RS}}(\gamma_m, D) = 2k_2m\langle c_1^{\mathrm{orb}}(\mathcal{Z}), (\Sigma)\rangle$. Moreover, for two-dimensional orbit spaces with action greater than 1, by the catenation property of the Robbin–Salamon index, we may decompose the orbit into a part with biggest possible integer action and a part with action smaller than 1. We then add the indices of these two orbits to get the Robbin–Salamon index. Note that for these two-dimensional orbit spaces, the tangential part of the flow is a loop, but the normal does not complete a loop; this explains the appearance of the summand 1 in the above formulae. Note that, here, the Robbin–Salamon index is nondecreasing with respect to action. Thus we may count the number of orbits with index less than 1. This will give a count of generators of contact homology of index less than $2(p_1+p_2+2)-2$. From the above discussion there are p_1-1+p_2-1 such orbits coming from lower-dimensional orbit spaces, and then one coming from $\breve{g}_1$. This gives exactly p_1+p_2-1 orbits in degree less than $2(p_1+p_2+1)$. This proves Proposition 3.11. □

Proposition 3.11 applies directly to the $Y^{p,q}$ manifolds. In this case the invariant is $2p-1$; note that it does not depend on q.

As an application let us use the preceding discussion to distinguish contact structures on the toric contact 5-manifolds corresponding to the 4-tuple $(1, 2k-1, l, l)$ for positive integers k and l such that the tuple $(1, 2k-1, l, l)$ is admissible. Then we see that $c_1(\mathcal{D}) = 2k-2l$. Let us fix the first Chern class of the contact distribution and see what happens. We see then that we must have

$$k = \frac{c_1(\mathcal{D})+2l}{2}.$$

Now using Proposition 3.11 we see that there are $2k-1$ generators in contact homology of degree less than $4k+2$.

A remark on the regular case. In the regular case the situation is somewhat simpler, but, on the other hand, there is less information available at first glance. In this case there is geometrically only one orbit space, $\mathcal{Z}$ itself. To get a handle on the contact homology let us look at the case $(k, k, k-c, k-c)$. This gives a regular contact manifold with $c_1(\mathcal{D}) = 2c$ times a generator. We choose a basis L_1, L_2 of $H_2(\mathcal{Z}, \mathbb{Z})$ so that $L_1 = xS_1 + yS_2$, and L_2 lifts to a class which evaluates to 0 under $\pi^*\phi$. Both x and y are chosen so that they give action 1 for a disk that projects to L_1. We define x and y as follows. Let m be the smallest number so that $mk \equiv -1$ mod c. Then we define $x = (km+1)/c$ and $y = x-m$. It is easy to see that x and y satisfy the above properties. With these choices the grading for contact homology for orbits of action N is given by

$$|\hat{\gamma}| = N(2x+2y)+2, \quad |\check{\gamma}| = N(2x+2y)-2, \quad |\gamma_N^{s_j}| = N(2x+2y).$$

In this picture, for $N = 1$, $\check{\gamma}$ gives the smallest possible grading. By varying k, we obtain infinitely many distinct contact structures whose contact distribution has the same Chern class.

Another way to distinguish contact structures. As described in [Eliashberg et al. 2000, §2.9.2], there is another situation where symplectic field theory can be used to distinguish toric contact structures. The following theorem is a generalization to smooth orbifolds of [Eliashberg et al. 2000, Proposition 2.9.4]:

Theorem 3.13. *Suppose we have two simply connected quasiregular toric contact manifolds of Reeb type in dimension five, such that each orbifold stratum is nonsingular in the sense that its underlying space is a smooth submanifold. Suppose that under the quotient of the Reeb action one of the base manifolds has an exceptional sphere while the other does not, then these two manifolds are not contactomorphic.*

Proof. We show that there is an odd element in the contact homology algebra of one manifold specialized at a class which is not in the other for any specialization. We assume here that all of the weights of the torus action are greater than 1 for the manifold containing no exceptional spheres. As in [Eliashberg et al. 2000] the potential specialized to the Poincaré dual of an exceptional divisor will give the potential for a standard S^3; but then for a chain which lifts to the volume form for this 3-form there is always a holomorphic curve to kill it as a generator for homology specialized at this three-class. Hence this homology contains no odd elements. Let us consider first the case where the base is a manifold. We look at the manifold containing no exceptional sphere. We must compute the Gromov–Witten potential (see the Appendix for a brief description). Unfortunately it does not vanish, but for *any* 2-classes the potential always vanishes. This is because the Gromov–Witten invariant, $\mathrm{GW}^0_{A,k}(\alpha, \dots, \alpha)$, is not equal to 0 for a two-dimensional class α only if

$$2k = 4 + 2c_1(A) + 2k - 6, \quad \text{i.e., } c_1(A) = 1.$$

But the weights make this impossible. Thus all coefficients for such curves vanish, and the potential vanishes on $\mathcal{Z}$, hence on M. So for a three-class in the contact manifold obtained from integration over the fiber of a two-class, there is no holomorphic curve to kill it. Hence specialized at such a three-class we have an odd generator which does not exist in the presence of exceptional spheres. The orbifold case is similar. The computation for the Gromov–Witten potential on the manifold with the exceptional sphere follows from the divisor axiom. To see that the coefficients for the Gromov–Witten potential vanish in the case where there is no exceptional sphere we note the Gromov–Witten invariant is nonzero only if the first Chern class evaluated on A is equal to one minus the degree shifting number of $\boldsymbol{x}$, which in the absence of exceptional spheres in the stratum in question is impossible. □

Remark 3.14. Since the base is four dimensional the results of [Hofer et al. 1997] tell us that we can indeed use the dimension formula above for computation of the Gromov–Witten invariants for the manifold case. To adjust for orbifold structure we use extra point conditions since all strata in this case are actually smooth manifolds additionally endowed with orbifold structure.

4. The contact equivalence problem

It is the purpose of this section to prove the contact equivalence of certain toric contact structures that are inequivalent as toric contact structures. To clarify this we say that two toric contact structures $\mathcal{D}_{\boldsymbol{p}}$ and $\mathcal{D}_{\boldsymbol{p}'}$ are equivalent as contact structures, but inequivalent as toric contact structures, if there is a contactomorphism $\varphi : M_{\boldsymbol{p}} \to M_{\boldsymbol{p}'}$ such that $\varphi_* \mathcal{D}_{\boldsymbol{p}} = \mathcal{D}_{\boldsymbol{p}'}$, but there is no T^3-equivariant contactomorphism. Then their 3-tori correspond to distinct conjugacy classes of maximal tori in the contactomorphism group [Lerman 2003b; Boyer 2013]. We remark that Theorems 1 and 2 are direct consequences of Theorem 4.11.

A complete answer to the equivalence problem appears to be quite difficult so we restrict ourselves to certain special cases of contact structures that are Seifert S^1-bundles (orbibundles) over Hirzebruch surfaces which generally have a nontrivial orbifold structure. In this case we show that certain T^3 equivariantly inequivalent contact structures are actually T^2 equivariantly equivalent for some subgroup $T^2 \subset T^3$.

Orbifold Hirzebruch surfaces. In this section we study a special class of toric contact structures on S^3 bundles over S^2 that can be realized as circle orbibundles over orbifold Hirzebruch surfaces. Since the reduction method gives all examples of such toric contact structures, it is important to make contact (no pun intended) with examples that are known in the literature. Here we shall always assume that the quadruple $\boldsymbol{p}$ is admissible.

When working with Hirzebruch surfaces, we often follow [Griffiths and Harris 1978] (but with slightly different notation) and represent S_n as the projectivized bundle $S_n = \mathbb{P}(\mathcal{O}(n) + \mathcal{O}) \to \mathbb{CP}^1$ with fibers $L = \mathbb{CP}^1$ and sections E and F with self-intersection numbers n and $-n$, respectively. The sections, which satisfy $E \cdot E = n$, $E \cdot L = 1$, and $L \cdot L = 0$, define divisors in S_n and determine a basis for the Picard group $\mathrm{Pic}(S_n) \approx H^2(S_n, \mathbb{Z}) \approx \mathbb{Z}^2$. However, when working with symplectic forms it is convenient to use a basis which appears for all admissible complex structures. Thus, it is convenient to treat the even and odd Hirzebruch surfaces separately. The even Hirzebruch surfaces S_{2n} are diffeomorphic to $S^2 \times S^2$, so we define $E_0 = E - nL$. Then we have

$$E_0 \cdot E_0 = (E - nL) \cdot (E - nL) = E \cdot E - 2nE \cdot L + L \cdot L = 2n - 2n = 0.$$

In this case the Poincaré duals α_L and α_{E_0} are the standard area forms for the two copies of S^2. Similarly, the odd Hirzebruch surfaces S_{2n+1} are diffeomorphic to $\mathbb{CP}^2$ blown up at a point which we denote by $\widetilde{\mathbb{CP}}^2$. In this case we define

$E_{-1} = E-(n+1)L$ which gives $E_{-1} \cdot E_{-1} = -1$. So E_{-1} is an exceptional divisor. Again the Poincaré duals α_L and $\alpha_{E_{-1}}$ represent the standard area forms on a fiber and exceptional divisor, respectively.

As mentioned previously the orbifolds that we encounter are of the form (X, Δ), where X is a smooth algebraic variety and Δ is a branch divisor. Specifically we are interested in the orbifolds (S_n, Δ), where $\Delta = \sum_i (1 - 1/m_j) D_j$ and D_i are Weil divisors on S_n. We refer to the pair (S_n, Δ) as an *orbifold Hirzebruch surface*. We now wish to compute the orbifold canonical divisor in this situation. Since we are working with $\mathbb{Q}$-divisors, we can express the result in terms of E_0 even though it is an honest divisor only on even Hirzebruch surfaces.

Lemma 4.1. *Let (S_n, Δ) be an orbifold Hirzebruch surface such that E and F are branch divisors, both with ramification index m. Then the orbifold canonical divisor of (S_n, Δ_m) is*

$$K^{\mathrm{orb}}_{(S_n,\Delta_m)} = -\frac{2}{m}E - \frac{2m-n}{m}L = -\frac{2}{m}E_0 - 2L.$$

Hence (S_n, Δ_m) is a log del Pezzo surface (Fano) if and only if $2m > n$.

Proof. We know [Griffiths and Harris 1978, p. 519] that the canonical divisor K_{S_n} of S_n is given by $K_{S_n} = -2E + (n-2)L = -2E_0 - 2L$, and the orbifold canonical divisor $K^{\mathrm{orb}}_{S_n,\Delta_m}$ satisfies (see [Boyer and Galicki 2008, p. 127])

$$K^{\mathrm{orb}}_{S_n,\Delta_m} = K_{S_n} + \Big(1 - \frac{1}{m}\Big)(E+F).$$

Now the divisor E has self-intersection n, and the divisor F has self-intersection $-n$, and since they both have intersection 1 with the fiber L, we have $(E+F) = 2E_0$. Putting this together gives the formula.

The orbifold (S_n, Δ_m) is log del Pezzo if and only if the orbifold anticanonical divisor $-K^{\mathrm{orb}}_{(S_n,\Delta_m)}$ is ample, and this happens if and only if $2m > n$ by Nakai's criterion since E and L are effective. □

Toric contact structures on $S^2 \times S^3$. The toric contact structures we describe here are not the most general, but are obtained by setting $\boldsymbol{p} = (j, 2k-j, l, l)$. That is, we consider contact structures of the form $\mathcal{D}_{j,2k-j,l,l}$ where the pair (k, l) is fixed with $k \geq l$, and $j = 1, \ldots, k$. Now since $\boldsymbol{p} \in \mathcal{A}$ we also have $\gcd(j, l) = \gcd(2k-j, l) = 1$. We denote the set of $j = 1, \ldots, k$ such that $\boldsymbol{p} = (j, 2k-j, l, l)$ is admissible by $\mathfrak{J}_{\mathcal{A}} = \mathfrak{J}_{\mathcal{A}}(k, l)$. The first Chern class of this contact structure is $c_1(\mathcal{D}_{j,2k-j,l,l}) = 2(k-l)\gamma$ where γ is a generator of $H^2(M_{\boldsymbol{p}}, \mathbb{Z}) \approx \mathbb{Z}$. So in this case $M_{\boldsymbol{p}}$ is $S^2 \times S^3$. The infinitesimal generator of the circle action is $L_{\boldsymbol{p}} = jH_1 + (2k-j)H_2 - lH_3 - lH_4$. Note that this case includes the $Y^{p,q}$ as a special case, namely, $p = k = l$ and $q = k - j$. So $Y^{p,q}$ is $\mathcal{D}_{p-q,p+q,p,p}$ with $p > q$ and $\gcd(p, q) = 1$.

We want to find a suitable Reeb vector field in the Sasaki cone, so we try $R_{j,k,l}=(2k-j)H_1+jH_2+lH_3+lH_4$ which clearly satisfies the positivity condition $\eta_0(R)>0$. The T^2 action generated by $L_{\boldsymbol{p}}$ and $R_{j,k,l}$ is

$$z\mapsto (e^{i((2k-j)\phi+j\theta)}z_1,\, e^{i(j\phi+(2k-j)\theta)}z_2,\, e^{il(\phi-\theta)}z_3,\, e^{il(\phi-\theta)}z_4).$$

Making the substitutions $\psi=\phi-\theta$ and $\chi=j\psi+2k\theta$ gives the action

$$z\mapsto (e^{i(2(k-j)\psi+\chi)}z_1,\, e^{i\chi}z_2,\, e^{il\psi}z_3,\, e^{il\psi}z_4). \tag{39}$$

We define $g_j=\gcd(l,2(k-j))$ and write $2(k-j)=n_jg_j$ and $l=m_jg_j$. Then $\gcd(m_j,n_j)=1$.

Theorem 4.2. *Consider the contact manifold $(S^2\times S^3,\mathcal{D}_{j,2k-j,l,l})$ where $1\le j\le k$ satisfies $\gcd(j,l)=\gcd(2k-j,l)=1$. Then we have*

(1) *The quotient space by the circle action generated by the Reeb vector field $R=(2k-j)H_1+jH_2+lH_3+lH_4$ is the Kähler orbifold $(S_{n_j},\Delta;\omega_{k,l,j})$ where S_{n_j} is a Hirzebruch surface, Δ is the branch divisor,*

$$\Delta=\Big(1-\frac{1}{m_j}\Big)(E+F), \tag{40}$$

and $\omega_{k,l,j}$ is an orbifold symplectic form satisfying $\pi^*\omega_{k,l,j}=d\eta_{k,l,j}$, where $\eta_{k,l,j}$ is the contact 1-form representing $\mathcal{D}_{j,2k-j,l,l}$ whose Reeb vector field is R.

(2) *The orbifold structure is trivial ($\Delta=\varnothing$) if and only if l divides $2(k-j)$.*

Proof. For (1) the idea, in the spirit of GIT quotient equals symplectic quotient [Kirwan 1984; Ness 1984], is to identify the symplectic quotient $\mu^{-1}(0)/T^2$ with a Hirzebruch surface as an analytic subspace of $\mathbb{CP}^1\times\mathbb{CP}^2$.

After shifting by a constant vector $\boldsymbol{a}=(a_1,a_2)$ the moment map of the T^2 action (39) is

$$\mu(z)=(2(k-j)|z_1|^2+l|z_3|^2+l|z_4|^2-a_1,\, |z_1|^2+|z_2|^2-a_2). \tag{41}$$

We need to choose the constant vector $\boldsymbol{a}$ so that 0 is a regular value of μ for all integers j and l such that $0<j\le k$ and $0<l\le k$. Alternatively, it suffices to show that the T^2 action on $\mu^{-1}(0)$ defined by (39) is locally free. This will be true if we choose $a_1,a_2>0$ and $a_1>2(k-j)a_2$. Following [Audin 1994] it is convenient to work with the corresponding $\mathbb{C}^*\times\mathbb{C}^*$ action on $\mathbb{C}^2\setminus\{0\}\times\mathbb{C}^2\setminus\{0\}$ given by

$$z\mapsto (\tau^{n_j}\zeta z_1,\, \zeta z_2,\, \tau^{m_j}z_3,\, \tau^{m_j}z_4), \tag{42}$$

where $\tau,\zeta\in\mathbb{C}^*$. From this we see that the action is free if $z_1z_2\neq 0$ and locally free with isotropy group $\mathbb{Z}_{m_j}$ on the two divisors obtained by setting $z_1=0$ and

$z_2 = 0$, respectively. It is not difficult to see [Lafontaine 1981] that $(z_1 = 0) = E$ and $(z_2 = 0) = F$.

We have a commutative diagram

$$\begin{tikzcd} \mathbb{C}^2 \setminus \{0\} \times \mathbb{C}^2 \setminus \{0\} \arrow[ddr, "\pi''"] & \mu^{-1}(0) \arrow[l, hook'] \arrow[dd, "\pi'"] \arrow[dr, "\rho"] & \\ & & M_{k,l,j}, \arrow[dl, "\pi"] \\ & \mathcal{Z}_{k,l,j} & \end{tikzcd} \tag{43}$$

and we want to identify the quotient space $\mathcal{Z}_{k,l,j}$. We know from the general theory (see [Boyer and Galicki 2008, Chapter 7]) that $\mathcal{Z}_{k,l,j}$ is a projective algebraic orbifold with an orbifold Kähler structure. Viewing $\mathcal{Z}_{k,l,j}$ as the $\mathbb{C}^* \times \mathbb{C}^*$ quotient by the map π'' of diagram (43), we can identify $\mathcal{Z}_{k,l,j}$ with a subvariety of $\mathbb{CP}^1 \times \mathbb{CP}^2$ as in [Hirzebruch 1951; Lafontaine 1981] as follows. If we define homogeneous coordinates in $\mathbb{CP}^1 \times \mathbb{CP}^2$ by setting $(w_1, w_2) = (z_3, z_4)$ and $(y_1, y_2, y_3) = (z_2^{m_j} z_3^{n_j}, z_2^{m_j} z_4^{n_j}, z_1^{m_j})$, we see that $\mathcal{Z}_{k,l,j}$ is represented by the equation

$$w_1^{n_j} y_2 = w_2^{n_j} y_1. \tag{44}$$

As an algebraic variety this identifies $\mathcal{Z}_{k,l,j}$ with the hypersurface in $\mathbb{CP}^1 \times \mathbb{CP}^2$ defined by (44) which is the original definition of the Hirzebruch surface S_{n_j}. However, the two divisors in S_{n_j} defined by $E = (y_3 = z_1^{m_j} = 0)$ and $F = (z_2 = 0(y_1 = y_2 = 0))$ are both m_j-fold branch covers with isotropy group $\mathbb{Z}_{m_j}$. Thus, we have an orbifold structure on S_{n_j} given by (40) which is trivial if and only if $m_j = 1$ which happens if and only if l divides $2(k-j)$.

Furthermore, it follows from the orbifold Boothby–Wang theorem [Boyer and Galicki 2000b] that $\mathcal{Z}_{k,l,j}$ has an orbifold Kähler form $\omega_{k,l,j}$ that satisfies $\pi^*\omega_{k,l,j} = d\eta_{k,l,j}$. This proves (1). For (2) we note that the orbifold structure is trivial if and only if $m_j = 1$ which happens if and only if $g_j = l$ divides $2(k-j)$. □

Notice that on subsets of $\mathfrak{J}_{\mathcal{A}}(k,l)$ where g_j is independent of j, the ramification index m_j is also independent of j, so the underlying orbifolds are the same. Thus, it is convenient to view g_j as a map $g : \mathfrak{J}_{\mathcal{A}}(k,l) \to \{1, \dots, l\}$, and we are interested in the level sets of this map. So we decompose $\mathfrak{J}_{\mathcal{A}}(k,l)$ into the level sets of g and then further decompose the level sets according to whether n_j is odd or even, that is, we define

$$g^{-1}(i)_{\text{even}} = \{j \in \mathfrak{J}_{\mathcal{A}}(k,l) \mid g_j = i,\ n_j \text{ is even}\},$$
$$g^{-1}(i)_{\text{odd}} = \{j \in \mathfrak{J}_{\mathcal{A}}(k,l) \mid g_j = i,\ n_j \text{ is odd}\}.$$

We can then decompose the admissible set as a disjoint union

$$\mathcal{J}_{\mathcal{A}}(k,l) = \bigsqcup_{i=1}^{l} g^{-1}(i)_{\text{even}} \sqcup g^{-1}(i)_{\text{odd}}. \tag{45}$$

We wish to compute the symplectic form on the base orbifold.

Lemma 4.3. *Let $j \in g^{-1}(i)_{\text{even}} \subset \mathcal{J}_{\mathcal{A}}(k,l)$ for a fixed $i \in \{1,\dots,l\}$ with $g^{-1}(i)_{\text{even}} \neq \varnothing$. Then the symplectic form $\omega_{k,l,i}$ on the quotient (S_{n_j}, Δ_i) is independent of j and satisfies*

$$[\omega_{k,l,i}] = i\alpha_{E_0} + k\alpha_L.$$

Proof. We know that $c_1^{\text{orb}}(S_{n_j}) = (2/m_i)\alpha_{E_0} + 2\alpha_L$ and this must pull back to $2(k-l)\gamma$. So $\pi^*\alpha_{E_0} = m_i k\gamma$ and $\pi^*\alpha_L = -l\gamma$. Now the class $[\omega_{k,l,i}]$ must transcend to 0 on $S^2 \times S^3$. So writing $[\omega] = a\alpha_{E_0} + b\alpha_L$ and using $im_i = l$, we see that

$$0 = a\pi^*\alpha_E + b\pi^*\alpha_L = (am_ik - bl)\gamma = m_i(ak - bi)\gamma.$$

So taking $a = i$ and $b = k$ gives the result. □

Lemma 4.4. *Let $j \in g^{-1}(i)_{\text{odd}} \subset \mathcal{J}_{\mathcal{A}}(k,l)$ for a fixed $i \in \{1,\dots,l\}$ with $g^{-1}(i)_{\text{odd}} \neq \varnothing$. Then i is even and the symplectic form $\omega_{k,l,i}$ on the quotient (S_{n_j}, Δ_i) is independent of j and satisfies*

$$[\omega_{k,l,i}] = i\alpha_{E_{-1}} + \Big(k + \frac{i}{2}\Big)\alpha_L.$$

Proof. First i must be even since $i = g_j = \gcd(l, 2(k-j))$ and $n_j = 2(k-j)/i$ is odd. The remainder of the proof is the same as that of Lemma 4.3, except for odd Hirzebruch surfaces we express the symplectic class in term of the exceptional divisor $E_{-1} = E_0 - \frac{1}{2}L$. □

Whenever possible we would like to determine the cardinalities $\#g^{-1}(i)_{\text{even}}$ and $\#g^{-1}(i)_{\text{odd}}$. First, as seen above, $g^{-1}(i)_{\text{odd}}$ is empty when i is odd. Moreover, if $g^{-1}(l)$ is not empty, then $\Delta_l = \varnothing$, so the orbifold structure is trivial.

The following lemma is taken from [Karshon 2003].

Lemma 4.5. (1) $\#g^{-1}(l)_{\text{even}} = \Big\lceil \dfrac{k}{l} \Big\rceil$. (2) $\#g^{-1}(l)_{\text{odd}} = \Big\lceil \dfrac{2k-l}{2l} \Big\rceil$.

Example 4.6. One obtains the $Y^{p,q}$ of [Gauntlett et al. 2004a] as a special case of Theorem 4.2 by putting $k = l = p$ and defining $q = k - j$. The contact structures are then $\mathcal{D}_{p-q,p+q,p,p}$, and the admissibility conditions boil down to $\gcd(q,p) = 1$. Clearly, we have $c_1(\mathcal{D}_{p-q,p+q,p,p}) = 0$. When p is odd we have $g_j = 1$, $m_j = p$, and $S_{n_j} = S_{2q}$, whereas if p is even we have $g_j = 2$, $m_j = p/2$ and $S_{n_j} = S_q$ with q

odd. So here we have only two nonempty level sets of the map g, namely,

$$\mathfrak{J}_{\mathcal{A}}(p,p)=\begin{cases} g^{-1}(1)_{\text{even}}, & \text{if } p \text{ is odd,}\\ g^{-1}(2)_{\text{odd}} & \text{if } p \text{ is even.}\end{cases}\tag{46}$$

Since the only admissibility condition is $\gcd(q,p)=1$ and $p>q$, we see that the cardinality $\#\mathfrak{J}_{\mathcal{A}}(p,p)=\phi(p)$, where ϕ is the well-known Euler phi function. For the orbifold canonical divisor Lemma 4.1 gives

$$K^{\text{orb}}_{(S_{2q},\Delta)}=-\frac{2}{p}E-\frac{2(p-q)}{p}L=-\frac{2}{p}E_0-2L,\tag{47}$$

$$K^{\text{orb}}_{(S_q,\Delta)}=-\frac{4}{p}E-\frac{2(p-q)}{p}L=-\frac{4}{p}E_{-1}-2\frac{p+1}{p}L,\tag{48}$$

so these are all log del Pezzo surfaces. The cohomology class of the corresponding symplectic forms is

$$[\omega_{p,p,1}]=\alpha_{E_0}+p\alpha_L$$

on the even orbifold Hirzebruch surface (S_{2q},Δ) with ramification index $m_1=p$ when p is odd. For even p we have

$$[\omega_{p,p,2}]=2\alpha_{E_{-1}}+(p+1)\alpha_L$$

on the odd orbifold Hirzebruch surface (S_q,Δ) with ramification index $m_2=p/2$. Note that in both cases there are precisely $\phi(p)$ values taken on by q. Note also that $p=2$ implies $q=1$ only, and that $m_2=1$, so we have a trivial orbifold structure on $(S_1,\varnothing)=\widetilde{\mathbb{CP}^2}$. A relation between the $Y^{p,q}$ toric contact structures and Hirzebruch surfaces was noted by Abreu [2010].

Except for the $Y^{p,q}$ case of Example 4.6, we do not have a general formula for the cardinalities $\#g^{-1}(i)$ for $i\neq l$. Specific cases, of course, are easy to work out.

Example 4.7. Consider the case $(k,l)=(9,8)$. We compute $\mathfrak{J}_{\mathcal{A}}(9,8)$. The possible values of j are all odd with $j\le 9$, and these all satisfy $\gcd(8,18-j)=1$. Next, we determine $g_j=\gcd(8,2(9-j))$ and $n_j=2(9-j)/g$. So we have $g^{-1}(8)_{\text{even}}=\{j=1,9\}$ with a trivial orbifold ($m_8=1$) on the Hirzebruch surfaces S_2 and S_0, respectively. We also have $g^{-1}(8)_{\text{odd}}=\{j=5\}$ with a trivial orbifold on S_1, and $g^{-1}(4)_{\text{odd}}=\{j=3,7\}$ with $m_4=2$ on the odd Hirzebruch surfaces S_3 and S_1, respectively. Notice that the cardinalities of $g^{-1}(8)_{\text{even}}$ and $g^{-1}(8)_{\text{odd}}$ agree with Lemma 4.5. In total we have $\#\mathfrak{J}_{\mathcal{A}}(9,8)=5$.

Toric contact structures on X_∞. For X_∞ we consider $\boldsymbol{p}=(j,2k-j+1,l,l)$ with $0<j\le k$. Here we have $c_1(\mathcal{D}_{j,2k-j+1,l,l})=\big(2(k-l)+1\big)\gamma$. We consider the Reeb vector field $R^\infty_{j,k,l}=(2k-j+1)H_1+jH_2+lH_3+lH_4$, which is clearly positive. The

T^2 action generated by this vector field and $L_{\boldsymbol{p}} = jH_1 + (2k-j+1)H_2 - lH_3 - lH_4$ is

$$z \mapsto (e^{i((2k-j+1)\phi+j\theta)}z_1, e^{i(j\phi+(2k-j+1)\theta)}z_2, e^{il(\phi-\theta)}z_3, e^{il(\phi-\theta)}z_4).$$

Making the substitutions $\psi = \phi - \theta$ and $\chi = j\psi + (2k+1)\theta$ gives the action

$$z \mapsto (e^{(i(2k-2j+1)\psi+\chi)}z_1, e^{i\chi}z_2, e^{il\psi}z_3, e^{il\psi}z_4). \tag{49}$$

Similarly to the previous section, we define $g_j = \gcd(l, 2k-2j+1)$ and write $2k-2j+1 = n_j g_j$ and $l = m_j g_j$. Then $\gcd(m_j, n_j) = 1$.

Theorem 4.8. *Consider the contact manifold* $(X_\infty, \mathcal{D}_{j,2k-j+1,l,l})$ *where* $1 \le j \le k$ *satisfies* $\gcd(j,l) = \gcd(2k-j+1, l) = 1$. *Then we have*

(1) *The quotient space by the circle action generated by the Reeb vector field* $R = (2k-j+1)H_1 + jH_2 + lH_3 + lH_4$ *is the Kähler orbifold* $(S_{n_j}, \Delta; \omega_{k,l,j})$ *where* S_{n_j} *is an odd Hirzebruch surface,* Δ *is the branch divisor, with*

$$\Delta = \Big(1 - \frac{1}{m_j}\Big)(z_1 = 0) + \Big(1 - \frac{1}{m_j}\Big)(z_2 = 0), \tag{50}$$

and $\omega_{k,l,j}$ *is an orbifold symplectic form satisfying* $\pi^*\omega_{k,l,j} = d\eta_{k,l,j}$, *where* $\eta_{k,l,j}$ *is the contact* 1*-form representing* $\mathcal{D}_{j,2k-j+1,l,l}$ *whose Reeb vector field is* R. *Here the integers* l, g_j, n_j, *and* m_j *are all odd.*

(2) *The orbifold structure is trivial* ($\Delta = \varnothing$) *if and only if* l *divides* $2k-j+1$.

Proof. The proof is essentially the same as that of Theorem 4.2. The details are left to the reader. □

We denote the set of $j = 1, \dots, k$ such that $\boldsymbol{p} = (j, 2k-j+1, l, l)$ is admissible by $\mathcal{J}^\infty_{\mathcal{A}} = \mathcal{J}^\infty_{\mathcal{A}}(k,l)$. Since the integers g_j and n_j are both odd for all $j \in \mathcal{J}_{\mathcal{A}}(k,l)$, the map g maps the set $\mathcal{J}_{\mathcal{A}}(k,l)$ to the set of positive odd integers less than or equal to l. Thus, we have

$$\mathcal{J}^\infty_{\mathcal{A}}(k,l) = \bigsqcup_{\text{odd } i=1}^{l} g^{-1}(i). \tag{51}$$

Similarly to Lemmas 4.3, 4.4, and 4.5 we find the following.

Lemma 4.9. *Let* $j \in g^{-1}(i) \subset \mathcal{J}^\infty_{\mathcal{A}}(k,l)$. *Then the symplectic form* $\omega_{k,l,i}$ *on the quotient* S_{n_j} *is independent of* j *and satisfies*

$$[\omega_{k,l,i}] = i\alpha_{E_{-1}} + \Big(k + \frac{i+1}{2}\Big)\alpha_L.$$

Furthermore, $\#g^{-1}(l) = \lceil (2k-l+1)/2l \rceil$.

Example 4.10. We consider the analogue on X_∞ of Example 4.6, so $(k, l) = (p, p)$ with p odd and $j = p - q$. The contact structure is $\mathcal{D}_{p-q,p+q+1,p,p}$ with $c_1(\mathcal{D}_{p-q,p+q+1,p,p}) = 1$. The admissibility conditions are $\gcd(q, p) = 1 = \gcd(q + 1, p)$. The function g satisfies $g_j = \gcd(2q + 1, p)$. If p is prime then the set of admissible q is $\{1, \ldots, p-2\}$, and $g = 1$ except when $q = (p-1)/2$ in which case $g = p$. The latter is smooth and corresponds to the trivial orbifold $(S_0, \varnothing)$. This has symplectic class

$$[\omega_{p,p}] = p\alpha_{E_{-1}} + \frac{3p+1}{2}\alpha_L,$$

whereas the $p-3$ elements in $g^{-1}(1)$ have symplectic class

$$[\omega_{p,p}] = \alpha_{E_{-1}} + (p+1)\alpha_L.$$

For a case when p is not prime consider $p = 9$. Then $\mathcal{J}^\infty_{\mathcal{A}}(9, 9) = \{2, 5, 8\}$ which corresponds to $q = 7, 4, 1$. We see that $g^{-1}(9) = \{5\}$ giving the smooth first Hirzebruch surface $(S_1, \varnothing)$ with symplectic class $[\omega_{9,9,9}] = 9\alpha_{E_{-1}} + 14\alpha_L$, while $g^{-1}(3) = \{2, 8\}$ giving the orbifold Hirzebruch surfaces (S_5, Δ) with $m_2 = 3$, and (S_1, Δ) with $m_8 = 3$, respectively, with symplectic class $[\omega_{9,9,3}] = 3\alpha_{E_{-1}} + 11\alpha_L$.

Equivalent contact structures. Here we show that certain inequivalent toric contact structures are equivalent as contact structures. The proof uses the fact that the identity map $(S_n, \Delta_m) \to (S_n, \varnothing)$ is a Galois cover, and combines this with the work of Karshon [1999; 2003] for the smooth case.

Theorem 4.11. *Consider the toric contact structures $(S^2 \times S^3, \mathcal{D}_{j,2k-j,l,l})$ and $(X_\infty, \mathcal{D}_{j,2k-j+1,l,l})$ of Theorems 4.2 and 4.8, respectively.*

(1) *For each fixed $1 \le i \le l$, the contact structures $\mathcal{D}_{j,2k-j,l,l}$ are T^2-equivariantly isomorphic for all $j \in g^{-1}(i)_{\mathrm{even}}$, and the contactomorphism group*

$$\mathfrak{Con}(\mathcal{D}_{j,2k-j,l,l})$$

has at least $\# g^{-1}(i)_{\mathrm{even}}$ conjugacy classes of maximal tori of dimension three.

(2) *For each fixed $1 \le i \le l$, the contact structures $\mathcal{D}_{j,2k-j,l,l}$ are T^2-equivariantly isomorphic for all $j \in g^{-1}(i)_{\mathrm{odd}}$, and the contactomorphism group*

$$\mathfrak{Con}(\mathcal{D}_{j,2k-j,l,l})$$

has at least $\# g^{-1}(i)_{\mathrm{odd}}$ conjugacy classes of maximal tori of dimension three.

(3) *For each fixed $1 \le i \le l$, the contact structures $\mathcal{D}_{j,2k-j+1,l,l}$ are T^2-equivariantly isomorphic for all $j \in g^{-1}(i) \subset \mathcal{J}^\infty_{\mathcal{A}}(k, l)$, and the contactomorphism group $\mathfrak{Con}(\mathcal{D}_{j,2k-j+1,l,l})$ has at least $\# g^{-1}(i)$ conjugacy classes of maximal tori of dimension three.*

(4) *The T^2-equivariantly isomorphic contact structures given in* (1)–(3) *are not T^3-equivariantly isomorphic.*

Proof. The proofs of (1)–(3) are quite analogous, so we give the details for (1) only. By Theorem 4.2 the contact structure is the orbifold Boothby–Wang construction over the symplectic orbifold $(S_{n_j}, \omega_{k,l,j})$ and by Lemma 4.3 the form $\omega_{k,l,j}$ only depends on $g_j = i$. Then for each $j \in g^{-1}(i)_{\text{even}}$ we consider the Galois cover $\mathbb{1}_{n_j} : (S_{n_j}, \Delta_{m_i}) \to (S_{n_j}, \varnothing)$ with n_j even, and both spaces having the same Kähler form, namely $\omega_{k,l,i}$ of Lemma 4.3. Now Karshon [2003] shows that $(S_{n_j}, \varnothing)$ and $(S_{n_{j'}}, \varnothing)$ are S^1-equivariantly symplectomorphic with the same symplectic form $\omega_{k,l,i}$ (but not the same Kähler structure) as long as $0 \le j' = j - 2r$ for some nonnegative integer r. We denote such a symplectomorphism by K. Now we have a commutative diagram,

$$
\begin{array}{ccc}
(S_{n_j}, \Delta_{m_i}) & \xrightarrow{K_i} & (S_{n_{j'}}, \Delta_{m_i}) \\
\Big\downarrow \mathbb{1}_{n_j} & & \mathbb{1}_{n_{j'}}^{-1} \Big\uparrow \\
(S_{n_j}, \varnothing) & \xrightarrow{K} & (S_{n_{j'}}, \varnothing),
\end{array}
\tag{52}
$$

which defines the upper horizontal arrow K_i and shows that it too is an S^1-equivariant symplectomorphism. We claim that K_i is also an orbifold diffeomorphism. This follows from Lemma 4.12. But then, as shown in [Lerman 2003b; Boyer 2013], this symplectomorphism lifts to a T^2-equivariant contactomorphism.

Here, and hereafter, by $g^{-1}(i)$ we mean any of the three sets $g^{-1}(i)_{\text{even}}$, $g^{-1}(i)_{\text{odd}}$, or $g^{-1}(i) \subset \mathfrak{J}^{\infty}_{\mathcal{A}}(k, l)$. Since our contact structures are independent of $j \in g^{-1}(i)$ up to isomorphism, we now denote the contact structures of items (1), (2), and (3) by $\mathcal{D}_{k,l,i,e}$, $\mathcal{D}_{k,l,i,o}$, and $\mathcal{D}_{k,l,i,\infty}$, respectively. To prove (4) we first notice that as in [Karshon 2003] the orbifold symplectomorphism K_i is only S^1-equivariant, not T^2-equivariant. So the corresponding 2-tori belong to different conjugacy classes in the group $\mathfrak{Ham}((\mathcal{B}, \Delta_{m_i}), \omega_{k,l,i})$ of Hamiltonian symplectomorphisms, where $\mathcal{B}$ is the symplectic orbifold $((S^2 \times S^2, \Delta_{m_i})\omega_{k,l,i})$ or $((X_\infty, \Delta_{m_i})\omega_{k,l,i})$ as the case may be. But then by [Boyer 2013, Theorem 6.4] these lift to nonconjugate 3-tori in $\mathfrak{Con}(\mathcal{D}_{k,l,i,e})$. Hence, the contact structures $\mathcal{D}_{j,2k-j,l,l}$ for different $j \in g^{-1}(i)$ are inequivalent as toric contact structures. The same holds for $\mathcal{D}_{j,2k-j+1,l,l}$. □

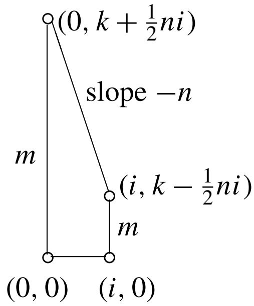
Lemma 4.12. *The Karshon symplectomorphism K of diagram* (52) *leaves the divisors* $(z_1 = 0)$ *and* $(z_2 = 0)$ *invariant.*

Proof. The T^2 action on any orbifold Hirzebruch surface (S_n, Δ_m) can be taken as

$$
([w_1, w_2] \times [y_1, y_2, y_3]) \mapsto ([\tau w_1, w_2] \times [\tau^n y_1, y_2, \rho y_3]),
\tag{53}
$$

where, as in the proof of Theorem 4.2, the coordinates are $(w_1, w_2) = (z_3, z_4)$ and $(y_1, y_2, y_3) = (z_2^m z_3^n, z_2^m z_4^n, z_1^m)$. By Proposition 4.1 of [Karshon 1999] K is an S^1-equivariant symplectomorphism, where the S^1 is that generated by ρ, and the fixed point set of this action is the disjoint union $(z_1 = 0) \sqcup (z_2 = 0)$. By Proposition 4.3 of the same work, K also intertwines the two S^1 moment maps. But in both cases these are represented by $\mu_S(z) = |z_1|^2$. So the divisors $(z_1 = 0)$ and $(z_2 = 0)$ are left invariant separately by K. □

Remark 4.13. Using the Delzant theorem for symplectic orbifolds in [Lerman and Tolman 1997], it is straightforward to construct the labeled polytope corresponding to the symplectic orbifold $((S_n, \Delta_m), \omega_{k,l,i})$. It is the labeled Hirzebruch trapezoid shown here (with label m on the two vertical axes):



The Galois cover $1\!\!1_{S_n} : (S_n, \Delta_m) \to (S_n, \varnothing)$ induces a map on this Hirzebruch trapezoid that simply removes the labels on the vertical edges. This implies that the corresponding Karshon graphs [1999] are the same. Hence, Theorem 4.1 of [Karshon 1999] easily generalizes to the types of orbifolds considered here, and the symplectomorphism K_i in diagram (52) can be constructed directly from this.

Inequivalence of contact structures. As discussed previously the inequivalence of contact structures is detected first by the first Chern class $c_1(\mathcal{D})$ and then by contact homology. The contact structures $\mathcal{D}_{j,2k-j,l,l}$ and $\mathcal{D}_{j,2k-j+1,l,l}$ are clearly inequivalent since they live on different manifolds, so adopting Proposition 3.11 to our current notation we have

Theorem 4.14. *The contact structures $\mathcal{D}_{j,2k-j,l,l}$ and $\mathcal{D}_{j',2k'-j',l',l'}$, and $\mathcal{D}_{j,2k-j+1,l,l}$ and $\mathcal{D}_{j',2k'-j'+1,l',l'}$ are inequivalent if $k' \neq k$.*

Remark 4.15. Unfortunately, combining Theorems 4.11 and 4.14 does not answer our equivalence problem completely even in our restrictive cases. For example, it would be nice to know that the $\mathcal{D}_{j,2k-j,l,l}$ are all contact equivalent as j runs through all admissible values from 1 to k. However, our equivalence statement in Theorem 4.11 only assures equivalence on the level sets of the map g, that is, if we fix $i \in \{1, \dots, l\}$, then $\mathcal{D}_{j,2k-j,l,l}$ are equivalent for all $j \in g^{-1}(i)$. Nevertheless, this is enough to give a complete answer to the equivalence problem for the $Y^{p,q}$ of [Gauntlett et al. 2004a] which in our notation is $\mathcal{D}_{p-q,p+q,p,p}$. See Corollary 5.5.

The general case with vanishing first Chern class $c_1(\mathcal{D})$ was studied in [Cvetič et al. 2005; Martelli and Sparks 2005] where it was shown that all the toric contact structures admit a compatible Sasaki–Einstein metric. These depend on three parameters a, b, and c with values in $\mathbb{Z}^+$ which in our notation is given by $\mathcal{D}_{a,b,c,a+b-c}$. Except for the subclass $Y^{p,q}$ these fall outside of the scope of our analysis.

Another statement of contact inequivalence is obtained as an immediate consequence of Theorem 3.13, namely:

Corollary 4.16. *The contact structures on $S^2 \times S^3$ described by items* (1) *and* (2) *of Theorem 4.11 are inequivalent.*

5. Applications to Sasakian geometry

In this final section we give some pertinent applications of our results to Sasakian geometry. Recall the contact Delzant-type result [Boyer and Galicki 2000a] that every toric contact structure of Reeb type admits a compatible Sasakian structure.

Sasaki cones and the Sasaki bouquet. Since we are dealing with toric geometry, the Sasaki cones in this paper all have dimension three. So it follows from [Boyer and Galicki 2008, Theorem 8.1.14] that all our Sasakian structures must be either positive or indefinite. Our first result says that all our Sasaki cones have a positive Sasakian structure.

Corollary 5.1. *Every toric contact structure on an S^3-bundle over S^2 can be realized as an orbifold fibration over $\mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4)$ for some quadruple of positive integers (p_1, p_2, p_3, p_4) satisfying* $\gcd(p_i, p_j) = 1$ *for $i = 1, 2$ and $j = 3, 4$ and $(p_1, p_2, p_3, p_4) = (k\bar{p}_1, k\bar{p}_2, l\bar{p}_3, l\bar{p}_4)$. Thus, every toric contact structure on an S^3-bundle over S^2 admits a ray of positive Sasakian structures in its Sasaki cone. Moreover, the subspace of positive Sasakian structures is open in the Sasaki cone.*

Proof. The first statement follows from Lemmas 2.2 and 2.5, while the second statement follows from the fact that the base orbifold $\mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4)$ is log del Pezzo. The last statement follows from the fact that having a positive representative of the basic first Chern class $c_1(\mathcal{F}_R)$ is an open condition. □

Proposition 5.2. *Consider the toric contact structure $\mathcal{D}_{j,2k-j,l,l}$ on $S^2 \times S^3$ and the toric contact structure $\mathcal{D}_{j,2k-j+1,l,l}$ on X_∞. Choose the Reeb vector field $R = (2k-j)H_1 + jH_2 + lH_3 + lH_4$ on $(S^2 \times S^3, \mathcal{D}_{j,2k-j+1,l,l})$ and $R = (2k-j+1)H_1 + jH_2 + lH_3 + lH_4$ on $(X_\infty, \mathcal{D}_{j,2k-j+1,l,l})$. Then the corresponding Sasakian structure is positive if and only if $l > k - j$ in the former case, and $2l > 2(k-j)+1$ in the latter case. In particular, if $l \leq k - j$ (or $2l > 2(k-j)+1$), then the toric contact structure $\mathcal{D}_{j,2k-j,l,l}$ (or $\mathcal{D}_{j,2k-j+1,l,l}$) has both a positive and indefinite Sasakian structure in its Sasaki cone.*

Proof. First we note as mentioned above that any Sasakian structure in a Sasaki cone of dimension greater than one is either positive or indefinite. By Corollary 5.1 any toric contact structure on an S^3-bundle over S^2 has a positive Sasakian structure, and we see that Lemma 4.1 states that (S_{n_j}, Δ_{m_j}) is log del Pezzo if and only if $2m > n$. For $\mathcal{D}_{j,2k-j,l,l}$ this becomes $2l = 2m_j g_j > n_j g_j = 2(k-j)$, whereas for $\mathcal{D}_{j,2k-j+1,l,l}$ it is $2l = 2m_j g_j > n g_j = 2k - 2j + 1$. The last result then follows from the fact that a quasiregular Sasakian structure is positive if and only if its orbifold quotient is log Fano [Boyer and Galicki 2008, Chapter 7]. □

Concerning Sasaki bouquets we note that actually more is true than is proven in Theorem 4.11. Since the base orbifolds (S_{n_j}, Δ_{m_i}) have the same symplectic form $\omega_{k,l,i}$ for all $j \in g^{-1}(i)$, this lifts to a contactomorphism $\varphi : \mathcal{D}_{j,2k-j,l,l} \to \mathcal{D}_{j',2k-j',l,l}$ such that $\varphi^*\eta_{j',2k-j',l,l} = \eta_{j,2k-j,l,l}$. Since the reduction process carries a preferred complex structure J along with it, the different indices j represent different transverse complex structures J. So by using the contactomorphisms φ for each admissible j, we have the following.

Corollary 5.3. *The contact structures $\mathcal{D}_{k,l,i,e}$ and $\mathcal{D}_{k,l,i,o}$ on $S^2 \times S^3$ and $\mathcal{D}_{k,l,i,\infty}$ on X_∞ each admit a Sasaki bouquet $\mathfrak{B}_N$ of toric Sasakian structures with $N = \#g^{-1}(i)$. Furthermore, the intersection $\bigcap_j \kappa(\mathcal{D}, J_j)$ of all the Sasaki cones is an open subset of the Lie algebra $\mathfrak{t}_2$ of a two-dimensional torus.*

Example 5.4. Consider the contact structure $\mathcal{D}_{12,8,2,o}$ on $S^2 \times S^3$. The admissible j's are $j = 1, 3, 5, 7$ and $g_j = 2$ for all j. So Theorem 4.11(ii) gives a T^2-equivariant contact equivalence of the T^3-equivariantly inequivalent toric contact structures

$$\mathcal{D}_{1,23,8,8} \approx \mathcal{D}_{3,21,8,8} \approx \mathcal{D}_{5,19,8,8} \approx \mathcal{D}_{7,17,8,8}.$$

This implies that the number of conjugacy classes $\mathfrak{n}(\mathcal{D}_{12,8,2,o}, 3)$ of 3-tori in $\mathfrak{Con}(\mathcal{D}_{12,8,2,o})$ is at least 4. Furthermore, by Theorem 4.2(3) the induced Sasakian metrics are positive for $\mathcal{D}_{5,19,8,8}$ and $\mathcal{D}_{7,17,8,8}$, whereas they are indefinite for the remaining two.

Another contact structure with the same first Chern class as $\mathcal{D}_{12,8,2,o}$, namely $c_1 = 8\gamma$, is $\mathcal{D}_{14,10,2,o}$. This consists of the two T^3-equivariantly inequivalent toric contact structures $\mathcal{D}_{1,27,10,10}$ and $\mathcal{D}_{7,21,10,10}$, but only for the latter is the induced Sasakian structure positive. In this case we have $\mathfrak{n}(\mathcal{D}_{14,10,2,o}, 3) \geq 2$. Moreover, it follows from Theorem 4.14 that the contact structures $\mathcal{D}_{12,8,2,o}$ and $\mathcal{D}_{14,10,2,o}$ are inequivalent.

As mentioned previously there is one subclass of contact structures on $S^2 \times S^3$ where a complete solution to the equivalence problem can be obtained, and they are all known to admit extremal (actually Sasaki–Einstein) metrics.

Corollary 5.5. *The contact structures $Y^{p,q}$ and $Y^{p',q'}$ on $S^2 \times S^3$ are inequivalent if and only if $p' \neq p$. Furthermore, the isotopy class of the contact structures defined by $Y^{p,1}$ admits a $\phi(p)$-bouquet $\mathfrak{B}_{\phi(p)}(Y^{p,q})$ such that each of the $\phi(p)$ Sasaki cones admits a unique Sasaki–Einstein metric. Moreover, these Einstein metrics are nonisometric as Riemannian metrics.*

Proof. Applying Theorem 4.11 to Example 4.6 shows that $Y^{p,q}$ is contactomorphic to $Y^{p,1}$ for all admissible q. But Abreu and Macarini [2012] show that the underlying contact structures of $Y^{p,1}$ and $Y^{p',1}$ are inequivalent if $p' \neq p$. (This also follows from Proposition 3.11.) By Corollary 5.3 there are precisely $\phi(p)$ Sasaki cones in the bouquet. The fact that there is a Sasaki–Einstein metric in the Sasaki cone for each $Y^{p,q}$ was first shown in [Gauntlett et al. 2004a] while its uniqueness in the Sasaki cone is proved in [Cho et al. 2008].

To prove the last statement, suppose to the contrary that the Sasaki–Einstein metrics g_q and $g_{q'}$ are isometric, that is, there is a diffeomorphism ψ of $S^2 \times S^3$ such that $\psi^* g_{q'} = g_q$. Then by a theorem of Tanno [1970] (see also [Boyer and Galicki 2008, Lemma 8.1.17]) the transformed Sasakian structure $\mathcal{S}^{\psi}$ is either $\mathcal{S}_q$ itself or its conjugate Sasakian structure $\mathcal{S}^c_q = (-R_q, -\eta_q, -\Phi_q, g_q)$. In either case ψ is a contactomorphism from $Y^{p,q}$ to $Y^{p,q'}$ satisfying

$$\psi^{-1} \circ \Phi_{q'} \circ \psi = \pm \Phi_q.$$

But this implies that the corresponding 3-tori are conjugate, which contradicts Theorem 4.11(4). □

Example 5.6. The analogues of the $Y^{p,q}$'s on the nontrivial bundle X_∞ are described in Example 4.10. For simplicity we consider only the case when p is an odd prime, in which case there are $p-3$ admissible values for q, namely $1, \dots, (p-1)/2-1, (p-1)/2+1, \dots, p-2$. These inequivalent toric structures are T^2-equivariantly equivalent contact structures by Theorem 4.11(3) and their induced Sasakian structures are all positive by Theorem 4.8(3). Moreover, the contactomorphism group of this contact structure has at least $p-3$ maximal tori of dimension three.

Some remarks concerning extremal Sasakian structures. As with Kähler geometry it is of interest to determine the most preferred Sasakian metrics, and as in Kähler geometry it seems reasonable to study the critical points of the (now transverse) Calabi functional [Boyer et al. 2008; 2009]. In [Boyer 2011b] the first author described bouquets of extremal Sasakian structures on S^3-bundles over S^2, and the existence of extremal Sasakian metrics on X_∞ was proven. It is not our intention here to delve much further into the existence of such extremal Sasakian structures, but rather to discuss briefly their relation to our current work.

Corollary 5.3 gives a partial generalization of Theorems 4.1 and 4.2 of [Boyer

2011b]. In this reference it was shown that when the quotient by the Reeb vector field is a smooth manifold, each Sasaki cone in a bouquet admits an extremal Sasakian metric. This follows from well-known work of Calabi. It would be interesting to generalize this to the orbifold case by generalizing the method of [Ghigi and Kollár 2007] to extremal metrics.

As with toric symplectic structures, all toric contact structures of Reeb type admit a compatible Sasakian metric [Boyer and Galicki 2000a]. Furthermore, in our present situation we have:

Corollary 5.7. *Every toric contact structure on an S^3 bundle over S^2 admits extremal Sasakian metrics with positive Ricci curvature.*

Proof. By Corollary 5.1 every toric contact structure on an S^3 bundle over S^2 can be realized as an orbifold fibration over a product of weighted projective spaces $\mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4)$ and have positive Sasakian structures. By a result of Bryant [2001] all weighted projective spaces admit Bochner-flat metrics and these are extremal [David and Gauduchon 2006], and the product of extremal Kähler metrics is extremal. So these extremal Kähler orbifold metrics lift to extremal Sasakian metrics [Boyer et al. 2008] which, since $\mathbb{CP}(\bar{p}_1, \bar{p}_2) \times \mathbb{CP}(\bar{p}_3, \bar{p}_4)$ is log del Pezzo, will have a deformation to a Sasakian metric with positive Ricci curvature by [Boyer and Galicki 2008, Theorem 7.5.31]. Moreover, it follows from a theorem of Calabi [1985] that the toric symmetry is retained by these metrics. □

Corollary 5.7 implies that each Sasaki cone in every Sasaki bouquet $\mathfrak{B}_N$ of toric contact structures on an S^3 bundle over S^2 admits extremal Sasakian metrics of positive Ricci curvature. Since the moment cone of any S^3 bundle over S^2 has exactly four facets, recent results of Legendre [2011a; 2011b] show that every toric contact structure on an S^3 bundle over S^2 admits at least one and at most seven distinct rays in the Sasaki cone consisting of Sasakian structures whose metrics have constant scalar curvature. Moreover, she shows that for the Wang–Ziller manifolds $M^{1,1}_{k,l}$ with $k > 5l$ there exist two distinct rays in the Sasaki cone whose Sasakian metrics have constant scalar curvature. This corresponds to the case $(\bar{p}_1, \bar{p}_2) = 1 = (\bar{p}_3, \bar{p}_4)$ of Lemma 2.5.

An interesting question which appears to be unanswered at this time is whether any Sasaki cones on these toric contact structures are exhausted by extremal Sasaki metrics. There are only a few known cases where this occurs, namely, the standard CR structure on the spheres S^{2n+1} [Boyer et al. 2008], the Heisenberg group [Boyer 2009], and T^2-invariant contact structures of Reeb type on S^3-bundles over Riemann surfaces [Boyer and Tønnesen-Friedman 2014; 2013] with genus $1 \leq g \leq 4$. However, when $g = 0$ we suspect that by using the admissible construction method of [Boyer and Tønnesen-Friedman 2014] the subclass of Sasaki structures considered here will each have an extremal representative.

Appendix: Orbifold Gromov–Witten invariants

In this appendix, for the convenience of the reader, we lay out some framework and definitions for Gromov–Witten invariants and the so-called Gromov–Witten potential for compact symplectic manifolds and orbifolds. In this paper we only consider the genus-0 invariants. The Gromov–Witten invariants that we are interested in occur in the base orbifold $\mathcal{Z}$ of an orbibundle $\pi : M \to \mathcal{Z}$ with $\dim(M) = 5$. Hence we are in the semipositive case and we can define the Gromov–Witten invariants as in [McDuff and Salamon 2004]. Our version of Gromov–Witten theory for symplectic orbifolds comes from [Chen and Ruan 2002]. The main difference here is that our marked points, and hence our cohomology classes taken as arguments for the invariant, have constraints determining in which orbifold stratum the curves in question lie. This is an issue since generally some homology classes may live in several strata.

Roughly speaking a Gromov–Witten invariant is a count of rigid J-holomorphic curves representing a homology class $A \in H_2(M, \mathbb{Z})/(\text{torsion})$ in general position with marked points in a symplectic manifold M for which the marked points are mapped into the Poincaré duals of certain cohomology classes. For example we may ask how many spheres (or lines) intersect two generic points in $\mathbb{CP}^n$. In this case we have two marked points, a top cohomology class, and for A the class of a line, $[L]$.

To make this precise let (M, ω) be a compact symplectic manifold, and let J be an ω-compatible almost-complex structure. Consider the moduli space

$$\mathcal{M}^A_{0,k}(M, J)$$

of genus-0 stable J-holomorphic curves into M representing the class A and assume here that we have regularity of the relevant linearized Cauchy–Riemann operator for the class A, either via some circumstances or by some sort of abstract perturbation argument. Note also that when we discuss Gromov–Witten theory for compact symplectic manifolds we will consider only *somewhere injective* curves. We define maps

$$ev_j : \mathcal{M}^A_{0,k}(M, J) \to M \quad \text{and} \quad ev : \mathcal{M}^A_{0,k}(M, J) \to M^{\times k}$$

by evaluation at the marked points.

By semipositivity the evaluation map represents a submanifold of $M^{\times k}$ of dimension

$$2n + \langle 2c_1(M), A\rangle + 2k + 6.$$

Now we define the Gromov–Witten invariant as a homomorphism

$$\mathrm{GW}^M_{A,k} : H^*(M)^{\otimes k} \otimes H_*(\mathcal{M}^A_{0,k}(M.J)) \to \mathbb{Q}$$

encoded formally as the integral

$$\mathrm{GW}^M_{A,k}(\alpha_1,\dots,\alpha_k) := \int_{\mathcal{M}^A_{0,k}(M.J)} ev_1^*\alpha_1 \cup \cdots \cup ev_k^*\alpha_k \cup \pi^*[\mathcal{M}^A_{0,k}(M.J)].$$

This is the definition for manifolds. This definition can be used without the semipositivity condition as long as there is a construction of an appropriate object on which to integrate. Since we will be working in dimension four this will not be an issue.

To extend this definition to orbifolds, there are issues with the definitions of J-holomorphic curves, since the idea of a map between orbifolds can be a rather sticky issue. We content ourselves here with knowing that we have a notion of a *good* map, and we will defer to [Chen and Ruan 2002; 2004] for the analytic setup. With that said, we still must extend the definition above so that it makes sense in a stratified space. We should also note that the orbifold cohomology of Chen and Ruan is *not* the same as the orbifold cohomology mentioned earlier. This cohomology is simply a way to organize how various classes interact with the stratification of the orbifold. As in the manifold case we start with a compact symplectic orbifold, $\mathcal{Z}$, and pick a compatible almost-complex structure, J. We then consider moduli spaces of (genus-0) J-holomorphic orbicurves into M representing a homology class $A \in H_2(\mathcal{Z},\mathbb{Q})$. But we now need to consider a new piece of data which organizes the intersection data so that it is compatible with the stratification. The extra data will be defined by a k-tuple $\boldsymbol{x}$, of orbifold strata, $(\mathcal{Z}_1,\dots,\mathcal{Z}_k)$. The length k of $\boldsymbol{x}$ should coincide with the number of marked points. We will write such a moduli space as

$$\mathcal{M}^A_{0,k}(\mathcal{Z}, J, \boldsymbol{x}),$$

and require that the evaluation takes the j-th marked point into $\mathcal{Z}_j$. The compactification is similar to the manifold case, and consists of stable maps with the obvious adjustments, the caveat being that we must choose our lift to an orbicurve. After an appropriate construction of cycles as in the manifold case, Chen and Ruan use a virtual cycle construction, so we can define this invariant as in the smooth case above, but we integrate over (the compactification of) $\mathcal{M}^A_{0,k}(\mathcal{Z}, J, \boldsymbol{x})$. We will write these invariants

$$\mathrm{GW}^{\mathcal{Z}}_{A,k,\boldsymbol{x}}(\alpha_1,\dots,\alpha_k).$$

Another key difference is that this moduli space differs from the predicted dimension in the smooth case by a factor of $-2\iota(\boldsymbol{x})$, the so-called *degree shifting number*. (Again, for the definition see [Chen and Ruan 2002].) The Gromov–Witten invariants satisfy a list of axioms developed by Kontsevich and Manin [1994; 1997]. We will not list all of the axioms, but will mention only some which are used in the text. We use the orbifold notation; for a manifold we would just delete $\boldsymbol{x}$ from the notation, setting $\iota(\boldsymbol{x}) = 0$.

(i) *Effective*: $\mathrm{GW}^{\mathcal{Z}}_{A,k,\boldsymbol{x}}(\alpha_1, \ldots, \alpha_n) = 0$ as long as $\omega(A) < 0$.

(ii) *Grading*: $\mathrm{GW}^{\mathcal{Z}}_{A,k,\boldsymbol{x}}(\alpha_1, \ldots, \alpha_n) \neq 0$ only if

$$\sum_j \deg(\alpha_j) = \dim(\mathcal{Z}) + 2c_1(A) + 2k - 6 - 2\iota(\boldsymbol{x}).$$

(iii) *Divisor*: Let $\boldsymbol{x}^j = \boldsymbol{x}$ with the j-th component removed. Suppose that for each component x_i of $\boldsymbol{x}$, if x_i is mapped into the orbifold singular locus, that stratum is nonsingular as a variety. If $\deg(\alpha_n) = 2$ then

$$\mathrm{GW}^{\mathcal{Z}}_{A,k,\boldsymbol{x}}(\alpha_1, \ldots, \alpha_n) = \left(\int_A \alpha_n \right) \mathrm{GW}^{\mathcal{Z}}_{A,k-1,\boldsymbol{x}^n}(\alpha_1, \ldots, \alpha_{n-1}).$$

Now we are in a position to define the Gromov–Witten potential. This is a generating function which gives a formal power series whose coefficients give Gromov–Witten invariants. It is a way to organize all the information from these invariants into one big package. We give the definition here for the manifold case. Pick a basis of $H^2(M)$, $a_1, \ldots, a_n$, for a vector t and a cohomology class a, and write $a := a_t = \sum_i t_i a_i$.

Definition A.8. Let (M, ω) and J be as above. Define the genus-0 *Gromov–Witten Potential* as

$$\boldsymbol{f}(a_t) = \sum_A \sum_k \frac{1}{k!} \mathrm{GW}^M_{A,k}(a_t, \ldots, a_t) z^{c_1(A)}.$$

The corresponding formula for orbifolds is obtained by accounting for the vector $\boldsymbol{x}$.

Acknowledgements

Boyer thanks Miguel Abreu, Vestislav Apostolov, David Calderbank, Paul Gauduchon, and Christina Tønnesen-Friedman for helpful discussions on toric geometry. Pati would like to thank Tobias Ekholm, Georgios Dimitroglou Rizell, and Clement Hyrvier for many useful discussions.

References

[Abreu 2010] M. Abreu, "Kähler–Sasaki geometry of toric symplectic cones in action-angle coordinates", *Port. Math.* **67**:2 (2010), 121–153. MR 2011f:53074 Zbl 1193.53109

[Abreu and Macarini 2012] M. Abreu and L. Macarini, "Contact homology of good toric contact manifolds", *Compos. Math.* **148**:1 (2012), 304–334. MR 2881318 Zbl 1238.53064

[Audin 1994] M. Audin, "Symplectic and almost complex manifolds", pp. 41–74 in *Holomorphic curves in symplectic geometry*, edited by M. Audin and J. Lafontaine, Progr. Math. **117**, Birkhäuser, Basel, 1994. MR 1274926

[Banyaga and Molino 1993] A. Banyaga and P. Molino, "Géométrie des formes de contact complètement intégrables de type toriques", pp. 1–25 in *Séminaire Gaston Darboux de Géométrie et Topologie Différentielle* (Montpellier, 1991–1992), edited by J. Lafontaine, Université Montpellier II, Montpellier, 1993. MR 94e:53029 Zbl 0783.58021

[Barden 1965] D. Barden, "Simply connected five-manifolds", *Ann. of Math.* (2) **82** (1965), 365–385. MR 32 #1714 Zbl 0136.20602

[Bourgeois 2002] F. Bourgeois, *A Morse–Bott approach to contact homology*, PhD thesis, Stanford University, Stanford, CA, 2002, Available at http://search.proquest.com/docview/305591502. MR 2703292

[Bourgeois 2003] F. Bourgeois, "A Morse–Bott approach to contact homology", pp. 55–77 in *Symplectic and contact topology: interactions and perspectives* (Toronto, ON/Montréal, QC, 2001), edited by Y. Eliashberg et al., Fields Inst. Commun. **35**, Amer. Math. Soc., Providence, RI, 2003. MR 2004a:53109 Zbl 1046.57017

[Bourgeois and Mohnke 2004] F. Bourgeois and K. Mohnke, "Coherent orientations in symplectic field theory", *Math. Z.* **248**:1 (2004), 123–146. MR 2005g:53173 Zbl 1060.53080

[Boyer 2009] C. P. Boyer, "The Sasakian geometry of the Heisenberg group", *Bull. Math. Soc. Sci. Math. Roumanie* (*N.S.*) **52(100)**:3 (2009), 251–262. MR 2011b:53098 Zbl 1199.53102

[Boyer 2011a] C. P. Boyer, "Completely integrable contact Hamiltonian systems and toric contact structures on $S^2 \times S^3$", *Symmetry Integrability Geom. Methods Appl.* **7** (2011), Paper ID #058. MR 2861218 Zbl 1242.53112 arXiv 1101.5587

[Boyer 2011b] C. P. Boyer, "Extremal Sasakian metrics on S^3-bundles over S^2", *Math. Res. Lett.* **18**:1 (2011), 181–189. MR 2012d:53132 Zbl 1246.53065

[Boyer 2013] C. P. Boyer, "Maximal tori in contactomorphism groups", *Differential Geom. Appl.* **31**:2 (2013), 190–216. MR 3032643

[Boyer and Galicki 2000a] C. P. Boyer and K. Galicki, "A note on toric contact geometry", *J. Geom. Phys.* **35**:4 (2000), 288–298. MR 2001h:53124 Zbl 0984.53032

[Boyer and Galicki 2000b] C. P. Boyer and K. Galicki, "On Sasakian–Einstein geometry", *Internat. J. Math.* **11**:7 (2000), 873–909. MR 2001k:53081 Zbl 1022.53038

[Boyer and Galicki 2008] C. P. Boyer and K. Galicki, *Sasakian geometry*, Oxford University Press, 2008. MR 2009c:53058 Zbl 1155.53002

[Boyer and Tønnesen-Friedman 2013] C. P. Boyer and C. W. Tønnesen-Friedman, "Extremal Sasakian geometry on $T^2 \times S^3$ and related manifolds", *Compos. Math.* **149**:8 (2013), 1431–1456.

[Boyer and Tønnesen-Friedman 2014] C. P. Boyer and C. W. Tønnesen-Friedman, "Extremal Sasakian geometry on S^3-bundles over Riemann surfaces", *Int. Math. Res. Not.* **2014**:8 (2014), rnt139.

[Boyer et al. 2005] C. P. Boyer, K. Galicki, and J. Kollár, "Einstein metrics on spheres", *Ann. of Math.* (2) **162**:1 (2005), 557–580. MR 2006j:53058 Zbl 1093.53044

[Boyer et al. 2007] C. P. Boyer, K. Galicki, and L. Ornea, "Constructions in Sasakian geometry", *Math. Z.* **257**:4 (2007), 907–924. MR 2008m:53103 Zbl 1134.53021

[Boyer et al. 2008] C. P. Boyer, K. Galicki, and S. R. Simanca, "Canonical Sasakian metrics", *Comm. Math. Phys.* **279**:3 (2008), 705–733. MR 2009a:53077 Zbl 1142.53036

[Boyer et al. 2009] C. P. Boyer, K. Galicki, and S. R. Simanca, "The Sasaki cone and extremal Sasakian metrics", pp. 263–290 in *Riemannian topology and geometric structures on manifolds* (Albuquerque, NM, 2006), edited by K. Galicki and S. R. Simanca, Progr. Math. **271**, Birkhäuser, Boston, 2009. MR 2010e:53078 Zbl 1171.53033

[Bryant 2001] R. L. Bryant, "Bochner–Kähler metrics", *J. Amer. Math. Soc.* **14**:3 (2001), 623–715. MR 2002i:53096 Zbl 1006.53019

[Calabi 1985] E. Calabi, "Extremal Kähler metrics, II", pp. 95–114 in *Differential geometry and complex analysis*, edited by I. Chavel and H. M. Farkas, Springer, Berlin, 1985. MR 86h:53067 Zbl 0574.58006

[Chen and Ruan 2002] W. Chen and Y. Ruan, "Orbifold Gromov–Witten theory", pp. 25–85 in *Orbifolds in mathematics and physics* (Madison, WI, 2001), edited by A. Adem et al., Contemp. Math. **310**, Amer. Math. Soc., Providence, RI, 2002. MR 2004k:53145 Zbl 1091.53058

[Chen and Ruan 2004] W. Chen and Y. Ruan, "A new cohomology theory of orbifold", *Comm. Math. Phys.* **248**:1 (2004), 1–31. MR 2005j:57036 Zbl 1063.53091

[Cho et al. 2008] K. Cho, A. Futaki, and H. Ono, "Uniqueness and examples of compact toric Sasaki–Einstein metrics", *Comm. Math. Phys.* **277**:2 (2008), 439–458. MR 2008j:53076 Zbl 1144.53058

[Cieliebak et al. 1996] K. Cieliebak, A. Floer, H. Hofer, and K. Wysocki, "Applications of symplectic homology, II: Stability of the action spectrum", *Math. Z.* **223**:1 (1996), 27–45. MR 97j:58045 Zbl 0869.58013

[Cvetič et al. 2005] M. Cvetič, H. Lü, D. N. Page, and C. N. Pope, "New Einstein–Sasaki spaces in five and higher dimensions", *Phys. Rev. Lett.* **95**:7 (2005), Article ID #071101. MR 2006f:53061

[David and Gauduchon 2006] L. David and P. Gauduchon, "The Bochner-flat geometry of weighted projective spaces", pp. 109–156 in *Perspectives in Riemannian geometry* (Montréal, QC, 2004), edited by V. Apostolov et al., CRM Proc. Lecture Notes **40**, Amer. Math. Soc., Providence, RI, 2006. MR 2007h:32032 Zbl 1109.32019

[Eliashberg et al. 2000] Y. Eliashberg, A. Givental, and H. Hofer, "Introduction to symplectic field theory", *Geom. Funct. Anal.* Special Volume, Part II (2000), 560–673. MR 2002e:53136 Zbl 0989.81114

[Gauduchon 2009] P. Gauduchon, "Hirzebruch surfaces and weighted projective planes", pp. 25–48 in *Riemannian topology and geometric structures on manifolds* (Albuquerque, NM, 2006), edited by K. Galicki and S. R. Simanca, Progr. Math. **271**, Birkhäuser, Boston, 2009. MR 2010h:32023 Zbl 1191.32009

[Gauntlett et al. 2004a] J. P. Gauntlett, D. Martelli, J. Sparks, and D. Waldram, "Sasaki–Einstein metrics on $S^2 \times S^3$", *Adv. Theor. Math. Phys.* **8**:4 (2004), 711–734. MR 2006m:53067 Zbl 1136.53317

[Gauntlett et al. 2004b] J. P. Gauntlett, D. Martelli, J. Sparks, and D. Waldram, "Supersymmetric AdS_5 solutions of M-theory", *Classical Quantum Gravity* **21**:18 (2004), 4335–4366. MR 2006b:83186 Zbl 1059.83038

[Gauntlett et al. 2005] J. P. Gauntlett, D. Martelli, J. Sparks, and D. Waldram, "Supersymmetric AdS backgrounds in string and M-theory", pp. 217–252 in *AdS/CFT correspondence: Einstein metrics and their conformal boundaries* (Strasbourg, 2003), edited by O. Biquard, IRMA Lect. Math. Theor. Phys. **8**, Eur. Math. Soc., Zürich, 2005. MR 2007b:83171 Zbl 1071.81556

[Ghigi and Kollár 2007] A. Ghigi and J. Kollár, "Kähler–Einstein metrics on orbifolds and Einstein metrics on spheres", *Comment. Math. Helv.* **82**:4 (2007), 877–902. MR 2008j:32027 Zbl 1135.53031

[Griffiths and Harris 1978] P. Griffiths and J. Harris, *Principles of algebraic geometry*, Wiley, New York, 1978. MR 80b:14001 Zbl 0408.14001

[Gromov 1985] M. Gromov, "Pseudoholomorphic curves in symplectic manifolds", *Invent. Math.* **82**:2 (1985), 307–347. MR 87j:53053 Zbl 0592.53025

[Haefliger 1984] A. Haefliger, "Groupoïdes d'holonomie et classifiants", pp. 70–97 in *Structure transverse des feuilletages* (Toulouse, 1982), edited by J. Pradines, Astérisque **116**, Soc. Math. France, Paris, 1984. MR 86c:57026a Zbl 0562.57012

[Hirzebruch 1951] F. Hirzebruch, "Über eine Klasse von einfach-zusammenhängenden komplexen Mannigfaltigkeiten", *Math. Ann.* **124** (1951), 77–86. MR 13,574e Zbl 0043.30302

[Hofer et al. 1996] H. Hofer, K. Wysocki, and E. Zehnder, "Properties of pseudoholomorphic curves in symplectisations, I: Asymptotics", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **13**:3 (1996), 337–379. MR 97e:58029 Zbl 0861.58018

[Hofer et al. 1997] H. Hofer, V. Lizan, and J.-C. Sikorav, "On genericity for holomorphic curves in four-dimensional almost-complex manifolds", *J. Geom. Anal.* **7**:1 (1997), 149–159. MR 2000d:32045 Zbl 0911.53014

[Karshon 1999] Y. Karshon, "Periodic Hamiltonian flows on four-dimensional manifolds", *Mem. Amer. Math. Soc.* **141**:672 (1999). MR 2000c:53113 Zbl 0982.70011

[Karshon 2003] Y. Karshon, "Maximal tori in the symplectomorphism groups of Hirzebruch surfaces", *Math. Res. Lett.* **10**:1 (2003), 125–132. MR 2004f:53101 Zbl 1036.53063

[Kirwan 1984] F. C. Kirwan, *Cohomology of quotients in symplectic and algebraic geometry*, Mathematical Notes **31**, Princeton University Press, 1984. MR 86i:58050 Zbl 0553.14020

[Kontsevich and Manin 1994] M. Kontsevich and Y. Manin, "Gromov–Witten classes, quantum cohomology, and enumerative geometry", *Comm. Math. Phys.* **164**:3 (1994), 525–562. MR 95i:14049 Zbl 0853.14020

[Kontsevich and Manin 1997] M. Kontsevich and Y. Manin, "Gromov–Witten classes, quantum cohomology, and enumerative geometry", pp. 607–653 in *Mirror symmetry, II*, edited by B. Greene and S.-T. Yau, AMS/IP Stud. Adv. Math. **1**, Amer. Math. Soc., Providence, RI, 1997. MR 1416351 Zbl 0931.14030

[Lafontaine 1981] J. Lafontaine, "Surfaces de Hirzebruch", pp. 61–76 in *Riemannian geometry in dimension* 4 (Paris, 1978–1979), edited by L. Bérard Bergery et al., Textes Math. **3**, CEDIC, Paris, 1981. MR 769131 Zbl 0482.14013

[Legendre 2011a] E. Legendre, "Existence and non-uniqueness of constant scalar curvature toric Sasaki metrics", *Compos. Math.* **147**:5 (2011), 1613–1634. MR 2834736 Zbl 1237.53039

[Legendre 2011b] E. Legendre, "Toric geometry of convex quadrilaterals", *J. Symplectic Geom.* **9**:3 (2011), 343–385. MR 2012g:53079 Zbl 1233.14032

[Lerman 2003a] E. Lerman, "Contact toric manifolds", *J. Symplectic Geom.* **1**:4 (2003), 785–828. MR 2004m:53147 Zbl 1079.53118

[Lerman 2003b] E. Lerman, "Maximal tori in the contactomorphism groups of circle bundles over Hirzebruch surfaces", *Math. Res. Lett.* **10**:1 (2003), 133–144. MR 2004g:53097 Zbl 1042.53059

[Lerman 2004] E. Lerman, "Homotopy groups of K-contact toric manifolds", *Trans. Amer. Math. Soc.* **356**:10 (2004), 4075–4083. MR 2005b:53136 Zbl 1057.53062

[Lerman and Tolman 1997] E. Lerman and S. Tolman, "Hamiltonian torus actions on symplectic orbifolds and toric varieties", *Trans. Amer. Math. Soc.* **349**:10 (1997), 4201–4230. MR 98a:57043 Zbl 0897.58016

[Martelli and Sparks 2005] D. Martelli and J. Sparks, "Toric Sasaki–Einstein metrics on $S^2 \times S^3$", *Phys. Lett. B* **621**:1-2 (2005), 208–212. MR 2006c:53041 Zbl 1247.53055

[McDuff and Salamon 2004] D. McDuff and D. Salamon, *J-holomorphic curves and symplectic topology*, American Mathematical Society Colloquium Publications **52**, Amer. Math. Soc., Providence, RI, 2004. MR 2004m:53154 Zbl 1064.53051

[Ness 1984] L. Ness, "A stratification of the null cone via the moment map", *Amer. J. Math.* **106**:6 (1984), 1281–1329. MR 86c:14010 Zbl 0604.14006

[Pati 2009] J. Pati, "Contact homology of S^1-bundles over some symplectically reduced orbifolds", preprint, 2009. arXiv 0910.5934

[Pati 2010] J. Pati, *Contact homology of toric contact manifolds of Reeb type*, PhD thesis, University of New Mexico, Albuquerque, NM, 2010, Available at http://search.proquest.com/docview/757731908. MR 2792890

[Robbin and Salamon 1993] J. Robbin and D. Salamon, "The Maslov index for paths", *Topology* **32**:4 (1993), 827–844. MR 94i:58071 Zbl 0798.58018

[Sparks 2011] J. Sparks, "Sasaki–Einstein manifolds", pp. 265–324 in *Geometry of special holonomy and related topics*, edited by N. C. Leung and S.-T. Yau, Surv. Differ. Geom. **16**, International Press, Somerville, MA, 2011. MR 2012k:53082 Zbl 1256.53037

[Tanno 1970] S. Tanno, "On the isometry groups of Sasakian manifolds", *J. Math. Soc. Japan* **22** (1970), 579–590. MR 42 #6755 Zbl 0197.48004

Received April 23, 2012. Revised July 4, 2013.

CHARLES P. BOYER
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEW MEXICO
222 SMLC
ALBUQUERQUE, NM 87131
UNITED STATES
cboyer@math.unm.edu

JUSTIN PATI
MATEMATISKA INSTITUTIONEN
UPPSALA UNIVERSITET
BOX 480
SE-751 06 UPPSALA
SWEDEN
justin@math.uu.se

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.325

AN ALMOST-SCHUR TYPE LEMMA FOR SYMMETRIC (2,0) TENSORS AND APPLICATIONS

XU CHENG

In a previous paper, we generalized the almost-Schur lemma of De Lellis and Topping for closed manifolds with nonnegative Ricci curvature to any closed manifolds. In this paper, we generalize the above results to symmetric (2, 0)-tensors and give the applications for r-th mean curvatures of closed hypersurfaces in space forms and k scalar curvatures for closed locally conformally flat manifolds.

1. Introduction

Recall that an n-dimensional Riemannian manifold (M, g) is said to be Einstein if its traceless Ricci tensor $\mathring{\mathrm{Ric}} = \mathrm{Ric} - (R/n)g$ is identically zero. Here Ric and R denote Ricci curvature and scalar curvature respectively. Schur's lemma states that the scalar curvature of an Einstein manifold of dimension $n \geq 3$ must be constant. De Lellis and Topping [2012] discussed the quantitative version, or the stability of Schur's lemma for closed manifolds, and proved the following almost-Schur lemma, as they called it.

Theorem 1.1 [De Lellis and Topping 2012]. *If (M, g) is a closed Riemannian manifold of dimension n with nonnegative Ricci curvature $n \geq 3$,*

$$\int_M (R - \bar{R})^2 \leq \frac{4n(n-1)}{(n-2)^2} \int_M \left| \mathrm{Ric} - \frac{R}{n} g \right|^2 \tag{1-1}$$

and, equivalently,

$$\int_M \left| \mathrm{Ric} - \frac{\bar{R}}{n} g \right|^2 \leq \frac{n^2}{(n-2)^2} \int_M \left| \mathrm{Ric} - \frac{R}{n} g \right|^2, \tag{1-2}$$

where $\bar{R} = (1/\operatorname{Vol} M) \int_M R \, dv$ is the average of R over M. Equality holds in (1-1) *or* (1-2) *if and only if M is Einstein.*

The author is partially supported by CNPq and Faperj of Brazil.
MSC2010: 53C21.
Keywords: traceless tensor, Ricci curvature, mean curvature.

B. Andrews also obtained the above inequalities in an unpublished paper under the assumption that the Ricci curvature is positive. De Lellis and Topping also proved their estimates are sharp. First, the constants are optimal in (1-1) and (1-2) [De Lellis and Topping 2012, Section 2]. Second, the curvature condition $\mathrm{Ric} \geq 0$ cannot simply be dropped (see the examples in the proof of Propositions 2.1 and 2.2 in their paper). Without the condition of nonnegativity of the Ricci curvature, the same type of inequalities as (1-1) and (1-2) cannot hold if the constants in these inequalities only depend on the lower bound of the Ricci curvature.

In the case of closed manifolds without the hypothesis of nonnegativity of Ricci curvature, we have:

Theorem 1.2 [Cheng 2013]. *If (M, g) is a closed Riemannian manifold of dimension $n \geq 3$, then*

$$\int_M (R-\bar{R})^2 \leq \frac{4n(n-1)}{(n-2)^2}\left(1+\frac{nK}{\lambda_1}\right)\int_M \left|\mathrm{Ric}-\frac{R}{n}g\right|^2; \tag{1-3}$$

equivalently,

$$\int_M \left|\mathrm{Ric}-\frac{\bar{R}}{n}g\right|^2 \leq \frac{n^2}{(n-2)^2}\left[1+\frac{4(n-1)K}{n\lambda_1}\right]\int_M \left|\mathrm{Ric}-\frac{R}{n}g\right|^2, \tag{1-4}$$

where λ_1 denotes the first nonzero eigenvalue of the Laplace operator on (M, g) and K is a nonnegative constant such that the Ricci curvature of (M, g) satisfies $\mathrm{Ric} \geq -(n-1)K$.

Equality holds in (1-3) *or* (1-4) *if and only if M is an Einstein manifold.*

Observe that Theorem 1.1 is a particular case of Theorem 1.2 ($K = 0$). After the work of De Lellis and Topping, in the case of dimension $n = 3, 4$, Y. Ge and G. Wang [2012; 2011] proved that Theorem 1.1 holds under the weaker condition of nonnegative scalar curvature. However, as pointed out in [De Lellis and Topping 2012], this is surely not possible for $n \geq 5$; this can be shown using constructions similar to the one in [De Lellis and Topping 2012, Section 3]. Also, Ge, Wang, and Xia [Ge et al. 2013] proved the case of equalities in (1-1) and (1-2) by a different method and generalized De Lellis and Topping's inequalities for k-Einstein tensors and Lovelock curvature.

On the other hand, there is a similar phenomenon in submanifold theory. In differential geometry, a classical theorem states that a closed totally umbilical surface in the Euclidean space $\mathbb{R}^3$ must be a round sphere $\mathbb{S}^2$ and thus its second fundamental form A is a constant multiple of its metric. This theorem is also true for hypersurfaces in $\mathbb{R}^{n+1}$. It is interesting to discuss the stability of this theorem. De Lellis and Müller [2005] obtained some L^2 inequalities for closed surfaces in $\mathbb{R}^3$ with universal constants. For convex hypersurfaces in $\mathbb{R}^{n+1}$ we have:

Theorem 1.3 [Perez 2011]. *Let Σ be a smooth, closed and connected hypersurface in $\mathbb{R}^{n+1}$, $n \geq 2$, with induced Riemannian metric g and nonnegative Ricci curvature. Then*

$$\int_\Sigma \left| A - \frac{\overline{H}}{n} g \right|^2 \leq \frac{n}{n-1} \int_\Sigma \left| A - \frac{H}{n} g \right|^2 \tag{1-5}$$

and, equivalently,

$$\int_\Sigma (H - \overline{H})^2 \leq \frac{n}{n-1} \int_\Sigma \left| A - \frac{H}{n} g \right|^2, \tag{1-6}$$

where A and $H = \operatorname{tr} A$ denote the second fundamental form and the mean curvature of Σ, respectively, and $\overline{H} = (1/\operatorname{Vol}_n \Sigma) \int_\Sigma H$. In particular, the above estimate holds for smooth, closed hypersurfaces which are the boundary of a convex set in $\mathbb{R}^{n+1}$.

As pointed out in [De Lellis and Topping 2012], Perez's theorem holds even for closed hypersufaces with nonnegative Ricci curvature when the ambient space is Einstein. Indeed, a slight modification of the proof of Theorem 1.3 gives the following.

Theorem 1.4. *Inequalities* (1-5) *and* (1-6) *hold under the same assumptions as in Theorem 1.3 except that the ambient space $(N^{n+1}, \tilde{g})$, $n \geq 2$, is supposed to be an Einstein manifold.*

Regarding the conditions for equality in (1-5) and (1-6), we have:

Theorem 1.5 [Cheng and Zhou 2012]. *Under the assumptions of Theorem 1.3, equality holds in* (1-5) *or* (1-6) *if and only if Σ is a totally umbilical hypersurface, that is, Σ is a distance sphere S^n in $\mathbb{R}^{n+1}$.*

We also studied the general case for hypersurfaces without a convexity hypothesis (that is, $A \geq 0$, which is equivalent to $\operatorname{Ric} \geq 0$ when Σ is a closed hypersurface in $\mathbb{R}^{n+1}$). We mention the following result (more details in the reference given):

Theorem 1.6 [Cheng and Zhou 2012]. *Let $(N^{n+1}, \tilde{g})$ be an Einstein manifold, $n \geq 2$. Let Σ be a smooth, connected, oriented and closed hypersurface immersed in N with induced metric g. Then*

$$\int_\Sigma \left| A - \frac{\overline{H}}{n} g \right|^2 \leq \frac{n}{n-1} \left(1 + \frac{K}{\lambda_1} \right) \int_\Sigma \left| A - \frac{H}{n} g \right|^2 \tag{1-7}$$

and, equivalently,

$$\int_\Sigma (H - \overline{H})^2 \leq \frac{n}{n-1} \left(1 + \frac{nK}{\lambda_1} \right) \int_\Sigma \left| A - \frac{H}{n} g \right|^2, \tag{1-8}$$

where λ_1 is the first nonzero eigenvalue of the Laplacian operator on Σ, $K \geq 0$ is a nonnegative constant so that the Ricci curvature of Σ satisfies $\mathrm{Ric} \geq -K$.

When N^{n+1} is the Euclidean space $\mathbb{R}^{n+1}$, the hyperbolic space $\mathbb{H}^{n+1}(-1)$, or the closed hemisphere $\mathbb{S}^{n+1}_+(1)$, equality holds in (1-7) *and* (1-8) *if and only if Σ is a totally umbilical hypersurface, that is, Σ is a distance sphere S^n in N^{n+1}.*

From [De Lellis and Topping 2012; Ge and Wang 2012; 2011; Ge et al. 2013; Cheng 2013; Perez 2011; Cheng and Zhou 2012], we observe that the inequalities mentioned above may be generalized to symmetric $(2, 0)$ tensor fields. Applying such unified inequalities for symmetric $(2, 0)$ tensors, we may obtain inequalities besides those in the papers mentioned above. For this purpose, we prove the following.

Theorem 1.7. *Let (M, g) be a closed Riemannian manifold of dimension $n \geq 2$. Let T be a symmetric $(2, 0)$-tensor field on M. If the divergence* $\operatorname{div} T$ *and the trace* $B = \operatorname{tr} T$ *satisfy* $\operatorname{div} T = c\nabla B$, *where c is a constant, then*

$$(nc-1)^2 \int_M (B-\bar{B})^2 \leq n(n-1)\left(1+\frac{nK}{\lambda_1}\right) \int_M \left|T-\frac{B}{n}g\right|^2 \tag{1-9}$$

and, equivalently,

$$(nc-1)^2 \int_M \left|T-\frac{\bar{B}}{n}g\right|^2 \leq \left[(nc-1)^2+(n-1)\left(1+\frac{nK}{\lambda_1}\right)\right] \int_M \left|T-\frac{B}{n}g\right|^2, \tag{1-10}$$

where $\bar{B} = (1/\operatorname{Vol} M)\int_M B\, dv$ denotes the average of B over M and λ_1 and the constant $K \geq 0$ are as in Theorem 1.2.

Assume the Ricci curvature Ric *of M is positive. If $c \neq 1/n$, statements* (i), (ii) *and* (iii) *below are equivalent. If $c = 1/n$, then* (i) *and* (ii) *are equivalent.*

(i) *Equality holds in* (1-9) *and in* (1-10).

(ii) *$T = (B/n)g$ on M.*

(iii) *$T = (\bar{B}/n)g$ on M.*

Take $K = 0$ in Theorem 1.7. We obtain corresponding inequalities with universal constants.

Theorem 1.8. *Let (M, g) be a closed Riemannian manifold of dimension $n \geq 2$ with nonnegative Ricci curvature. With the same notation as in Theorem 1.7, we have*

$$(nc-1)^2 \int_M (B-\bar{B})^2 \leq n(n-1) \int_M \left|T-\frac{B}{n}g\right|^2 \tag{1-11}$$

and, *equivalently*,

$$(nc-1)^2 \int_M \left| T - \frac{\bar{B}}{n} g \right|^2 \le [(nc-1)^2+1] \int_M \left| T - \frac{B}{n} g \right|^2. \tag{1-12}$$

Assume the Ricci curvature Ric *of* M *is positive. If* $c \neq 1/n$, *statements* (i), (ii) *and* (iii) *below are equivalent. If* $c = 1/n$, *then* (i) *and* (ii) *are equivalent.*

(i) *Equality holds in* (1-11) *and* (1-12).

(ii) $T = (B/n)g$ *on* M.

(iii) $T = (\bar{B}/n)g$ *on* M.

It is a known fact that, for (M^n, g), $n \ge 2$, if $T = (B/n)g$ and $\operatorname{div} T = c\nabla B$ with constant $c \neq 1/n$, then B is constant on M and thus T is a constant multiple of its metric g (see Proposition 2.1). Theorems 1.7 and 1.8 discuss the stability and rigidity of this fact for closed manifolds. Especially, take $T = \mathrm{Ric}$, A, etc. in Theorems 1.7 and 1.8. We obtain the corresponding inequalities mentioned before 1.7. In this paper, we obtain two other applications as follows.

First we deal with r-th mean curvatures of closed hypersurfaces in space forms. Assume (Σ, g) is a connected oriented closed hypersurface immersed in a space form with induced metric g. Associated with the second fundamental form A of Σ, we have r-th mean curvatures H_r of Σ and the Newton transformations P_r, $0 \le r \le n$, (see their definition and related notation in Section 4). Since Reilly [1973] introduced them, there has been much work in studying high-order r-mean curvatures (see, for instance, [Rosenberg 1993; Barbosa and Colares 1997; Cheng and Rosenberg 2005; Alías et al. 2006]). It can be verified that if the Newton transformations P_r satisfy $P_r = (\operatorname{tr} P_r/n)g$ on Σ, Σ has constant r-th mean curvature and thus P_r is a constant multiple of its metric g (see Proposition 2.1 and Section 4). In this paper, we discuss the stability of this fact.

In addition, although it is true that a closed immersed totally umbilical hypersurface Σ (that is, Σ satisfies $P_1 = (\operatorname{tr} P_1/n)g$) in $\mathbb{R}^{n+1}$ must be a round sphere $\mathbb{S}^n$, it is unknown, to the best of our knowledge, if it is true that a closed immersed hypersurface Σ satisfying $P_r = (\operatorname{tr} P_r/n)g$ in $\mathbb{R}^{n+1}$ must be a round sphere $\mathbb{S}^n$ for $r \ge 2$. When Σ is embedded, Ros [1988; 1987] showed that the round spheres are the only closed embedded hypersurfaces with constant r-th mean curvature in $\mathbb{R}^{n+1}$, $2 \le r \le n$ (recall that the Alexandrov theorem says [Aleksandrov 1958] that the round spheres are the only closed embedded hypersurfaces in $\mathbb{R}^{n+1}$ with constant mean curvature). Hence the round spheres are the only closed embedded hypersurfaces in $\mathbb{R}^{n+1}$ with $P_r = (\operatorname{tr} P_r/n)g$, $2 \le r \le n$.

In Section 4, we prove the following.

Theorem 1.9. *Let $(N_a^{n+1}, \tilde{g})$ be a space form with constant sectional curvature a, $n \geq 2$. Assume that Σ is a smooth connected oriented closed hypersurface immersed in N with induced metric g. Then, for $2 \leq r \leq n$,*

$$(n-r)^2 \int_\Sigma (s_r - \bar{s}_r)^2 \leq n(n-1)\left(1 + \frac{nK}{\lambda_1}\right) \int_\Sigma \left| P_r - \frac{(n-r)s_r}{n} g \right|^2 \tag{1-13}$$

and, equivalently,

$$\int_\Sigma \left| P_r - \frac{(n-r)\bar{s}_r}{n} g \right|^2 \leq n\left[1 + \frac{(n-1)K}{\lambda_1}\right] \int_\Sigma \left| P_r - \frac{(n-r)s_r}{n} g \right|^2, \tag{1-14}$$

where $s_r = \operatorname{tr} P_r = \binom{n}{r} H_r$, $\bar{s}_r = (1/\operatorname{Vol}\Sigma)\int_\Sigma s_r \, dv$, and λ_1 and the constant $K \geq 0$ are as in Theorem 1.6. Moreover:

(1) *If the Ricci curvature* Ric *of Σ is positive, these three statements are equivalent*:
  - (i) *Equality holds in* (1-13) *and* (1-14).
  - (ii) *$P_r = ((n-r)s_r/n)g$ holds on Σ.*
  - (iii) *$P_r = ((n-r)\bar{s}_r/n)g$ holds on Σ.*

(2) *If Σ is embedded in the Euclidean space $\mathbb{R}^{n+1}$ and the Ricci curvature* Ric *of Σ is positive, equality holds in* (1-13) *and* (1-14) *if and only if Σ is a round sphere $\mathbb{S}^{n+1}$ in $\mathbb{R}^{n+1}$.*

Taking $K = 0$ in Theorem 1.9, we obtain the following inequalities.

Theorem 1.10. *Besides the same assumptions as in Theorem 1.9, assume that Σ has nonnegative Ricci curvature. Then, for $2 \leq r \leq n$,*

$$(n-r)^2 \int_\Sigma (s_r - \bar{s}_r)^2 \leq n(n-1) \int_\Sigma \left| P_r - \frac{(n-r)s_r}{n} g \right|^2 \tag{1-15}$$

and, equivalently,

$$\int_\Sigma \left| P_r - \frac{(n-r)\bar{s}_r}{n} g \right|^2 \leq n \int_\Sigma \left| P_r - \frac{(n-r)s_r}{n} g \right|^2. \tag{1-16}$$

Second, we consider the k-scalar curvatures of locally conformally flat closed manifolds (see their definition in Section 5). Since they were first introduced in [Viaclovsky 2000], k-scalar curvatures have been much studied; see, for instance, [Guan 2002; Viaclovsky 2006]. When M is locally conformally flat, we obtain an almost-Schur type lemma for k-scalar curvatures, $k \geq 2$, as follows.

Theorem 1.11. *Let (M^n, g) be an n-dimensional closed locally conformally flat manifold, $n \geq 3$. Then, for $2 \leq k \leq n$, the k-scalar curvature $\sigma_k(S_g)$ and the Newton transformation T_k associated with the Schouten tensor S_g satisfy*

$$(1\text{-}17)\quad (n-k)^2\int_M(\sigma_k(S_g)-\overline{\sigma}_k(S_g))^2 \le n(n-1)\left(1+\frac{nK}{\lambda_1}\right)\int_M\left|T_k-\frac{(n-k)\sigma_k(S_g)}{n}g\right|^2$$

and, *equivalently*,

$$(1\text{-}18)\quad \int_M\left|T_k-\frac{(n-k)\overline{\sigma}_k(S_g)}{n}g\right|^2\le n\left[1+\frac{(n-1)K}{\lambda_1}\right]\int_M\left|T_k-\frac{(n-k)\sigma_k(g)}{n}g\right|^2,$$

where $\overline{\sigma}_k(S_g)=(1/\operatorname{Vol}M)\int_M\sigma_k(S_g)\,dv$ *and* λ_1 *and the constant* $K\ge 0$ *are as in Theorem 1.2.*

If the Ricci curvature Ric *of* M *is positive*, *these three statements are equivalent*:

(i) *Equality holds in* (1-17) *and* (1-18).

(ii) $T_k=((n-k)\sigma_k(S_g)/n)g$ *on* M.

(iii) $T_k=((n-k)\overline{\sigma}_k(S_g)/n)g$ *on* M.

As for Theorem 1.10, taking $K=0$ in Theorem 1.11, one obtains the corresponding inequalities with the universal constants.

The rest of this paper is organized as follows. In Section 2, we prove Theorems 1.7 and 1.8. In Section 3, we recall the definitions of the Newton transformation and the r-th symmetric function associated with a symmetric endomorphism of an n-dimensional vector space. In Section 4, we prove Theorem 1.9 by applying Theorem 1.7. In Section 5, we prove Theorem 1.11 by applying Theorem 1.7.

2. Proof of theorems on symmetric (2, 0)-tensors

First we give some notation. Assume (M, g) is an n-dimensional closed, that is, compact and without boundary, Riemannian manifold. Let ∇ denote the Levi-Civita connection on (M, g) and also the induced connections on tensor bundles on M. Let T denote a symmetric $(2,0)$-tensor field on M. Let tr denote the trace of a tensor. $B=\operatorname{tr}T=T_i^i=g^{ij}T_{ij}$ denotes the trace of T. Hereafter we use the Einstein summation convention. Denote by $\overline{B}=(1/\operatorname{Vol}M)\int_M B$ the average of B over M and set $\mathring{T}=T-(B/n)g$. Denote by div the divergence of a tensor field. For T, $\operatorname{div}T=\operatorname{tr}\nabla T$ is a $(1,0)$-tensor. Under the local coordinates $\{x_i\}$ on M, $\operatorname{div}T=g^{ij}(\nabla_iT_{jk})\,dx^k$, where $\nabla_iT_{jk}=(\nabla_{\partial_i}T)(\partial_j,\partial_k)$.

The following fact, already mentioned in the introduction, can be proved directly by noting that $T=(B/n)g$ implies $\operatorname{div}T=\nabla B/n$.

Proposition 2.1. *Assume* (M^n, g), $n\ge 2$, *is a connected Riemannian manifold of dimension* n. *If* $T=(B/n)g$ *and* $\operatorname{div}T=c\nabla B$, *where* $c\ne 1/n$ *is a constant*, *then* $B=\text{const}$ *on* M *and* T *is a constant multiple of its metric* g.

The argument of Theorem 1.7 is similar to that of Theorem 1.2 (that is, [Cheng 2013, Theorem 1.2]) and, in the case of $K = 0$, that of Theorem 1.1 (that is, [De Lellis and Topping 2012, Theorem 0.1]).

Proof of Theorem 1.7. Obviously, it suffices to prove the case $c \neq 1/n$. By the assumption $\operatorname{div} T = c\nabla B$,

$$\operatorname{div} \mathring{T} = \operatorname{div} T - \operatorname{div}\Big(\frac{B}{n} g\Big) = \operatorname{div} T - \frac{\nabla B}{n} = \frac{nc-1}{n}\nabla B. \tag{2-1}$$

Let f be the unique solution of the following Poisson equation on M:

$$\Delta f = B - \bar{B}, \quad \int_M f = 0. \tag{2-2}$$

By (2-1), (2-2), and Stokes' formula,

$$\begin{aligned}\int_M (B-\bar{B})^2 &= \int_M (B-\bar{B})\Delta f = -\int_M \langle \nabla B, \nabla f\rangle \\ &= -\frac{n}{nc-1}\int_M \langle \operatorname{div} \mathring{T}, \nabla f\rangle \\ &= \frac{n}{nc-1}\int_M \langle \mathring{T}, \nabla^2 f\rangle \\ &= \frac{n}{nc-1}\int_M \langle \mathring{T}, \nabla^2 f - \frac{1}{n}(\Delta f) g\rangle \\ &\leq \frac{n}{|nc-1|}\Big(\int_M |\mathring{\mathrm{T}}|^2\Big)^{1/2}\Big[\int_M |\nabla^2 f - \frac{1}{n}(\Delta f) g|^2\Big]^{1/2} \\ &= \frac{n}{|nc-1|}\Big(\int_M |\mathring{\mathrm{T}}|^2\Big)^{1/2}\Big[\int_M |\nabla^2 f|^2 - \frac{1}{n}\int_M (\Delta f)^2\Big]^{1/2}.\end{aligned} \tag{2-3}$$

Recall the Bochner formula

$$\tfrac{1}{2}\Delta|\nabla f|^2 = |\nabla^2 f|^2 + \operatorname{Ric}(\nabla f, \nabla f) + \langle \nabla f, \nabla(\Delta f)\rangle,$$

and integrate it. By Stokes' formula, we have

$$\int_M |\nabla^2 f|^2 = \int_M (\Delta f)^2 - \int_M \operatorname{Ric}(\nabla f, \nabla f). \tag{2-4}$$

By (2-3) and (2-4),

$$\int_M (B-\bar{B})^2 \leq \frac{n}{|nc-1|}\Big(\int_M |\mathring{\mathrm{T}}|^2\Big)^{1/2}\Big[\frac{n-1}{n}\int_M (\Delta f)^2 - \int_M \operatorname{Ric}(\nabla f, \nabla f)\Big]^{1/2}. \tag{2-5}$$

By (2-2), $f \equiv 0$ if and only if $B - \bar{B} \equiv 0$ on M. In this case, (1-9) and (1-10) obviously hold. In the following we only consider that f is not identically zero.

Since the Ricci curvature has $\mathrm{Ric} \geq -(n-1)K$ on M,

$$\int_M \mathrm{Ric}(\nabla f, \nabla f) \geq -(n-1)K \int_M |\nabla f|^2. \tag{2-6}$$

By (2-6), (2-5) turns into

$$\int_M (B-\bar{B})^2 \leq \frac{n}{|nc-1|}\left(\int_M |\mathring{T}|^2\right)^{1/2}\left[\frac{n-1}{n}\int_M (\Delta f)^2 + (n-1)K\int_M |\nabla f|^2\right]^{1/2}. \tag{2-7}$$

Since the first nonzero eigenvalue λ_1 of the Laplace operator on M satisfies

$$\lambda_1 = \inf\left\{\frac{\int_M |\nabla\varphi|^2}{\int_M \varphi^2} : \varphi \in C^\infty(M) \text{ is not identically zero and } \int_M \varphi = 0\right\}$$

and

$$\begin{aligned}\int_M |\nabla f|^2 &= -\int_M f\Delta f = -\int_M f(B-\bar{B}) \\ &\leq \left(\int_M f^2\right)^{1/2}\left[\int_M (B-\bar{B})^2\right]^{1/2} \\ &\leq \left(\frac{\int_M |\nabla f|^2}{\lambda_1}\right)^{1/2}\left[\int_M (B-\bar{B})^2\right]^{1/2},\end{aligned}$$

we have

$$\int_M |\nabla f|^2 \leq \frac{1}{\lambda_1}\int_M (B-\bar{B})^2. \tag{2-8}$$

Substitute (2-8) into (2-7) and note that $K \geq 0$. We have

$$\begin{aligned}&\int_M (B-\bar{B})^2 \\ &\leq \frac{n}{|nc-1|}\left(\int_M |\mathring{\mathrm{T}}|^2\right)^{1/2}\left[\frac{n-1}{n}\int_M (B-\bar{B})^2 + \left(\frac{(n-1)K}{\lambda_1}\right)\int_M (B-\bar{B})^2\right]^{1/2} \\ &= \frac{n^{1/2}(n-1)^{1/2}}{|nc-1|}\left(1+\frac{nK}{\lambda_1}\right)^{1/2}\left[\int_M |\mathring{T}|^2\right]^{1/2}\left[\int_M (B-\bar{B})^2\right]^{1/2},\end{aligned} \tag{2-9}$$

which implies that

$$\int_M (B-\bar{B})^2 \leq \frac{n(n-1)}{(nc-1)^2}\left(1+\frac{nK}{\lambda_1}\right)\int_M |\mathring{T}|^2. \tag{2-10}$$

Thus we have inequality (1-9):

$$(nc-1)^2\int_M (B-\bar{B})^2 \leq n(n-1)\left(1+\frac{nK}{\lambda_1}\right)\int_M \left|T - \frac{B}{n}g\right|^2.$$

From the identity

$$|T-(\bar{B}/n)g|^2=|T-(B/n)g|^2+(1/n)(B-\bar{B})^2,$$

we obtain (1-10):

$$(nc-1)^2\int_M\left|T-\frac{\bar{B}}{n}g\right|^2\le\left[(nc-1)^2+(n-1)\left(1+\frac{nK}{\lambda_1}\right)\right]\int_M\left|T-\frac{B}{n}g\right|^2.$$

Now, assuming positivity of the Ricci curvature Ric of M, we may prove the case of equalities in (1-9) and (1-10). Obviously, if $T=(B/n)g$ on M, the equalities in (1-9) and (1-10) hold. On the other hand, suppose the equality in (1-9) (or, equivalently, (1-10)) holds. If $c=1/n$, it is obvious that $T=(B/n)g$ on M. If $c\neq 1/n$, we may take $K=0$. By the proof of (1-9), the equality in (1-9) holds if and only if

(1) $\mathrm{Ric}(\nabla f,\nabla f)=0$ on M and

(2) $T-B/ng$ and $\nabla^2 f-1/n(\Delta f)g$ are linearly dependent.

Note that Ric >0 and (1) holds. $\nabla f\equiv 0$ on M must hold. Then $f\equiv 0$. Thus $B=\bar{B}$ on M. By (1-9), we obtain that $T=(B/n)g$ on M. Hence conclusions (i) and (ii) are equivalent. Obviously (iii) implies (ii). When $c\neq 1/n$, if (ii) holds, by the above argument, (ii) implies $B=\bar{B}$ on M. Thus (iii) also holds. □

Corollary 2.2. *Besides the assumptions and notation of Theorem 1.7, suppose the constant c satisfies $c\neq 1/n$. Then*

$$\int_M(B-\bar{B})^2\le C_{(Kd^2)}\int_M\left|T-\frac{B}{n}g\right|^2 \tag{2-11}$$

and

$$\int_M\left|T-\frac{\bar{B}}{n}g\right|^2\le\bar{C}_{(Kd^2)}\int_M\left|T-\frac{B}{n}g\right|^2, \tag{2-12}$$

where d denotes the diameter of M and $C_{(Kd^2)}$ and $\bar{C}_{(Kd^2)}$ are constants only depending on Kd^2.

Proof. When Ric $\ge -(n-1)K$, where the constant $K>0$, Li and Yau [1980] proved that the first nonzero eigenvalue λ_1 has the lower bound

$$\lambda_1\ge\alpha=\frac{1}{(n-1)d^2\exp[1+\sqrt{1+4(n-1)^2Kd^2}]},$$

where d denotes the diameter of M. So

$$\frac{K}{\lambda_1} \le \frac{K}{\alpha} = (n-1)Kd^2 \exp[1+\sqrt{1+4(n-1)^2Kd^2}\,].$$

By Theorem 1.7, we obtain inequality (2-11) with the constant

$$C_{(Kd^2)} = \frac{n(n-1)}{(nc-1)^2}\big(1+n(n-1)Kd^2 \exp[1+\sqrt{1+4(n-1)^2Kd^2}\,]\big).$$

Inequality (2-11) implies (2-12). □

Remark 2.3. There are other lower estimates α of λ_1 using the diameter d and negative lower bound $-(n-1)K$ of the Ricci curvature (see, for example, [Kalka et al. 1997]). Hence we may have other values of constants $C_{(Kd^2)}$ and $\overline{C}_{(Kd^2)}$.

3. Newton transformations and the r-th elementary symmetric function

Let $\sigma_r : \mathbb{R}^r \to \mathbb{R}$ denote the elementary symmetric function in $\mathbb{R}^n$ given by

$$\sigma_r(x_1, \dots, x_n) = \sum_{i_1 < \cdots < i_r} x_{i_1} \cdots x_{i_r}, \quad 1 \le r \le n.$$

Let V be an n-dimensional vector space and $A : V \to V$ be a symmetric linear transformation. If $\eta_1, \dots, \eta_n$ are the eigenvalues of A corresponding to the orthonormal eigenvectors $\{e_i\}$, $i = 1, \dots, n$, respectively, define the r-th symmetric functions $\sigma_r(A)$ associated with A by

$$\begin{aligned} &\sigma_0(A) = 1, \\ &\sigma_r(A) = \sigma_r(\eta_1, \dots, \eta_n),\ 1 \le r \le n. \end{aligned} \tag{3-1}$$

For convenience of notation, we simply denote $\sigma_r(A)$ by σ_r if there is no confusion. The Newton transformations $P_r : V \to V$ associated with A, $0 \le r \le n$ are defined by

$$\begin{aligned} &P_0 = I, \\ &P_r = \sum_{j=0}^{r} (-1)^j \sigma_{r-j} A^j = \sigma_r I - \sigma_{r-1} A + \cdots + (-1)^r A^r, \quad r = 1, \dots, n. \end{aligned}$$

By definition, $P_r = \sigma_r I - AP_{r-1}$, $P_n = 0$. It was proved in [Reilly 1973] that P_r has the following basic properties:

(i) $P_r(e_i) = \dfrac{\partial \sigma_{r+1}}{\partial \eta_i} e_i$.

(ii) $\mathrm{tr}(P_r) = (n-r)\sigma_r$.

(iii) $\mathrm{tr}(AP_r) = (r+1)\sigma_{r+1}$.

Clearly, each P_r corresponds to a symmetric $(2,0)$-tensor on V, still denoted by P_r.

4. High-order mean curvatures of hypersurfaces in space forms

Assume $(N, \tilde{g})$ is an $(n+1)$-dimensional Riemannian manifold, $n \geq 2$. Suppose (Σ, g) is a smooth connected oriented closed hypersurface immersed in $(N, \tilde{g})$ with induced metric g. Let ν denote the outward unit normal to Σ, and $A = (h_{ij})$, the second fundamental form $A : T_p\Sigma \otimes_s T_p\Sigma \to \mathbb{R}$, defined by $A(X, Y) = -\langle \widetilde{\nabla}_X Y, \nu \rangle$, where $X, Y \in T_p\Sigma$, $p \in \Sigma$, and $\widetilde{\nabla}$ denotes the Levi-Civita connection of $(N, \tilde{g})$. A determines an equivalent $(1, 1)$-tensor, called the shape operator A of

$$\Sigma : T_p\Sigma \to T_p\Sigma,$$

given by $AX = \widetilde{\nabla}_X \nu$. Σ is called totally umbilical if A is a multiple of its metric g at every point of Σ, that is, $A = (\operatorname{tr} A/n)g$ on Σ. Now we recall the definition of r-th mean curvatures of a hypersurface, which was introduced in [Reilly 1973]; compare [Rosenberg 1993].

Let η_i, $i = 1, \ldots, n$ denote the principle curvatures of Σ at p, which are the eigenvalues of A at p corresponding to the orthonormal eigenvectors $\{e_i\}$, $i = 1. \ldots, n$, respectively. By Section 3, we have the r-th symmetric functions $\sigma_r(A)$ associated with A, denoted by $s_r = \sigma_r(A)$, and the Newton transformations P_r associated with A at p, $0 \leq r \leq n$.

Definition 4.1. The r-th mean curvature H_r of Σ at p is defined by $s_r = \binom{n}{r} H_r$, $0 \leq r \leq n$.

For instance, $H_1 = s_1/n = H/n$ (in this paper, we also call $H = \operatorname{tr} A$ the mean curvature of Σ, consistent with earlier papers [Perez 2011; Cheng and Zhou 2012], among others). H_n is the Gauss–Kronecker curvature. When the ambient space N is a space form N_a^{n+1} with constant sectional curvature a,

$$\begin{aligned} \operatorname{Ric} &= (n-1)aI + HA - A^2, \\ R = \operatorname{tr} \operatorname{Ric} &= n(n-1)c + H^2 - |A|^2 = n(n-1)a + 2s_2. \end{aligned}$$

Hence H_2 is, modulo a constant, the scalar curvature of Σ.

Lemma 4.2 ([Reilly 1973]; cf. [Rosenberg 1993; Alías et al. 2006]). *When the ambient space is a space form N_a^{n+1}, we have* $\operatorname{div} P_r = 0$, *for* $0 \leq r \leq n$.

Proof of Theorem 1.9. By Section 3, $\operatorname{tr} P_r = (n-r)s_r$. Denote $\bar{s}_r = (1/\operatorname{Vol}\Sigma)\int_\Sigma s_r$. By Lemma 4.2, $\operatorname{div} P_r = 0$. Take $T = P_r$ and $B = (n-r)s_r$ in Theorem 1.7. Then

$$(n-r)^2 \int_\Sigma (s_r - \bar{s}_r)^2 \leq n(n-1)\left(1 + \frac{nK}{\lambda_1}\right) \int_\Sigma \left| P_r - \frac{(n-r)s_r}{n} g \right|^2;$$

equivalently,

$$\int_\Sigma \left| P_r - \frac{(n-r)\bar{s}_r}{n} g \right|^2 \leq n\left(1 + \frac{(n-1)K}{\lambda_1}\right) \int_\Sigma \left| P_r - \frac{(n-r)s_r}{n} g \right|^2,$$

which are (1-13) and (1-14), respectively.

Now we prove conclusions (1) and (2) in Theorem 1.9. If the Ricci curvature of Σ is positive, by Theorem 1.7, conclusion (1) holds and $s_r = \bar{s}_r$ is constant on Σ. If Σ is also embedded in $\mathbb{R}^{n+1}$, by a theorem of Ros [1987] stating that a closed embedded hypersurface in $\mathbb{R}^{n+1}$ with constant r-th mean curvature must be a distance sphere $\mathbb{S}^{n+1}$, $2 \le r \le n$, we obtain conclusion (2). □

Remark 4.3. If $r = 1$, $P_1 = s_1 I - A = HI - A$. P_1 is equivalent to the symmetric $(2,0)$-tensor $P_1 = Hg - A$. So (1-13) turns into

$$\begin{aligned}\int_\Sigma (H-\overline{H})^2 &\le \frac{n}{n-1}\left(1+\frac{nK}{\lambda_1}\right)\int_\Sigma \left|Hg - A - \frac{(n-1)H}{n}g\right|^2 \\ &= \frac{n}{n-1}\left(1+\frac{nK}{\lambda_1}\right)\int_\Sigma \left|A - \frac{H}{n}g\right|^2.\end{aligned} \tag{4-1}$$

In particular, if $K = 0$,

$$\int_\Sigma (H-\overline{H})^2 \le \frac{n}{n-1}\int_\Sigma \left|A - \frac{H}{n}g\right|^2. \tag{4-2}$$

Equations (4-1) and (4-2) are (1-8) and (1-6), respectively, which were proved in [Cheng and Zhou 2012] and [Perez 2011], respectively, if Σ is a closed hypersurface immersed in an Einstein manifold. This is because $\operatorname{div} P_1 = 0$ even if the ambient space is Einstein.

When $r = 2$, we have $2s_2 = R - n(n-1)a$,

$$P_2 = s_2 I - s_1 A + s_0 A^2 = \frac{R-(n-2)(n-1)a}{2} I - \mathrm{Ric},$$

and, by direct computation,

$$P_2 - \frac{(n-2)s_2}{n} g = \frac{R}{n} I - \mathrm{Ric}\,.$$

As a symmetric $(2,0)$-tensor, $P_2 = (R/n)g - \mathrm{Ric}$. Hence (1-13) turns into

$$\int_\Sigma (s_2 - \bar{s}_2)^2 \le \frac{n(n-1)}{(n-2)^2}\left(1+\frac{nK}{\lambda_1}\right)\int_\Sigma \left|P_2 - \frac{(n-2)s_2}{n}g\right|^2,$$

which is

$$\int_\Sigma (R-\overline{R})^2 \le \frac{4n(n-1)}{(n-2)^2}\left(1+\frac{nK}{\lambda_1}\right)\int_\Sigma \left|\mathrm{Ric} - \frac{R}{n}g\right|^2. \tag{4-3}$$

Equation (4-3) was proved in [Cheng 2013], and, in the case of $K = 0$, in [De Lellis and Topping 2012].

If $r = n$, (1-13) is trivial.

5. *k*-scalar curvature of locally conformal flat manifolds

We first recall the definition of the k-scalar curvatures of a Riemannian manifold, introduced in [Viaclovsky 2000]. If (M^n, g) is an n-dimensional Riemannian manifold, $n \geq 3$, the Schouten tensor of M is

$$S_g = \frac{1}{n-2}\left(\mathrm{Ric} - \frac{1}{2(n-1)} Rg\right).$$

By definition, $S_g : TM \to TM$ is a symmetric (1, 1)-tensor field. By Section 3, we have the symmetric k-th function $\sigma_k(S_g)$ and the Newton transformations $T_k(S_g) = T_k$ associated with S_g, $1 \leq k \leq n$. We call $\sigma_k(S_g)$ the k-scalar curvatures of M

Lemma 5.1 [Viaclovsky 2000]. *If* (M, g) *is locally conformally flat*, *then*, *for* $1 \leq k \leq n$, $\operatorname{div} T_k(S_g) = 0$.

Because of Lemma 5.1, we can applying Theorem 1.7 to $T_k(S_g)$ to obtain Theorem 1.11.

Remark 5.2. When $k = 1$, $\sigma_1(S_g) = \operatorname{tr} S_g = R/(2(n-1))$ and $T_1 = \sigma_1(S_g)I - S_g$. As a symmetric (2, 0)-tensor, $T_1 = -(1/(n-2))(\mathrm{Ric} - Rg/2)$. Hence (1-17) turns into (1-3),

$$\int_M (R - \bar{R})^2 \leq \frac{4n(n-1)}{(n-2)^2}\left(1 + \frac{nK}{\lambda_1}\right)\int_M \left|\mathrm{Ric} - \frac{R}{n}g\right|^2,$$

and, in particular, if $K = 0$, (1-17) turns into (1-1),

$$\int_M (R - \bar{R})^2 \leq \frac{4n(n-1)}{(n-2)^2}\int_M \left|\mathrm{Ric} - \frac{R}{n}g\right|^2.$$

Equations (1-3) and (1-1) were proved in [Cheng 2013] and [De Lellis and Topping 2012], respectively, without the hypothesis that M is locally conformally flat. The reason is that $\operatorname{div} T_1 = 0$ (the contracted second Bianchi identity) holds on any Riemannian manifold.

Acknowledgements

The author would like to thank the referee for some suggestions.

References

[Aleksandrov 1958] A. D. Aleksandrov, "Uniqueness theorems for surfaces in the large. V", *Am. Math. Soc.* **21** (1958), 412–416. Zbl 0119.16603

[Alías et al. 2006] L. J. Alías, J. H. S. de Lira, and J. M. Malacarne, "Constant higher-order mean curvature hypersurfaces in Riemannian spaces", *J. Inst. Math. Jussieu* **5**:4 (2006), 527–562. MR 2007i:53062 Zbl 1118.53038

[Barbosa and Colares 1997] J. L. M. Barbosa and A. G. Colares, "Stability of hypersurfaces with constant r-mean curvature", *Ann. Global Anal. Geom.* **15**:3 (1997), 277–297. MR 98h:53091 Zbl 0891.53044

[Cheng 2013] X. Cheng, "A generalization of almost-Schur lemma for closed Riemannian manifolds", *Ann. Global Anal. Geom.* **43**:2 (2013), 153–160. MR 3019161 Zbl 06147899

[Cheng and Rosenberg 2005] X. Cheng and H. Rosenberg, "Embedded positive constant r-mean curvature hypersurfaces in $M^m \times \mathbf{R}$", *An. Acad. Brasil. Ciênc.* **77**:2 (2005), 183–199. MR 2006e:53105 Zbl 1074.53049

[Cheng and Zhou 2012] X. Cheng and D. Zhou, "Rigidity for nearly umbilical hypersurfaces in space forms", *J. Geom. Anal* (2012).

[De Lellis and Müller 2005] C. De Lellis and S. Müller, "Optimal rigidity estimates for nearly umbilical surfaces", *J. Differential Geom.* **69**:1 (2005), 75–110. MR 2006e:53078 Zbl 1087.53004

[De Lellis and Topping 2012] C. De Lellis and P. M. Topping, "Almost-Schur lemma", *Calc. Var. Partial Differential Equations* **43**:3-4 (2012), 347–354. MR 2012k:53062 Zbl 1236.53036

[Ge and Wang 2011] Y. Ge and G. Wang, "A new conformal invariant on 3-dimensional manifolds", preprint, 2011.

[Ge and Wang 2012] Y. Ge and G. Wang, "An almost Schur theorem on 4-dimensional manifolds", *Proc. Amer. Math. Soc.* **140**:3 (2012), 1041–1044. MR 2869088 Zbl 1238.53026

[Ge et al. 2013] Y. Ge, G. Wang, and X. Chao, "On problems related to an inequality of Andrews, De Lellis, and Topping", 2013.

[Guan 2002] P. Guan, "Topics in geometric fully nonlinear equations", lecture notes, 2002, Available at www.math.mcgill.ca/guan/zheda0508.pdf.

[Kalka et al. 1997] M. Kalka, E. Mann, D. Yang, and A. Zinger, "The exponential decay rate of the lower bound for the first eigenvalue of compact manifolds", *Internat. J. Math.* **8**:3 (1997), 345–355. MR 98e:58172 Zbl 0884.58096

[Li and Yau 1980] P. Li and S. T. Yau, "Estimates of eigenvalues of a compact Riemannian manifold", pp. 205–239 in *Geometry of the Laplace operator* (Honolulu, 1979), edited by R. Osserman and A. Weinstein, Proc. Sympos. Pure Math. **36**, Amer. Math. Soc., Providence, R.I., 1980. MR 81i:58050 Zbl 0441.58014

[Perez 2011] D. Perez, *On nearly umbilical hypersurfaces*, Dr.sc.nat. thesis, Uniersität Zürich, 2011.

[Reilly 1973] R. C. Reilly, "Variational properties of functions of the mean curvatures for hypersurfaces in space forms", *J. Differential Geometry* **8** (1973), 465–477. MR 49 #6102 Zbl 0277.53030

[Ros 1987] A. Ros, "Compact hypersurfaces with constant higher order mean curvatures", *Rev. Mat. Iberoamericana* **3**:3-4 (1987), 447–453. MR 90c:53160 Zbl 0673.53003

[Ros 1988] A. Ros, "Compact hypersurfaces with constant scalar curvature and a congruence theorem", *J. Differential Geom.* **27**:2 (1988), 215–223. MR 89b:53096 Zbl 0638.53051

[Rosenberg 1993] H. Rosenberg, "Hypersurfaces of constant curvature in space forms", *Bull. Sci. Math.* **117**:2 (1993), 211–239. MR 94b:53097 Zbl 0787.53046

[Viaclovsky 2000] J. A. Viaclovsky, "Conformal geometry, contact geometry, and the calculus of variations", *Duke Math. J.* **101**:2 (2000), 283–316. MR 2001b:53038 Zbl 0990.53035

[Viaclovsky 2006] J. Viaclovsky, "Conformal geometry and fully nonlinear equations", pp. 435–460 in *Inspired by S. S. Chern*, edited by P. A. Griffiths, Nankai Tracts Math. **11**, World Sci. Publ., Hackensack, NJ, 2006. MR 2008c:53030 Zbl 1142.53030

Received September 21, 2012. Revised February 19, 2013.

Xu Cheng
Instituto de Matemática
Universidade Federal Fluminense
24020-140 Niterói, RJ
Brazil
xcheng@impa.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.341

ALGEBRAIC INVARIANTS, MUTATION, AND COMMENSURABILITY OF LINK COMPLEMENTS

ERIC CHESEBRO AND JASON DEBLOIS

We construct a family of hyperbolic link complements by gluing tangles along totally geodesic four-punctured spheres, then investigate the commensurability relation among its members. Those with different volume are incommensurable, distinguished by their scissors congruence classes. Mutation produces arbitrarily large finite subfamilies of nonisometric manifolds with the same volume and scissors congruence class. Depending on the choice of mutation, these manifolds may be commensurable or incommensurable, distinguished in the latter case by cusp parameters. All have trace field $\mathbb{Q}(i, \sqrt{2})$; some have integral traces while others do not.

1. Introduction

Manifolds are *commensurable* if they have a common cover, of finite degree over each. W. P. Thurston first studied the commensurability relation among hyperbolic knot and link complements in S^3, describing commensurable and incommensurable examples in his notes [Thurston 1979, Chapter 6]. The families of chain link complements [Neumann and Reid 1992], two-bridge knot complements [Reid and Walsh 2008], and certain pretzel knot complements [Macasieb and Mattman 2008] have since been further explored. Here we construct another infinite family of hyperbolic link complements and explore the commensurability relation among its members.

We compute the following invariants on members of our family. For $\Gamma < \mathrm{PSL}_2(\mathbb{C})$ the *trace field* of $M = \mathbb{H}^3/\Gamma$ is the smallest field containing the traces of elements of Γ. If each such trace is an algebraic integer we say M has *integral traces*. The *cusp parameters* of M, used in [Thurston 1979; Neumann and Reid 1992], are algebraic invariants of the Euclidean structures on horospherical cross sections of the cusps of M. The *Bloch invariant* [Neumann and Yang 1999] is determined by a polyhedral decomposition.

DeBlois was partially supported by NSF grant DMS-1240329.
MSC2010: primary 57M10, 57M50; secondary 20F55, 22E40.
Keywords: commensurability, mutation, Bloch invariant, link, trace field.

Theorem 1. *For each $n \in \mathbb{N}$, there is a link $L_n \subset S^3$ such that $M_n = S^3 - L_n$ is hyperbolic with trace field $\mathbb{Q}(i, \sqrt{2})$ and integral traces. If $m \neq n$ then M_m and M_n are incommensurable, distinguished by their Bloch invariants and cusp parameters.*

Having integral traces is commensurability-invariant [Maclachlan and Reid 2003, §5.2], and the trace field is a commensurability invariant of link complements [Maclachlan and Reid 2003, Corollary 4.2.2]. Commensurable manifolds have $\mathbb{Q}$-dependent Bloch invariants and $\mathrm{PGL}_2(\mathbb{Q})$-dependent cusp parameters, but the M_n have neither (see Proposition 4.7 and Lemma 4.18).

Figure 1 depicts L_2. The gray lines there indicate the presence of 2-spheres that each meet L_2 in four points, separating it left-to-right into a tangle S in the three-ball B^3, two copies of a tangle $T \subset S^2 \times I$, and the mirror image $\bar{S}$ of S. For arbitrary $n \in \mathbb{N}$, the link L_n is constructed analogously, using S, $\bar{S}$, and n copies of T. We number the corresponding 2-spheres for L_n as $S^{(i)}$ for $0 \leq i \leq n$, so that $S^{(0)}$ bounds S, $S^{(n)}$ bounds $\bar{S}$, and $S^{(i)}$ bounds a copy of T with $S^{(i-1)}$ for $0 < i \leq n$.

We also describe the commensurability relation among the complements of links related to the L_n by *mutation* along the $S^{(i)}$: cutting along $S^{(i)}$ and regluing by an order-two mapping class that preserves $S^{(i)} \cap L_n$ and acts on it as an even permutation. With L_n projected as in Figure 1, for each i we mark the points of $S^{(i)} \cap L_n$ by 2, 3, 4, and 1, reading top to bottom, and refer to a mutation homeomorphism of $S^{(i)}$ by its permutation representation.

Below, for $n \in \mathbb{N}$ and $I \in \{0, 1, 2\}^{n+1}$ let L_I be the link obtained from L_n by the mutation $(13)(24)$ (respectively, $(12)(34)$) along $S^{(i)}$, for each i such that the i-th entry of I is 1 (respectively, 2). Write $M_I = S^3 - L_I$ for each such I.

Theorem 2. *For $n \in \mathbb{N}$ and $I = (a_0, \dots, a_n) \in \{0, 1\}^{n+1}$, M_I is commensurable to M_n. For $J = (b_0, \dots, b_n) \in \{0, 1\}^{n+1}$, M_J is isometric to M_I if and only if either $b_i = a_i$ for each $i \in \{1, \dots, n-1\}$ or $b_i = a_{n-i}$ for each such i.*

We will show in a future paper that Theorem 2 reflects the fact that M_n has a *hidden symmetry* (see, for example, [Neumann and Reid 1992]) arising from a *hidden extension* of the mutation $(13)(24)$, an extension of a lift of $(13)(24)$ over a finite cover of $(S^2 \times I) - T$.

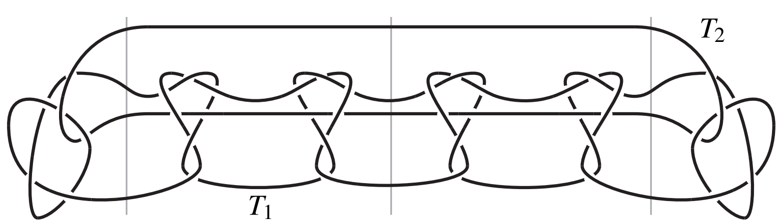


Figure 1. The link L_2.

Here we prove Theorem 2 more directly, identifying an orbifold O_n jointly covered by M_n and the M_I; see Proposition 6.4. The key advantage of this approach is that we can also prove the isometry classification above (see Proposition 6.6) using the fact that O_n is minimal in the commensurability class of M_n (Corollary 6.5).

Corollary 6.5 is proved following an idea of Goodman, Heard, and Hodgson [Goodman et al. 2008]. The key step, for each n, is to construct a tiling $\mathcal{T}_n$ of $\mathbb{H}^3$ by convex polyhedra that is *canonical* in the sense of [Goodman et al. 2008, §2]. See Theorem 4. This is of independent interest, as there are few infinite families for which canonical tilings have been identified.

The mutation (12)(34) has a very different effect than (13)(24).

Theorem 3. *For $n \in \mathbb{N}$, let $\mathcal{L}_n = \{L_I \mid I \in \{0, 2\}^{n+1}\}$. Then*:

(1) *For each $I \in \{0, 2\}^{n+1} - \{(0, \dots, 0)\}$, M_I has the same volume, Bloch invariant, and trace field as M_n, but has a nonintegral trace.*

(2) *There is a subfamily of $\mathcal{L}_n$ with at least $n/2$ mutually incommensurable members, distinguished by their cusp parameters.*

(3) *There is a subfamily of $\mathcal{L}_n$ with n members which all share cusp parameters.*

Remarks. 1. Mutation along 4-punctured spheres preserves hyperbolic volume [Ruberman 1987], the trace field [Neumann and Reid 1991], and the Bloch invariant [Neumann 2011, Theorem 2.13]. While unaware of the Bloch invariant reference we proved our case directly; see Proposition 7.2.

2. $L_n = L_{(0,\dots,0)}$, which accounts for the gap in statement (1) of the theorem.

3. Corollaries 7.4 and 7.5 describe the subfamilies from (2) and (3) above. We do not know the commensurability relation among members of the latter subfamily.

Theorems 2 and 3 comprise the first study (to our knowledge) of commensurability among an infinite family of link complements related by mutation. Mutants have a longstanding reputation for being difficult to distinguish, although the algorithm of [Goodman et al. 2008] can now be used to check particular examples. (For instance, the complement of the Kinoshita–Terasaka knot, 11n42 in the knot tables, is incommensurable with that of its mutant, the Conway knot, 11n34.)

Theorem 2 further gives some evidence counter to the following conjecture of Reid and Walsh [2008]: the commensurability class of a hyperbolic knot complement in S^3 contains at most two others. This implies in particular that any hyperbolic knot complement is incommensurable with all but two of its (nonisometric) mutants.

We now outline the remainder of the paper. We name the tangle complements $M_S \doteq B^3 - S$ and $M_T \doteq S^2 \times I - T$, and note that M_T is the double of $M_{T_0} \doteq M_T \cap (S^2 \times [0, 1/2])$ across a single boundary component. Section 2 describes hyperbolic structures with totally geodesic boundary on M_S and M_{T_0} as identification spaces of the regular ideal octahedron and the right-angled ideal cuboctahedron, respectively.

The totally geodesic boundary ∂M_S is isometric to the component of ∂M_{T_0} contained in ∂M_T, and the reflective symmetry of M_T ensures that its totally geodesic boundary components are orientation-reversing isometric. In forming M_n we glue $\partial B^3 - S$ to $S^2 \times \{0\} - T$ by a map isotopic to an isometry, so that the separating four-punctured spheres $F^{(i)} = S^{(i)} - T$ are totally geodesic in M_n for $0 \le i \le n$. Section 3 describes this assembly.

Because the $F^{(i)}$ are totally geodesic, each copy of M_S and M_T in M_n inherits its structure with totally geodesic boundary from the ambient hyperbolic structure. This in turn makes it possible to compute the commensurability invariants of Theorem 1 on the M_n. We carry this out in Section 4. Few other link complements are known to contain a surface that is totally geodesic without some topological constraint forcing it so; see, for example, [Maclachlan and Reid 1991; Aitchison and Rubinstein 1997]. For related results see [Menasco and Reid 1992; Adams and Schoenfeld 2005; Leininger 2006; Adams et al. 2008].

Our method of construction owes a debt to one that Adams [1985] and Neumann and Reid used to produce families of hyperbolic 3-manifolds, gluing together manifolds with 3-punctured sphere boundary. (However unlike the 4-punctured sphere, a 3-punctured sphere is totally geodesic in any hyperbolic 3-manifold that contains it [Adams 1985, Theorem 3.1].) The work of Neumann and Reid (see [Maclachlan and Reid 2003, §5.6]) can be can be combined with an argument like the one in Proposition 4.2 to show that for each imaginary biquadratic extension k of $\mathbb{Q}$, there are infinitely many commensurability classes of hyperbolic 3-manifolds with trace field k.

In every hyperbolic 4-punctured sphere, each mutation determines a homeomorphism properly isotopic to an isometry [Ruberman 1987]. In Section 5 we describe the isometries determined by $(13)(24)$ and $(12)(34)$ and the hyperbolic structures on mutants of the M_n. We prove Theorem 2 in Section 6 and Theorem 3 in Section 7.

2. A pair of tangles

This section is devoted to describing hyperbolic structures with totally geodesic boundary on the complements of the tangles S, in B^3, and T_0, in $S^2 \times I$, depicted in Figure 2. For a manifold M with boundary, we refer by a *tangle* in M to a pair (M, T), where T is the image of a disjoint union of circles and closed intervals, embedded in M by a map taking each circle into the interior of M and restricting on each interval to a proper embedding.

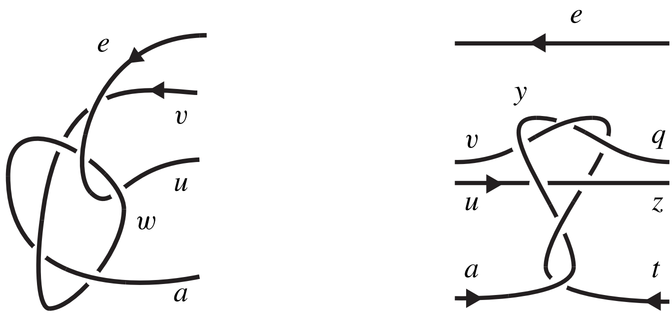


Figure 2. Tangles S and T_0, labeled with Wirtinger generators.

We will prove there is a homeomorphism taking $B^3 - S$ to a hyperbolic manifold with totally geodesic boundary which is an identification space of an ideal octahedron by pairing certain faces. This was previously known, and it follows from results in [Paoluzzi and Zimmermann 1996] upon taking a geometric limit, but we do not know a reference for a direct proof. We also prove there is a homeomorphism taking $S^2 \times I - T_0$ to a certain identification space of the right-angled ideal cuboctahedron. As far as we are aware, this description was not previously known.

We prove existence of homeomorphisms using faithful representations, from the fundamental groups of tangle complements onto Kleinian groups generated by face pairings. Our main tools drawing connections between the geometric, algebraic, and topological objects involved are Lemma 2.1, which relates a hyperbolic 3-manifold with totally geodesic boundary produced by pairing some faces of a right-angled polyhedron to the Kleinian group generated by the face-pairing isometries, and Lemma 2.6, which describes a homeomorphism from a pared manifold M and the convex core of $\mathbb{H}^3/\Gamma$, where Γ is a Kleinian group isomorphic to $\pi_1(M)$.

In the remainder of the paper, we will let $\mathbb{H}^3 = \{(z, t) \mid z \in \mathbb{C}, t \in (0, \infty)\}$, the upper half-space model of hyperbolic space, equipped with the complete Riemannian metric of constant sectional curvature -1. In this model, the group of orientation-preserving isometries, $\mathrm{PSL}_2(\mathbb{C})$, acts by extending its action by Möbius transformations on the *ideal boundary* or *sphere at infinity* $\mathbb{C} \cup \{\infty\}$.

The *horosphere of height t centered at* ∞ is $\mathbb{C} \times \{t\} \subset \mathbb{H}^3$. This inherits the Euclidean metric, scaled by $1/t$, from the ambient hyperbolic metric. For $v \in \mathbb{C} \times \{0\}$, a *horosphere centered at* v is a Euclidean sphere in $\mathbb{C} \times \mathbb{R}$ centered at a point in $\mathbb{H}^3$ and tangent to $\mathbb{C} \times \{0\}$ at $(v, 0)$. It is a standard fact that isometries of $\mathbb{H}^3$ take horospheres to horospheres.

A *hyperplane* of $\mathbb{H}^3$ is a totally geodesic subspace of the form $\ell \times \mathbb{R}^+$ for a line $\ell \subset \mathbb{C}$, or the intersection with $\mathbb{H}^3$ of a Euclidean sphere centered at a point in $\mathbb{C} \times \{0\}$. A *half-space* is the closure of a component of the complement in $\mathbb{H}^3$ of a hyperplane, and a *polyhedron* is the nonempty intersection of a collection of half-spaces with the property that the corresponding collection of defining hyperplanes is locally finite. A *face* of a polyhedron is its intersection with one of its defining hyperplanes. A polyhedron is *right angled* if its defining hyperplanes meet at right angles (if at all) and *ideal* if any point at which more than two of its defining hyperplanes meet is on the sphere at infinity. Such points are *ideal vertices*.

We say a polyhedron $\mathcal{P} \subset \mathbb{H}^3$ is *checkered* if its set of faces is partitioned into sets $\mathcal{S}_i$ and $\mathcal{S}_e$ of *internal* and *external* faces, respectively, so that each $f \in \mathcal{S}_i$ intersects only faces in $\mathcal{S}_e$ and vice versa. For a face f of a checkered, right-angled ideal polyhedron $\mathcal{P}$, let $\mathcal{H}_f$ be the geodesic hyperplane in $\mathbb{H}^3$ containing f and let U_f be the half-space bounded by $\mathcal{H}_f$ that contains $\mathcal{P}$. Let the *expansion* of $\mathcal{P}$ be

$$E(\mathcal{P}) = \bigcap_{f \in \mathcal{S}_i} U_f.$$

The expansion $E(\mathcal{P})$ is a polyhedron of infinite volume that contains $\mathcal{P}$, and the components of the frontier of $\mathcal{P}$ in $E(\mathcal{P})$ are the external faces of $\mathcal{P}$.

An *internal face pairing* for a checkered polyhedron $\mathcal{P} \subset \mathcal{H}^3$ is a collection $\{\phi_f \mid f \in \mathcal{S}_i\}$ of isometries, such that for each $f \in \mathcal{S}_i$ there exists $f' \in \mathcal{S}_i$ with $\phi_f(f) = f'$, $\phi_f(\mathcal{P}) \cap \mathcal{P} = f'$, and $\phi_{f'} = \phi_f^{-1}$. It is *proper* if $f' \neq f$ for all $f \in \mathcal{S}_i$. A proper internal face pairing determines a *proper* Isom($\mathbb{H}^3$)*-side-pairing* of the expansion $E(\mathcal{P})$, in the sense of [Ratcliffe 1994, §10.1]. (In [Ratcliffe 1994], *faces* are called *sides*.)

Given a proper internal face pairing $\{\phi_f\}$ of a checkered polyhedron $\mathcal{P}$, [Ratcliffe 1994, Theorem 10.1.2] implies the identification space $E(\mathcal{P})/\{\phi_f\}$, determined by setting $x \sim \phi_f(x)$ for all $f \in \mathcal{S}_i$ and $x \in f$, is a hyperbolic manifold. The inclusion $\mathcal{P} \hookrightarrow E(\mathcal{P})$ induces an inclusion from $M_{\mathcal{P}} \doteq \mathcal{P}/\{\phi_f\}$ to $E(\mathcal{P})/\{\phi_f\}$. For each edge e of each $g \in \mathcal{S}_e$, there is a unique $f \in \mathcal{S}_i$ such that $e \subset f \cap g$. Since $f' = \phi_f(f)$ intersects a unique $g' \in \mathcal{S}_i$ along $\phi_f(e)$, the internal face pairing for $\mathcal{P}$ determines an edge pairing for the disjoint union of external faces of $\mathcal{P}$. Thus $M_{\mathcal{P}} \doteq \mathcal{P}/\{\phi_f\}$ is an isometrically embedded submanifold of $E(\mathcal{P})/\{\phi_f\}$, where $\partial M_{\mathcal{P}}$ is the quotient of the disjoint union of the external faces by the edge pairing induced by $\{\phi_f\}$.

Given an edge pair $\{e, e'\}$ for $\partial M_{\mathcal{P}}$, the total angle around this edge in $M_{\mathcal{P}}$ is the sum of the dihedral angles for e and e' in $\mathcal{P}$. Therefore, if $\mathcal{P}$ is right angled, $\partial M_{\mathcal{P}}$ is totally geodesic.

If Γ is a Kleinian group, we refer to the convex core of $\mathbb{H}^3/\Gamma$ as $C(\Gamma)$. This is the convex submanifold of $\mathbb{H}^3/\Gamma$, minimal with respect to inclusion, with the property that the inclusion-induced homomorphism $\pi_1 C(\Gamma) \to \mathbb{H}^3/\Gamma$ is surjective. (See [Morgan 1984] for background on Kleinian groups. The beginning of §6 therein covers convex cores.)

Lemma 2.1. *Let $\mathcal{P} \subset \mathbb{H}^3$ be a finite-sided, checkered right-angled ideal polyhedron, with a proper internal face pairing $\{\phi_f \mid f \in \mathcal{S}_i\}$. Then $\Gamma \doteq \langle \phi_f \mid f \in \mathcal{S}_i \rangle$ is a free*

Kleinian group, and the inclusion $\mathcal{P} \hookrightarrow \mathbb{H}^3$ *induces an isometry* $p\colon M_{\mathcal{P}} \to C(\Gamma)$. *If* $\mathcal{H}$ *is the hyperplane containing* $g \in \mathcal{S}_e$ *then* $\mathcal{H} \to \mathbb{H}^3$ *induces an isometric embedding of* $\mathcal{H}/\operatorname{Stab}_\Gamma(\mathcal{H})$ *to the component of* $\partial C(\Gamma)$ *containing* $p(g)$.

Proof. We will continue to use some terminology and results from [Ratcliffe 1994]. With these hypotheses the inclusion $\mathcal{P} \to E(\mathcal{P})$ induces an isometric embedding $M_{\mathcal{P}} \to E(\mathcal{P})/\{\phi_f\}$, and $\partial M_{\mathcal{P}}$ is totally geodesic. If $E(\mathcal{P})/\{\phi_f\}$ is complete as a hyperbolic 3-manifold, then by Poincaré's polyhedron theorem (see, for example, [Ratcliffe 1994, Theorem 11.2.2]), $\Gamma = \langle \phi_f \mid f \in \mathcal{S}_i \rangle$ is discrete and $E(\mathcal{P})$ is a fundamental domain for Γ.

By [Ratcliffe 1994, Theorem 11.1.6], to show completeness it suffices to check that the link of any cusp is a complete Euclidean surface. Let $\lfloor v \rfloor = \{v_0, v_1, \dots, v_{n-1}\}$ be an equivalence class of ideal vertices of $\mathcal{P}$ under the relation generated by $x \sim \phi_f(x)$, $f \in \mathcal{S}_i$, enumerated so that for each j there exists $f_j \in \mathcal{S}_i$ with $\phi_{f_j}(v_j) = v_{j+1}$ (taken modulo n). In particular, v_j is an ideal vertex of f_j and also of $f_j' \doteq \phi_{j-1}(f_{j-1})$.

For each j, let $\mathcal{B}_j$ be a horosphere centered at v_j, chosen small enough that $\mathcal{B}_j \cap \mathcal{B}_{j'} = \varnothing$ for $j \neq j'$. Since $\mathcal{P}$ is right angled, $\mathcal{B}_j \cap \mathcal{P}$ is a Euclidean rectangle for each j. We may assume, by renumbering if necessary, that $\mathcal{B}_0 \cap f_0$ has the shortest length of all the arcs $\mathcal{B}_j \cap f_j$. Then since $\phi_0(\mathcal{B}_0) \cap f_1'$ is parallel to $\phi_0(\mathcal{B}_0) \cap f_1$ in $\phi_0(\mathcal{B}_0) \cap \mathcal{P}$, they have the same length: that of $\mathcal{B}_0 \cap f_0$. Since this is less than the length of $\mathcal{B}_1 \cap f_1$, we have $\phi_0(\mathcal{B}_0) \subset \mathcal{B}_1$.

We may replace $\mathcal{B}_1$ by $\phi_0(\mathcal{B}_0)$, then replace $\mathcal{B}_2$ with $\phi_1(\mathcal{B}_1)$, and so on, yielding a new collection of horospheres which are pairwise disjoint and have the additional property that they are interchanged by the face pairings of $\mathcal{P}$. Equivalence classes of ideal vertices of $E(\mathcal{P})$ are the same as those of $\mathcal{P}$; thus this collection satisfies the hypotheses of [Ratcliffe 1994, Theorem 11.1.4], and the link of $\lfloor v \rfloor$ is complete. It follows that $E(\mathcal{P})/\{\phi_f\}$ is a complete hyperbolic 3-manifold.

Now by the polyhedron theorem, Γ is discrete and $E(\mathcal{P})$ is a fundamental domain for Γ. It follows from a ping-pong argument that Γ is free, since the fact that $\mathcal{P}$ is right angled implies that the hyperplanes containing its internal faces are mutually disjoint. The inclusion $E(\mathcal{P}) \to \mathbb{H}^3$ induces an isometry $E(\mathcal{P})/\{\phi_f\} \to \mathbb{H}^3/\Gamma$, so the inclusion $\mathcal{P} \to \mathbb{H}^3$ induces an isometric embedding $p\colon M_{\mathcal{P}} \to \mathbb{H}^3/\Gamma$.

That $p(M_{\mathcal{P}}) \subseteq C(\Gamma)$ will follow from the fact that $\mathcal{P}$ is contained in the convex hull of the limit set $\operatorname{Hull}(\Gamma)$ of Γ, since this is well known to be the universal cover of $C(\Gamma)$. Fixed points of parabolic elements of Γ lie in $\operatorname{Hull}(\Gamma)$, so since $\mathcal{P}$ is the convex hull of its ideal vertices we show that it is in $\operatorname{Hull}(\Gamma)$ by observing that each such vertex is a parabolic fixed point of Γ. Indeed, if $\{v_0, v_1, \dots, v_{n-1}\}$ is an equivalence class of ideal vertices enumerated as we described above, then v_0 is fixed by $\phi_{f_{n-1}} \circ \cdots \circ \phi_{f_1} \circ \phi_{f_0} \in \Gamma$.

Since $p(M_{\mathcal{P}})$ has totally geodesic boundary it is convex (see [Canary et al. 2006, Corollary I.1.3.7]). Thus if $p(M_{\mathcal{P}})$ carries $\pi_1(\mathbb{H}^3/\Gamma)$ then $C(\Gamma) \subseteq p(M_{\mathcal{P}})$. To show this we use the nearest-point retraction $r\colon E(\mathcal{P}) \to \mathcal{P}$ to produce a homeomorphism $M_{\mathcal{P}} \cup_{\partial M_{\mathcal{P}}} (\partial M_{\mathcal{P}} \times [0, \infty)) \to \mathbb{H}^3/\Gamma$ that restricts to p on $M_{\mathcal{P}}$. The closure of each component of $E(\mathcal{P}) - \mathcal{P}$ intersects $\mathcal{P}$ in a unique $g \in \mathcal{G}_e$, and the map $x \mapsto (r(x), d(x, r(x)))$ determines a homeomorphism to $g \times [0, \infty)$. The inverses of these homeomorphisms, taken over the disjoint union of all $g \in \mathcal{G}_e$, combine to induce the map in question.

The two paragraphs above combine to prove that $C(\Gamma) = p(M_{\mathcal{P}})$. In particular, $C(\Gamma)$ has totally geodesic boundary, so its preimage in $\mathbb{H}^3$ under the universal cover is a disjoint union of geodesic planes. Since p takes $g \in \mathcal{G}_e$ to $\partial C(\Gamma)$, the hyperplane $\mathcal{H}$ containing g is a component of the preimage of $\partial C(\Gamma)$. The final claim of the lemma follows. $\square$

Corollary 2.2. *Let $\mathcal{P}_1$ be the regular ideal octahedron in $\mathbb{H}^3$, embedded as indicated in Figure 3, and checkered by declaring the face A to be external. The collection $\{\mathsf{s}^{\pm 1}, \mathsf{t}^{\pm 1}\}$ is an internal face pairing for $\mathcal{P}_1$, where*

$$\mathsf{s} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \quad \textit{and} \quad \mathsf{t} = \begin{pmatrix} 2i & 2-i \\ i & 1-i \end{pmatrix}.$$

Let $M_S = \mathcal{P}_1/\{\mathsf{s}^{\pm 1}, \mathsf{t}^{\pm 1}\}$, and let $\Gamma_S = \langle \mathsf{s}, \mathsf{t} \rangle$. Then the inclusion $\mathcal{P}_1 \to \mathbb{H}^3$ induces an isometry $p_S\colon M_S \to C(\Gamma_S)$.

Proof. With the indicated embedding, $\mathcal{P}_1$ is a tile of the $\mathrm{PSL}_2(\mathcal{O}_1)$-invariant tessellation $\mathcal{T}_1$ constructed in [Hatcher 1983]. Here $\mathcal{O}_1 = \mathbb{Z}[i]$ is the ring of integers of the

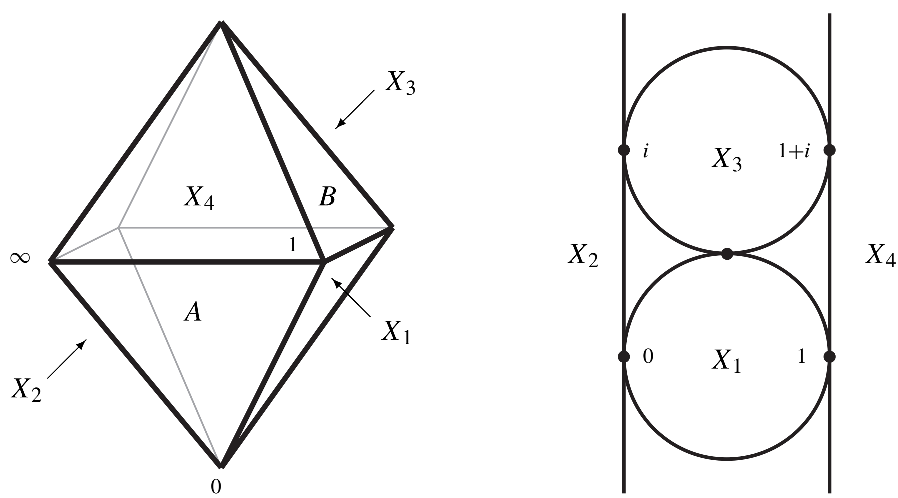


Figure 3. The regular ideal octahedron $\mathcal{P}_1$, and its expansion $E(\mathcal{P}_1)$.

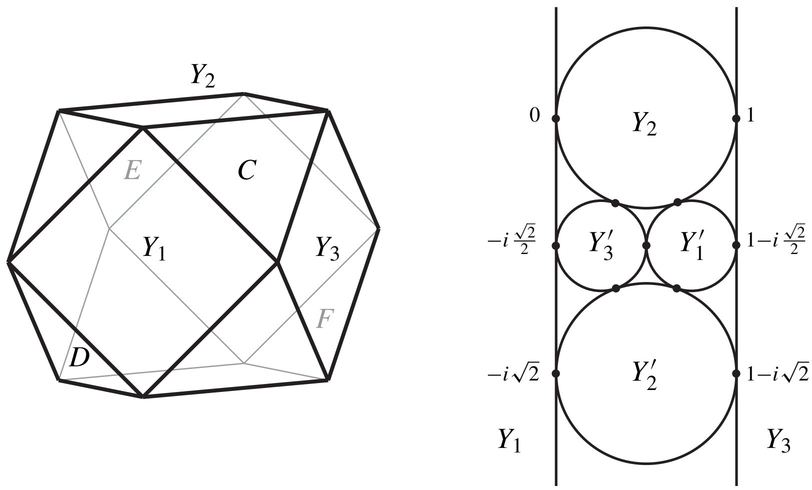


Figure 4. The right-angled ideal cuboctahedron $\mathcal{P}_2$, and $E(\mathcal{P}_2)$.

field $\mathbb{Q}(i)$. In particular, the face A shown on the left in Figure 3 has ideal vertices 0, 1, and ∞, and all other ideal vertices of $\mathcal{P}_1$ have positive imaginary part.

Since A is external, the faces X_1, X_2, X_3, and X_4 of $\mathcal{P}_1$ indicated on the left in Figure 3 are internal. Direct computation reveals that s takes X_1 to X_2, fixing the ideal vertex they share, and t takes X_3 to X_4 so that the vertex they share goes to the vertex shared by X_4 and X_2. Hence $\{\mathsf{s}^{\pm 1}, \mathsf{t}^{\pm 1}\}$ is an internal face pairing for $\mathcal{P}_1$. The corollary now follows from Lemma 2.1. □

The external faces of $\mathcal{P}_1$ triangulate ∂M_S, and their images under p_S determine a triangulation of $\partial C(\Gamma_S)$, which we will denote by Δ_S.

Corollary 2.3. *Let $\mathcal{P}_2$ be the right-angled ideal cuboctahedron in $\mathbb{H}^3$, embedded as indicated in Figure 4, and checkered by declaring triangular faces external. The collection $\{\mathsf{f}^{\pm 1}, \mathsf{g}^{\pm 1}, \mathsf{h}^{\pm 1}\}$ is an internal face pairing for $\mathcal{P}_2$, where*

$$\mathsf{f} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad \mathsf{g} = \begin{pmatrix} -1+i\sqrt{2} & 1-2i\sqrt{2} \\ -2 & 3-i\sqrt{2} \end{pmatrix}, \quad \mathsf{h} = \begin{pmatrix} 2i\sqrt{2} & -3-i\sqrt{2} \\ -3+i\sqrt{2} & -3i\sqrt{2} \end{pmatrix}.$$

Let $M_{T_0} = \mathcal{P}_2/\{\mathsf{f}^{\pm 1}, \mathsf{g}^{\pm 1}, \mathsf{h}^{\pm 1}\}$, and let $\Gamma_{T_0} = \langle \mathsf{f}, \mathsf{g}, \mathsf{h} \rangle$. The inclusion $\mathcal{P}_2 \to \mathbb{H}^3$ induces an isometry $p_{T_0} \colon M_{T_0} \to C(\Gamma_{T_0})$.

Proof. With the indicated embedding, $\mathcal{P}_2$ is a tile of the $\mathrm{PSL}_2(\mathbb{O}_2)$-invariant tessellation $\mathcal{T}_2$ of $\mathbb{H}^3$ defined in [Hatcher 1983], where $\mathbb{O}_2 = \mathbb{Z}[i\sqrt{2}]$ is the ring of integers of $\mathbb{Q}(i\sqrt{2})$. In particular, the face C labeled in the figure has ideal vertices 0, 1, and ∞.

Label the internal faces Y_i as indicated on the left in Figure 4, and label the square face opposite Y_i as Y_i'. Direct computation reveals that f takes Y_2 to Y_1, fixing the ideal vertex they share, g takes Y_3 to Y_1', fixing the ideal vertex they share, and h takes Y_2' to Y_3', taking the vertex they share to the opposite vertex on Y_3'. Hence $\{\mathsf{f}^{\pm 1}, \mathsf{g}^{\pm 1}, \mathsf{h}^{\pm 1}\}$ is an internal face pairing for $\mathcal{P}_2$. The corollary now follows from Lemma 2.1. □

The external faces of $\mathcal{P}_2$ triangulate ∂M_{T_0}. This has two components that we will call $\partial_+ M_{T_0}$ and $\partial_- M_{T_0}$, with the latter triangulated by the letter-labeled faces of Figure 4. Let $\partial_\pm C(\Gamma_{T_0}) = p_{T_0}(\partial_\pm M_{T_0})$ and let $\Delta^\pm_{T_0}$ refer to the triangulation for $\partial_\pm C(\Gamma_{T_0})$ determined by the images under p_{T_0} of the external faces of $\mathcal{P}_2$.

In the remainder of the paper, if g and h are elements of a group and G is a subgroup, we let g^{h} denote the conjugate of g by h, hgh^{-1}, and $G^{\mathsf{h}} = \mathsf{h}G\mathsf{h}^{-1}$. We describe parabolic isometries p_1, p_2, and p_3 which lie in $\Gamma_S \cap \Gamma_{T_0}$:

$$\begin{aligned} \mathsf{p}_1 &= \mathsf{s}^{-1} &&= \mathsf{f}^{-1} &&= \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \\ \mathsf{p}_2 &= \mathsf{stst}^{-2} &&= \mathsf{fg}^{-1}\mathsf{f}^{-1}\mathsf{h}^{-1}\mathsf{g} &&= \begin{pmatrix} -1 & 5 \\ 0 & -1 \end{pmatrix}, \\ \mathsf{p}_3 &= (\mathsf{s}^{-1})^{\mathsf{tst}} &&= (\mathsf{g}^{-1})^{\mathsf{g}^{-1}\mathsf{f}^{-1}\mathsf{h}} &&= \begin{pmatrix} -14 & 25 \\ -9 & 16 \end{pmatrix}. \end{aligned}$$

Since these are in $\mathrm{PSL}_2(\mathbb{R})$, they stabilize the hyperplane $\mathcal{H}$ with boundary $\mathbb{R} \cup \{\infty\}$.

Lemma 2.4. *The polygon $\mathcal{F}$ of Figure 5 is a fundamental domain for the action of $\Lambda \doteq \langle \mathsf{p}_1, \mathsf{p}_2, \mathsf{p}_3 \rangle < \mathrm{PSL}_2(\mathbb{R})$ on $\mathcal{H}$, and $F^{(0)} = \mathcal{H}/\Lambda$ is a 4-punctured sphere. Also*:

(1) $\Lambda = \mathrm{Stab}_{\Gamma_S}(\mathcal{H}) = \mathrm{Stab}_{\Gamma_{T_0}}(\mathcal{H})$,

(2) *the inclusion $\mathcal{H} \hookrightarrow \mathbb{H}^3$ induces an isometry $\iota_-^{(0)} \colon F^{(0)} \to \partial C(\Gamma_S)$ and an isometry $\iota_+^{(0)} \colon F^{(0)} \to \partial_- C(\Gamma_{T_0})$, and*

(3) *the triangulation of $\mathcal{F}$ pictured in Figure 5 projects to a triangulation Δ_F of $F^{(0)}$ taken by $\iota_-^{(0)}$ and $\iota_+^{(0)}$, respectively, to Δ_S and $\Delta^-_{T_0}$.*

Proof. With $\mathcal{P}_1$ and $\mathcal{P}_2$ embedded as prescribed in Figures 3 and 4, respectively, their faces A and C coincide and lie in $\mathcal{H}$ as described in Figure 5. As noted in the proofs of Corollaries 2.2 and 2.3, Γ_S-translates of $\mathcal{P}_1$ lie in the tessellation $\mathcal{T}_1$ described in [Hatcher 1983], and Γ_{T_0}-translates of $\mathcal{P}_2$ lie in $\mathcal{T}_2$. The Farey tessellation is $\mathcal{T}_1 \cap \mathcal{H} = \mathcal{T}_2 \cap \mathcal{H}$, so this contains any Γ_S-translate of any face of $\mathcal{P}_1$ and any Γ_{T_0}-translate of any face of $\mathcal{P}_2$.

Let A' be the external face of $\mathcal{P}_1$ which shares the ideal vertex 0 with A, and let B' be the external face which shares the vertex ∞ with A and $1+i$ with B. Since t takes X_3 to X_4, with the edge $X_3 \cap B'$ taken to $X_4 \cap A$, it follows that $\mathsf{t}(B')$ lies in $\mathcal{H}$, abutting A along the geodesic between 1 and ∞. Since $\mathsf{t}(B')$ is a Farey triangle it has its other ideal vertex at 2. It follows similarly that $\mathsf{g}^{-1}(E) = \mathsf{t}(B')$ (where E

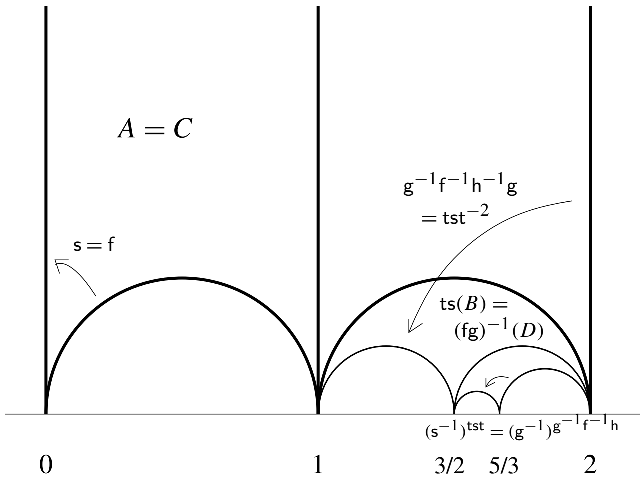


Figure 5. A triangulated fundamental domain $\mathcal{F}$ for the action of Λ on $\mathcal{H}$, and side pairings.

is as labeled in Figure 4), that $\mathsf{ts}(B) = (\mathsf{fg})^{-1}(D)$, as indicated in Figure 5, and that $\mathsf{tst}(A') = \mathsf{g}^{-1}\mathsf{f}^{-1}\mathsf{h}(F)$ has vertices at $3/2$, 2, and $5/3$.

Combinatorial considerations or direct calculation establish that $\mathsf{s} = \mathsf{f}$, $\mathsf{tst}^{-2} = \mathsf{g}^{-1}\mathsf{f}^{-1}\mathsf{h}^{-1}\mathsf{g}$, and $(\mathsf{s}^{-1})^{\mathsf{tst}} = (\mathsf{g}^{-1})^{\mathsf{g}^{-1}\mathsf{f}^{-1}\mathsf{h}}$, and that each stabilizes $\mathcal{H}$ and pairs edges of $\mathcal{F}$ as indicated in Figure 5. By inspection the quotient is a 4-punctured sphere $F^{(0)}$. By the polyhedron theorem $\mathcal{F}$ is a fundamental domain for the group that they generate, which acts on $\mathcal{H}$ with quotient $F^{(0)}$. Since p_1, p_2, and p_3 are easily obtained from the edge pairings above and vice versa, it follows that

$$\Lambda = \langle \mathsf{s}, \mathsf{tst}^{-2}, (\mathsf{s}^{-1})^{\mathsf{tst}} \rangle = \langle \mathsf{f}, \mathsf{g}^{-1}\mathsf{f}^{-1}\mathsf{h}^{-1}\mathsf{g}, (\mathsf{g}^{-1})^{\mathsf{g}^{-1}\mathsf{f}^{-1}\mathsf{h}} \rangle.$$

Therefore $\mathcal{F}$ is a fundamental domain for Λ, and $\mathcal{H}/\Lambda = F^{(0)}$.

It is easy to see from its combinatorics that ∂M_S is a four-punctured sphere, as is the component of ∂M_{T_0} containing C. Thus by Corollaries 2.2 and 2.3 the same holds true for $\partial C(\Gamma_S)$ and the component of $\partial C(\Gamma_{T_0})$ containing the image of C. Lemma 2.1 implies that $\partial C(\Gamma_S)$ is the image of $\mathcal{H}/\operatorname{Stab}_{\Gamma_S}(\mathcal{H})$ under the inclusion-induced map. Since it is clear from the above that $\Lambda < \operatorname{Stab}_{\Gamma_S}(\mathcal{H})$, and since $\mathcal{H}/\Lambda$ is itself a four-punctured sphere, the conclusions of assertions (1) and (2) above follow for Γ_S. A similar argument implies the same for Γ_{T_0}. The conclusions of (3) follow from the description above of the triangulation of $\mathcal{F}$. □

Remarks. 1. The parabolic elements of Λ fixing the ideal points 0, ∞, and $5/3$ of $\mathcal{H}$ are p_1, p_2, and p_3. The final conjugacy class of parabolic elements in Λ

is represented by

$$\mathsf{p}_4 = \mathsf{p}_1\mathsf{p}_2\mathsf{p}_3^{-1} = (\mathsf{stst}^{-2})^{\mathsf{tst}^{-1}} = \begin{pmatrix} 29 & -45 \\ 20 & -31 \end{pmatrix}.$$

Evidently p_1 and p_3 are conjugate in Γ_S, as are p_2 and p_4. The combinatorial considerations of Lemma 4.13 will show that $C(\Gamma_S)$ has exactly two cusps, each of rank one, so every parabolic element of Γ_S is conjugate to one of p_1 or p_2.

2. There exists $\mathsf{k} \in \mathrm{PSL}_2(\mathbb{C})$, with order 2, which normalizes Γ_{T_0}:

$$\mathsf{k} = \begin{pmatrix} i & i-\sqrt{2} \\ 0 & -i \end{pmatrix}. \tag{1}$$

The action of k on the generators f, g, and h is given by

$$\mathsf{f}^{\mathsf{k}} = \mathsf{g}^{\mathsf{fg}^{-1}}, \quad \mathsf{g}^{\mathsf{k}} = \mathsf{f}^{\mathsf{fg}^{-1}}, \quad \text{and} \quad \mathsf{h}^{\mathsf{k}} = (\mathsf{h}^{-1})^{\mathsf{fg}^{-1}}.$$

If Γ is a Kleinian group and $\mathsf{u} \in \mathrm{Isom}(\mathbb{H}^3)$, we write $\phi_{\mathsf{u}} \colon C(\Gamma) \to C(\Gamma^{\mathsf{u}})$ for the restriction to $C(\Gamma)$ of the isometry $\mathbb{H}^3/\Gamma \to \mathbb{H}^3/\Gamma^{\mathsf{u}}$ induced by u. Since k normalizes Γ_{T_0}, $\phi_{\mathsf{k}} \colon C(\Gamma_{T_0}) \to C(\Gamma_{T_0})$ is an orientation-preserving involution. The elements $\mathsf{p}_i^{\mathsf{k}}$, $i \in \{1, 2, 3\}$, all preserve the geodesic hyperplane $\mathsf{k}(\mathcal{H})$, which lies over the line $\mathbb{R} - i\sqrt{2}$ and contains an external face of $\mathcal{P}_2$ projecting to $\partial_+ C(\Gamma_{T_0})$. The lemma below follows and, together with Lemma 2.4, completely describes $\partial C(\Gamma_{T_0})$.

Lemma 2.5. $\Lambda^{\mathsf{k}} = \mathrm{Stab}_{\Gamma_T}(\mathsf{k}(\mathcal{H}))$, *and the inclusion* $\mathsf{k}(\mathcal{H}) \to \mathbb{H}^3$ *induces an isometry from* $F' \doteq \mathsf{k}(\mathcal{H})/\Lambda^{\mathsf{k}}$ *to* $\partial_+ C(\Gamma_{T_0})$.

It is easy to see that $\mathsf{p}_1^{\mathsf{k}}$ is conjugate in Γ_{T_0} to p_3 and that $\mathsf{p}_2^{\mathsf{k}} = \mathsf{p}_2^{-1}$. The combinatorial considerations of Lemma 4.14 will imply that M_{T_0} has four cusps. Hence, by Lemma 2.5, each of the cusps of $C(\Gamma_{T_0})$ joins $\partial_- C(\Gamma_{T_0})$ to $\partial_+ C(\Gamma_{T_0})$, and each parabolic in Γ_{T_0} is conjugate to exactly one p_i, $i \in \{1, 2, 3, 4\}$.

Our second main tool in this section is Lemma 2.6. We refer to [Morgan 1984, Definition 4.8] for the definition of a pared manifold.

Lemma 2.6. *Let* (M, P) *be a pared manifold, and suppose that* $\rho : \pi_1 M \to \Gamma < \mathrm{PSL}_2(\mathbb{C})$ *is a faithful representation onto a non-Fuchsian geometrically finite Kleinian group* Γ, *where* $C(\Gamma)$ *has totally geodesic boundary. If* ρ *determines a one-to-one correspondence between conjugacy classes of subgroups of* $\pi_1(M)$ *corresponding to components of* P *and conjugacy classes of maximal parabolic subgroups of* Γ, *then* ρ *is induced by a homeomorphism of* $M - P$ *to* $C(\Gamma)$.

This is well known to experts in Kleinian groups, but we do not know of a reference for a written proof. It seems worth writing down as it may fail if $C(\Gamma)$

does not have totally geodesic boundary (see [Canary and McCullough 2004] for a thorough exploration of this phenomenon). The proof follows easily from results in [Canary and McCullough 2004], for example, but requires introduction of the characteristic submanifold machinery. Since this falls outside the scope of the rest of the paper, we defer the proof to the Appendix.

Let (B^3, S) be the tangle pictured on the left in Figure 2. Take a base point for $\pi_1(B^3 - S)$ on $\partial(B^3 - S)$ high above the projection plane, and let its Wirtinger generators correspond in the usual way to labeled arcs of the diagram.

Proposition 2.7. *There is a homeomorphism* $f_S\colon B^3 - S \to C(\Gamma_S)$, *such that*

$$f_{S*}\colon \pi_1(B^3 - S) \to \Gamma_S$$

is given by $f_{S*}(a) = \mathsf{p}_1^{-1}$, $f_{S*}(e) = \mathsf{p}_2$, *and* $f_{S*}(v) = \mathsf{p}_3^{-1}$.

Proof. Reducing a standard Wirtinger presentation for $\pi_1(B^3 - S)$, we obtain

$$\langle a, w, e \mid ewe^{-1}a = awaw^{-1}\rangle = \langle a, w, e \mid w(e^{-1}aw) = (e^{-1}aw)a\rangle.$$

Thus, taking $b = e^{-1}aw$, one finds that $\pi_1(B^3 - S)$ is freely generated by a and b.

By Lemma 2.1 and Corollary 2.2, Γ_S is free on the generators s and t. Hence, the map $f_{S*}\colon \pi_1(B^3 - S) \longrightarrow \Gamma_S$ given by $a \mapsto \mathsf{s}$ and $b \mapsto \mathsf{t}$ is an isomorphism. Notice that the subgroup of $\pi_1(B^3 - S)$ corresponding to the 4-punctured sphere $\partial B^3 - \partial S$ is freely generated by a, v, and e. It is easily checked that

$$f_{S*}(v) = \begin{pmatrix} 16 & -25 \\ 9 & -14 \end{pmatrix} = \mathsf{p}_3^{-1} \quad \text{and} \quad f_{S*}(e) \;=\; \begin{pmatrix} -1 & 5 \\ 0 & -1 \end{pmatrix} = \mathsf{p}_2.$$

The map f_{S*} takes $\pi_1(\partial B^3 - S)$ isomorphically to Λ. Since a, v, and e generate $\pi_1(\partial B^3 - S)$ and their images in Γ_S generate Λ, Since any meridian of S is conjugate in $\pi_1(B^3 - S)$ to either a or e, and these are taken to p_1 and p_2, respectively, meridians are taken to parabolic elements of Γ_S.

Now let $N(S)$ be a small open tubular neighborhood of S in B^3. Then $B^3 - N(S)$ is a compact manifold with genus-2 boundary, and the pair $(B^3 - N(S), \partial N(S))$ is a pared manifold. The proposition follows from Lemma 2.6, after noting that $(B^3 - N(S)) - \partial N(S)$ is homeomorphic to $B^3 - S$. □

Let $(S^2 \times I, T_0)$ be the tangle pictured on the right side of Figure 2, where I is oriented so that $\partial_- T_0 \doteq T_0 \cap S^2 \times \{0\}$ contains the endpoints labeled a, u, and v. Take a base point for $\pi_1(S^2 \times I - T_0)$ on $S^2 \times \{0\}$ high above the projection plane and let Wirtinger generators correspond to the labeled arcs of Figure 2.

The next proposition is the analog of Proposition 2.7 for T_0.

Proposition 2.8. *There is a homeomorphism* $f_{T_0}\colon S^2 \times I - T_0 \longrightarrow C(\Gamma_{T_0})$ *such that*

$$f_{T_0*}\colon \pi_1(S^2 \times I - T_0) \longrightarrow \Gamma_{T_0}$$

is given by $f_{T_0*}(a) = \mathsf{p}_1^{-1}$, $f_{T_0*}(e) = \mathsf{p}_2$, *and* $f_{T_0*}(v) = \mathsf{p}_3^{-1}$.

Proof. $(S^2 \times I - N(T_0))$ may be isotoped in S^3 to a standard embedding of a genus-3 handlebody. Thus $\pi_1(S^2 \times I - T_0)$ is free on three generators. We claim that the group is generated by a, e, and t. This follows after noticing that $v = y^{-1}xy$, where $y = (ta)^{-1}a(ta)$ and $x = (azq)^{-1}t(azq) = (ate)^{-1}t(ate)$. (The relation $zq = te$ used in the last equality comes from the relationship between four peripheral elements in a 4-punctured sphere group.) So far, we have established that $v, y \in \langle a, e, t\rangle$. Now using the other punctured sphere relation, we have $u = a^{-1}ev \in \langle a, e, t\rangle$. Finally, $z = yuy^{-1}$ and $q = z^{-1}te$. Therefore a, e, and t generate the group as claimed.

By Lemma 2.1 and Corollary 2.3, Γ_{T_0} is freely generated by f, g, and h. For our purposes, a more convenient free generating set for Γ_{T_0} is $\{\mathsf{f}, \mathsf{fgf}^{-1}, \mathsf{p}_2\}$. Note that all of these generators are parabolic and peripheral, and conjugation by k interchanges the first two and takes the third to its inverse. The representation of $\pi_1(S^2 \times [0, 1/2] - T_0)$ given by

$$a \mapsto \mathsf{f}, \quad t \mapsto \mathsf{fgf}^{-1}, \quad e \mapsto \mathsf{p}_2$$

is clearly faithful, and it is easily checked that v maps to p_3^{-1}. Because $u = a^{-1}ev$ is mapped to $\mathsf{p}_1\mathsf{p}_2\mathsf{p}_3^{-1} = \mathsf{p}_4$, we conclude that meridians are mapped to parabolic elements and that $\pi_1(S^2 \times \{0\} - \partial_- T_0)$ is taken to Λ. The result now follows from Lemma 2.6 as previously. $\square$

There is a visible involution of $S^2 \times I - T_0$ which is a rotation by π around a circle in $S^2 \times \{1/2\}$. This involution exchanges the two boundary components. With a proper choice of path between our basepoint and its image under this involution, the corresponding action on $\pi_1(S^2 \times I - T_0)$ is given by

$$a \leftrightarrow t, \quad e \leftrightarrow e^{-1}.$$

This commutes with the action of the element k defined in (1) on Γ_{T_0}, under the representation f_{T_0*}. Hence this involution is isotopic to the pullback of ϕ_{k} by f_{T_0}.

Recall from Lemma 2.4 that $\Lambda = \mathrm{Stab}_{\Gamma_{T_0}}(\mathcal{H})$, and from Lemma 2.5 that $\Lambda^{\mathsf{k}} = \mathrm{Stab}_{\Gamma_{T_0}}(\mathsf{k}(\mathcal{H}))$. By its definition in Proposition 2.8, it is clear that f_{T_0*} maps $\pi_1(S^2 \times \{0\} - \partial_- T_0)$ isomorphically to Λ. Since $\mathcal{H}$ projects to $\partial_- C(\Gamma_{T_0})$, using the involution equivariance of f_{T_0} we obtain the corollary below.

Corollary 2.9. *Let* $\partial_+ T_0 = T_0 \cap S^2 \times \{1\}$. *Then* $f_{T_0}(S^2 \times \{0\} - \partial_- T_0) = \partial_- C(\Gamma_{T_0})$, *and* $f_{T_0}(S^2 \times \{1\} - \partial_+ T_0) = \partial_+ C(\Gamma_{T_0})$.

3. Combination

In this section, we will describe how to join copies of the tangles S and T_0 to construct links in S^3 whose complements are uniformized by combinations of

Γ_S and Γ_{T_0}. The main tool in this section is a corollary of Maskit's combination theorem for free products with amalgamation [Maskit 1971]. Denote the convex hull of the limit set for a Kleinian group Γ by $\mathrm{Hull}(\Gamma)$.

Definition 3.1. Kleinian groups Γ_0 and Γ_1 *meet cute* along a hyperplane $\mathcal{H} \subset \mathbb{H}^3$ if $\mathcal{H} = \mathrm{Hull}(\Gamma_0) \cap \mathrm{Hull}(\Gamma_1)$ and $\mathrm{Stab}_{\Gamma_0}(\mathcal{H}) = \mathrm{Stab}_{\Gamma_1}(\mathcal{H})$.

The fact below follows easily from this definition, and accounts for its utility.

Fact. *If* Γ_0 *and* Γ_1 *meet cute along* $\mathcal{H}$ *then* $\mathrm{Stab}_{\Gamma_0}(\mathcal{H}) = \Gamma_0 \cap \Gamma_1 = \mathrm{Stab}_{\Gamma_1}(\mathcal{H})$. *Furthermore*, $\mathcal{H}$ *divides* $\mathbb{H}^3$ *into* $\mathcal{B}_0$ *and* $\mathcal{B}_1$ *such that for* $i \in \{0, 1\}$, *if* $g_i \in \Gamma_i$ *satisfies* $g_i(\mathcal{B}_{1-i}) \cap \mathcal{B}_{1-i} \neq \varnothing$, *then* $g_i \in \Gamma_0 \cap \Gamma_1$.

In general, if Θ is a subgroup of Γ, the limit set of Θ is contained in that of Γ, and so the covering map $\mathbb{H}^3/\Theta \to \mathbb{H}^3/\Gamma$ maps $C(\Theta)$ into $C(\Gamma)$ — we will call this restriction the *natural map* $C(\Theta) \to C(\Gamma)$. When Γ_0 and Γ_1 meet cute along $\mathcal{H}$ then the natural map $C(\Gamma_0 \cap \Gamma_1) \to C(\Gamma_i)$ restricts to an embedding of the 2-orbifold $\mathcal{H}/(\Gamma_0 \cap \Gamma_1)$.

The lemma below is a geometric combination theorem for Kleinian groups which meet cute along a hyperplane. It follows from Maskit's combination theorem and observations on the geometry of Kleinian groups that go back at least to J. Morgan's [1984] account of geometrization for Haken manifolds.

Lemma 3.2. *Suppose* Γ_0 *and* Γ_1 *meet cute along a plane* $\mathcal{H}$. *Let* $E = \mathcal{H}/\Theta$, *where* $\Theta = \Gamma_0 \cap \Gamma_1$, *and for* $i = 0, 1$ *let* $\iota_i : E \to C(\Gamma_i)$ *be the natural embedding. Then* $\langle \Gamma_0, \Gamma_1 \rangle$ *is a Kleinian group, and the inclusions* $\Gamma_i \to \langle \Gamma_0, \Gamma_1 \rangle$ *determine an isomorphism* $\Gamma_0 *_\Theta \Gamma_1 \to \langle \Gamma_0, \Gamma_1 \rangle$ *as abstract groups. The natural maps* $C(\Gamma_i) \to C(\langle \Gamma_0, \Gamma_1 \rangle)$ *determine an isometry* $C(\Gamma_0) \cup_{\iota_1 \iota_0^{-1}} C(\Gamma_1) \to C(\langle \Gamma_0, \Gamma_1 \rangle)$.

In using Lemma 3.2, we often write $C(\Gamma_0) \cup_E C(\Gamma_1)$ when the maps ι_i are clear.

Proof. We will use Theorem 8.2 of [Morgan 1984], a version of Maskit's combination theorem. Its hypotheses are satisfied by any Γ_0 and Γ_1 that meet cute along a hyperplane $\mathcal{H}$, as a consequence of the Fact above; the group-theoretic conclusions follow. That the desired isometry exists follows from the remarks given below Theorem 8.2 of [Morgan 1984], which have since been considerably fleshed out in [Anderson and Canary 2001].

The function $\tilde{f} : \mathbb{H}^3 \to [0, 1]$ described in [Morgan 1984] is the harmonic extension of the characteristic function of $\mathcal{B}_0$: $\tilde{f}(y)$ is the visual measure of the set of vectors pointing from y toward $\mathcal{B}_0$. See [Anderson and Canary 2001, §2] for a precise analytic definition. It is not hard to see that here $\mathcal{H} = \tilde{f}^{-1}\left(\frac{1}{2}\right)$ (see [Anderson and Canary 2001, Proposition 2.2]), whence our E is Morgan's $X = f^{-1}\left(\frac{1}{2}\right)$.

Our ι_i is Morgan's p_i, mapping to $N_i = \mathbb{H}^3/\Gamma_i$ for $i \in \{0, 1\}$. For each i, $\iota_i(E)$ is a convex core boundary component of N_i, so $p_0(N_-) \cap C(N_0) = p_0(E)$ and $p_1(N_+) \cap C(N_1) = p_1(E)$, and the result follows from the equation at the bottom of [Morgan 1984, p. 76]. See [Anderson and Canary 2001, Proposition 5.2 and remarks after 5.3] for related results. □

We first apply Lemma 3.2 to join $C(\Gamma_{T_0})$ to a copy of itself across $\partial_+ C(\Gamma_{T_0})$. Recall from the discussion above Lemma 2.4 that we have defined $\mathcal{H}$ to be the geodesic hyperplane of $\mathbb{H}^3$ with ideal boundary $\mathbb{R}\cup\{\infty\}$. Let $\mathsf{r}\in\mathrm{Isom}(\mathbb{H}^3)$ be the reflection through $\mathcal{H}$. This acts on $\mathbb{C}\cup\{\infty\}$ by complex conjugation; thus if $\mathsf{q}\in\Gamma<\mathrm{PSL}_2(\mathbb{C})$, then $\mathsf{q}^{\mathsf{r}}=\bar{\mathsf{q}}$, where $\bar{\mathsf{q}}\in\mathrm{PSL}_2(\mathbb{C})$ is the element whose entries are the conjugates of the entries of q. Hence, we let $\overline{\Gamma}$ denote Γ^{r}.

Lemma 3.3. *Define*

$$\mathsf{c}=\begin{pmatrix}1 & i\sqrt{2}\\ 0 & 1\end{pmatrix}.$$

Then $\Gamma_T \doteq \langle \Gamma_{T_0}, \overline{\Gamma}_{T_0}^{\mathsf{c}^{-2}}\rangle$ *is a Kleinian group, and there is an inclusion-induced isomorphism* $\Gamma_{T_0} *_{\Lambda^{\mathsf{k}}} \overline{\Gamma}_{T_0}^{\mathsf{c}^{-2}} \to \Gamma_T$ *and an isometry* $C(\Gamma_{T_0})\cup_{F'} C(\overline{\Gamma}_{T_0}^{\mathsf{c}^{-2}})\to C(\Gamma_T)$ *determined by the natural maps. Furthermore,* $\mathsf{c}^{-2}\mathsf{r}$ *normalizes* Γ_T, *and* $\phi_{\mathsf{c}^{-2}\mathsf{r}}\colon C(\Gamma_T)\to C(\Gamma_T)$ *is an orientation-reversing involution fixing* F' *and exchanging its complementary components.*

Proof. Recall from Lemmas 2.4 and 2.5 that Λ and Λ^{k} are the stabilizers in Γ_{T_0} of the geodesic planes $\mathcal{H}$ and $\mathsf{k}(\mathcal{H})$, respectively, and that these planes project to the components of $\partial C(\Gamma_{T_0})$. It follows that $\mathcal{H}$ and $\mathsf{k}(\mathcal{H})$ are components of the boundary of $\mathrm{Hull}(\Gamma_{T_0})$, so $\mathrm{Hull}(\Gamma_{T_0})$ is contained in the region between them.

With c as defined in the statement of the lemma, note that $\mathsf{c}(\mathcal{H})=(\mathbb{R}+i\sqrt{2})\times\mathbb{R}^+$ and that $\mathsf{ck}(\mathcal{H})=\mathcal{H}$. Since $\mathrm{Hull}(\Gamma_{T_0}^{\mathsf{c}})$ has boundary components $\mathsf{c}(\mathcal{H})$ and $\mathsf{ck}(\mathcal{H})=\mathcal{H}$, and $\Lambda^{\mathsf{ck}}=\mathrm{Stab}_{\Gamma_{T_0}^{\mathsf{c}}}(\mathcal{H})$ is invariant under conjugation by r, $\Gamma_{T_0}^{\mathsf{c}}$ and $\overline{\Gamma_{T_0}^{\mathsf{c}}}$ meet cute along $\mathcal{H}$. Applying Lemma 3.2, we obtain an isomorphism $\Gamma_{T_0}^{\mathsf{c}} *_{\Lambda^{\mathsf{ck}}} \overline{\Gamma_{T_0}^{\mathsf{c}}} \to \langle \Gamma_{T_0}^{\mathsf{c}}, \overline{\Gamma_{T_0}^{\mathsf{c}}}\rangle$ and an isometry

$$C(\Gamma_{T_0}^{\mathsf{c}})\cup_{\phi_{\mathsf{c}}(F')} C(\overline{\Gamma_{T_0}^{\mathsf{c}}}) \to C(\langle \Gamma_{T_0}^{\mathsf{c}}, \overline{\Gamma_{T_0}^{\mathsf{c}}}\rangle)$$

induced by the natural maps. It is clear that r normalizes $\langle \Gamma_{T_0}^{\mathsf{c}}, \overline{\Gamma_{T_0}^{\mathsf{c}}}\rangle$, exchanging amalgamands, hence ϕ_{r} acts as an orientation-reversing involution of $C(\langle \Gamma_{T_0}^{\mathsf{c}}, \overline{\Gamma_{T_0}^{\mathsf{c}}}\rangle)$, fixing F' and exchanging $C(\Gamma_{T_0}^{\mathsf{c}})$ with $C(\overline{\Gamma_{T_0}^{\mathsf{c}}})$.

Observe that $\bar{\mathsf{c}}=\mathsf{c}^{-1}$. It follows that $\overline{\Gamma_{T_0}^{\mathsf{c}}}=\overline{\Gamma}_{T_0}^{c^{-1}}$, and hence $\Gamma_T=\langle \Gamma_{T_0}^{\mathsf{c}}, \overline{\Gamma_{T_0}^{\mathsf{c}}}\rangle^{\mathsf{c}^{-1}}$. Conjugating the groups of the paragraph above by c^{-1}, we obtain an inclusion-induced isomorphism $\Gamma_{T_0} *_{\Lambda^k} \overline{\Gamma}_{T_0}^{\mathsf{c}^{-2}} \to \Gamma_T$, and an isometry

$$C(\Gamma_{T_0})\cup_{F'} C(\overline{\Gamma}_{T_0}^{\mathsf{c}^{-2}}) \to C(\Gamma_T)$$

induced by the natural maps. Furthermore, $\mathsf{c}^{-1}\mathsf{r}\mathsf{c}=\mathsf{c}^{-2}\mathsf{r}$ normalizes Γ_T and induces an orientation-reversing involution $\phi_{\mathsf{c}^{-2}\mathsf{r}}$, fixing F' and exchanging its sides. □

If M is an oriented manifold with a boundary component F, the *double* of M across F is $M \cup_F \overline{M}$, where $\overline{M}$ is a copy of M with orientation reversed, and the gluing map $F \to \overline{F} \subset \overline{M}$ is the identity map.

Corollary 3.4. *There is an isometry $p_T\colon M_T \to C(\Gamma_T)$, where M_T is the double of M_{T_0} across F', that is the natural map following p_{T_0} from Corollary 2.3 on M_{T_0}.*

The advantage that Γ_T has over Γ_{T_0} for our purposes is that the components of $\partial C(\Gamma_T)$ are naturally orientation-reversing isometric, since they are exchanged by $\phi_{\mathsf{c}^{-2}\mathsf{r}}$. Recall from Lemma 2.4 that $\Lambda = \mathrm{Stab}_{\Gamma_{T_0}}(\mathcal{H})$; thus by Lemma 3.3, $\Lambda = \mathrm{Stab}_{\Gamma_T}(\mathcal{H})$. We will again refer by $i_+^{(0)}$ to the natural map $F^{(0)} \to C(\Gamma_T)$. Then the lemma below follows from Lemma 3.3.

Lemma 3.5. *Let $F^{(1)} = \mathsf{c}^{-2}\mathcal{H}/\Lambda^{\mathsf{c}^{-2}}$, and let $\phi_{\mathsf{c}^{-2}}\colon F^{(0)} \to F^{(1)}$ and $\iota_-^{(1)}\colon F^{(1)} \to C(\Gamma_T)$ be the natural maps. Then $\partial C(\Gamma_T) = \partial_- C(\Gamma_T) \sqcup \partial_+ C(\Gamma_T)$, where $\partial_- C(\Gamma_T) \doteq \iota_+^{(0)}(F^{(0)})$ and $\partial_+ C(\Gamma_T) \doteq \iota_-^{(1)}(F^{(1)})$, and $\iota_-^{(1)}\phi_{\mathsf{c}^{-2}} = \phi_{\mathsf{c}^{-2}\mathsf{r}}\iota_+^{(0)}$.*

$C(\Gamma_T)$ is a geometric model for the double of $(S^2 \times I, T_0)$ across $(S^2 \times \{1\}, \partial_+ T_0)$. Note that the double of $S^2 \times I$ across $S^2 \times \{1\}$ is again homeomorphic to $S^2 \times I$, by a map taking $(p, t) \in S^2 \times I$ to $(p, t/2)$ and $(p, t) \in \overline{S^2 \times I}$ to $(p, 1 - t/2)$.

Definition 3.6. Let $(S^2 \times I, T)$ be the double of $(S^2 \times I, T_0)$ across $(S^2 \times \{1\}, \partial_+ T_0)$. We will identify $(S^2 \times I, T_0) \subset (S^2 \times I, T)$ with its image under the map discussed above, so that $T_0 = T \cap S^2 \times [0, 1/2]$. In particular, we have $\partial_- T = \partial_- T_0 = T \cap S^2 \times \{0\}$ and $\partial_+ T_0 = T \cap S^2 \times \{1/2\}$, and we will take $\partial_+ T = T \cap S^2 \times \{1\}$.

The tangle $(S^2 \times I, T)$ is pictured in Figure 6, with $T_0 \subset T$ visible to the left of the gray vertical line representing $S^2 \times \{1/2\}$. There is a mirror symmetry of $(S^2 \times I, T)$, visible in the figure as reflection through the gray vertical line:

$$r_T\colon\ (S^2 \times I, T) \to (S^2 \times I, T)$$

given by $r_T(p, x) = (p, 1 - x)$, hence fixing $(S^2 \times \{1/2\}, \partial_+ T_0)$.

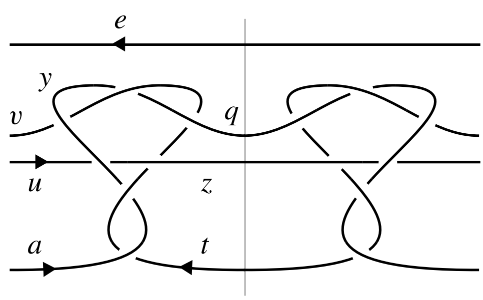


Figure 6. The tangle $T \subset S^2 \times I$ with labeled Wirtinger generators for T_0.

Proposition 3.7. *There is a homeomorphism* $f_T\colon S^2 \times I - T \to C(\Gamma_T)$, *which restricts on* $S^2 \times [0, 1/2] - T_0$ *to* f_{T_0} *followed by the natural map*, *such that the following diagram commutes*:

$$\begin{tikzcd} S^2 \times I - T \arrow[r, "f_T"] \arrow[d, "r_T"'] & C(\Gamma_T) \arrow[d, "\phi_{\mathsf{c}^{-2}\mathsf{r}}"] \\ S^2 \times I - T \arrow[r, "f_T"] & C(\Gamma_T). \end{tikzcd}$$

Furthermore, f_T *takes* $S^2 \times \{0\} - \partial_- T$ *to* $\partial_- C(\Gamma_T)$ *and* $S^2 \times \{1\} - \partial_+ T$ *to* $\partial_+ C(\Gamma_T)$.

Proof. We define f_T using the properties described in the statement of the proposition. Namely, we first require f_T to restrict on $S^2 \times [0, 1/2] - T_0$ to the homeomorphism f_{T_0} defined in Proposition 2.8, followed by the natural map $C(\Gamma_{T_0}) \to C(\Gamma_T)$. For $x \in S^2 \times [1/2, 1] - T$, we define $f_T(x) = \phi_{\mathsf{c}^{-2}\mathsf{r}} f_T r_T(x)$. The resulting map is well defined, since r_T fixes $S^2 \times \{1/2\} - \partial_+ T_0$ and $\phi_{\mathsf{c}^{-2}\mathsf{r}}$ fixes F'. It is a homeomorphism, since r_T, f_{T_0}, and $\phi_{\mathsf{c}^{-2}\mathsf{r}}$ are. By Corollary 2.9, f_{T_0} takes $S^2 \times \{0\} - \partial_- T_0$ to $\partial_- C(\Gamma_{T0})$; it therefore follows from the definitions and Lemma 3.5 that $f_T(S^2 \times \{0\} - \partial_- T) = \partial_- C(\Gamma_T)$. The conclusion thus follows from the reflection equivariance of f_T. □

Definitions 3.8. (1) Let $j\colon (\partial B^3, \partial S) \to (S^2 \times \{0\}, \partial_- T)$ be the homeomorphism such that $(B^3, S) \cup_j (S^2 \times I, T)$ is the tangle pictured in Figure 7.

(2) Define $h\colon S^2 \times \mathbb{R} \to S^2 \times \mathbb{R}$ by $h(p, x) = (p, x + 1)$, and with $T \subset S^2 \times I \subset S^2 \times \mathbb{R}$, let $T^{(i)} = h^{i-1}(T)$ (so $T^{(1)} = T$ in particular). For $n \in \mathbb{N}$, define

$$(S^3, L_n) = (B^3, S) \cup_j \Big(S^2 \times [0, n], \bigcup_{i=1}^{n} T^{(i)}\Big) \cup_{j_n} (\bar{B}^3, \bar{S}).$$

For $i \in \{0, 1, \dots, n\}$, let $S^{(i)}$ be the image in (S^3, L_n) of $S^2 \times \{i\} \subset S^2 \times [0, n]$. Above, $(\bar{B}^3, \bar{S}) = r_S(B^3, S)$, where r_S is the reflective involution of S^3 fixing the boundary of an embedding of B^3 and exchanging its sides, and $j_n = r_S j^{-1} r_T h^{-n+1}\colon (S^{(n)}, \partial_+ T^{(n)}) \longrightarrow (\partial \bar{B}^3, \partial \bar{S})$.

(3) Using Figure 6 and taking $T \subset S^2 \times I \subset S^2 \times \mathbb{R}$, label the points of $S^{(0)} \cap L_n = S^2 \times \{0\} \cap T$ by 2, 3, 4, and 1 top-to-bottom, so that, for example, 2 is the terminal point of the tangle string labeled e and 1 is the initial point of the string labeled a. Label each point of $S^{(i)} \cap L_n$ by its image under h^{-i}.

Remark. With Wirtinger generators for $\pi_1(B^3 - S)$ and $\pi_1(S^2 \times I - T)$ as labeled in Figures 2 and 6, we have $j_*(a) = a$, $j_*(u) = u$, and $j_*(v) = v$.

We now construct a geometric model of $S^3 - L_n$.

Definitions 3.9. (1) For $i \geq 0$, let $\Lambda^{(i)} = \Lambda^{\mathsf{c}^{-2i}}$ and $F^{(i)} = \mathsf{c}^{-2i}(\mathcal{H})/\Lambda^{(i)}$.

(2) For $i \geq 1$, let $\Gamma_T^{(i)} = \Gamma_T^{\mathsf{c}^{-2(i-1)}}$, and define $\phi_i = \phi_{\mathsf{c}^{-2(i-1)}}\colon C(\Gamma_T) \to C(\Gamma_T^{(i)})$.

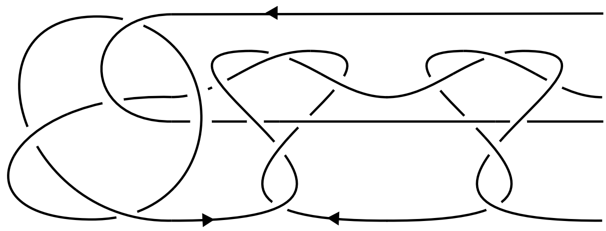


Figure 7. $S \cup T$.

The definitions above of $F^{(0)}$ and $F^{(1)}$ above agree with our previous definitions. Also, $\Gamma_T^{(1)} = \Gamma_T$, and since $\Gamma_T^{(i)} = \mathsf{c}^{-2(i-1)}\Gamma_T \mathsf{c}^{2(i-1)}$, Lemma 3.5 implies that

$$\Lambda^{(i-1)} = \mathrm{Stab}_{\Gamma_T^{(i)}}(\mathsf{c}^{-2(i-1)}(\mathcal{H})) \quad \text{and} \quad \Lambda^{(i)} = \mathrm{Stab}_{\Gamma_T^{(i)}}(\mathsf{c}^{-2i}(\mathcal{H})),$$

and the resulting natural maps $\iota_+^{(i-1)} \colon F^{(i-1)} \to C(\Gamma_T^{(i)})$ and $\iota_-^{(i)} \colon F^{(i)} \to C(\Gamma_T^{(i)})$ map to the components of its totally geodesic boundary.

Proposition 3.10. *For $n \in \mathbb{N}$, define $M_n = C(\Gamma_S) \cup C(\Gamma_T^{(1)}) \cup \cdots \cup C(\Gamma_T^{(n)}) \cup C(\overline{\Gamma}_S)$, using gluing maps defined as follows*:

$$\begin{aligned} \iota_+^{(0)}(\iota_-^{(0)})^{-1} \colon \quad & \partial C(\Gamma_S) \to \partial_- C(\Gamma_T^{(1)}), \\ \iota_+^{(i)}(\iota_-^{(i)})^{-1} \colon \; & \partial_+ C(\Gamma_T^{(i)}) \to \partial_- C(\Gamma_T^{(i+1)}) \quad \textit{for } 1 \le i < n, \\ \phi_{\mathrm{r}}\iota_-^{(0)}\phi_{n+1}^{-1}(\iota_-^{(n)})^{-1} \colon \; & \partial_+ C(\Gamma_T^{(n)}) \to \partial C(\overline{\Gamma}_S). \end{aligned}$$

There is a homeomorphism $f_n \colon S^3 - L_n \to M_n$ which restricts on $B^3 - S$ to f_S, on $S^2 \times [i-1, i] - T^{(i)}$ to $\phi_i f_T h^{-i+1}$ for $1 \le i \le n$, and on $\bar{B}^3 - \bar{S}$ to $\phi_{\mathrm{r}} f_S r_S$.

Proof. We will use the description of f_n above as its definition. Then, by Proposition 2.7 and the definitions, f_n restricts on $B^3 - S$ and $\bar{B}^3 - \bar{S}$ to homeomorphisms to $C(\Gamma_S)$ and $C(\overline{\Gamma}_S)$, respectively. By Proposition 3.7 and definitions, for each i between 1 and n it restricts on $S^2 \times [i-1, i] - T^{(i)}$ to a homeomorphism to $C(\Gamma_T^{(i)})$. Thus in order to show that f_n is a homeomorphism, we must only show that it is well defined on the spheres $S^{(i)} - \{1, 2, 3, 4\}$ that separate these tangle complements.

We first check the case $i = 1$, showing that f_n is well defined on $S^{(0)} - \{1, 2, 3, 4\}$. Since $T^{(1)} = T$ and $\Gamma_T^{(1)} = \Gamma_T$, and h^0 and ϕ_1 are each the identity map, in this case we must only show that on $\partial B^3 - \partial S$, $f_T \circ j = \iota_+^{(0)}(\iota_-^{(0)})^{-1} \circ f_S$.

By their definitions above, f_S and $f_T \circ j$ induce the same isomorphism from $\pi_1(\partial B^3 - \partial S)$ to $\Lambda = \Gamma_S \cap \Gamma_T$. Recall from Lemma 2.4 and the remarks above Lemma 3.5 that the $\iota_\pm^{(0)}$ are induced by the inclusions of Λ into Γ_S and Γ_T. Therefore at the level of fundamental group, $(\iota_+^{(0)}(\iota_-^{(0)})^{-1} \circ f_S)_* = (f_T \circ j)_*$. Since any two homeomorphisms between 4-punctured spheres that induce the same map on fundamental groups are properly isotopic, we may isotope j so that f_S and $f_T j$ agree on $S^{(0)}$. The conclusion thus follows in this case.

For $1 \leq i < n$, we may use the fact that $\Gamma_T^{(i)}$ and $\Gamma_T^{(i+1)}$ are conjugates of Γ_T to obtain the following model descriptions for $\iota_+^{(i)}$ and $\iota_-^{(i)}$:

$$\iota_+^{(i)} = \phi_{i+1}\iota_+^{(0)}\phi_{i+1}^{-1}, \quad \iota_-^{(i)} = \phi_i \iota_-^{(1)}\phi_i^{-1}. \tag{2}$$

Here $\iota_+^{(0)}\colon F^{(0)} \to \partial_- C(\Gamma_T)$ and $\iota_-^{(1)}\colon F^{(1)} \to \partial_+ C(\Gamma_T)$ are the natural maps of Lemma 3.5. Using the reflection-invariance property described there, we thus obtain

$$\iota_+^{(i)}(\iota_-^{(i)})^{-1} = \phi_{i+1}\iota_+^{(0)}(\iota_-^{(1)}\phi_2)^{-1}\phi_i^{-1} = \phi_{i+1}\phi_{\mathsf{c}^{-2}\mathsf{r}}^{-1}\phi_i^{-1}. \tag{3}$$

Then, by the reflection-equivariance property of f_T from Proposition 3.7, we have

$$\iota_+^{(i)}(\iota_-^{(i)})^{-1} \circ \phi_i f_T h^{-i+1} = \phi_{i+1}\phi_{\mathsf{c}^{-2}\mathsf{r}}^{-1} f_T h^{-i+1} = \phi_{i+1} f_T r_T h^{-i+1}.$$

It follows directly from the definitions that $r_T h^{-i+1} = h^{-i}$ on $S^{(i)}$, whence f_n is well defined on $S^{(i)} - \{1, 2, 3, 4\}$ for $1 \leq i < n$.

To show f_n is well defined on $S^{(n)}$ requires another definition chase, this time to check

$$\phi_{\mathsf{r}} f_S r_S \circ j_n = \phi_{\mathsf{r}}\iota_-^{(0)}\phi_{n+1}^{-1}(\iota_-^{(n)})^{-1} \circ \phi_n f_T h^{-n+1}.$$

By Definitions 3.8(2), $j_n = r_S j^{-1} r_T h^{-n+1}$; therefore, simplifying the left-hand side above yields $\phi_{\mathsf{r}} f_S j^{-1} r_T h^{-n+1}$. On the other hand, using the model description of $\iota_-^{(n)}$ from (2), the right-hand side above simplifies to $\phi_{\mathsf{r}}\iota_-^{(0)}\phi_2^{-1}(\iota_-^{(1)})^{-1} f_T h^{-n+1}$. The reflection-invariance property of Lemma 3.5 and an appeal to the case $i = 0$ establish the desired equation. □

Corollary 3.11. *For $0 \leq i < n$, refer again by $F^{(i)}$ to the image of $\iota_+^{(i)}(F^{(i)}) \subset C(\Gamma_T^{(i)})$ under its inclusion into M_n, and refer by $F^{(n)}$ to the image of $\iota_-^{(n)}(F^{(n)})$. For each i, $F^{(i)}$ is totally geodesic in M_n and $f_n(S^{(i)} - \{1, 2, 3, 4\}) = F^{(i)}$.*

This follows immediately from Proposition 3.10, since the maps $\iota_\pm^{(i)}$ are isometric embeddings. The following proposition describes an algebraic model for M_n.

Proposition 3.12. *For $n \in \mathbb{N}$, define $\Gamma_n = \langle \Gamma_S, \Gamma_T^{(1)}, \ldots, \Gamma_T^{(n)}, \overline{\Gamma}_S^{\mathsf{c}^{-2n}} \rangle$. There is an isometry $M_n \to \mathbb{H}^3/\Gamma_n$ which restricts on $C(\Gamma_S)$ and each $C(\Gamma_T^{(i)})$ to the natural map, and on $C(\overline{\Gamma}_S)$ to ϕ_{n+1} followed by the natural map.*

Proof. We first recall from Lemma 2.4 that the plane $\mathcal{H}$ with ideal boundary $\mathbb{R} \cup \{\infty\}$ projects to $\partial C(\Gamma_S)$ under the quotient map $\mathbb{H}^3 \to \mathbb{H}^3/\Gamma_S$, so it is a component of $\partial\,\mathrm{Hull}(\Gamma_S)$. Because the octahedron $\mathcal{P}_1$ is contained in $\mathrm{Hull}(\Gamma_S)$ and all its ideal vertices have nonnegative imaginary part, it follows that

$$\mathrm{Hull}(\Gamma_S) \subset \{z \in \mathbb{C} \mid \Im z \geq 0\} \cup \{\infty\}.$$

Similarly, from Lemma 3.5 and the positioning of $\mathcal{P}_2$ we find that

$$\mathrm{Hull}(\Gamma_T) \subset \{z \in \mathbb{C} \mid 0 \geq \Im z \geq -2\sqrt{2}\} \cup \{\infty\}.$$

Then, inspecting the action of c on $\mathbb{C} \cup \{\infty\}$, we find that for each $i \in \mathbb{N}$, any point of $\mathrm{Hull}(\Gamma_T^{(i)})$ has imaginary part between $-2(i-1)\sqrt{2}$ and $-2i\sqrt{2}$ for $i \in \mathbb{N}$.

The following claim builds an inductive picture of a family of isometrically embedded, codimension-0 submanifolds of M_n with totally geodesic boundary.

Claim. *For $1 \leq i \leq n$, define $\Gamma_-^{(i)} = \langle \Gamma_S, \Gamma_T^{(1)}, \dots, \Gamma_T^{(i)} \rangle$. There is an isometry $C(\Gamma_S) \cup C(\Gamma_T^{(1)}) \cup \cdots \cup C(\Gamma_T^{(i)}) \to C(\Gamma_-^{(i)})$, where the gluing maps for the domain are as in Proposition 3.10, which restricts on $C(\Gamma_S)$ and each $C(\Gamma^{(j)})$, $j < i$, to the natural map. Furthermore*:

(1) $\Lambda^{(i)} = \mathrm{Stab}_{\Gamma_-^{(i)}}(\mathsf{c}^{-2n}(\mathcal{H}))$, *and the resulting natural map $F^{(i)} \to \partial C(\Gamma_-^{(i)})$ factors as $\iota_-^{(i)}$ followed by the natural map $C(\Gamma_T^{(i)}) \to C(\Gamma_-^{(i)})$.*

(2) $\mathrm{Hull}(\Gamma_-^{(i)}) \subset \{z \in \mathbb{C} \mid \Im z \geq -i\sqrt{2}\} \cup \{\infty\}$.

Proof of claim. We will prove the claim by induction. If it holds for some $i < n$, then (1) and (2) above, together with the observations above the claim, imply that $\Gamma_-^{(i)}$ and $\Gamma_T^{(i+1)}$ meet cute along $\mathsf{c}^{-2i}(\mathcal{H})$. Then, by Lemma 3.2, the natural maps determine an isometry $C(\Gamma_-^{(i)}) \cup C(\Gamma_T^{(i+1)}) \to C(\Gamma_-^{(i+1)})$, where by the inductive hypothesis and the observation above Proposition 3.10, the gluing map for the domain is $\iota_+^{(i)}(\iota_-^{(i)})^{-1}$ following the inverse of the natural map.

Furthermore, since $C(\Gamma_-^{(i)})$ has a unique totally geodesic boundary component, which is isometrically identified with $\partial_- C(\Gamma_T^{(i+1)})$ in the isometry to $C(\Gamma_-^{(i+1)})$ described above, the unique totally geodesic boundary component of $C(\Gamma_-^{(i+1)})$ is the isometric image of $\partial_+ C(\Gamma_T^{(i+1)})$. Therefore the observations above Proposition 3.10 imply that this boundary component is the image of $\iota_-^{(i+1)}(F^{(i+1)})$ under the natural map. Assertion (1) of the claim thus follows for $\Gamma_-^{(i+1)}$. It follows that $\mathrm{Hull}(\Gamma_-^{(i+1)})$ is entirely on one side or the other of the boundary at infinity of $\mathsf{c}^{-2(i+1)}(\mathcal{H})$. Since $\Gamma_T^{(i+1)} < \Gamma_-^{(i+1)}$, assertion (2) now follows.

By our definition of "natural map" above Lemma 3.2, the composition of the natural map $C(\Gamma_T^{(j)}) \to C(\Gamma_-^{(i)})$ with the natural map $C(\Gamma_-^{(i)}) \to C(\Gamma_-^{(i+1)})$ is itself natural, for $j \leq i$. Hence if the claim holds for $\Gamma_-^{(i)}$, $i < n$, it holds for $\Gamma_-^{(i+1)}$. The claim will therefore hold by induction if it is true in the base case $i = 1$. But this follows from the fact that Γ_S and $\Gamma_T^{(1)}$ meet cute along $\mathcal{H}$. This follows in turn from Lemmas 2.4 and 3.5, which establish that $\Lambda^{(0)} = \mathrm{Stab}_{\Gamma_S}(\mathcal{H}) = \mathrm{Stab}_{\Gamma_T}(\mathcal{H})$, and the first paragraph of the proof. □

Using the claim, it now follows that $\Gamma_-^{(n)}$ and $\overline{\Gamma}_S^{\mathsf{c}^{-2n}}$ meet cute along $\mathsf{c}^{-2n}(\mathcal{H})$; hence a final application of Lemma 3.2 implies that the natural maps determine an isometry $C(\Gamma_-^{(n)}) \cup C(\overline{\Gamma}_S^{\mathsf{c}^{-2n}}) \to C(\Gamma_n)$. Since each of $C(\Gamma_-^{(n)})$ and $C(\overline{\Gamma}_S^{\mathsf{c}^{-2n}})$ has a unique boundary component, $C(\Gamma_n)$ is boundaryless and hence equal to $\mathbb{H}^3/\Gamma_n$. The conclusion of the proposition follows. □

The result below follows from Proposition 3.12, or, really, its proof.

Corollary 3.13. *For fixed n and* $0 \le i \le n$, *define*

$$\Gamma_-^{(i)} = \langle \Gamma_S, \Gamma_T^{(1)}, \dots, \Gamma_T^{(i)} \rangle, \quad \Gamma_+^{(i)} = \langle \Gamma_T^{(i+1)}, \dots, \Gamma_T^{(n)}, \overline{\Gamma}_S^{\mathsf{c}^{-2n}} \rangle.$$

Then $\Gamma_+^{(i)}$ *and* $\Gamma_-^{(i)}$ *meet cute along* $\mathsf{c}^{-2i}(\mathcal{H})$ *and the natural maps determine an isometry* $C(\Gamma_-^{(i)}) \cup C(\Gamma_+^{(i)}) \to \mathbb{H}^3/\Gamma_n$. *The isometry of Proposition 3.12 factors through this map, so that the component of* $M_n - F^{(i)}$ *containing* $C(\Gamma_S)$ *is taken isometrically to its image in* $C(\Gamma_-^{(i)})$.

In the remainder of the paper, we will frequently take the isometry above for granted and refer to the components obtained by splitting M_n along $F^{(i)}$ by $C(\Gamma_\pm^{(i)})$.

4. Invariants

4.1. ***Traces.*** If $\Gamma \subset \mathrm{PSL}_2(\mathbb{C})$ is a discrete group, its trace field $\mathbb{Q}(\operatorname{tr}\Gamma)$ is obtained by adjoining to $\mathbb{Q}$ the traces of elements of Γ. If the hyperbolic 3-manifold $M = \mathbb{H}^3/\Gamma$ has finite volume, Mostow rigidity implies that this is a topological invariant of M. It follows from the local rigidity theorems of Garland and Prasad that in this case the trace field is a number field; that is, a finite extension of $\mathbb{Q}$. The trace field is not generally an invariant of the commensurability class of M, however, and to obtain one we pass to the *invariant trace field* $k\Gamma$. This is obtained by adjoining to $\mathbb{Q}$ the traces of squares of elements of Γ. When M is the complement of a link in a $\mathbb{Z}_2$-homology sphere, its trace field and invariant trace field coincide (see [Maclachlan and Reid 2003]).

Proposition 4.1. *We have* $k(\Gamma_S) = \mathbb{Q}(i)$, $k(\Gamma_T) = \mathrm{Q}(i\sqrt{2})$, *and* $k(M_n) = \mathbb{Q}(i, i\sqrt{2})$ *for all* $n \in \mathbb{N}$. *In particular*, M_n *is not arithmetic for any* $n \in \mathbb{N}$.

Proof. Its definition in Corollary 2.2 immediately implies $\Gamma_S < \mathrm{PSL}_2(\mathbb{Q}(i))$. The description in Corollary 2.3, of Γ_{T_0}, and Lemma 3.3 imply that $\Gamma_T < \mathrm{PSL}_2(\mathbb{Q}(i\sqrt{2}))$. Thus $k\Gamma_S \subseteq \mathbb{Q}(i)$, and $k\Gamma_T \subseteq \mathbb{Q}(i\sqrt{2})$. That equality holds is clear upon noting that $\mathrm{Tr}(\mathsf{h}) = \pm i\sqrt{2}$ and $\mathrm{Tr}(\mathsf{t}) = \pm(1+i)$. Since Γ_S and Γ_T are in Γ_n we have $\mathbb{Q}(i, i\sqrt{2}) \subseteq k(\Gamma_n)$. For the other containment we note that c from Lemma 3.3 lies in $\mathrm{PSL}_2(\mathbb{Q}(i\sqrt{2}))$, and Γ_n is contained in the group generated by Γ_S, Γ_T, and c.

It is well known that any noncompact arithmetic manifold M has $k(M) \subset \mathbb{Q}(i\sqrt{d})$ for some $d \in \mathbb{N}$ (see, for example, [Maclachlan and Reid 2003, Theorem 8.2.3]), so M_n is not arithmetic. □

We say $M = \mathbb{H}^3/\Gamma$ has *integral traces* if for each $\gamma \in \Gamma$, $\operatorname{tr}\gamma$ is an algebraic integer. Otherwise we say M_n has a nonintegral trace. M has integral traces if and only if all manifolds commensurable to M do as well (see [Maclachlan and Reid 2003]).

Proposition 4.2. *For each n, M_n has integral traces.*

Proof. As in the proposition above, this follows from the fact that each Γ_n is contained in the group generated by Γ_S, Γ_T, and c. It is easy to see that the *entries* of the generators for Γ_S and Γ_{T_0} are algebraic integers. Since c has integral entries as well, all elements of Γ_n have integral entries, and hence integral traces. □

Remark. Bass [1984] showed that if $M = \mathbb{H}^3/\Gamma$ where Γ has an element with a nonintegral trace, there are closed essential surfaces in M associated to this trace. We say that such surfaces are *detected by the trace ring*. For fixed n and $1 \le i \le n$, closed essential surfaces in M_n can be obtained by "tubing" $S^{(i)}$ through $B^3 - L_-^{(i)}$. More precisely, let $\mathcal{N}_i$ be a regular neighborhood of $L_-^{(i)}$ in $(B^3, L_-^{(i)}) \subset (S^3, L_n)$, let $A_i = \mathcal{N}_i \cap \overline{B^3 - \mathcal{N}_i}$, and let

$$\hat{S}_i = \overline{S^{(i)} - (S^{(i)} \cap \mathcal{N}_i)} \cup A_i .$$

Then $\hat{S}_i$ is a closed surface of genus 2 which is incompressible in M_n. We will show below that certain mutants have nonintegral traces, and one easily finds surfaces analogous to $\hat{S}_i$ in the mutants. It is interesting to note that although these surfaces are present in all of these link complements, the trace ring does not detect any closed surfaces in the M_n.

4.2. ***Scissors congruence and the Bloch invariant.*** In Proposition 4.7 we will prove that the *Bloch invariant* distinguishes the commensurability class of M_m from that of M_n for $m \neq n$. This is an invariant of a polyhedral decomposition which by construction is invariant under *scissors congruence*: cutting the constituent polyhedra apart and reassembling them in new ways. Its deep connection to algebraic k-theory is what makes the Bloch invariant useful, though. For background and an account of the connection to scissors congruence we refer the reader to [Dupont 2001] and [Neumann 1998], our main source for the expository material here.

Definition 4.3. For a field $k \subset \mathbb{C}$, define the *pre-Bloch group* $\mathcal{P}(k)$ to be the quotient of the free $\mathbb{Z}$-module on $k - \{0, 1\}$ by all instances of the following relations:

$$(4)\quad [x] - [y] + \left[\frac{y}{x}\right] - \left[\frac{1-x^{-1}}{1-y^{-1}}\right] + \left[\frac{1-x}{1-y}\right] = 0, \qquad x \neq y \in k - \{0, 1\},$$

$$(5)\quad [z] = \left[1 - \frac{1}{z}\right] = \left[\frac{1}{1-z}\right] = -\left[\frac{1}{z}\right] = -\left[\frac{z}{z-1}\right] = -[1-z], \quad z \in k - \{0, 1\}.$$

There is a map $\delta \colon \mathcal{P}(k) \to k^* \wedge k^*$ given by $[z] \mapsto 2(z \wedge (1 - z))$. (Here k^* is considered a $\mathbb{Z}$-module with multiplication as the group operation and $\mathbb{Z}$-action given by $a.x = x^a$, $a \in \mathbb{Z}$.) The *Bloch group* is $\mathcal{B}(k) = \ker \delta$.

Remark. If k is algebraically closed, relation (4) above, called the *five term relation*, implies (5). For instance, taking $\sqrt{z}$ and $\sqrt{z^{-1}}$ as x and y, respectively, in (4), then interchanging their roles and summing the results yields $[z]+[1/z]=0$.

For any ideal tetrahedron T in $\mathbb{H}^3$, there is an orientation-preserving isometry of $\mathbb{H}^3$ taking its ideal vertices to 0, 1, ∞, and a complex number z with nonnegative imaginary part. Let the *cross ratio parameter* of T be $[z] \in \mathcal{P}(\mathbb{C})$. This is well defined because any other isometry answering the description above fixes z or replaces it by one of $1-1/z$ or $1/(1-z)$.

For $k' \subset k$, inclusion induces a map $\mathcal{P}(k') \to \mathcal{P}(k)$. Although this is not injective in general, a theorem of Borel that we record below implies that if k' is a number field then $\mathcal{B}(k')$ does inject, modulo torsion. We offer this observation to excuse occasional imprecision about the precise location of our invariants.

Definition 4.4. Let $M = T_1 \cup \cdots \cup T_n$ be a triangulated complete, orientable hyperbolic 3-manifold of finite volume (with or without boundary); that is, with each T_i isometric to an ideal hyperbolic tetrahedron and $T_i \cap T_j$ either empty, an edge of each, or a face of each for $i \neq j$. Define the *Bloch invariant* of M as

$$\beta(M) = [z_1] + [z_2] + \cdots + [z_n] \in \mathcal{P}(\mathbb{C}),$$

where $[z_i]$ is the cross ratio parameter of T_i for each i in $\{1, \ldots, n\}$.

Remark. If $\partial M = \varnothing$ then $\beta(M) \in \mathcal{B}(\mathbb{C})$ by a geometric interpretation of the Bloch invariant, and by work of Neumann and Yang [1999] it does not vary with triangulation.

We will obtain a triangulation of M_n by subdividing the decomposition below.

Lemma 4.5. *The members of* $\mathcal{S}=\{\mathcal{P}_1, \mathcal{P}_2, \mathsf{c}^{-1}\mathcal{P}_2, \ldots, \mathsf{c}^{-2n+1}\mathcal{P}_2, \mathsf{c}^{-2n}\mathsf{r}\mathcal{P}_1\}$ *project under* $\mathbb{H}^3 \to \mathbb{H}^3/\Gamma_n$ *to the cells of an ideal polyhedral decomposition of* M_n.

Proof. By Corollary 2.2, $\mathcal{P}_1$ projects under $\mathbb{H}^3 \to \mathbb{H}^3/\Gamma_S$ to an ideal polyhedral decomposition of $C(\Gamma_S)$: it maps onto $C(\Gamma_S)$ with internal faces identified in pairs. Corollary 2.3 implies the same for $\mathcal{P}_2 \to C(\Gamma_{T_0})$ under $\mathbb{H}^3 \to \mathbb{H}^3/\Gamma_{T_0}$, and hence also for $\mathsf{c}^{-2}\mathsf{r}\mathcal{P}_2 \to C(\overline{\Gamma}_{T_0}^{c^{-2}})$ (see the paragraph above Lemma 3.3).

It is easy to see that $\mathsf{r}\mathcal{P}_2 = \mathsf{c}\mathcal{P}_2$, for instance, by comparing sets of ideal vertices, so $\mathsf{c}^{-2}\mathsf{r}\mathcal{P}_2 = \mathsf{c}^{-1}\mathcal{P}_2$. Therefore Lemma 3.3 implies that $\mathcal{P}_2 \cup \mathsf{c}^{-1}\mathcal{P}_2$ projects to an ideal polyhedral decomposition of $C(\Gamma_T)$ under $\mathbb{H}^3 \to \mathbb{H}^3/\Gamma_T$. In particular, this projection identifies the external faces of $\mathcal{P}_2$ that map to F' with external faces of $\mathsf{c}^{-1}\mathcal{P}_2$ pairwise, since their images are fixed by the doubling involution $\phi_{\mathsf{c}^{-2}\mathsf{r}}$.

It follows from the above that $\mathsf{c}^{-2(i-1)}\mathcal{P}_2 \cup \mathsf{c}^{-2i+1}\mathcal{P}_2$ projects to a decomposition of $C(\Gamma_T^{(i)})$ for any $i \in \mathbb{N}$ (recall Definitions 3.9), and from the first paragraph that the same holds for $\mathsf{c}^{-2n}\mathsf{r}\mathcal{P}_1 \to C(\overline{\Gamma}_S^{\mathsf{c}^{-2n}})$. By Proposition 3.12, it remains only to show that the gluings producing M_n preserve induced triangulations of boundaries. These are defined in Proposition 3.10. Lemma 2.4(3) implies that $\iota_+^{(0)}(\iota_-^{(0)})^{-1}$ preserves triangulations, and (3) does the same for $\iota_+^{(i)}(\iota_-^{(i)})^{-1}$ for $1 \le i \le n$. They combine to imply that the final map does as well. $\square$

Lemma 4.6. *M_n has Bloch invariant $\beta_1 - \bar{\beta}_1 + n\beta_2 \in \mathcal{B}(\mathbb{Q}(i, i\sqrt{2}))$ for any n, where $\beta_1 = 4[(1+i)/2] \in \mathcal{P}(\mathbb{Q}(i))$, $\bar{\beta}_1 = 4[(1-i)/2]$, and $\beta_2 \in \mathcal{P}(\mathbb{Q}(i\sqrt{2}))$.*

Proof. We will produce a triangulation of M_n by subdividing the polyhedral decomposition from Lemma 4.5. $\mathcal{P}_1$ divides into a collection of 4 tetrahedra by the addition of a single edge γ joining the ideal vertices $(1+i)/2$ and ∞, and four ideal triangular faces that share γ. One has ideal vertices 0, 1, ∞, and $(1+i)/2$ and thus a parameter of $[(1+i)/2]$. Since the others are its image under rotation about γ they have identical cross ratio parameters. Their union projects to a triangulation of $C(\Gamma_S)$ with Bloch invariant $\beta_1 = 4[(1+i)/2]$.

Any ideal tetrahedron with its vertex set contained in that of $\mathcal{P}_2$ has cross ratio parameter in $\mathcal{P}(\mathbb{Q}(i\sqrt{2}))$, since $\mathcal{P}_2$ has ideal vertices in $\mathbb{Q}(i\sqrt{2}) \cup \{\infty\}$. We leave it to the reader to divide $\mathcal{P}_2$ into ideal tetrahedra in such a way that the resulting division of square faces, each into two ideal triangles, is preserved by the face-pairings that produce M_{T_0}. Such a triangulation projects to one of $C(\Gamma_{T_0})$, and its image under $\mathsf{c}^{-2}\mathsf{r}$ projects to one of $C(\Gamma_{T_0})$.

Above it is important to use $\mathsf{c}^{-2}\mathsf{r}$ and not c^{-1}, since the face pairings of $\mathsf{c}^{-1}\mathcal{P}_2$ project it to $C(\bar{\Gamma}_{T_0}^{\mathsf{c}^{-2}})$. Recall that r is a reflection, extending to $\mathbb{C}$ as complex conjugation. One checks using (5) that if a tetrahedron has cross ratio parameter $[z]$ then its mirror image has parameter $-[\bar{z}]$. Since $\mathbb{Q}(i\sqrt{2})$ is preserved by complex conjugation, using the triangulations from the paragraph above gives $C(\Gamma_T)$ a Bloch invariant $\beta_2 \in \mathcal{P}(\mathbb{Q}(i\sqrt{2}))$.

For each i with $1 \le i \le n$, $C(\Gamma_T^{(i)})$ inherits a triangulation with Bloch invariant β_2 from $\mathsf{c}^{-2(i-1)}\mathcal{P}_2 \cup \mathsf{c}^{-2i+1}\mathcal{P}_2 = \mathsf{c}^{-2(i-1)}(\mathcal{P}_2 \cup \mathsf{c}^{-1}\mathcal{P}_2)$, and $C(\bar{\Gamma}_S)$ inherits one with invariant $\bar{\beta}_1$ from $\mathsf{r}(\mathcal{P}_1)$. Lemma 4.5 implies that these combine to triangulate M_n, so its Bloch invariant is as described above. $\square$

Below we record a standard formula for the *Bloch–Wigner dilogarithm function* $D_2 \colon \mathbb{C} - \{0, 1\} \to \mathbb{R}$ in terms of the *dilogarithm*, $\psi(z) = \sum_{i=1}^{\infty}(z^n/n^2)$ (for $|z| < 1$):

$$D_2(z) = \Im\psi(z) + \log|z| \arg(1-z).$$

For z in the upper half plane, the ideal tetrahedron with ideal vertices 0, 1, ∞, and z has volume $D_2(z)$; note also that $D_2(\bar{z}) = -D_2(z)$. D_2 determines a well-defined functional on $\mathcal{P}(\mathbb{C})$, and this in turn produces the *Borel regulator*, B_k.

Theorem [Borel 1977]. *For a number field k fix embeddings $\sigma_1, \ldots, \sigma_{r_2}$ to $\mathbb{C}$, one representing each complex-conjugate pair. The map $\mathrm{B}_k \colon \mathcal{P}(k) \to \mathbb{R}^{r_2}$ extending $[z] \mapsto (D_2(\sigma_1(z)), \ldots, D_2(\sigma_{r_2}(z)))$ takes $\mathcal{B}(k)$ onto a lattice in $\mathbb{R}^{r_2}$, with kernel consisting entirely of torsion elements.*

We use the Borel regulator B_k to show that Bloch invariants distinguish the commensurability class of M_m from that of M_n for $m \neq n$.

Proposition 4.7. *For $m \neq n$, M_m is not commensurable with M_n.*

Remark. We thank the referee on an earlier version of this paper for describing the argument below. (Our original proof used cusp parameters; see Lemma 4.18.)

Proof. It is clear that $k = \mathbb{Q}(i, i\sqrt{2})$ has two pairs of complex conjugate embeddings, each determined by its action on i and $i\sqrt{2}$. We will take $\sigma_1 = id_k$, and $\sigma_2(i) = i$, $\sigma_2(i\sqrt{2}) = -i\sqrt{2}$, in defining the Borel regulator B_k on k. Since each σ_i restricts on $\mathbb{Q}(i)$ to the identity, B_k takes each of β_1 and $-\bar{\beta}_1$ to $(v_1, v_1) \in \mathbb{R}^2$, where v_1 is the volume of $\mathcal{P}_1$. On the other hand, $\mathrm{B}_k(\beta_2) = 2(v_2, -v_2)$, where $v_2 = \mathrm{vol}(\mathcal{P}_2)$.

For any n, a covering space $\widetilde{M} \to M_n$ of degree k has $\beta(\widetilde{M}) = k\beta(M_n)$. This is because the preimage in $\widetilde{M}$ of each tetrahedron T from the triangulation of M_n described in Lemma 4.6 is a nonoverlapping union of k isometric copies of T. Thus if $\widetilde{M} \to M_m$ with degree p and $\widetilde{M} \to M_n$ with degree q it would follow that

$$p[\beta_1 - \bar{\beta}_1 + m\beta_2] = q[\beta_1 - \bar{\beta}_1 + n\beta_2].$$

Applying B_k to each side of the equation above, we find that since (v_1, v_1) and $(v_2, -v_2)$ are linearly independent in $\mathbb{R}^2$ we must have $p = q$ and $m = n$. $\square$

4.3. ***Cusp parameters.*** Following Neumann and Reid [1992, §2.3], for a cusp of a complete hyperbolic 3-manifold M we will call the *cusp parameter* the complex modulus (or the conformal parameter) of a horospherical cusp cross section, a Euclidean torus. Thurston [1979, Chapter 6] also used this invariant to distinguish hyperbolic manifolds.

Definition 4.8. Let $T = \mathbb{C}/\Lambda$ be a Euclidean torus, where $\Lambda \subset \mathbb{C}$ is a lattice. Define the *complex modulus* of T as $m(T) = \alpha/\beta$, where $\Lambda = \langle \alpha, \beta \rangle$.

Remark. The complex modulus is not really an invariant of a Euclidean torus; rather, it is an invariant of a particular basis for π_1. However, we have:

Lemma 4.9. *The* $\mathrm{PGL}_2(\mathbb{Z})$*-orbit of the complex modulus is a similarity invariant of Euclidean tori. The* $\mathrm{PGL}_2(\mathbb{Q})$*-orbit is a commensurability invariant.*

Here we say T and T' are *commensurable* if T has a finite cover which is similar to a cover of T'.

Proof. The complex modulus is clearly scale-invariant.

Let $T = \mathbb{C}/\Lambda$ be a Euclidean torus, where $\Lambda = \langle \alpha, \beta \rangle$. For a different generating pair $\gamma = p\alpha + q\beta$, $\delta = r\alpha + s\beta$ the change-of-basis matrix

$$\mathsf{m} = \begin{pmatrix} p & r \\ q & s \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z})$$

has an inverse there as well, since α and β are linear combinations of γ and δ. Computing the modulus with γ and δ yields

$$\frac{p\alpha+q\beta}{r\alpha+s\beta}=\frac{p(\alpha/\beta)+q}{r(\alpha/\beta)+s}=\mathsf{m}^{\mathrm{T}}(m(T)).$$

If γ and λ generate a finite-index sublattice then, since they are linearly independent, m has a nonzero determinant. This implies the commensurability-invariance assertion. □

It will prove useful here to understand the complex modulus of a torus by decomposing it into annuli using a family of parallel geodesics.

Definition 4.10. For a Euclidean annulus A with core of length ℓ and distance d between geodesic boundary components, let the *real modulus* of A be $m(A)=d/\ell$.

If $T=\mathbb{C}/\Lambda$, and $\Lambda=\langle\alpha,\beta\rangle$, then α and β determine isotopy classes of simple closed geodesics on T with representatives which intersect once. These are the projections to T of the line segments in $\mathbb{C}$ joining 0 to α and β, respectively. Below let A_β denote the Euclidean annulus with geodesic boundary obtained as the path completion of the metric on $T-\beta$ inherited from T.

Lemma 4.11. *Let $T=C/\Lambda$ be a Euclidean torus, and suppose α, β is a generating pair for Λ. Decompose $m(T)$ into real and imaginary parts*:

$$m(T)=\tau_\beta+i\cdot\mu_\beta,$$

where $\tau_\beta=\Re(\alpha/\beta)$ and $\mu_\beta=\Im(\alpha/\beta)\in\mathbb{R}$. Then $\tau_\beta=(\|\alpha\|/\|\beta\|)\cos\theta$, where θ is the angle between the geodesics α and β on T, and $|\mu_\beta|=m(A_\beta)$.

Proof. Write $\alpha=\|\alpha\|e^{i\theta_1}$ and $\beta=\|\beta\|e^{i\theta_2}$. Then $\theta=\theta_1-\theta_2$ is the angle between the geodesics corresponding to α and β, and $\alpha/\beta=(\|\alpha\|/\|\beta\|)e^{i\theta}$. Writing $e^{i\theta}=\cos\theta+i\sin\theta$ yields the first assertion immediately.

To establish the second, consider the strip $\tilde{A}_\beta$ in $\mathbb{C}$ bounded by the line containing 0 and β and its translate by α, containing α and $\alpha+\beta$. The quotient of $\tilde{A}_\beta$ induced by the action of β is the universal covering $\tilde{A}_\beta\to A_\beta$. The distance between boundary components of $\tilde{A}_\beta$ is $\|\alpha\||\sin\theta|$, and the length of the core of A_β is the translation length of β, which is $\|\beta\|$. □

Lemma 4.11 provides a convenient means for understanding the modulus of a Euclidean torus in terms of "Fenchel–Nielsen" coordinates (μ_β,τ_β) associated to a simple closed geodesic β. We regard μ_β as a length parameter for the annulus A_β, and τ_β as a twist parameter.

Lemma 4.12. *Suppose T is a Euclidean torus decomposed into annuli $A_1,\dots,A_n$ by simple closed geodesics parallel to β. Then*

$$|\mu_\beta|=m(A_1)+m(A_2)+\cdots+m(A_n).$$

Proof. By isotoping β if necessary, we may assume that it is one of the geodesics determining the A_i; hence $A_\beta = A_1 \cup A_2 \cup \cdots \cup A_n$. Then if α_0 is an arc perpendicular to ∂A_β, joining one component to the other, for each i, $\alpha_0 \cap A_i$ is an arc perpendicular to ∂A_i joining one component to the other. This is because ∂A_i is parallel to β. Since $\ell(\alpha_0) = \sum_i \ell(\alpha_0 \cap A_i)$ and the core of each A_i has length $\ell(\beta)$, the result follows. $\square$

The annuli we are concerned with arise as horospherical cross sections of the cusps of M_S and M_T. Recall from Lemma 2.4 that $\mathrm{Stab}_{\Gamma_S}(\mathcal{H})$ is a group Λ generated by parabolic isometries p_1, p_2, and p_3. Furthermore, as pointed out in Remark 1 on page 351, p_1 and p_3 are conjugate in Γ_S, as are p_2 and $\mathsf{p}_4 = \mathsf{p}_1\mathsf{p}_2\mathsf{p}_3^{-1}$. We asserted there that $C(\Gamma_S)$ has two cusps, one corresponding to p_1 and one to p_2. This follows from the lemma below.

In what follows, we let $\mathcal{V}_1 = \{\infty, 0, 1, i, 1+i, (1+i)/2\}$, the set of ideal vertices of the ideal octahedron $\mathcal{P}_1$. Let $\{h_v \mid v \in \mathcal{V}_1\}$ be a collection of horospheres invariant under the action of the symmetry group of $\mathcal{P}_1$, such that h_v is centered at v for each $v \in \mathcal{V}_1$ and h_∞ is at height 2.

Lemma 4.13. *The projection to M_S of $\bigcup(h_v \cap \mathcal{P}_1)$ is a disjoint union of Euclidean annuli A_1 and A_2 with geodesic boundary, such that $p_S(A_1)$ is a horospherical cross section of the cusp of $C(\Gamma_S)$ corresponding to p_1, $p_S(A_2)$ is a cross section of the cusp corresponding to p_2, and $m(A_1) = 1$, $m(A_2) = 1/5$.*

Proof. Since h_∞ is at height 2 and our embedding of $\mathcal{P}_1$ is as in Figure 3, $h_\infty \cap \mathcal{P}_1$ is a square with sides of length $1/2$. Since the symmetry group of $\mathcal{P}_1$ acts transitively on vertices, this holds for all $h_v \cap \mathcal{P}_1$, $v \in \mathcal{V}_1$. We will call a side of $h_v \cap \mathcal{P}_1$ *internal* if it is contained in an internal face of $\mathcal{P}_1$ and *external* otherwise. The face-pairing s has the property that if v and v' are ideal vertices of $\mathcal{P}_1$ and $\mathsf{s}(v) = v'$, then $\mathsf{s}(h_v) = h_{v'}$, and $\mathsf{s}(h_v \cap \mathcal{P}_1)$ abuts $h_{v'} \cap \mathcal{P}_1$ along an internal side. The analogous property holds for t.

Each of s and t identifies a pair of internal faces of $\mathcal{P}_1$, yielding M_S. The isometry p_S of Corollary 2.2 is induced by the inclusion $\mathcal{P}_1 \to \mathbb{H}^3$. Since $\mathsf{p}_1 = \mathsf{s}^{-1}$ fixes the ideal vertex of $\mathcal{P}_1$ at 0, it identifies the opposite internal sides of $h_0 \cap \mathcal{P}_1$. This square thus projects to a cusp cross section A_1 of M_S, mapped by p_S to one of the cusp of $C(\Gamma_S)$ corresponding to p_1. This is depicted on the left side of Figure 8.

The other cusp cross section of M_S, the annulus A_2, is the identification space of the collection

$$\{h_v \cap \mathcal{P}_1 \mid v \in \mathcal{V}_1 - \{0\}\},$$

shown on the right side of Figure 8. In this figure, each square is the projection to M_S of $h_v \cap \mathcal{P}_1$ for the ideal vertex v, by which it is labeled. The combinatorics can be verified by considering the action of s and t on $\mathcal{V}_1$.

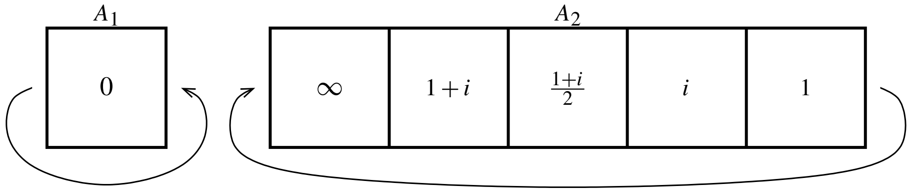


Figure 8. Cross sections of the cusps of M_S.

By assumption each square in Figure 8 has side length $1/2$, and so the cores of A_1 and A_2 have lengths $1/2$ and $5/2$, respectively. For any square in Figure 8, a vertical side projects to an arc joining the distinct boundary components of the corresponding A_i, hence the distance between them is $1/2$. Thus it follows directly from the definition that $m(A_1) = 1$ and $m(A_2) = 1/5$. $\square$

The following lemma describes the moduli of the cusps of $C(\Gamma_{T_0})$. We asserted below Lemma 2.5 that $C(\Gamma_{T_0})$ has four cusps, one corresponding to each p_i, $i \in \{1, 2, 3, 4\}$, and each joining $\partial_- C(\Gamma_{T_0})$ to $\partial_+ C(\Gamma_{T_0})$. This follows from Lemma 4.14. Let $\mathcal{V}_2$ be the set of ideal vertices of $\mathcal{P}_2$, and consider a collection of horospheres $\{h_v \mid v \in \mathcal{V}_2\}$, invariant under the symmetry group of $\mathcal{P}_2$, such that h_v is centered at v for each $v \in \mathcal{V}$ and h_∞ is at height 2.

Lemma 4.14. *The projection of* $\bigcup(h_v \cap \mathcal{P}_2)$ *to* M_{T_0} *is a collection of disjoint Euclidean annuli* B_j *with geodesic boundary*, $j \in \{1, 2, 3, 4\}$, *such that* $p_{T_0}(B_j)$ *is a cross section of the cusp of* $C(\Gamma_{T_0})$ *corresponding to* $\mathsf{p}_j \in \Lambda$, *and* $m(B_1) = m(B_3) = \sqrt{2}$ *and* $m(B_2) = m(B_4) = \sqrt{2}/5$.

Proof. For $v \in \mathcal{V}_2$, we again call a side of $h_v \cap \mathcal{P}_2$ external if it is contained in an external face of $\mathcal{P}_2$ and internal otherwise. Each cusp cross section of M_{T_0} is the projection of a subcollection of the $h_v \cap \mathcal{P}_2$, identified along their internal faces. From Figure 4, we find that $h_\infty \cap \mathcal{P}_2$ is a Euclidean rectangle with two opposite internal sides and two external. Since the symmetry group of $\mathcal{P}_2$ is transitive on its set of ideal vertices, this holds for the other h_v as well. It follows that each cusp cross section of M_{T_0} is a Euclidean annulus with geodesic boundary.

In Figure 9, the lower rectangles of each annulus DB_j are labeled by vertices v such that $h_v \cap \mathcal{P}_2$ projects to a subrectangle of the cross section of the cusp of M_{T_0} whose image under p_{T_0} corresponds p_j. Then B_j is the lower half of DB_j. The reasons for this picture will become clear after the current proof.

The isometries f, g, and h defined in Corollary 2.3 identify the internal sides of $\mathcal{P}_2$ in pairs, yielding the manifold M_{T_0} with totally geodesic boundary. The parabolic $\mathsf{p}_1 = \mathsf{f}^{-1}$ fixes 0, identifying the internal sides of $\mathcal{P}_2$ sharing this ideal vertex. Thus in M_{T_0}, B_1 consists of $h_0 \cap \mathcal{P}_2$ with its internal sides identified. The description of the p_i in terms of f, g, and h above Lemma 2.4 shows that p_3 is a

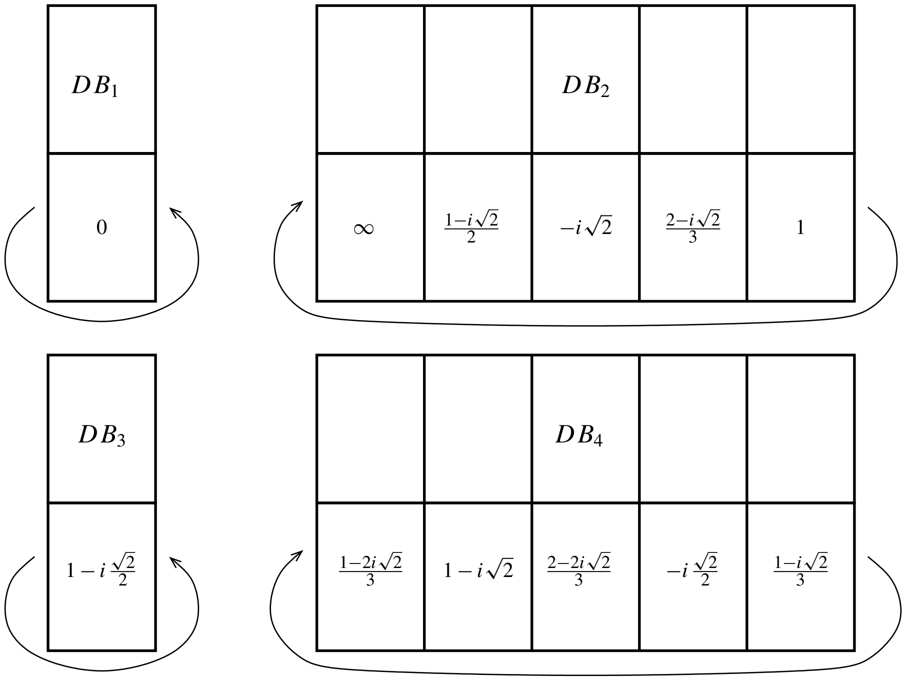


Figure 9. Cross sections of the cusps of M_T.

conjugate of g^{-1}. Since g^{-1} fixes $1-i\sqrt{2}/2$, identifying the internal edges of $\mathcal{P}_2$ which abut it, $h_{1-i\sqrt{2}/2}$ projects to B_3 in M_{T_0}. This justifies the depictions of B_1 and B_3 in Figure 9.

Since p_2 fixes ∞, $h_\infty \cap \mathcal{P}_2$ projects to a subrectangle of B_2. Since g takes the internal side Y_3 to Y_1' and ∞ to $(1-i\sqrt{2})/2$, in B_2 the projection of $h_\infty \cap \mathcal{P}_2$ meets the projection of $h_{(1-i\sqrt{2})/2} \cap \mathcal{P}_2$ along a side contained in the projection of Y_3 to M_{T_0}. Since the internal face of $\mathcal{P}_2$ meeting Y_1' at $(1-i\sqrt{2})/2$ is Y_3', and this is taken to Y_2' by h^{-1}, the rectangle meeting the projection of $h_{(1-i\sqrt{2})/2}$ in B_2 on the internal side opposite its intersection with h_∞ is $h_{-i\sqrt{2}}$. Carrying this line of argument to completion yields the depictions of B_2 and B_4 in the figure.

From Figure 4, we find that the internal sides of $h_\infty \cap \mathcal{P}_2$ have length $\sqrt{2}/2$ and the external sides length $1/2$. Since the symmetry group of $\mathcal{P}_2$ is transitive on its ideal vertices, the same holds for each rectangle $h_v \cap \mathcal{P}_2$. Thus the cores of B_1 and B_3 have length $1/2$, and the cores of B_2 and B_4 have length $5/2$. For any square $h_v \cap \mathcal{P}_2$, an internal side projects to a perpendicular arc joining opposite sides of the cusp cross section in M_{T_0} containing $h_v \cap \mathcal{P}_2$. The moduli are thus as described. $\square$

By Corollary 3.4, $p_{T_0} : M_{T_0} \to C(\Gamma_{T_0})$ determines a reflection-invariant map from the double M_T of M_{T_0} across $\partial_+ M_{T_0}$ to $C(\Gamma_T)$. Furthermore, as we remarked below Lemma 2.5, each cusp of $C(\Gamma_{T_0})$ joins one component of $\partial C(\Gamma_{T_0})$ to the other. Therefore taking $DB_j \subset M_T$, $j \in \{1, 2, 3, 4\}$, to be the double of B_j across its component of intersection with $\partial_+ M_{T_0}$, we have:

Lemma 4.15. *For each $j \in \{1, 2, 3, 4\}$, the image in $C(\Gamma_T)$ of DB_j is a cross section of the cusp corresponding to p_j, and $m(DB_1) = m(DB_3) = 2\sqrt{2}$ and $m(DB_2) = m(DB_4) = 2\sqrt{2}/5$.*

It is a well-known consequence of the Margulis lemma that each cusp C of a hyperbolic manifold $M = \mathbb{H}^3/\Gamma$ of finite volume is foliated by similar Euclidean tori, the projections to M of horospheres in $\mathbb{H}^3$ centered at the fixed point of a parabolic subgroup of Γ corresponding to C.

Definition 4.16. The *parameter* of a cusp C of a finite-volume complete hyperbolic manifold is the complex modulus of a horospherical cross section of C.

By Lemma 4.9 the $\mathrm{PSL}_2(\mathbb{Z})$ orbit of the cusp parameter is an invariant of the *cusp shape*, the Euclidean similarity class of a cross section.

Proposition 4.17. *For $j = 1, 2$, let T_j be a cusp cross section of M_n containing the annular cusp cross section A_j of M_S (see Lemma 4.13). Then $m(T_1) = i(2+4n\sqrt{2})$, and $m(T_2)$ is $\mathrm{PGL}_2(\mathbb{Q})$-equivalent to $m(T_1)$.*

Remarks. 1. It is not hard to show that $m(T_2) = i(2 + 4n\sqrt{2})/5$, but this is not necessary for our purposes and requires more work.

2. The cusps T_1 and T_2 are labeled in Figure 1.

Proof. By Proposition 3.10 $M_n = C(\Gamma_S) \cup C(\Gamma_T^{(1)}) \cup \cdots \cup C(\Gamma_T^{(n)}) \cup C(\overline{\Gamma}_S)$, with gluing maps that factor through the inclusion-induced isometries $\iota_\pm^{(i)}$ defined on $F^{(i)}$ for $0 \le i \le n-1$, and final gluing $\phi_{\mathsf{r}}\iota_-^{(0)}\phi_{n+1}^{-1}(\iota_-^{(n)})^{-1}\colon \partial_+ C(\Gamma_T^{(n)}) \to \partial C(\overline{\Gamma}_S)$.

For $j \in \{1, 2, 3, 4\}$ and $i \in \mathbb{N}$, define $DB_j^{(i)} = \phi_i \circ p_T(DB_j) \subset C(\Gamma_T^{(i)})$, with ϕ_i as in Definitions 3.9. We also refer by $DB_j^{(i)}$ to its image in $C(\Gamma_n)$ under the natural map, or in M_n under inclusion. Let $\partial_\pm DB_j^{(i)} = DB_j^{(i)} \cap \partial_\pm C(\Gamma_T^{(i)})$.

By Lemma 4.13, $p_S(A_1)$ is a cross section of the cusp of $C(\Gamma_S)$ corresponding to p_1, and by Lemma 4.15, $DB_1^{(1)}$ is a cross section of the cusp of $C(\Gamma_T)$ corresponding to p_1. Lemma 2.4 thus implies that $\iota_+^{(0)}(\iota_-^{(0)})^{-1}$ takes one component of $p_S(\partial A_1)$ to $\partial_- DB_1^{(0)}$. In Remark 1 on page 351, we note that p_1 and p_3 are conjugate in Γ_S. It follows that the other component of $p_S(\partial A_1)$ is a cross section of the cusp of $\partial C(\Gamma_S)$ corresponding to p_3, so $\iota_+^{(0)}(\iota_-^{(0)})^{-1}$ takes this component to $\partial_- DB_3^{(1)}$.

The doubling involution of M_T preserves DB_j by construction, exchanging its boundary components. Therefore by Corollary 3.4, $\phi_{\mathsf{c}^{-2}\mathsf{r}}$ preserves $p_T(DB_j)$ and exchanges boundary components. It follows that $\iota_+^{(i)}(\iota_-^{(i)})^{-1}$ takes $\partial_+ DB_j^{(i)}$ to $\partial_- DB_j^{(i+1)}$ for each i between 1 and $n-1$, upon recalling the identity (3):

$$\iota_+^{(i)}(\iota_-^{(i)})^{-1} = \phi_{i+1}\iota_+^{(0)}\phi_2^{-1}(\iota_-^{(1)})^{-1}\phi_i^{-1} = \phi_{i+1}\phi_{\mathsf{c}^{-2}\mathsf{r}}\phi_i^{-1}.$$

One finds that $\phi_{\mathsf{r}}\iota_-^{(0)}\phi_{n+1}^{-1}(\iota_-^{(n)})^{-1}$ takes $\partial_+ DB_1^{(n)} \sqcup \partial_+ DB_3^{(n)}$ to the components of $\phi_{\mathsf{r}} \circ p_S(\partial A_1)$, arguing as above and applying (2). Therefore T_1 is decomposed by its intersection in M_n with the separating spheres $F^{(i)}$ into the following collection of Euclidean annuli with geodesic boundary:

$$p_S(A_1) \cup DB_1^{(1)} \cup \cdots \cup DB_1^{(n)} \cup \phi_{\mathsf{r}} \circ p_S(A_1) \cup DB_3^{(n)} \cup \cdots \cup DB_3^{(1)}.$$

Similarly, we find that T_2 decomposes into the union of $p_S(A_2)$, $\phi_{\mathsf{r}} \circ p_S(A_2)$, and $DB_j^{(i)}$ for $1 \le i \le n$ and $j = 2, 4$. We may take β_1 to be the geodesic $\partial_- DB_1^{(1)}$ on T_1 and $\beta_2 = \partial_- DB_2^{(1)} \subset T_2$. Then we obtain the following from Lemma 4.12, applying Lemmas 4.13 and 4.15:

$$\Im(m(T_1)) = \pm(2 + 4n\sqrt{2}), \quad \Im(m(T_2)) = \pm\frac{2 + 4n\sqrt{2}}{5}.$$

We will show $m(T_1)$ and $m(T_2)$ have real part equal to 0 by describing geodesics α_j, $j = 1, 2$, which meet the β_j once, perpendicularly. Let a_1 be the arc in A_1 which is the projection of the internal edges of $h_0 \cap \mathcal{P}_1$ (the vertical arcs on the left-hand square in Figure 8). Recall that the internal edges of $h_0 \cap \mathcal{P}_1$. In particular, $p_S(\partial a_1)$ is the intersection of $p_S(A_1)$ with the one-skeleton of the triangulation Δ_S defined below Corollary 2.2.

Let $b_1 \subset B_1$ and $b_3 \subset B_3$ similarly be projections of internal edges of $h_0 \cap \mathcal{P}_2$ and $h_{1-i\sqrt{2}/2} \cap \mathcal{P}_2$, respectively (see Figure 9), and let db_1 and db_3 be the geodesic arcs of DB_1 and DB_3 containing them. Let $db_j^{(i)} = \phi_i \circ p_T(db_j)$, and let $\partial_\pm db_j^{(i)} = db_j^{(i)} \cap \partial_\pm DB_j^{(i)}$, $j = 1, 3$ and $i \in \mathbb{N}$. Let Δ_T^- be the image of the triangulation $\Delta_{T_0}^-$ defined below Corollary 2.3 under the inclusion $M_{T_0} \to M_T$, and let Δ_T^+ be its image under the doubling involution of M_T. Then $\partial_\pm db_j^{(i)}$ is the intersection of $\partial DB_j^{(i)}$ with the one-skeleton of $\phi_i(\Delta_T^\pm)$.

By Lemma 2.4, $\iota_+^{(0)}(\iota_-^{(0)})^{-1}$ preserves triangulations, and the discussion above implies that the other gluing maps do as well. From Figure 5 it is apparent that the cusps of $F^{(0)}$ corresponding to p_1 and p_3 each contain only one end of an edge of the triangulation that $F^{(0)}$ inherits from the pictured fundamental domain $\mathcal{F}$. Therefore $\iota_0(\partial a_1) = \partial_- db_1^{(1)} \cup \partial_- db_3^{(1)}$. It then follows as before that

$$\alpha_1 = p_S(a_1) \cup db_1^{(1)} \cup \cdots \cup db_1^{(n)} \cup \phi_{\mathsf{r}} \circ p_S(a_1) \cup db_3^{(n)} \cup \cdots \cup db_3^{(1)}$$

is a closed geodesic on T_1 which meets β_1 once, at right angles. Therefore by Lemma 4.11 $\Re(m(T_1)) = 0$, so $m(T_1) = i(2 + 4n\sqrt{2})$.

A similar argument will give $m(T_2)$. Let $\mathcal{A}_2$ be the collection of arcs in A_2 which are the projections of internal edges of the squares h_v, $v \in \mathcal{V}_1 - \{0\}$. From Figure 8, $\mathcal{A}_2$ consists of five arcs evenly spaced around A_2, each joining one component of ∂A_2 to the other and perpendicular to ∂A_2 at each endpoint. For $j = 2, 4$, we define a collection of arcs $D\mathcal{B}_j \subset DB_j$ analogously, and take $D\mathcal{B}_j^{(i)} = \phi_i \circ p_T(D\mathcal{B}_j)$. Let

$\partial_\pm D\mathcal{B}_j^{(i)} = D\mathcal{B}_j^{(i)} \cap \partial_\pm DB_j^{(i)}$, and note that the points of $\partial_\pm D\mathcal{B}_j^{(i)}$ are the points of intersection of $\partial DB_j^{(i)}$ with the one-skeleton of $\phi_i(\Delta_T^\pm)$.

For the same reasons as above, $i_+^{(0)}(i_-^{(0)})^{-1}$ takes $p_S(\partial\mathcal{A}_2)$ to $\partial_- D\mathcal{B}_2^{(1)} \cup \partial_- D\mathcal{B}_4^{(1)}$, and the other gluing maps take the $\partial_+ D\mathcal{B}_j^{(i)}$ to $\partial_- D\mathcal{B}_j^{(i+1)}$ for the appropriate i and j. Then the collection

$$p_S(\mathcal{A}_2) \cup D\mathcal{B}_2^{(1)} \cup \cdots \cup D\mathcal{B}_2^{(n)} \cup \phi_{\mathsf{r}} \circ p_S(\mathcal{A}_2) \cup D\mathcal{B}_4^{(n)} \cup \cdots \cup D\mathcal{B}_4^{(1)}$$

consists of a disjoint union of up to five closed geodesics, each meeting β_2 perpendicularly in at most five points.

Fix a component α_2 of the collection above, let k be the intersection number of α_2 with β_2, and let $\tilde{T}_2$ be the k-fold cover of T_2 dual to α_2. Then β_2 lifts to $\tilde{T}_2$, and any lift intersects the preimage $\tilde{\alpha}_2$ of α once, perpendicularly. Computing the modulus of $\tilde{T}_2$ using this pair, we obtain $\pm k \cdot i(2+4n\sqrt{2})/5$. This is $\mathrm{PGL}_2(\mathbb{Q})$-equivalent to $m(T_1)$, so the result follows from Lemma 4.9. □

Lemma 4.18. *Suppose* $z = i(m+n\sqrt{2})$ *is* $\mathrm{PGL}_2(\mathbb{Q})$*-equivalent to* $z' = i(m+n'\sqrt{2})$, *where* $m, n, n' \in \mathbb{Q}$ *and* $m \neq 0$*. Then* $n' = \pm n$*.*

Remark. Since commensurable hyperbolic manifolds have commensurable cusps, the collection of $\mathrm{PGL}_2(\mathbb{Q})$-orbits of cusp parameters is a commensurability invariant (see Lemma 4.9). Thus Proposition 4.17 and Lemma 4.18 imply Proposition 4.7.

Proof. Suppose $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PGL}_2(\mathbb{Q})$ takes z to z'. After clearing denominators (which does not change the action by Möbius transformations), we may assume that $a, b, c, d \in \mathbb{Z}$. We have

$$\frac{ai(m+n\sqrt{2})+b}{ci(m+n\sqrt{2})+d} = i(m+n'\sqrt{2}).$$

Multiplying by the denominator on the left, and collecting the real and imaginary parts, we find

$$m(a-d)+(an-dn')\sqrt{2}=0, \quad b+c(m^2+2nn')+cm(n'+n)\sqrt{2}=0.$$

Since 1 and $\sqrt{2}$ are linearly independent over $\mathbb{Q}$, the left-hand equation above implies that $m(a-d)=0$ and $an-dn'=0$. Since $m \neq 0$, the first equation implies $a=d$. Then the second equation implies $n=n'$ unless $a=d=0$. But in this case, $c \neq 0$ since $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PGL}_2(\mathbb{Q})$. Hence, using the coefficient of $\sqrt{2}$ in the right-hand equation above, we find $n' = -n$. □

5. Mutants

In the remaining sections, we will consider links obtained from L_n by mutation along the separating spheres $S^{(i)}$, $0 \leq i \leq n$, from Definitions 3.8(3).

Definition 5.1. For marked points $1, 2, 3, 4 \in S^2$, a *mutation* of $(S^2, \{1, 2, 3, 4\})$ is a mapping class of order 2 which acts on $\{1, 2, 3, 4\}$ by an even permutation.

Above a *mapping class* is the isotopy class, rel $\{1, 2, 3, 4\}$, of an orientation-preserving self-homeomorphism of the pair $(S^2, \{1, 2, 3, 4\})$. The set $\mathrm{Mod}_{0,4}$ of such classes inherits the structure of a group from its bijection with the quotient of the group of orientation-preserving homeomorphisms by its identity component. See, for example, [Farb and Margalit 2012] for an introduction to the study of mapping class groups; here we need only the following fact on recognizing mutations using the symmetric group S_4:

Proposition 5.2. *The homomorphism* $\theta\colon \mathrm{Mod}_{0,4} \to S_4$ *that records the action on* $\{1, 2, 3, 4\}$ *takes the set of mutations bijectively to* $\{(12)(34), (13)(24), (14)(23)\}$.

Proof. Here we will embed S^2 as the unit sphere in $\mathbb{R}^3$ and take

$$\{1, 2, 3, 4\} = \left\{ \left(\frac{\pm 1}{\sqrt{2}}, \frac{\pm 1}{\sqrt{2}}, 0 \right), \left(\frac{\pm 1}{\sqrt{2}}, \frac{\mp 1}{\sqrt{2}}, 0 \right) \right\}.$$

The definitions imply that θ takes any mutation into the subset of S_4 listed above, and the 180-degree rotations m_x, m_y, and m_z in the three coordinate axes of $\mathbb{R}^3$ determine mutations of $(S^2, \{1, 2, 3, 4\})$ taken by θ to each of its distinct elements.

The kernel of θ is the *pure mapping class group* $\mathrm{PMod}_{0,4}$. This group is free on two generators: Dehn twists in essential simple closed curves $\alpha, \beta \subset S^2 - \{1, 2, 3, 4\}$ that intersect exactly twice. See the beginning of [Farb and Margalit 2012, §4.2.4] for a proof of this fact, and for the definition of a Dehn twist see [Farb and Margalit 2012, §3.1.1]. With S^2 as above we can take α to be its intersection with the xz-plane and β the intersection with the yz-plane; then it is clear that each of m_x, m_y, and m_z takes each of α and β to itself. It follows that m_x, m_y, and m_z centralize $\mathrm{PMod}_{0,4}$ (see [Farb and Margalit 2012, Fact 3.8]).

For an arbitrary mutation $m \in \mathrm{Mod}_{0,4}$ we have $\theta(m) = \theta(m_x)$, $\theta(m) = \theta(m_y)$, or $\theta(m) = \theta(m_z)$. Assuming (without loss of generality) that the first case holds, it follows that $m = m_x h$ for some $h \in \mathrm{PMod}_{0,4}$. Since m has order 2 we have:

$$id = m^2 = (m_x h)^2 = m_x^2 h^2 = h^2.$$

Thus since $\mathrm{PMod}_{0,4}$ is a free group, $h = id$ and $m = m_x$. □

It is easy to see that every mutation of $(S^2, \{1, 2, 3, 4\})$ is isotopic to the identity as a self-homeomorphism of S^2, so cutting S^3 along a smoothly embedded copy and regluing by a mutation recovers S^3. This motivates:

Definition 5.3. For a link $L \subset S^3$ and a smoothly embedded two-sphere $S \subset S^3$ intersecting L in four points, let $B^{\pm}$ be the closures of the components of $S^3 - S$ and $T^{\pm} = L \cap B^{\pm}$. For a mutation m of $(S, S \cap L)$, we define $(S^3, L') = (B^-, T^-) \cup_m (B^+, T^+)$ and say L' is *obtained from L by mutation along S*.

The lemma below describes the change in projection from L to a link L' obtained from it by mutation. Below we refer to mutations by their images under θ.

Lemma 5.4. *For a link L projected to $\mathbb{R}^2$, if a two-sphere $S \subset S^3$ intersects $\mathbb{R}^2$ in a vertical line and L in four points, label them* 2, 3, 4, *and* 1, *reading top-to-bottom. The link obtained by the mutation* (13)(24) (*respectively*, (12)(34)) *along S has projection obtained by cutting L along S and inserting the braid on the left* (*resp. right*) *of Figure 10.*

Proof. We may assume L is arranged so that there is an axis in $\mathbb{R}^2$ intersecting S perpendicularly, midway between points 3 and 4 and so that points 2 and 1 are also equidistant from it. The 180-degree rotation in this axis restricts on S to an involution acting on the marked points by the permutation (12)(34).

There is a homeomorphism $R\colon S \times I \to S \times I$, which preserves slices $S \times \{t\}$ and restricts on each to rotation by $-180 \cdot t$ degrees in the horizontal axis. This interpolates between the identity, on $S \times \{0\}$, and the inverse of (12)(34) on $S \times \{1\}$, although it does not preserve marked points for $0 < t < 1$.

Let $(B^{\pm}, T^{\pm})$ be as in Definition 5.3, and let C be a collar of S in B^- that is small enough that it intersects T^- in the collection of horizontal arcs $\{\{j\} \times I \mid j \in \{1, 2, 3, 4\}\}$. There is a homeomorphism $h\colon B^- \cup_{(12)(34)} B^+ \to S^3$ defined as the identity on B^+ and the complement of C in B^-, and as R on C. By the definition of R, the image of $T^- \cap C$ under R is as pictured on the right-hand side of Figure 10, thus the image of L' in S^3 under h is as stated in the lemma.

Note that the braid on the left-hand side of Figure 10 is the conjugate of the braid on the right by a left-handed half-twist exchanging the points 2 and 3. This reflects the fact that the conjugate of (12)(34), by any homeomorphism of $(S, \{1, 2, 3, 4\})$ which exchanges 2 and 3 and fixes 1 and 4, is a mapping class of order 2 acting on the marked points as (13)(24); hence such a conjugate is (13)(24). The conjugating braid in Figure 10 tracks the marked points under an isotopy $S \times I \to S$ taking the simplest such conjugator to the identity. The conclusion for (13)(24) thus follows as it did above for (12)(34). $\square$

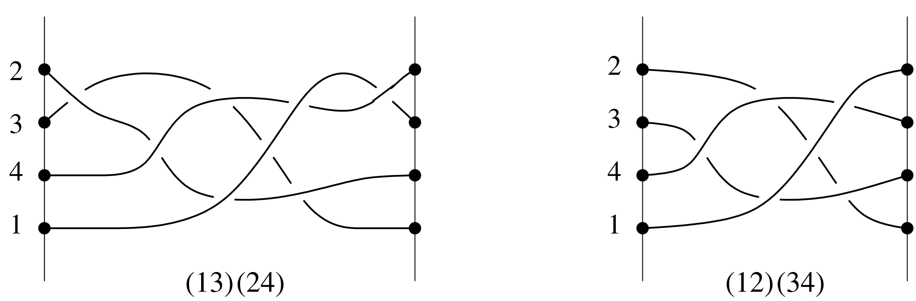


Figure 10. The mutations as braids.

The numbering of marked points from Lemma 5.4 and Figure 10 agrees with the numbering of the $S^{(i)} \cap L_n$ from Definitions 3.8(3). This, in turn, was chosen to agree with the numbering of parabolics of Λ from Lemma 2.4. To be more precise:

Let S be the sphere obtained by compactifying each cusp of $F^{(0)} = \mathcal{H}/\Lambda$ with a single point. Label each new point with a number between 1 and 4, according to the parabolic p_i corresponding to the cusp it compactifies. With the points of $S^{(0)} \cap L_n$ numbered as in Definitions 3.8(3), it follows from Proposition 2.8 that the restriction of f_T (as in Proposition 3.7) to $S^{(0)} - T$ extends to a map $S^{(0)} \to S$ that preserves numbering. Corollary 3.11 and the definition of $F^{(i)}$ (see Definitions 3.9) now imply that for each i between 0 and n, $\phi_{\mathsf{c}^{2i}} \circ \iota_i^{-1} \circ f_n$ extends to a homeomorphism $S^{(i)} \to S$ that takes marked points to marked points preserving numbering.

By [Ruberman 1987, Theorem 2.2], each mutation of $(S, \{1, 2, 3, 4\})$ is *realized* by an isometry of $F^{(0)}$; that is, there exists an isometry of $F^{(0)}$ whose extension to $(S, \{1, 2, 3, 4\})$ represents the mutation mapping class. The following lemma identifies lifts to $\mathrm{PSL}_2(\mathbb{R})$ of the isometries realizing $(13)(24)$ and $(12)(34)$.

Lemma 5.5. *Define*

$$\mathsf{m}_1 = \begin{pmatrix} -3 & 5 \\ -2 & 3 \end{pmatrix}, \quad \mathsf{m}_2 = \begin{pmatrix} 0 & \sqrt{5} \\ -1/\sqrt{5} & 0 \end{pmatrix}.$$

Each of m_1 *and* m_2 *normalizes* Λ (*from Lemma 2.4*), *and the induced isometries* ϕ_{m_1} *and* ϕ_{m_2} *of* $F^{(0)}$ *realize* $(13)(24)$ *and* $(12)(34)$, *respectively.*

Proof. Since each of m_1 and m_2 has trace equal to zero, it has order 2 in $\mathrm{PSL}_2(\mathbb{C})$. Their actions by conjugation on the generators p_1, p_2, and p_3 for Λ defined above Lemma 2.4 are given by

$$\mathsf{p}_1^{\mathsf{m}_1} = \mathsf{p}_3^{-1}, \quad \mathsf{p}_2^{\mathsf{m}_1} = \mathsf{p}_4^{-1}, \quad \mathsf{p}_1^{\mathsf{m}_2} = \mathsf{p}_2, \quad \mathsf{p}_3^{\mathsf{m}_2} = \mathsf{p}_4^{\mathsf{p}_1^{-1}},$$

as may be verified by direct computation. Here $\mathsf{p}_4 = \mathsf{p}_1\mathsf{p}_2\mathsf{p}_3^{-1}$ is as described in Remark 1 on page 351. Therefore m_1 and m_2 normalize Λ and induce isometries ϕ_{m_1} and ϕ_{m_2}, respectively, of $F^{(0)} = C(\Lambda)$.

Each of ϕ_{m_1} and ϕ_{m_2} has order 2, since m_1 and m_2 have order 2, and their extensions to S act on the set of marked points as described in the statement of the lemma. Its conclusion therefore follows from Proposition 5.2. □

Corollary 5.6. *For* $j = 1, 2$ *and* $i \in \mathbb{Z}$, *let* $\mathsf{m}_j^{(i)} = \mathsf{c}^{-2i}\mathsf{m}_j\mathsf{c}^{2i}$. *Each of* $\mathsf{m}_1^{(i)}$ *and* $\mathsf{m}_2^{(i)}$ *normalizes* $\Lambda^{(i)}$ (*from Definitions 3.9*), *and the induced isometries of* $F^{(i)}$ *realize* $(13)(24)$ *and* $(12)(34)$, *respectively.*

Lemma 5.4 gives a prescription for describing links obtained from L_n by the mutations $(13)(24)$ and $(12)(34)$. The result below describes hyperbolic manifolds to which their complements are homeomorphic, analogous to Proposition 3.10.

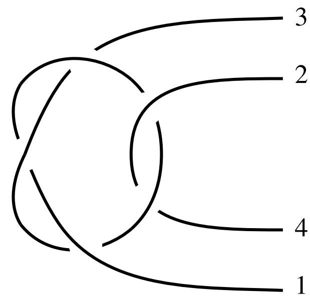


Figure 11. The tangle S admits an order-2 rotational symmetry which restricts to the mutation $(13)(24)$ on its boundary.

Proposition 5.7. *For $I = (a_0, a_1, \ldots, a_n) \in \{0, 1, 2\}^{n+1}$, let L_I be the link obtained from L_n by the following prescription: for $0 \le i \le n$, if $a_i = 0$, do not mutate along $S^{(i)}$; if $a_i = 1$, mutate by $(13)(24)$; and if $a_i = 2$, mutate by $(12)(34)$. Let $M_I = C(\Gamma_S) \cup C(\Gamma_T^{(1)}) \cup \cdots \cup C(\Gamma_T^{(n)}) \cup C(\overline{\Gamma}_S)$, where for each i such that $a_i = 0$ the gluing is as in Proposition 3.10, and otherwise is given by*

$$\iota_+^{(i)} \phi_{\mathsf{m}_j^{(i)}} (\iota_-^{(i)})^{-1} \qquad \text{for } 0 \le i < n, \text{ where } a_i = j \in \{1, 2\}, \text{ and}$$

$$\phi_{\mathsf{r}} \iota_-^{(0)} \phi_{\mathsf{c}^{2n}} \phi_{\mathsf{m}_j^{(n)}} (\iota_-^{(n)})^{-1} \qquad \text{if } a_n = j \in \{1, 2\}.$$

Then there is a homeomorphism $f_I \colon S^3 - L_I \to M_I$ whose restriction to each complementary component of the collection $\{S^{(i)}\}$ agrees with that of f_n.

Proposition 5.7 follows immediately from Proposition 3.10 and Corollary 5.6. Below we note a couple of "obvious" isometry relations on the $\{M_I\}$.

Lemma 5.8. *For fixed $(a_1, \ldots, a_n) \in \{0, 1, 2\}^n$ let $I_0 = (0, a_1, \ldots, a_n)$ and $I_1 = (1, a_1, \ldots, a_n)$. M_{I_0} is isometric to M_{I_1}.*

Proof. It is evident from Figure 11 that the mutation $(13)(24)$ extends to a homeomorphism on $B^3 - S$. Thus (S^3, L_{I_0}) is homeomorphic to (S^3, L_{I_1}), and the result follows from Mostow rigidity. □

Lemma 5.9. *For $I = (a_0, a_1, \ldots, a_n) \in \{0, 1, 2\}^{n+1}$, let $\bar{I} = (a_n, a_{n-1}, \ldots, a_0)$. There is an orientation-reversing isometry $M_I \to M_{\bar{I}}$ that, for each $i \in \{1, \ldots, n\}$, takes the image of $C(\Gamma_T^{(i)})$ in M_I to the image of $C(\Gamma_T^{(n-i)})$ in $M_{\bar{I}}$.*

Proof. It is straightforward to check that the braids in Figure 10 are isotopic (in $S^2 \times I$) to their mirror images. Therefore, there is an orientation-reversing homeomorphism $L_I \to L_{\bar{I}}$. By composing this homeomorphism with f_I^{-1} and $f_{\bar{I}}$ we get a homeomorphism $M_I \to M_{\bar{I}}$. The result follows by Mostow rigidity. □

Below we describe the change effected at the level of Kleinian groups by cutting a hyperbolic manifold along an embedded, separating totally geodesic surface and regluing by an isometry.

Lemma 5.10. *Suppose Γ_0 and Γ_1 meet cute along a plane $\mathcal{H}$, and take $\Theta = \Gamma_0 \cap \Gamma_1$, $E = \mathcal{H}/\Theta$, and ι_0 and ι_1 as in Lemma 3.2. If n normalizes Θ and preserves components of $\mathbb{H}^3 - \mathcal{H}$, then $\langle \Gamma_0, \Gamma_1^{\mathsf{n}} \rangle$ is a Kleinian group, and there is an isometry*

$$C(\Gamma_0) \cup_{\iota_1 \phi_{\mathsf{n}}^{-1} \iota_0^{-1}} C(\Gamma_1) \to C(\langle \Gamma_0, \Gamma_1^{\mathsf{n}} \rangle)$$

which restricts on $C(\Gamma_0)$ to the natural map, and on $C(\Gamma_1)$ to $\phi_{\mathsf{n}} \colon C(\Gamma_1) \to C(\Gamma_1^{\mathsf{n}})$ followed by the natural map.

Proof. Since n normalizes Θ, it preserves $\mathcal{H}$; hence our hypotheses ensure that Γ_0 and Γ_1^{n} meet cute along $\mathcal{H}$, and Lemma 3.2 applies. Thus $\langle \Gamma_0, \Gamma_1^{\mathsf{n}} \rangle$ is a Kleinian group, and in particular, the natural maps $C(\Gamma_0) \to C(\langle \Gamma_0, \Gamma_1^{\mathsf{n}} \rangle)$ and $C(\Gamma_1^{\mathsf{n}}) \to C(\langle \Gamma_0, \Gamma_1^{\mathsf{n}} \rangle)$ determine an isometry

$$C(\Gamma_0) \cup_{n\iota_1 \iota_0^{-1}} C(\Gamma_1^{\mathsf{n}}) \to C(\langle \Gamma_0, \Gamma_1^{\mathsf{n}} \rangle).$$

Here we are using $n\iota_1 \colon E \to C(\Gamma_1^{\mathsf{n}})$ to refer to the natural map. It is now an exercise in definition-chasing to show that $n\iota_1 \circ \phi_{\mathsf{n}} = \phi_{\mathsf{n}} \circ \iota_1$, whence the map

$$C(\Gamma_0) \cup_{\iota_1 \phi_{\mathsf{n}}^{-1} \iota_0^{-1}} C(\Gamma_1) \to C(\Gamma_0) \cup_{n\iota_1 \iota_0^{-1}} C(\Gamma_1^{\mathsf{n}}),$$

defined as the identity on $\mathbb{C}(\Gamma_0)$ and ϕ_{n} on $C(\Gamma_1)$, is well defined. The lemma follows. □

Since m_1 and m_2 have order 2, $\phi_{\mathsf{m}_i} = \phi_{\mathsf{m}_i}^{-1}$ for $i = 1, 2$. Lemma 5.10 thus yields the result below, which describes how the algebraic model for M_n from Proposition 3.12 changes under mutation.

Proposition 5.11. *For $I = (a_0, a_1, \ldots, a_n) \in \{0, 1, 2\}^{n+1}$ let $\mathsf{q}_{i+1} = \mathsf{m}_{a_0}^{(0)} \cdots \mathsf{m}_{a_i}^{(i)}$ for $0 \le i \le n$, with $\mathsf{m}_0^{(j)} := id$, and $\mathsf{m}_1^{(j)}$ and $\mathsf{m}_2^{(j)}$ as in Corollary 5.6 for every j. Define*

$$\Gamma_I = \left\langle \Gamma_S, \left(\Gamma_T^{(1)}\right)^{\mathsf{q}_1}, \ldots, \left(\Gamma_T^{(n)}\right)^{\mathsf{q}_n}, \left(\overline{\Gamma}_S^{\mathsf{c}^{-2n}}\right)^{\mathsf{q}_{n+1}} \right\rangle.$$

There is an isometry $M_I \to C(\Gamma_I)$ that restricts on $C(\Gamma_S)$ to the natural map, and on $C(\Gamma_T^{(i)})$ to $\phi_{\mathsf{q}_i} \circ \phi_i$ followed by the natural map, for $1 \le i \le n$.

The proof of Proposition 5.11 follows the inductive approach of that of Proposition 3.12, but at each stage appeals to Lemma 5.10 for instructions on how to change the construction. We will not write the details, as it is very similar.

6. Commensurable mutants

Here we show that M_n is commensurable to each of its mutants by (13)(24) and with this fact classify the M_I up to isometry for $I \in \{0, 1\}^{n+1}$, proving Theorem 2. In the process, we show that our polyhedral decomposition of M_n is "canonical" in the sense of [Goodman et al. 2008, §2]; that is, produced by a construction of Epstein and Penner [1988]. This allows us to identify the commensurator for Γ_n and the minimal orbifold quotient of M_n. In practice, it is a challenge to find Epstein–Penner decompositions, commensurators, and commensurator quotients. Infinite families where these are known are rare.

Below, let $\mathcal{B}_0$ be the open half-ball in $\mathbb{H}^3$ bounded by the Euclidean hemisphere of unit radius centered at $0 \in \mathbb{C}$, and let $\mathcal{B}_j = \mathsf{c}^{-j}(\mathcal{B}_0)$, where c is as defined in Lemma 3.3. Recall that we have defined $\mathcal{H}$ as the geodesic hyperplane of $\mathbb{H}^3$ with ideal boundary $\mathbb{R} \cup \{\infty\}$. If w and z are complex numbers, we will take $w\mathcal{H} + z$ to be the hyperplane with ideal boundary $(w\mathbb{R} + z) \cup \{\infty\}$.

Definitions 6.1. (1) Let f_0 be obtained by first reflecting in $i\mathcal{H}$ and then in $i\mathcal{H}+1/2$.

(2) Let b_0 be obtained by first reflecting in $\mathcal{H} + i/2$ and then in $\partial\mathcal{B}_0$.

(3) For $j \geq 0$, let a_j be obtained by reflecting in $i\mathcal{H} + 1/2$ and then in $\partial\mathcal{B}_j$.

Since $i\mathcal{H}$ and $i\mathcal{H}+1/2$ are parallel and share the ideal point ∞, f_0 is a parabolic isometry fixing ∞. $\mathcal{H} + i/2$ meets $\partial\mathcal{B}_0$ at an angle of $\pi/3$, so b_0 is an elliptic isometry of order 3 rotating around the geodesic of intersection. For the same reason, a_i is elliptic of order 3, rotating around the geodesic $i\mathcal{H} + 1/2 \cap \partial\mathcal{B}_i$, for each $i \geq 0$.

Lemma 6.2. *Let G_n be the group generated by reflections in the face of P_n, where*

$$P_n = \left\{(z, t) \in \mathbb{H}^3 \mid 0 \leq \Re(z) \leq 1/2, -n\sqrt{2} \leq \Im(z) \leq 1/2\right\} - \bigcup_{k=0}^{n} \mathcal{B}_k.$$

Then G_n contains a_i for $0 \leq i \leq 2n$, as well as f_0 and b_0, and $O_n \doteq \mathbb{H}^3/G_n$ is a one-cusped hyperbolic orbifold.

Proof. By its definition, P_n is cut out by $\mathcal{H} + i/2$, $i\mathcal{H}$, $i\mathcal{H} + 1/2$, $\mathcal{H} - n \cdot i\sqrt{2}$, and the $\partial\mathcal{B}_k$, $0 \leq k \leq n$. It is not hard to show directly that the dihedral angle between any two of these planes that intersect is an integer submultiple of π, whence by the Poincaré polyhedron theorem G_n is discrete and $\mathbb{H}^3/G_n$ is an orbifold isometric to P_n with mirrored sides (see [Ratcliffe 1994, Theorem 13.5.1]). In particular, since P_n has a single ideal point $\mathbb{H}^3/G_n$ has one cusp.

One finds that G_n contains f_0, b_0, and the a_i, for $0 \leq i \leq n$, by direct appeal to Definitions 6.1. It remains to establish that G_n contains a_i for $n < i \leq 2n$. Note that $\mathcal{H} - n \cdot i\sqrt{2}$ is the image of $\mathcal{H}$ under c^{-n}, so reflection in $\mathcal{H} - n \cdot i\sqrt{2}$ is given by $\mathsf{c}^{-n}\mathsf{r}\mathsf{c}^{n}$, where r is the reflection through $\mathcal{H}$. By the property of r observed above Lemma 3.3, conjugating an element $\mathsf{x} \in \mathrm{PSL}_2(\mathbb{C})$ by reflection in $\mathcal{H} - n \cdot i\sqrt{2}$ gives

$$\mathsf{c}^{-n}\mathsf{r}\mathsf{c}^{n}\mathsf{x}\mathsf{c}^{-n}\mathsf{r}\mathsf{c}^{n} = \mathsf{c}^{-2n}\bar{\mathsf{x}}\mathsf{c}^{2n}. \tag{6}$$

We further observe that c conjugates a_i to a_{i-1} for $i \geq 1$, since $\mathsf{c}(i\mathcal{H}+1/2) = i\mathcal{H}+1/2$ and $\mathsf{c}(\mathcal{B}_i) = \mathcal{B}_{i-1}$, and we note that $\bar{\mathsf{a}}_0 = \mathsf{a}_0$. Thus:

$$\overline{\mathsf{c}^{i}\mathsf{a}_i\mathsf{c}^{-i}} = \bar{\mathsf{a}}_0 = \mathsf{a}_0 = \mathsf{c}^{i}\mathsf{a}_i\mathsf{c}^{-i} \quad \Rightarrow \quad \mathsf{c}^{-2i}\bar{\mathsf{a}}_i\mathsf{c}^{2i} = \mathsf{a}_i. \tag{7}$$

For $0 \leq i \leq n$, it follows that the conjugate of a_i by reflection in $\mathcal{H} - n \cdot i\sqrt{2}$ is

$$\mathsf{c}^{-2n}\bar{\mathsf{a}}_i\mathsf{c}^{2n} = \mathsf{c}^{-2(n-i)}\mathsf{a}_i\mathsf{c}^{2(n-i)} = \mathsf{a}_{2n-i} \in G_n.$$

Therefore G_n contains a_i for $n \leq i \leq 2n$ as well, and the lemma is proved. □

Since $\mathcal{H}$ meets both $\partial\mathcal{B}_0$ and $i\mathcal{H} + 1/2$ at right angles, it does the same for the fixed geodesic of a_0 and is therefore preserved by a_0. In fact, the following description of $\mathsf{a}_0 \in \mathrm{PSL}_2(\mathbb{C})$ is easily obtained from its definition:

$$\mathsf{a}_0 = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}.$$

In particular, a_0 acts on the ideal points of $\mathcal{P}_1 \cap \mathcal{P}_2$ by $0 \mapsto 1 \mapsto \infty \mapsto 0$. Similarly, it is easy to see that $f_0(z, t) = (z+1, t)$.

Then the face pairings f (defined in Corollary 2.3) and s (defined in Corollary 2.2), which are equal, are obtained from f_0 by conjugating by a_0:

$$\mathsf{s} = \mathsf{f} = \mathsf{a}_0\mathsf{f}_0\mathsf{a}_0^{-1}. \tag{8}$$

One may use similar analyses to establish the following:

$$\mathsf{t} = \mathsf{f}_0\mathsf{a}_0\mathsf{b}_0, \quad \mathsf{g} = (\mathsf{a}_0^{-1}\mathsf{a}_1)\mathsf{f}_0^{-1}(\mathsf{a}_0^{-1}\mathsf{a}_1)^{-1}, \quad \mathsf{h} = \mathsf{a}_1\mathsf{a}_0\mathsf{f}_0^{-1}\mathsf{a}_1. \tag{9}$$

The main group-theoretic fact of this section extends these observations.

Proposition 6.3. *For each $n \in \mathbb{N}$, G_n contains Γ_n and $\mathsf{m}_1^{(i)}$ for $0 \leq i \leq n$.*

Proof. We recall from Proposition 3.12 that $\Gamma_n = \langle \Gamma_S, \Gamma_T^{(1)}, \ldots, \Gamma_T^{(n)}, \bar{\Gamma}_S^{\mathsf{c}^{-2n}} \rangle$, where by Definitions 3.9(2), $\Gamma_T^{(i)} \doteq \Gamma_T^{\mathsf{c}^{-2(i-1)}}$ for each i between 1 and n. Furthermore, by Lemma 3.3, $\Gamma_T = \langle \Gamma_{T_0}, \bar{\Gamma}_{T_0}^{\mathsf{c}^{-2}} \rangle$.

It is a direct consequence of the descriptions (8) and (9) that $\Gamma_S < G_n$ and $\Gamma_{T_0} < G_n$. Furthermore, since f_0 commutes with c and $\bar{\mathsf{f}}_0 = \mathsf{f}_0$, (7) implies, for instance, that

$$\mathsf{c}^{-2}\bar{\mathsf{f}}\mathsf{c}^{2} = \mathsf{c}^{-2}(\bar{\mathsf{a}}_0\bar{\mathsf{f}}_0\bar{\mathsf{a}}_0^{-1})c^{2} = \mathsf{a}_2\mathsf{f}_0\mathsf{a}_2^{-1} \in \Gamma_n,$$

since $\bar{\mathsf{a}}_0 = \mathsf{a}_0$ and $\mathsf{c}^{-2}\mathsf{a}_0 c^2 = \mathsf{a}_2$. Using the same strategy, we find

$$\mathsf{c}^{-2}\bar{\mathsf{g}}c^{2} = (\mathsf{a}_2^{-1}\mathsf{a}_1)\mathsf{f}_0^{-1}(\mathsf{a}_2^{-1}\mathsf{a}_1)^{-1} \in G_n \quad \text{and} \quad \mathsf{c}^{-2}\bar{\mathsf{h}}\mathsf{c}^{2} = \mathsf{a}_1\mathsf{a}_2\mathsf{f}_0^{-1}\mathsf{a}_1 \in G_n.$$

Thus G_n contains $\Gamma_T = \Gamma_T^{(1)}$. Since conjugation by c^{-1} takes a_i to a_{i+1}, and $\mathsf{a}_i \in \Gamma_n$ for each i between 0 and $2n$, it follows from the descriptions above and in (8) and (9) that $\Gamma_T^{(i)} < G_n$, for $1 \le i \le n$. Finally the relation (6) immediately implies that $\overline{\Gamma}_S^{\mathsf{c}^{-2n}} < G_n$, and we have established that $\Gamma_n < G_n$.

To show that G_n contains the elements $\mathsf{m}_1^{(j)}$ for each j between 0 and n, we observe that the element obtained by reflecting first in $\partial\mathcal{B}_0$ and then in $i\mathcal{H}$ is the rotation of order 2 described by $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. This is well known to generate $\mathrm{PSL}_2(\mathbb{Z})$, along with a_0. Since $\mathsf{m}_1 \in \mathrm{PSL}_2(\mathbb{Z})$, it follows that $\mathsf{m}_1 \in G$.

We note that c^{-2j} preserves $i\mathcal{H}$ and takes $\mathcal{B}_0$ to $\mathcal{B}_{2j}$, and that $\mathcal{B}_{2j}$ intersects P_n, for $j \le n/2$, and intersects its image under reflection in $\mathcal{H} - n \cdot i\sqrt{2}$ for $n/2 \le j \le n$. Thus for each j between 0 and n, the rotation obtained by reflecting first in $\partial\mathcal{B}_{2j}$ and then in $i\mathcal{H}$ is contained in G_n. If m_1 is expressed as a word in the two elements described in the paragraph above, then $\mathsf{c}^{-2j}\mathsf{m}_1\mathsf{c}^{2j}$ is expressed as the same word in a_{2j} and the rotation obtained from $\partial\mathcal{B}_{2j}$ as above. The lemma follows. □

It is now easy to prove the first part of Theorem 2, that the complement of each link obtained from L_n using only the mutation (13)(24) is commensurable to M_n.

Proposition 6.4. *M_n branched covers O_n, as does M_I for any $I \in \{0, 1\}^{n+1}$. Hence these are commensurable.*

Proof. Since G_n is a discrete reflection group, it is enough to show that $\Gamma_I \subset G_n$. This is immediate from Propositions 5.11 and 6.3. □

To finish the proof of Theorem 2 we need an isometry classification of the link complements that fall under the purview of Proposition 6.4. Our first step is to show that G_n is the commensurator of Γ_n.

The *commensurator* of a Kleinian group Γ is the group

$$\mathrm{Comm}(\Gamma) = \{\mathrm{g} \in \mathrm{Isom}(\mathbb{H}^3) \mid [\Gamma : \mathrm{g}\Gamma\mathrm{g}^{-1}] < \infty\}.$$

It follows easily from the definition that since Γ_n is a finite-index subgroup of G_n, G_n is contained in $\mathrm{Comm}(\Gamma_n)$. Since Γ_n is nonarithmetic (by Proposition 4.1), by a famous theorem of Margulis [1991, (1) Theorem] $\mathrm{Comm}(\Gamma_n)$ is discrete.

Let O_n' be the hyperbolic orbifold $\mathbb{H}^3/\mathrm{Comm}(\Gamma_n)$. Since $G_n < \mathrm{Comm}(\Gamma_n)$, O_n' is finitely covered by O_n. Recall from Lemma 6.2 that O_n, and therefore also O_n', has exactly one cusp. It is our goal to show that $G_n = \mathrm{Comm}(\Gamma_n)$; hence $O_n = O_n'$.

We use the strategy of [Goodman et al. 2008]. Recall the *hyperboloid* model for $\mathbb{H}^3$. The *Lorentz inner product* on $\mathbb{R}^4$ is the indefinite bilinear form

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = v_1 w_1 + v_2 w_2 + v_3 w_3 - v_4 w_4.$$

We let $\mathbb{H}^3 = \{\boldsymbol{v} \mid \langle \boldsymbol{v}, \boldsymbol{v} \rangle = -1,\ v_4 > 0\}$ and equip $T_{\boldsymbol{v}}\mathbb{H}^3$ with the Riemannian metric determined by the Lorentz inner product. The *positive light cone* is the set $L^+ = \{\boldsymbol{v} \mid \langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0,\ v_4 \ge 0\}$. The *ideal point* of $\mathbb{H}^3$ represented by $\boldsymbol{v} \in L^+$ is the set $[\boldsymbol{v}]$ of scalar multiples of $\boldsymbol{v}$ in $L^+ - \{\mathbf{0}\}$. $\mathrm{Isom}(\mathbb{H}^3)$ is the group of matrices in $\mathrm{GL}_4(\mathbb{R})$ which act on $\mathbb{R}^4$ preserving the Lorentz inner product and the sign of the last coordinate, hence acting on $\mathbb{H}^3$ by isometries. Those in $\mathrm{Isom}^+(\mathbb{H}^3) \subset \mathrm{Isom}(\mathbb{H}^3)$ preserve orientation on $\mathbb{H}^3$.

For $\boldsymbol{v} \in L^+ - \{0\}$ the set $H_{\boldsymbol{v}} = \{\boldsymbol{w} \in \mathbb{H}^3 \mid \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -1\}$ is a horosphere centered at the ideal point $[\boldsymbol{v}]$. If $\alpha \in \mathbb{R}^+$ then $H_{\alpha\boldsymbol{v}}$ is a horosphere centered at $[\alpha\boldsymbol{v}] = [\boldsymbol{v}]$, and if $\alpha \leq 1$ then $H_{\boldsymbol{v}}$ is contained in the horoball $\{\boldsymbol{w} \mid \langle \alpha\boldsymbol{v}, \boldsymbol{w} \rangle \geq -1\}$ determined by $\alpha\boldsymbol{v}$. This determines a bijective correspondence between vectors in L^+ and horospheres in $\mathbb{H}^3$, so we call the vectors in L^+ *horospherical vectors*.

We use the hyperboloid model to construct certain canonical tilings of $\mathbb{H}^3$ associated to M_n as in [Epstein and Penner 1988]. First, choose a horospherical vector $\boldsymbol{v} \in L^+$ fixed by a peripheral element of Γ_n, so that under the covering map $\mathbb{H}^3 \to O'_n$ the horosphere $H_{\boldsymbol{v}}$ projects to a cross section of the cusp. Then $V_n = \mathrm{Comm}(\Gamma_n) \cdot \boldsymbol{v}$ is $\mathrm{Comm}(\Gamma_n)$-invariant and determines a $\mathrm{Comm}(\Gamma_n)$-invariant set of horospheres. The convex hull of V_n in $\mathbb{R}^4$ is called the *Epstein–Penner convex hull*; we denote it as C_n. Epstein and Penner show that ∂C_n consists of a countable set of 3-dimensional faces F_i, where each F_i is a finite-sided Euclidean polyhedron in $\mathbb{R}^4$. Furthermore, this decomposition of ∂C_n projects along straight lines through the origin to a $\mathrm{Comm}(\Gamma_n)$-invariant tiling $\mathcal{T}_n$ of $\mathbb{H}^3$ by ideal polyhedra [Epstein and Penner 1988, Proposition 3.5 and Theorem 3.6]. We refer to the tiling so obtained as a *canonical tiling*. (It is easy to see that a different choice for the vector $\boldsymbol{v}$ yields a convex hull which differs from C_n by multiplication by a positive scalar. Therefore it projects to the same canonical tiling as C_n.)

Consider the group of symmetries $\mathrm{Sym}(\mathcal{T}_n) < \mathrm{Isom}(\mathbb{H}^3)$. Given that $\mathcal{T}_n$ is $\mathrm{Comm}(\Gamma_n)$-invariant we have that $\mathrm{Comm}(\mathcal{T}_n) < \mathrm{Sym}(\mathcal{T}_n)$. On the other hand, $\mathrm{Sym}(\mathcal{T}_n)$ is discrete [Goodman et al. 2008, Lemma 2.1] and since Γ_n is nonarithmetic $\mathrm{Comm}(\Gamma_n)$ is the maximal discrete group containing Γ_n. Therefore $\mathrm{Sym}(\mathcal{T}_n) = \mathrm{Comm}(\Gamma_n)$. Below we will first identify the tiling $\mathcal{T}_n$ and then show that $G_n = \mathrm{Sym}(\mathcal{T}_n)$.

Theorem 4. *With $\mathcal{S}$ as in Lemma 4.5, $\mathcal{T}_n = \Gamma_n \cdot \bigcup\{\mathcal{P} \in \mathcal{S}\}$ is the canonical tiling for* $\mathrm{Comm}(\Gamma_n)$.

Proof. We choose matrices

$$M = \begin{pmatrix} 2 & 1 & 0 & 1 & 0 & -1 & -2 & -1 & 1 & -1 & -1 & 1 \\ 0 & 1 & 2 & 1 & -2 & -1 & 0 & -1 & -1 & 1 & 1 & -1 \\ 0 & \sqrt{2} & 0 & -\sqrt{2} & 0 & \sqrt{2} & 0 & -\sqrt{2} & -\sqrt{2} & -\sqrt{2} & \sqrt{2} & \sqrt{2} \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \end{pmatrix}, \quad N = \begin{pmatrix} \sqrt{2} & 0 & 0 & -\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 & -\sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} & 0 & 0 & -\sqrt{2} \\ \sqrt{2} & \sqrt{2} & \sqrt{2} & \sqrt{2} & \sqrt{2} & \sqrt{2} \end{pmatrix}.$$

For $X = M, N$, let x_i be the i-th column of X. Each x_i below lies in L^+ and so represents an ideal point of $\mathbb{H}^3$. We will call $\mathcal{P}_X$ the convex hull in $\mathbb{H}^3$ of the $[x_i]$.

Our M and N are such that $\mathcal{P}_M$ is a right-angled ideal cuboctahedron and $\mathcal{P}_N$ a regular ideal octahedron, and, furthermore:

- For $X = M, N$, each member of $\mathrm{Isom}(\mathcal{P}_X)$ fixes $(0, 0, 0, 1)^T \in \mathbb{H}^3$ and the set of columns of X is $\mathrm{Isom}(\mathcal{P}_X)$-invariant.
- There exists $\mathsf{h} \in \mathrm{Isom}^+(\mathbb{H}^3)$ with $\mathsf{h}(n_1) = m_1$, $\mathsf{h}(n_2) = m_9$, and $\mathsf{h}(n_3) = m_4$, so that $\mathsf{h}(\mathcal{P}_N) \cap \mathcal{P}_M$ is the face (m_1, m_9, m_4) with ideal vertices at $[m_1]$, $[m_9]$, and $[m_4]$.

Let $\mathcal{P}_1 = \mathsf{h}(\mathcal{P}_N)$ and $\mathcal{P}_2 = \mathcal{P}_M$. There is an isometry p from the upper half-space to the hyperboloid model taking $\mathcal{P}_i$ (as in Corollaries 2.2 and 2.3) to $\mathcal{P}_i$ as above for $i = 1, 2$, and ∞ to the center $[m_1]$ of the horosphere H_{m_1}. We again refer by $\mathcal{S}$ to the image under p of the set $\mathcal{S}$ from Lemma 4.5. Also, p conjugates each of the isometries we've used thus far in our constructions to elements of $\mathrm{GL}_4(\mathbb{R})$, to which we'll refer by the same names.

From the explicit description in Lemma 6.2 it is clear that $[m_1]$ is a parabolic fixed point of G_n. Since G_n is discrete, each element fixing $[m_1]$ actually fixes m_1, so the orbit $V_n = G_n.m_1$ is a G_n-invariant collection of horospherical vectors bijective to the set of parabolic fixed points of G_n. Since $O_n = \mathbb{H}^3/G_n$ has one cusp and the same holds for $O'_n = \mathbb{H}^3/\,\mathrm{Comm}(\Gamma_n)$, it follows that V_n is also $\mathrm{Comm}(\Gamma_n)$-invariant.

Lemma 4.5 implies that $\Gamma_n . \bigcup\{\mathcal{P} \in \mathcal{S}\}$ is a Γ_n-invariant tiling of $\mathbb{H}^3$. We claim that it is identical to the canonical tiling $\mathcal{T}_n$, the projection to $\mathbb{H}^3$ of the boundary of the convex hull of V_n in $\mathbb{R}^4$. Note that $\mathcal{T}_n$ is also Γ_n-invariant, since it is G_n-invariant by construction and $\Gamma_n < G_n$.

We will use [Goodman et al. 2008, Proposition 6.1] to prove the claim. The proposition requires for each element of $\mathcal{S}$ that the horospherical vectors representing its vertices be coplanar in $\mathbb{R}^4$, and that the angle between this plane and the plane determined by each neighboring tile be convex. Equivalently, if $v_1, \dots, v_k \in V_n$ represent the ideal vertices of an element of $\mathcal{S}$ and $w \in V_n - \{v_1, \dots, v_k\}$ represents a vertex of a neighboring tile, then there exists a vector $\boldsymbol{n} \in \mathbb{R}^4$ such that

(1) (coplanarity) $\boldsymbol{n} \cdot v_i = 1$ for every $i = 1, \dots, k$, and

(2) (convex angles) $\boldsymbol{n} \cdot w > 1$.

(See the proof of [Goodman et al. 2008, Proposition 6.1].) Observe that these conditions are invariant under $\mathrm{Isom}(\mathbb{H}^3)$, for if $\boldsymbol{n} \cdot v = \alpha$ and $A \in \mathrm{Isom}(\mathbb{H}^3)$ then $(\boldsymbol{n}A^{-1}) \cdot Av = \alpha$.

For each member $\mathcal{P}$ of $\mathcal{S}$, we note that the subset of V_n representing the set of ideal points of $\mathcal{P}$ contains m_1 and is $\mathrm{Isom}(\mathcal{P})$-invariant. This is because the members of $\mathcal{S}$ all share the ideal vertex $[m_1]$, and the stabilizer in G_n of any $\mathcal{P} \in \mathcal{S}$ acts transitively on its set of ideal vertices. (The latter assertion can be proved by directly examining $P_n \cap \mathcal{P}$, for P_n as in Lemma 6.2.) In particular, the ideal vertices of $\mathcal{P}_1$ are represented in V_n by $\{\mathsf{h}(n_i)\}_{i=1}^{6}$ and those of $\mathcal{P}_2$ by $\{m_i\}_{i=1}^{12}$, by the properties bulleted above.

Take $\boldsymbol{n} = (0, 0, 0, 1/2)^T$. Then $\boldsymbol{n} \cdot m_i = 1$ for $i = 1, \dots, 12$, so the m_i are coplanar. The n_i (and hence also the $\mathsf{h}(n_i)$) are also coplanar, since $\sqrt{2}\boldsymbol{n} \cdot n_i = 1$ for $i = 1, \dots, 6$ and the same $\boldsymbol{n}$. Coplanarity follows for the other elements of $\mathcal{S}$, since, for example, $\{\mathsf{c}^{-1}(m_i)\}_{i=1}^{12}$ is an $\mathrm{Isom}(\mathsf{c}^{-1}(\mathcal{P}_2))$-invariant collection of horospherical vectors containing $m_1 = \mathsf{c}^{-1}(m_1)$ and representing the ideal vertices of $\mathsf{c}^{-1}(\mathcal{P}_2)$.

Consider all pairs $(\mathcal{Q}, \mathcal{P}_X)$ where $X \in \{M, N\}$ and $\mathcal{Q}$ is a regular ideal octahedron or cuboctahedron which meets $\mathcal{P}_X$ in a face. Choose horospherical vectors for $\mathcal{Q}$ to agree with those chosen for $\mathcal{P}_X$ and to be $\mathrm{Isom}(\mathcal{Q})$-invariant. Since the convexity condition (2) is invariant under isometries, to finish the proof it suffices to check this condition for each possible pair $(\mathcal{Q}, \mathcal{P}_X)$.

If $\mathcal{Q}$ is a cuboctahedron adjacent to $\mathcal{P}_M$ sharing the triangular face (m_1, m_9, m_4) then $w = (7, 1, -5\sqrt{2}, 10)^T$ is a horospherical vector for $\mathcal{Q}$ which is not shared by $\mathcal{P}_M$. We have $\boldsymbol{n} \cdot w = 5 > 1$. If $\mathcal{Q}$ is a cuboctahedron adjacent to $\mathcal{P}_M$ sharing the square face (m_1, m_2, m_3, m_4) then $w = (3, 5, -\sqrt{2}, 6)^T$ is a horospherical vector for $\mathcal{Q}$ which is not shared by $\mathcal{P}_M$. We have $\boldsymbol{n} \cdot w = 3 > 1$. If $\mathcal{Q}$ is an octahedron adjacent to $\mathcal{P}_N$ sharing the face (n_1, n_2, n_3) then $w = \sqrt{2}(1, 2, 2, 3)^T$ is a horospherical vector for $\mathcal{Q}$ which is not shared by $\mathcal{P}_N$. We have $\sqrt{2}\boldsymbol{n} \cdot w = 3 > 1$. By construction, $\mathcal{P}_1 = \mathsf{h}(\mathcal{P}_N)$ is an octahedron intersecting $\mathcal{P}_M$ in (m_1, m_9, m_4). For $w = \mathsf{h}(n_1) = (2+2\sqrt{2}, 0, -2-2\sqrt{2}, 4+4\sqrt{2})^T$ we have $\boldsymbol{n} \cdot w = 2+\sqrt{2} > 1$.

With coplanarity and convex angles thus established, [Goodman et al. 2008, Proposition 6.1] implies that $\Gamma_n . \bigcup\{\mathcal{P} \in \mathcal{S}\}$ implies the claim; hence the result. □

By construction G_n is a subgroup of the symmetry group for $\mathcal{T}_n$. We complete the proof that $G_n = \mathrm{Comm}(\Gamma_n)$ below, showing that it is the full symmetry group.

Corollary 6.5. *G_n is the commensurator of Γ_n and O_n is the minimal orbifold quotient of M_n. If $I \in \{0, 1\}^{n+1}$, then $\mathrm{Comm}(\Gamma_I) = G_n$ and O_n is the minimal quotient of M_I.*

Proof. Proposition 6.4 implies $\mathrm{Comm}(\Gamma_I) = \mathrm{Comm}(\Gamma_n)$. Take $\mathsf{x} \in \mathrm{Comm}(\Gamma_n)$. We want to show that $\mathsf{x} \in G_n$. Recall that $\mathsf{c}^{-n}\mathsf{r}\mathsf{c}^{n} \in G_n$ exchanges $\mathcal{P}_1$ and $\mathsf{c}^{-2n}\mathsf{r}\mathcal{P}_1$. Therefore the octahedral tiles of $\mathcal{T}_n$ lie in a single G_n-orbit, and we may assume that x fixes $\mathcal{P}_1$.

Recall, for instance from Corollary 2.2, that $\mathcal{P}_1$ is checkered and its face A spanned by the vertices 0, 1, and ∞ is external, with $A = \mathcal{P}_1 \cap \mathcal{P}_2$. We have that $\mathsf{a}_0, \mathsf{b}_0 \in \mathrm{Isom}(\mathcal{P}_1) \cap G_n$. The internal faces of $\mathcal{P}_1$ are paired by elements of Γ_S, so every internal face of $\mathcal{P}_1$ meets an octahedron in $\mathcal{T}_n$. Since $\mathcal{P}_2$ is a cuboctahedron, $\mathsf{x}(A)$ must be an external face of $\mathcal{P}_1$.

It follows immediately from the definitions of a_0 and b_0 that $\langle \mathsf{a}_0, \mathsf{b}_0\rangle$ acts transitively on the external faces of $\mathcal{P}_1$. Hence after multiplying by an element of $\langle \mathsf{a}_0, \mathsf{b}_0\rangle < G_n$, we may assume that $\mathsf{x}(A) = A$. By construction it is clear that G_n contains the stabilizer of A in $\mathrm{Isom}(\mathcal{P}_1)$, so we have $\mathsf{x} \in G_n$ as desired. □

The second half of Theorem 2 follows from the isometry classification below.

Proposition 6.6. *Suppose* $I = (0, a_1 \dots, a_{n-1}, 0)$ *and* $J = (0, b_1, \dots, b_{n-1}, 0)$ *are elements of* $\{0, 1\}^{n+1}$. M_I *is isometric to* M_J *if and only if* $J = I$ *or* $J = \bar{I}$.

Remark. We have assumed that the first and last entries of I and J are all zero to make the proposition easier to state. By Lemma 5.8, changing the first or last entry of either I or J to "1" yields another isometric manifold.

Proof. Any two distinct tiles of $\mathcal{T}_n$ which meet the interior of the fundamental domain P_n from Lemma 6.2 have distinct G_n-orbits. On the other hand, any tile that does not is contained in the orbit of one that does. It follows that G_n has a unique orbit of octahedral tiles (that of $\mathcal{P}_1$) and exactly n of cuboctahedral tiles, those of $\mathcal{P}_2, \mathsf{c}^{-1}(\mathcal{P}_2), \dots, \mathsf{c}^{-n+1}(\mathcal{P}_2)$, since P_n has an open subset in each of these and is contained in their union.

The planes $\mathsf{c}^{-i}(\mathcal{H})$ meet the interior of P_n for $i \in \{0, 1, \dots, n-1\}$, so their G_n-orbits are also distinct. We note that the G_n-orbit of $\mathcal{H}$ is distinct from that of $i\mathcal{H}$ since $\mathcal{H}$ contains points of the interior of P_n but $i\mathcal{H}$ contains a face. Since $i\mathcal{H} \cap P_n$ is contained in an internal face of $\mathcal{P}_1$ and $\mathcal{H} \cap P_n$ in an external face, it follows that the G_n-orbit of an internal face of $\mathcal{P}_1$ is distinct from that of an external face.

For I as in the hypothesis, it follows, as in Lemma 4.5, that the members of

$$\mathcal{S}_I = \{\mathcal{P}_1, \mathcal{P}_2, \mathsf{c}^{-1}\mathcal{P}_2, \mathsf{q}_2\mathsf{c}^{-2}\mathcal{P}_2, \dots, \mathsf{q}_n\mathsf{c}^{-2n+1}\mathcal{P}_2, \mathsf{q}_n\mathsf{c}^{-2n}\mathsf{r}\mathcal{P}_1\}$$

project to a polyhedral decomposition of M_I, where the q_i are as defined in Proposition 5.11. (In particular, $\mathsf{q}_1 = 1$ and $\mathsf{q}_{n+1} = \mathsf{q}_n$, since I has first and last entries equal to 0.) This is because $\mathsf{q}_i(\mathsf{c}^{-2(i-1)}\mathcal{P}_2 \cup \mathsf{c}^{-2i+1}\mathcal{P}_2)$ projects to a decomposition of $C((\Gamma_T^{(i)})^{\mathsf{q}_i})$ for each i (see the proof of Lemma 4.5), and ϕ_{m_1} preserves the triangulation $\Delta_{\mathcal{F}}$ of Lemma 2.4. Therefore $\mathcal{S}_I$ is in bijective correspondence with the set of Γ_I-orbits of the top-dimensional tiles of $\mathcal{T}_n$.

Clearly $\mathsf{q}_i(\mathsf{c}^{-2(i-1)}\mathcal{P}_2)$ is G_n-equivalent to $\mathsf{c}^{-2(i-1)}\mathcal{P}_2$ for each i between 1 and n, and $\mathsf{q}_i(\mathsf{c}^{-2i+1}\mathcal{P}_2)$ to $\mathsf{c}^{-2i+1}\mathcal{P}_2$. The reflection u through $\mathcal{H} - n \cdot i\sqrt{2}$, also in G_n, exchanges $\mathcal{P}_1$ with $\mathsf{c}^{-2n}\mathsf{r}\mathcal{P}_1$ and $\mathsf{c}^{-i}\mathcal{P}_2$ with $\mathsf{c}^{-2n+i+1}\mathcal{P}_2$ for each i between 0 and $2n-1$. It follows that each G_n-orbit of top-dimensional tiles of $\mathcal{T}_n$ is the union of exactly two Γ_I-orbits.

Now suppose for some J as in the hypothesis that there is an isometry $M_J \to M_I$. This lifts to $\mathsf{x} \in \mathrm{Isom}(\mathbb{H}^3)$ with the property that $\Gamma_J^{\mathsf{x}} = \Gamma_I$. Since Γ_I and Γ_J are each finite-index subgroups of G_n they are commensurable, by definition $\mathsf{x} \in \mathrm{Comm}(\Gamma_J) = G_n$. By the above, $\mathsf{x}\mathcal{P}_1$ is Γ_I-equivalent to one of $\mathcal{P}_1$ or $\mathsf{q}_n\mathsf{c}^{-2n}\mathsf{r}\mathcal{P}_1$.

The reflection isometry of Lemma 5.9 determines $\rho \in \mathrm{Isom}(\mathbb{H}^3)$ that conjugates Γ_I to $\Gamma_{\bar{I}}$ and takes $\mathsf{q}_n\mathsf{c}^{-2n}\mathsf{r}\mathcal{P}_1$ into the $\Gamma_{\bar{I}}$-orbit of $\mathcal{P}_1$, so replacing x by $\rho\mathsf{x}$ (and I by $\bar{I}$) if necessary, we may ensure that there exists $\gamma \in \Gamma_I$ with $\gamma\mathsf{x}\mathcal{P}_1 = \mathcal{P}_1$. By the above $\gamma\mathsf{x}$ takes internal faces of $\mathcal{P}_1$ to internal faces. Because it conjugates Γ_J to Γ_I and $\mathcal{P}_1$ is contained in a fundamental domain for each, $\gamma\mathsf{x}$ preserves the internal face-pairings induced by the projections to M_I and to M_J.

It follows from Proposition 5.11 that each of these is the pairing described in Corollary 2.2. The combinatorial description there implies that $\gamma\mathsf{x}$ preserves the pairs $\{X_1, X_2\}$ and $\{X_3, X_4\}$ (see Figure 3), so it is either the identity or 180-degree rotation in the axis joining the ideal vertex 0 (the "intersection" $X_1 \cap X_2$ on the sphere at infinity) to $1+i = X_3 \cap X_4$. However the latter map does not preserve equivalence classes of the ideal vertices of X_3 and X_4 under face pairing, so $\gamma\mathsf{x} = 1$. It follows that $\mathsf{x} \in \Gamma_I$, so $\Gamma_J = \Gamma_I$.

We claim, however, that if $J \neq I$ then $\Gamma_J \neq \Gamma_I$. The key fact here is that $\Gamma_{T_0}^{\mathsf{m}_1} \neq \Gamma_{T_0}$: for instance, the face $(\mathsf{fg})^{-1}(Y_2')$ of $(\mathsf{fg})^{-1}(\mathcal{P}_2)$ is taken by $(\mathsf{fg})^{-1}\mathsf{hfg} \in \Gamma_{T_0}$ to $(\mathsf{fg})^{-1}(Y_3')$ (see the proof of Corollary 2.3), but $(\mathsf{fg})^{-1}(Y_2') = \mathsf{m}_1\mathsf{g}^{-1}(Y_1')$ is taken by $\mathsf{m}_1\mathsf{g}^{-1}\mathsf{m}_1^{-1} \in \Gamma_{T_0}^{\mathsf{m}_1}$ to $\mathsf{m}_1\mathsf{g}^{-1}(Y_3)$. (This description follows from the fact that m_1 preserves the polygon $\mathcal{F}$ from Lemma 2.4, acting on it as a rotation exchanging $\mathsf{g}^{-1}(E)$ with $(\mathsf{fg})^{-1}(D)$.) In fact, this further implies that no group Γ containing Γ_{T_0} also contains $\Gamma_{T_0}^{\mathsf{m}_1}$, as long as the natural map $C(\Gamma_{T_0}) \to C(\Gamma)$ is embedding.

If $J \neq I$ then for the minimal i such that $b_i \neq a_i$ we have $\Gamma_{T_0}^{\mathsf{w}} < \Gamma_I$ and $(\Gamma_{T_0}^{\mathsf{m}_1})^{\mathsf{w}} < \Gamma_J$, where $\mathsf{w} = \mathsf{q}_i\mathsf{c}^{-2(i-1)}$ (see Proposition 5.11). The claim, and hence also the result, thus follows from Proposition 5.11 and Lemma 3.3. □

7. Incommensurable mutants

Lemma 5.5 might lead one to suspect that the mutations (13)(24) and (12)(34) of $F^{(0)}$ act very differently at the level of Kleinian groups. Indeed, it follows from Proposition 7.1 below, together with Proposition 4.2, that $S^3 - L_n$ is incommensurable with the complement of any link obtained from it by the mutation (12)(34) along a subcollection of the $S^{(i)}$. In fact, we consider it likely that no two such mutants are commensurable unless they are isometric.

We lack the tools to fully prove this assertion — mutants are notoriously difficult to distinguish — but in this section we will describe large families of mutants whose members have different cusp parameters and are mutually incommensurable. We begin with traces, however. By [Neumann and Reid 1991] the M_I all have trace field $\mathbb{Q}(i, \sqrt{2})$.

Proposition 7.1. *For fixed n and any $I = (a_0, \ldots, a_n) \in \{0, 1, 2\}^{n+1}$ such that $a_i = 2$ for some i, Γ_I has a nonintegral trace.*

Proof. Suppose $I = (a_0, \ldots, a_n) \in \{0, 1, 2\}^{n+1}$ satisfies the hypothesis, and fix i_0

with $a_{i_0} = 2$. By Proposition 5.11, if $i_0 = 1$ then Γ_I contains the following matrix:

$$\mathsf{tm}_2\mathsf{gm}_2^{-1} = \begin{pmatrix} \frac{1}{5}(-2+12\sqrt{2}+31i+4i\sqrt{2}) & -2+\sqrt{2}+21i+2i\sqrt{2} \\ \frac{1}{5}(-1+7\sqrt{2}+16i+2i\sqrt{2}) & -1+\sqrt{2}+11i+i\sqrt{2} \end{pmatrix}.$$

(Recall from Corollary 5.6 that $\mathsf{m}_2^{(0)} = \mathsf{m}_2$.) The trace of $\mathsf{tm}_2\mathsf{gm}_2^{-1}$ is not an algebraic integer, since the ring of integers of $\mathbb{Q}(i, \sqrt{2})$ is $\mathbb{Z}[i, \sqrt{2}]$. If $i_0 = n$ then Γ_I contains a conjugate of $\bar{\mathsf{g}}\mathsf{m}_2\bar{\mathsf{t}}\mathsf{m}_2^{-1} = (\mathsf{m}_2\bar{g})^{-1}(\overline{\mathsf{tm}_2\mathsf{gm}_2^{-1}})\mathsf{m}_2\bar{g}$.

In all other cases, Proposition 5.11 implies that Γ_I contains an element with the same trace as the following matrix:

$$\begin{aligned} \bar{\mathsf{h}}(\mathsf{m}_2\mathsf{hm}_2^{-1}) &= \begin{pmatrix} -2i\sqrt{2} & -3+i\sqrt{2} \\ -3-i\sqrt{2} & 3i\sqrt{2} \end{pmatrix} \begin{pmatrix} -3i\sqrt{2} & 15-5i\sqrt{2} \\ \frac{1}{5}(3+i\sqrt{2}) & 2i\sqrt{2} \end{pmatrix} \\ &= \begin{pmatrix} -71/5 & -20-30i\sqrt{2} \\ \frac{18}{5}(-2+3i\sqrt{2}) & 55 \end{pmatrix}. \end{aligned}$$

The trace of this matrix is evidently not an algebraic integer. □

For fixed n and any $I \in \{0, 1, 2\}^{n+1}$, since M_n and M_I decompose along totally geodesic surfaces into isometric pieces, they have the same volume. (In fact, [Ruberman 1987, Theorem 1.3] asserts that hyperbolic volume is always invariant under mutation.) It would follow from the classical "Dehn invariant sufficiency" conjecture that any two hyperbolic manifolds with the same volume are scissors congruent (again see [Neumann 1998], for instance). In our situation we will verify this explicitly.

Proposition 7.2. *For fixed n and any $I \in \{0, 1, 2\}^{n+1}$, M_n and M_I have the same Bloch invariant.*

Proof. Recall from Lemma 2.4 that $F^{(0)}$ inherits a triangulation Δ_F from the fundamental domain $\mathcal{F}$ for the action of Λ on $\mathcal{H}$ pictured in Figure 5. From the figure, one finds that Δ_F has six edges, each a geodesic arc joining cusps of F. For example, the geodesic joining 0 and ∞ projects to an edge which joins cusp 1 to cusp 2. Of the other five edges, one joins 3 to 4, two join 2 to 4, and for each of 2 and 4 there is an edge joining it to itself.

Since $\mathsf{m}_1 \in \mathrm{PSL}_2(\mathbb{Z})$ it preserves the Farey tessellation of $\mathcal{H}$, which restricts on $\mathcal{F}$ to the triangulation pictured in Figure 5. Therefore ϕ_{m_1} preserves Δ_F. On the other hand, since ϕ_{m_2} exchanges 1 with 2 and 3 with 4 it does not preserve Δ_F. For instance, if e is the edge joining 2 to itself then $\phi_{\mathsf{m}_2}(e)$ joins 1 to itself.

Fix $I = (a_0, \ldots, a_n) \in \{0, 1, 2\}^{n+1}$ and suppose $a_i = 2$ for some $0 < i < n$. The gluing map $C(\Gamma_T^{(i)}) \to C(\Gamma_T^{(i+1)})$ factors through $\phi_{\mathsf{m}_2^{(i)}} : F^{(i)} \to F^{(i)}$ by Proposition 5.7. This is conjugate to ϕ_{m_2} by the inverse of ϕ_{i+1} from Definitions 3.9, so the gluing does not preserve the triangulations of $F^{(i)}$ induced by its intersections with external faces of the cuboctahedra on either side (see Lemma 2.4(3)). The cases

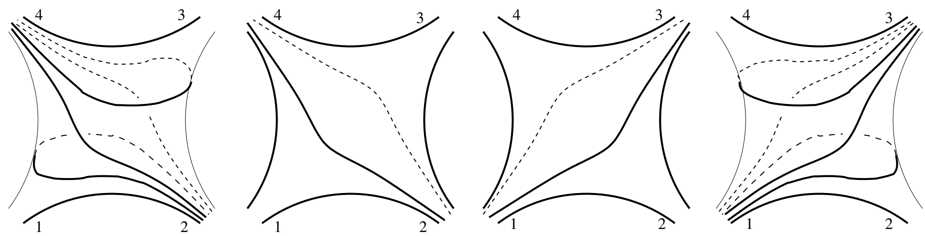


Figure 12. Interpolating between Δ_F, on the left, and $\phi_{\mathsf{m}_2}(\Delta_F)$.

$i = 0$ and $i = n$ are analogous, and show that if $a_i = 2$ for any i then the division of M_I into octahedra and cuboctahedra is not a true ideal polyhedral decomposition.

It is possible to rectify this by gluing "flat" tetrahedra between copies of $C(\Gamma_S)$ and/or $C(\Gamma_T)$ joined by the mutation ϕ_{m_2}. If $\mathcal{T}$ is a flat tetrahedron glued to, say, $C(\Gamma_S)$ along two adjacent triangles in $\partial C(\Gamma_S)$, then $C(\Gamma_S) \cup \mathcal{T}$ is homeomorphic to $C(\Gamma_S)$, but in the induced triangulation of the boundary, the edge separating the triangles along which $\mathcal{T}$ is glued has been replaced by an edge joining their two opposite vertices. See [Neumann and Yang 1999, §4] for a more thorough exposition.

Figure 12 illustrates a process by which Δ_F may be changed to its image under ϕ_{m_2} by a sequence of moves on edges. The edges of Δ_F are pictured on the left in bold. Moving left to right, at each stage two edges are replaced by edges transverse to them and disjoint from the remaining edges. After three such moves, the original triangulation has been changed to its image under ϕ_{m_2}.

Now suppose $I = (a_1, \ldots, a_n) \in \{0, 1, 2\}^{n+1}$. For each $i < n$ such that $a_i = 2$, replace $C(\Gamma_T^{(i+1)})$ by its union with 6 flat tetrahedra, glued successively along $\partial_- C(\Gamma_T^{(i+1)})$, to realize the change of triangulations illustrated in Figure 12. The result is homeomorphic to $C(\Gamma_T^{(i+1)})$, since adding a flat tetrahedron does not change the homeomorphism type, but the gluing induced by $\phi_{\mathsf{m}_2^{i+1}}$ now preserves the triangulation. The case $i = n$ is similar, but $C(\overline{\Gamma}_S)$ is changed instead.

It follows from the above that the Bloch invariant $\beta(M_I)$ may be calculated using the resulting polyhedral decomposition. This differs from the original by the addition of the cross ratio parameters of the flat tetrahedra. Each of these is equal to 2, since the triangulation of F is a projection of the Farey tessellation of $\mathbb{H}$. But in the Bloch group, $2 \cdot [2] = 0$ is a consequence of the relation $[z] = [z/(z-1)]$. Since the number of flat tetrahedra is a multiple of 6, the sum of their cross ratio parameters contributes nothing to the Bloch invariant. □

The following proposition tracks the change of cusp parameters under mutation. To simplify our task, we restrict our attention to complements of links obtained by mutating only with (12)(34) along a subcollection of the $S^{(i)}$ and note in passing that since those obtained by mutating only with (13)(24) are commensurable with M_n, their cusp parameters are $\mathrm{PGL}_2(\mathbb{Q})$-equivalent to those of M_n.

Proposition 7.3. *For $I = (t_0, t_1, \ldots, t_n) \in \{0, 2\}^{n+1}$ and $j \in \{0, 1, \ldots, n\}$, define*

$$c_j = \sum_{k=0}^{j} \frac{t_k}{2} \pmod 2.$$

Let T_1 be a cross section of the cusp of M_I such that $T_1 \cap C(\Gamma_S) = p_S(A_1)$ (as defined in Lemma 4.13), and let T_2 be a cross section of the cusp of M_I with $T_2 \cap C(\Gamma_S) = p_S(A_2)$. Up to the action of $\mathrm{PGL}_2(\mathbb{Q})$, *their complex moduli are*:

$$m(T_1) = i\left[1 + 2\sum_{j=1}^{n} \frac{2\sqrt{2}}{5^{c_{j-1}}} + \frac{1}{5^{c_n}}\right],$$

$$m(T_2) = i\left[\frac{1}{5} + 2\sum_{j=1}^{n} \frac{2\sqrt{2}}{5^{(1-c_{j-1})}} + \frac{1}{5^{(1-c_n)}}\right].$$

Proof. To simplify notation, we will identify A_k with $p_S(A_k)$ and view $A_k \subset C(\Gamma_S)$ for $k = 1, 2$. Recall the decomposition of M_I, along the surfaces $F^{(j)}$, into a union of isometric copies of $C(\Gamma_S)$ and $C(\Gamma_T)$ as described in Proposition 5.7:

$$C(\Gamma_S) \cup C(\Gamma_T^{(1)}) \cup \cdots \cup C(\Gamma_T^{(n)}) \cup C(\overline{\Gamma}_S) \to M_I.$$

We will denote by l_j the gluing map supplied by Proposition 5.7, taking $\partial_+ C(\Gamma_T^{(j)})$ to $\partial_- C(\Gamma_T^{(j+1)})$ when $1 \le j < n$. The map l_0 takes $\partial C(\Gamma_S)$ to $\partial_- C(\Gamma_T^{(1)})$, and $l_n \colon \partial_+ C(\Gamma_T^{(n)}) \to \partial C(\overline{\Gamma}_S)$.

For $1 \le j \le n$ and $k \in \{1, 2, 3, 4\}$ we take $DB_k^{(j)} = \phi_j \circ p_T(DB_k)$ as in the proof of Proposition 4.17. DB_k is defined above Lemma 4.15, which implies that $DB_k^{(j)}$ is an annular cross section of the cusp of $C(\Gamma_T^{(j)})$ corresponding to $\mathsf{p}_k^{\mathsf{c}^{-2j}}$. Each of T_1 and T_2 meets each of the $C(\Gamma_T^{(j)})$ in a collection of cusp cross sections parallel to a subcollection of the $DB_k^{(j)}$, $k \in \{1, 2, 3, 4\}$. Similarly, each of $T_1 \cap C(\overline{\Gamma}_S)$ and $T_2 \cap C(\overline{\Gamma}_S)$ is parallel to one of the cross sections $\bar{A}_1$ or $\bar{A}_2$.

By the proof of Proposition 4.17, for $1 \le j < n$, if $t_j = 0$ then $l_j = \iota_+^{(j)}(\iota_-^{(j)})^{-1}$ takes $\partial_+ DB_k^{(j)}$ to $\partial_- DB_k^{(j+1)}$ for each $k \in \{1, 2, 3, 4\}$. However if $t_j = 2$ then l_j acts on the indices k by the permutation (12)(34), since it uses $\phi_{\mathsf{m}_2}^{(j)}$. Likewise if $t_0 = 0$ then $l_0(\partial A_k) = \partial_- DB_k^{(1)} \sqcup \partial_- DB_{k+2}^{(1)}$ for $k = 1, 2$, by the proof of Proposition 4.17; hence if $t_0 = 2$, then $l_0(\partial A_k) = \partial_- DB_{3-k}^{(1)} \sqcup \partial_- DB_{5-k}^{(1)}$. A similar dichotomy holds for l_n.

Remark. The definitions of the annular cusp cross sections in Lemmas 4.13 and 4.14 depended on a particular collection of horospheres centered at the ideal vertices of $\mathcal{P}_1$ and $\mathcal{P}_2$. These give rise to a particular collection of horospherical cross sections of the cusps of $F^{(0)}$, which is not preserved by ϕ_{m_2}.

It is more accurate to say, for example, that when $t_j = 2$ and $1 \le j < n$, $l_j(\partial_+ DB_1^{(j)})$ is a cusp cross section of $\partial_- C(\Gamma_T^{(j+1)})$ *parallel* (and therefore similar) to $\partial_- DB_2^{(j+1)}$. Since the modulus is unaffected by similarities, we have largely ignored this distinction above and will continue to do so below.

Claim. *For each $j \in \{1, \dots, n\}$,*

$$T_1 \cap C(\Gamma_T^{(j)}) = \begin{cases} DB_1^{(j)} \cup DB_3^{(j)} & \text{if } c_{j-1} = 0, \\ DB_2^{(j)} \cup DB_4^{(j)} & \text{if } c_{j-1} = 1. \end{cases}$$

Furthermore, $T_1 \cap C(\overline{\Gamma}_S) = \bar{A}_1$ if $c_n = 0$ and $\bar{A}_2$ if $c_n = 1$.

Proof of claim. This is proved by induction on j. In the base case $j = 1$, since $c_0 = t_0/2$ and $T_1 \cap C(\Gamma_S) = A_1$, the conclusion in this case follows directly from the dichotomy in the behavior of l_0 recorded above the claim.

Suppose now that the claim holds for some $j < n$, and note that therefore $T_1 \cap M_T^{(j)}$ has components $DB_k^{(j)}$ and $DB_{k'}^{(j)}$, where $k, k' \in \{1, 2, 3, 4\}$ have the same parity, which is opposite that of c_{j-1}. By definition, c_j has the opposite parity from c_{j-1} if and only if $t_j = 2$. Writing $l_j(\partial_+ DB_k^{(j)}) = \partial_- DB_{k''}^{(j+1)}$, the above implies that k'' has parity opposite that of k if and only if $t_j = 2$. A similar assertion holds for k', and the claim follows for $j + 1$.

By induction, the claim holds for each $j \le n$. The final statement in the claim follows by an argument that mimics the one used in the inductive step. □

The moduli of A_1, A_2, $\bar{A}_1$, and $\bar{A}_2$ are described in Lemma 4.13, and those of the $DB_j^{(i)}$ are described in Lemma 4.15. Using these descriptions and Lemma 4.12, the claim above shows that the imaginary part of $m(T_1)$ is as described in the statement of the proposition. The description of the imaginary part of $m(T_2)$ follows similarly.

Now recall the definitions of the arcs a_1 and $db_k^{(j)}$ for $1 \le j \le n$ and $k = 1, 3$, and the collections of arcs $\mathcal{A}_2$ and $D\mathcal{B}_k^{(j)}$ for $1 \le j \le n$ and $k = 2, 4$, from the proof of Proposition 4.17. For our purposes here, we additionally define $\mathcal{A}_1$ to be a collection of five arcs evenly spaced around A_1, each perpendicular to ∂A_1 at each of its endpoints, such that $a_1 \in \mathcal{A}_1$. We analogously define $D\mathcal{B}_k^{(j)}$ as a collection of evenly spaced arcs in $DB_k^{(j)}$ containing $db_k^{(j)}$ for $1 \le j \le n$ and $k = 1, 3$.

Claim. *If $t_0 = 0$ then $l_0(\partial \mathcal{A}_k) = \partial_- D\mathcal{B}_k^{(1)} \cup \partial_- D\mathcal{B}_{k+2}^{(1)}$ for $k = 1, 2$, and if $t_0 = 2$ then $l_0(\partial \mathcal{A}_k) = \partial_- D\mathcal{B}_{3-k}^{(1)} \cup \partial_- D\mathcal{B}_{5-k}^{(1)}$. Similarly, for $1 \le j \le n - 1$,*

$$l_j(\partial_+ D\mathcal{B}_k^{(j)}) = \partial_- D\mathcal{B}_k^{(j+1)} \text{ for } k = 1, 2, 3, 4, \ \text{ if } t_j = 0,$$

$$l_j(\partial_+ D\mathcal{B}_k^{(j)}) = \begin{cases} \partial_- D\mathcal{B}_{3-k}^{(j+1)} & \text{for } k = 1, 2, \\ \partial_- D\mathcal{B}_{7-k}^{(j+1)} & \text{for } k = 3, 4, \end{cases} \quad \text{if } t_j = 2.$$

Also, if $t_n = 0$ *then* $l_n^{-1}(\partial\overline{\mathcal{A}}_k) = \partial_+ D\mathcal{B}_k^{(n)} \cup \partial_+ D\mathcal{B}_{k+2}^{(n)}$ *for* $k = 1, 2$, *and if* $t_n = 2$ *then* $l_n^{-1}(\partial\overline{\mathcal{A}}_k) = \partial_+ D\mathcal{B}_{3-k}^{(n)} \cup \partial_- D\mathcal{B}_{5-k}^{(n)}$.

In the discussion above the first claim, we recorded the analogous dichotomy to that of the claim above for the action of the gluing maps l_j on boundaries of annular cusp cross sections. The substance of this claim is thus that the gluing maps preserve arc endpoints.

Proof of claim. Suppose first that $t_j = 0$, so by its definition $l_j = \iota_+^{(j)}(\iota_-^{(j)})^{-1}$. The proof of Proposition 4.17 directly addresses the cases of $\mathcal{A}_2$, $\overline{\mathcal{A}}_2$, and $D\mathcal{B}_k^{(j)}$, where $k = 2$ or 4. In the remaining case of $\mathcal{A}_1$, the definition implies that $\partial\mathcal{A}_1$ consists of ten points, five evenly spaced around each component of ∂A_1, with each such collection containing a point of ∂a_1. Also by definition, $\partial_- D\mathcal{B}_k^{(1)}$ is a collection of five points spaced evenly around $\partial_- DB_k^{(1)}$, one of which is $\partial_- db_k^{(1)}$ for $k = 1, 3$. By the proof of Proposition 4.17, $\iota_+^{(0)}(\iota_-^{(0)})^{-1}$ takes ∂a_1 to $\partial_- db_1^{(1)} \cup \partial_- db_3^{(1)}$; hence the entire collection $\partial\mathcal{A}_1$ is taken to $\partial_- D\mathcal{B}_1^{(1)} \cup \partial_- D\mathcal{B}_3^{(1)}$ since $\iota_+^{(0)}(\iota_-^{(0)})^{-1}$ is an isometry. The remaining cases when $t_j = 0$, $j \geq 1$, follow similarly.

To illustrate the case $t_j = 2$ we focus on the subcase $1 \leq j < n$. When $t_0 = 2$, l_j takes $\partial_+ DB_1^{(j)}$ to $\partial_- DB_2^{(j+1)}$, for example. The crucial observation here is that $l_0(\partial_+ db_1^{(j)})$ is in $\partial_- D\mathcal{B}_2^{(j+1)}$. This holds because by definition, $\partial_+ db_1^{(j)}$ is a point in the edge of the triangulation Δ_T which exits the ideal vertex 1. (This is the top edge in Figure 12.) Although ϕ_{m_2} does not preserve Δ_T, it preserves this edge, exchanging its endpoints at 1 and 2. Since $\partial_- D\mathcal{B}_2^{(j+1)}$ has a point in each edge which exits 2, it contains $\phi_{\mathsf{m}_2}(\partial_+ db_1^{(j)})$. Since the points of $\partial_+ D\mathcal{B}_1^{(j)}$ are evenly spaced around $\partial_+ DB_1^{(j)}$ and the same is true for $\partial_- D\mathcal{B}_2^{(j+1)}$, it follows that $l_0(\partial_+ D\mathcal{B}_1^{(j)}) = \partial_- D\mathcal{B}_2^{(j+1)}$.

Since ϕ_{m_2} takes the edge of Δ_T to itself and exchanges its endpoints, we have $l_0(\partial_+ db_3^{(j)}) \in \partial_- D\mathcal{B}_4^{(j+1)}$ in this case. Then it follows from "even-spacedness" that $l_0(\partial_+ D\mathcal{B}_3^{(j)}) = \partial_- D\mathcal{B}_4^{(j+1)}$. The same argument implies that $\partial_- db_1^{(j+1)}$ lies in $l_0(\partial_+ D\mathcal{B}_2^{(j+1)})$ and therefore that $l_0(\partial_+ D\mathcal{B}_2^{(j)}) = \partial_- D\mathcal{B}_1^{(j+1)}$, and similarly that $l_0(\partial_+ D\mathcal{B}_4^{(j)}) = \partial_- D\mathcal{B}_3^{(j+1)}$. The same sequence of observations, applied to $\partial\mathcal{A}_k$ and $\partial\overline{\mathcal{A}}_k$, $k = 1, 2$, completes the claim. □

The second claim implies that the set

$$\mathcal{A}_1 \cup \mathcal{A}_2 \cup \bigcup_{j,k} D\mathcal{B}_k^{(j)} \cup \overline{\mathcal{A}}_1 \cup \overline{\mathcal{A}}_2$$

consists of a disjoint union of closed geodesics, some in T_1 and some in T_2, each meeting any of the geodesics $F^{(j)} \cap T_1$ or $F^{(j)} \cap T_2$ perpendicularly in up to five points. That $m(T_1)$ and $m(T_2)$ have real part equal to 0 (up to the action of $\mathrm{PGL}_2(\mathbb{Q})$) now follows as in the proof of Proposition 4.17. □

Proposition 7.3 allows us to describe arbitrarily large subfamilies of the M_I which have $\mathrm{PGL}_2(\mathbb{Q})$-inequivalent cusp parameters and hence are not commensurable.

Corollary 7.4. *For $0 \le k \le n$, let $I_k = (t_0, t_1, \dots, t_n)$ be defined by $t_i = 0$ for $i \neq k$, and $t_k = 2$. The cusp parameters of M_{I_k} are not $\mathrm{PGL}_2(\mathbb{Q})$-equivalent to those of $M_{I_{k'}}$ for $k \neq k'$, when both are less than $(n+1)/2$.*

Proof. By Proposition 7.3, the cusps of M_{I_k} have moduli described as

$$m(T_1) = i\left[\tfrac{6}{5} + \tfrac{4}{5}(n+4k)\sqrt{2}\right], \quad m(T_2) = i\left[\tfrac{6}{5} + \tfrac{4}{5}(5n-4k)\sqrt{2}\right].$$

Since $m(T_1)$ and $m(T_2)$ are both of the form described in Lemma 4.18 for any k, if the cusp parameters of M_{I_k} are equivalent to those of $M_{I_{k'}}$, then one of the two following cases holds:

$$\begin{aligned} n+4k &= n+4k', &&\text{and}\quad 5n-4k = 5n-4k', \\ n+4k &= 5n-4k', &&\text{and}\quad 5n-4k = n+4k', \end{aligned}$$

In the first case, $k = k'$, and in the second, $k' = n - k$. Thus as long as k and $k' < (n+1)/2$ are unequal, their cusp parameters are as well. □

There are also arbitrarily large subfamilies which share cusp parameters, even among complements of links obtained by mutating only with $(12)(34)$. We do not know if these are commensurable, although we suspect they are not.

Corollary 7.5. *For $0 \le k < n$, let $I_k = (t_0, \dots, t_n)$ be defined by $t_i = 0$ for $i \neq k, k+1$, and $t_k = t_{k+1} = 2$. For each k, the cusp parameters of M_{I_k} are*

$$m(T_1) = i\left[2 + 4\left(n - \tfrac{4}{5}\right)\sqrt{2}\right], \quad m(T_2) = i\left[\tfrac{2}{5} + \tfrac{4}{5}(n+4)\sqrt{2}\right],$$

up to the action of $\mathrm{PGL}_2(\mathbb{Q})$.

Corollaries 7.4 and 7.5 prove parts (2) and (3), respectively, of Theorem 3.

Appendix: Proof of Lemma 2.6

Following Morgan [1984], we define a *pared manifold* to be a pair (M, P), where M is a compact, orientable, irreducible 3-manifold with nonempty boundary which is not a 3-ball, and $P \subseteq \partial M$ is the union of a collection of disjoint incompressible annuli and tori satisfying the following properties:

- Every noncyclic abelian subgroup of $\pi_1 M$ is conjugate into the fundamental group of a component of P.
- Every map $\phi\colon (S^1 \times I, S^1 \times \partial I) \to (M, P)$ which induces an injection on fundamental groups is homotopic as a map of pairs to a map ψ such that $\psi(S^1 \times I) \subset P$.

This definition describes the topology of the compact manifold obtained by truncating the cusps of the convex core of a geometrically finite hyperbolic 3-manifold by open horoball neighborhoods. Indeed, Corollary 6.10 of [Morgan 1984] asserts that if (M, P) is obtained in this way, where P consists of the collection of boundaries of the truncating horoball neighborhoods, then (M, P) is a pared manifold.

Lemma 2.6 asserts that if (M, P) has the pared homotopy type of a geometrically finite hyperbolic manifold $\mathbb{H}^3/\Gamma$ where Γ is not Fuchsian and $\partial C(\Gamma)$ is totally geodesic, then $M - P$ is homeomorphic to $C(\Gamma)$. The key point of the proof is that the geometric conditions on Γ ensure that (M, P) is an acylindrical pared manifold. Then Johannson's theorem [Johannson 1979], that pared homotopy equivalences between acylindrical pared manifolds are homotopic to pared homeomorphisms, applies. We expand on this below.

It is worth noting that Lemma 2.6 fails in more general circumstances. The memoir [Canary and McCullough 2004] gives examples of this; Example 1.4.5, for instance, describes homotopy-equivalent non-Fuchsian geometrically finite manifolds with incompressible convex core boundary which are not homeomorphic. That work is devoted to understanding the ways in which homotopy equivalences of hyperbolic 3-manifolds can fail to be homotopic to homeomorphisms, and Lemma 2.6 follows quickly from results therein.

The treatment of Canary and McCullough itself uses the theory of *characteristic submanifolds* of manifolds with *boundary pattern* developed in [Johannson 1979]. The characteristic submanifold of a manifold with boundary pattern is a maximal collection of disjoint codimension-zero submanifolds, each an interval bundle or Seifert-fibered space embedded reasonably with respect to the boundary pattern. Rather than attempting to establish all of the notation necessary to define this formally, we refer the interested reader to the two works just cited. Here we simply transcribe the relevant theorem of [Canary and McCullough 2004], which strongly restricts the topology of the characteristic submanifold of a pared manifold with boundary pattern determined by the pared locus.

For the purposes of Lemma 2.6 we exclude from consideration certain pared manifolds which never arise from convex cores of geometrically finite hyperbolic 3-manifolds. We say (M, P) is *elementary* if it is homeomorphic to one of $(T^2 \times I, T^2 \times \{0\})$, $(A^2 \times I, A^2 \times \{0\})$, or $(A^2 \times I, \varnothing)$, where T^2 and A^2 denote the torus and annulus, respectively; otherwise (M, P) is nonelementary. Define $\partial_0 M := \overline{M - P}$. We say an annulus properly embedded in $M - P$ is *essential* in (M, P) if it is incompressible and boundary-incompressible in $M - P$. For a codimension-0 submanifold V embedded in M, we denote by $\mathrm{Fr}(V)$ the *frontier* of V (that is, its topological boundary in M), and note that $\mathrm{Fr}(V) = \overline{\partial V - (V \cap \partial M)}$. With notation thus established, the following theorem combines the definition of the characteristic submanifold with [Canary and McCullough 2004, Theorem 5.3.4].

Theorem. *Let (M, P) be a nonelementary pared manifold with $\partial_0 M$ incompressible. Select the fibering of the characteristic submanifold so that no component is an I-bundle over an annulus or Möbius band.*

(1) *Suppose V is a component of the characteristic submanifold which is an I-bundle over a surface B. Then each component of the associated ∂I-bundle is contained in $\partial_0 M$, each component of the associated I-bundle over ∂B is either a component of P or a properly embedded essential annulus, and B has negative Euler characteristic.*

(2) *A Seifert-fibered component V of the characteristic submanifold is homeomorphic either to $T^2 \times I$ or to a solid torus. If V is $T^2 \times I$ then one component of $T^2 \times \partial I$ lies in P and the other components of $V \cap \partial M$ are annuli in $\partial_0 M$. If V is a solid torus, then $V \cap \partial M$ has at least one component, each an annulus either containing a component of P or contained in $\partial_0 M$. In either case, each component of the frontier* $\mathrm{Fr}(V)$ *of V in M is a properly embedded essential annulus.*

The characteristic submanifold contains regular neighborhoods of all components of P.

The key claim in the proof of Lemma 2.6 is a further restriction on the characteristic submanifold of (M, P), in the case that M is obtained from the convex core of a non-Fuchsian geometrically finite manifold with totally geodesic convex core boundary by removing horoball neighborhoods of the cusps. P is the union of the boundaries of these neighborhoods.

Claim. (M, P) *as above is nonelementary, and $\partial_0 M$ is incompressible. The characteristic submanifold of* (M, P) *consists only of* (*Seifert-fibered*) *regular neighborhoods of the components of P, each of whose boundary has a unique component of intersection with ∂M.*

We prove the claim below, but assuming it for now, the proof of Lemma 3 proceeds as follows. A representation as given in the statement of the lemma induces a pared homotopy equivalence between (M, P) and the pared manifold (N, Q) obtained by truncating $C(\Gamma)$ with open horoball neighborhoods. Since $C(\Gamma)$ has totally geodesic convex core boundary, (N, Q) is as described by the claim; hence (M, P) is as well (see [Canary and McCullough 2004, Theorem 2.11.1], for example). Johansson's classification theorem (see [Canary and McCullough 2004, Theorem 2.9.10]) implies that the original pared homotopy equivalence is homotopic to one which maps the complement of the characteristic submanifold of (M, P) homeomorphically to the complement of the characteristic submanifold of (N, Q). It follows from the claim that these are homeomorphic to $M - P$ and $N - Q$, respectively, and the lemma follows.

Proof of claim. As was mentioned above, the elementary pared manifolds do not arise from geometrically finite hyperbolic manifolds. Since (M, P) is obtained from the convex core of a geometrically finite manifold with totally geodesic convex core boundary, the following are known not to occur:

(1) A compressing disk for $\partial_0 M$. (By definition $\partial_0 M$ lifts to a geodesic hyperplane in $\mathbb{H}^3$, hence the induced map $\pi_1 \partial M_0 \to \pi_1 M$ is injective.)

(2) An *accidental parabolic*: an essential annulus properly embedded in M with one boundary component in P and one in $\partial_0 M$, which is not parallel to P. (Every essential curve on $\partial_0 M$ that is not boundary-parallel is homotopic to a geodesic, but an element of $\pi_1(M)$ corresponding to an accidental parabolic has translation length 0.)

(3) A *cylinder*; that is, a properly embedded essential annulus in $M - P$. (The double DM of M across $\partial_0 M$ is a hyperbolic manifold, but the double of a cylinder in M would be an essential torus in DM.)

We show that if the characteristic manifold has any components other than those listed in the claim then at least one of the above facts cannot hold.

For a component V of the characteristic submanifold which is an I-bundle over a surface B, at least one component of the associated I-bundle over ∂B must be properly embedded, since otherwise we would have $M = V$ and it is well known that an I-bundle over a surface does not admit a hyperbolic structure with totally geodesic convex core boundary unless the convex core is a Fuchsian surface. But this annulus violates (2) or (3). Thus there are no I-bundle components of the characteristic submanifold.

If V is a Seifert-fibered component of the characteristic submanifold homeomorphic to $T^2 \times I$, then one component of ∂V is a torus $P_1 \subset P$, and all other components of $\partial V \cap \partial M$ are annuli in $\partial_0 M$. If this second class is nonempty then each component of $\mathrm{Fr}(V)$ is an essential annulus properly embedded in $M - P$, contradicting fact (3). Thus $\partial V \cap \partial M = P_1$ and V is a regular neighborhood of P_1.

If V is a solid torus and $V \cap \partial M$ contains a component of P, then a similar argument shows that this is the unique component of $\partial V \cap \partial M$, so in this case V is a regular neighborhood of an annular component of P. If, on the other hand, $V \cap \partial M$ does *not* contain any components of P, then it has at least two components, for otherwise a meridional disk of V determines a boundary compression of the annulus $\mathrm{Fr}(V)$ in $M - P$. But then any component of $\mathrm{Fr}(V)$ violates fact (3). □

Acknowledgements

The authors thank Ian Agol, Richard Kent, Chris Leininger, Peter Shalen, and Christian Zickert for helpful conversations, Joe Masters for suggesting the cusp parameter, and Dick Canary for helping us with Lemma 2.6. A referee on an earlier version of this paper pointed us to the Bloch invariant and motivated several major changes in this paper. We also appreciate the thoughtful editorial comments from a second referee. We want to especially thank Alan Reid for suggesting these questions to us and for many helpful conversations and suggestions. The second author is grateful to the Clay Mathematics Institute for support during part of this project. The authors also thank the University of Montana's Faculty Development Committee for their support.

References

[Adams 1985] C. C. Adams, "Thrice-punctured spheres in hyperbolic 3-manifolds", *Trans. Amer. Math. Soc.* **287**:2 (1985), 645–656. MR 86k:57008 Zbl 0527.57002

[Adams and Schoenfeld 2005] C. C. Adams and E. Schoenfeld, "Totally geodesic Seifert surfaces in hyperbolic knot and link complements, I", *Geom. Dedicata* **116** (2005), 237–247. MR 2006j:57008 Zbl 1092.57003

[Adams et al. 2008] C. C. Adams, H. Bennett, C. Davis, M. Jennings, J. Kloke, N. Perry, and E. Schoenfeld, "Totally geodesic Seifert surfaces in hyperbolic knot and link complements, II", *J. Differential Geom.* **79**:1 (2008), 1–23. MR 2009b:57032 Zbl 1158.57004

[Aitchison and Rubinstein 1997] I. R. Aitchison and J. H. Rubinstein, "Geodesic surfaces in knot complements", *Experiment. Math.* **6**:2 (1997), 137–150. MR 98h:57011 Zbl 0891.57017

[Anderson and Canary 2001] J. W. Anderson and R. D. Canary, "The visual core of a hyperbolic 3-manifold", *Math. Ann.* **321**:4 (2001), 989–1000. MR 2002i:57014 Zbl 0992.30028

[Bass 1984] H. Bass, "Finitely generated subgroups of GL_2", Chapter VI, pp. 127–136 in *The Smith conjecture* (New York, 1979), edited by J. W. Morgan and H. Bass, Pure Appl. Math. **112**, Academic Press, Orlando, FL, 1984. MR 758465 Zbl 0599.57003

[Borel 1977] A. Borel, "Cohomologie de SL_n et valeurs de fonctions zeta aux points entiers", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **4**:4 (1977), 613–636. MR 58 #22016 Zbl 0382.57027

[Canary and McCullough 2004] R. D. Canary and D. McCullough, "Homotopy equivalences of 3-manifolds and deformation theory of Kleinian groups", *Mem. Amer. Math. Soc.* **172**:812 (2004). MR 2005j:57027 Zbl 1062.57027

[Canary et al. 2006] R. D. Canary, D. B. A. Epstein, and P. L. Green, "Notes on notes of Thurston", pp. 1–115 in *Fundamentals of hyperbolic geometry: selected expositions*, edited by R. D. Canary et al., London Math. Soc. Lecture Note Ser. **328**, Cambridge Univ. Press, 2006. MR 2235710 Zbl 0612.57009

[Dupont 2001] J. L. Dupont, *Scissors congruences, group homology and characteristic classes*, Nankai Tracts in Mathematics **1**, World Scientific, River Edge, NJ, 2001. MR 2002g:52013 Zbl 0977.52020

[Epstein and Penner 1988] D. B. A. Epstein and R. C. Penner, "Euclidean decompositions of noncompact hyperbolic manifolds", *J. Differential Geom.* **27**:1 (1988), 67–80. MR 89a:57020 Zbl 0611.53036

[Farb and Margalit 2012] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton Mathematical Series **49**, Princeton Univ. Press, 2012. MR 2012h:57032 Zbl 1245.57002

[Goodman et al. 2008] O. Goodman, D. Heard, and C. Hodgson, "Commensurators of cusped hyperbolic manifolds", *Experiment. Math.* **17**:3 (2008), 283–306. MR 2009h:57029 Zbl 05500548 arXiv 0801.4815

[Hatcher 1983] A. Hatcher, "Hyperbolic structures of arithmetic type on some link complements", *J. London Math. Soc.* (2) **27**:2 (1983), 345–355. MR 84m:57005 Zbl 0516.57001

[Johannson 1979] K. Johannson, *Homotopy equivalences of* 3*-manifolds with boundaries*, Lecture Notes in Mathematics **761**, Springer, Berlin, 1979. MR 82c:57005 Zbl 0412.57007

[Leininger 2006] C. J. Leininger, "Small curvature surfaces in hyperbolic 3-manifolds", *J. Knot Theory Ramifications* **15**:3 (2006), 379–411. MR 2007a:57025 Zbl 1090.57012

[Macasieb and Mattman 2008] M. L. Macasieb and T. W. Mattman, "Commensurability classes of $(-2, 3, n)$ pretzel knot complements", *Algebr. Geom. Topol.* **8**:3 (2008), 1833–1853. MR 2009g: 57011 Zbl 1162.57005

[Maclachlan and Reid 1991] C. Maclachlan and A. W. Reid, "Parametrizing Fuchsian subgroups of the Bianchi groups", *Canad. J. Math.* **43**:1 (1991), 158–181. MR 92d:11040 Zbl 0739.20020

[Maclachlan and Reid 2003] C. Maclachlan and A. W. Reid, *The arithmetic of hyperbolic* 3*-manifolds*, Graduate Texts in Mathematics **219**, Springer, New York, 2003. MR 2004i:57021 Zbl 1025.57001

[Margulis 1991] G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **17**, Springer, Berlin, 1991. MR 92h:22021 Zbl 0732.22008

[Maskit 1971] B. Maskit, "On Klein's combination theorem, III", pp. 297–316 in *Advances in the theory of Riemann surfaces* (Stony Brook, NY, 1969), edited by L. V. Ahlfors et al., Ann. of Math. Studies **66**, Princeton Univ. Press, 1971. MR 44 #6955 Zbl 0222.30029

[Menasco and Reid 1992] W. Menasco and A. W. Reid, "Totally geodesic surfaces in hyperbolic link complements", pp. 215–226 in *Topology '90* (Columbus, OH, 1990), edited by B. Apanasov et al., Ohio State Univ. Math. Res. Inst. Publ. **1**, de Gruyter, Berlin, 1992. MR 94g:57016 Zbl 0769.57014

[Morgan 1984] J. W. Morgan, "On Thurston's uniformization theorem for three-dimensional manifolds", pp. 37–125 in *The Smith conjecture* (New York, 1979), edited by J. W. Morgan and H. Bass, Pure Appl. Math. **112**, Academic Press, Orlando, FL, 1984. MR 758464 Zbl 0599.57002

[Neumann 1998] W. D. Neumann, "Hilbert's 3rd problem and invariants of 3-manifolds", pp. 383–411 in *The Epstein birthday schrift*, edited by I. Rivin et al., Geom. Topol. Monogr. **1**, Geom. Topol. Publ., Coventry, 1998. MR 2000h:57023 Zbl 0902.57013

[Neumann 2011] W. D. Neumann, "Realizing arithmetic invariants of hyperbolic 3-manifolds", pp. 233–246 in *Interactions between hyperbolic geometry, quantum topology and number theory* (New York, 2009), edited by A. Champagnerkar et al., Contemp. Math. **541**, Amer. Math. Soc., Providence, RI, 2011. MR 2012e:57030 Zbl 1237.57013

[Neumann and Reid 1991] W. D. Neumann and A. W. Reid, "Amalgamation and the invariant trace field of a Kleinian group", *Math. Proc. Cambridge Philos. Soc.* **109**:3 (1991), 509–515. MR 92b:30053 Zbl 0728.57009

[Neumann and Reid 1992] W. D. Neumann and A. W. Reid, "Arithmetic of hyperbolic manifolds", pp. 273–310 in *Topology '90* (Columbus, OH, 1990), edited by B. Apanasov et al., Ohio State Univ. Math. Res. Inst. Publ. **1**, de Gruyter, Berlin, 1992. MR 94c:57024 Zbl 0777.57007

[Neumann and Yang 1999] W. D. Neumann and J. Yang, "Bloch invariants of hyperbolic 3-manifolds", *Duke Math. J.* **96**:1 (1999), 29–59. MR 2000h:57022 Zbl 0943.57008

[Paoluzzi and Zimmermann 1996] L. Paoluzzi and B. Zimmermann, "On a class of hyperbolic 3-manifolds and groups with one defining relation", *Geom. Dedicata* **60**:2 (1996), 113–123. MR 97e: 57014 Zbl 0857.57009

[Ratcliffe 1994] J. G. Ratcliffe, *Foundations of hyperbolic manifolds*, Graduate Texts in Mathematics **149**, Springer, New York, 1994. MR 95j:57011 Zbl 0809.51001

[Reid and Walsh 2008] A. W. Reid and G. S. Walsh, "Commensurability classes of 2-bridge knot complements", *Algebr. Geom. Topol.* **8**:2 (2008), 1031–1057. MR 2009i:57020 Zbl 1154.57001

[Ruberman 1987] D. Ruberman, "Mutation and volumes of knots in S^3", *Invent. Math.* **90**:1 (1987), 189–215. MR 89d:57018 Zbl 0634.57005

[Thurston 1979] W. P. Thurston, "The geometry and topology of 3-manifolds", 1979, Available at http://library.msri.org/books/gt3m. Mimeographed lecture notes.

Received March 18, 2012. Revised August 13, 2013.

ERIC CHESEBRO
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF MONTANA
MATHEMATICS BUILDING 102
MISSOULA, MT 59812-0864
UNITED STATES
Eric.Chesebro@mso.umt.edu

JASON DEBLOIS
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF PITTSBURGH
301 THACKERAY HALL
PITTSBURGH, PA 15260
UNITED STATES
jdeblois@pitt.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.399

TAUT FOLIATIONS AND THE ACTION OF THE FUNDAMENTAL GROUP ON LEAF SPACES AND UNIVERSAL CIRCLES

YOSUKE KANO

Let $\mathcal{F}$ be a leafwise hyperbolic taut foliation of a closed 3-manifold M and let L be the leaf space of the pullback of $\mathcal{F}$ to the universal cover of M. We show that if $\mathcal{F}$ has branching, then the natural action of $\pi_1(M)$ on L is faithful. We also show that if $\mathcal{F}$ has a finite branch locus B whose stabilizer acts on B nontrivially, then the stabilizer is an infinite cyclic group generated by an indivisible element of $\pi_1(M)$.

1. Introduction

Unless otherwise specified, we assume throughout this article that M is a closed oriented 3-manifold and $\mathcal{F}$ a codimension-one transversely oriented, leafwise hyperbolic, taut foliation of M. Here we say that $\mathcal{F}$ is *leafwise hyperbolic* if there is a transversely continuous leafwise Riemannian metric on M where the leaves are locally isometric to the hyperbolic plane, and that $\mathcal{F}$ is *taut* if there is a loop in M which intersects every leaf of $\mathcal{F}$ transversely. Note that by [Candel 1993], if M is irreducible and atoroidal, then every taut foliation of M is leafwise hyperbolic.

Leafwise hyperbolic taut foliations have been extensively investigated by many people in connection with the theory of 3-manifolds (see, for example, Calegari's book [2007]). One of the most powerful methods of analyzing the structure of such foliations is to consider canonical actions of $\pi_1(M)$ on 1-manifolds naturally associated with $\mathcal{F}$. Two kinds of such 1-manifolds are known. The first one, denoted L, is the leaf space of $\widetilde{\mathcal{F}}$, where $\widetilde{\mathcal{F}}$ is the pullback of $\mathcal{F}$ to the universal cover $\widetilde{M}$ of M. The action of $\pi_1(M)$ on $\widetilde{M}$ induces an action of $\pi_1(M)$ on L. In the sequel we refer to it as the *natural action*. The second one is a universal circle. By unifying circles at infinity of all the leaves of a given $\widetilde{\mathcal{F}}$, Thurston [1998] (see also [Calegari and Dunfield 2003]) constructs a universal circle with a canonical $\pi_1(M)$ action.

This work is partially supported by the AGSST support program for young researchers, 2011, Chiba University, Japan.
MSC2010: 57M05, 57M60, 57R30.
Keywords: foliation, leaf space, universal circle.

We say that $\mathcal{F}$ has *branching* if L is non-Hausdorff. The first result of this article is the following:

Theorem 3.2. *If $\mathcal{F}$ has branching, then the natural action on L is faithful.*

This result is obtained from an investigation of both actions of $\pi_1(M)$ on the leaf space and on the universal circle (see Section 3). Notice that the hypothesis that $\mathcal{F}$ has branching is indispensable. In fact, just consider a surface bundle over S^1 foliated by fibers. Notice also that, by Theorem 7.10 of [Calegari and Dunfield 2003], any taut foliation can be modified by suitable Denjoy-like insertions so that the natural action associated with the resulting foliation becomes faithful. In the case where the foliation is leafwise hyperbolic and has branching, our result is stronger in that we assure faithfulness without performing any modifications.

Next we consider the stabilizer of a branch locus of $\mathcal{F}$. We call a subset B of L a *branch locus* if B contains at least two points and can be expressed in the form $B = \lim_{t\to 0} \nu_t$ for some interval $\{\nu_t \in L \mid 0 < t < \epsilon\}$ embedded in L. Furthermore, if the parameter t of the interval is incompatible (resp. compatible) with the orientation of L, we call B a *positive* (resp. *negative*) branch locus. (Note that L has a natural orientation induced from the transverse orientation of $\widetilde{\mathcal{F}}$.) Branch loci have been studied, for example, in [Fenley 1998; Shields 2002]. For a branch locus B we define the *stabilizer* of B by $\mathrm{Stab}(B) = \{\alpha \in \pi_1(M) \mid \alpha(B) = B\}$.

In the case where a branch locus B is finite, we obtain the following results about the action of $\mathrm{Stab}(B)$ on B (see Section 5 for details).

Theorem 5.2. *Let B be a finite branch locus of L. If an element of* $\mathrm{Stab}(B)$ *fixes some point of B then it fixes all the points of B.*

We remark that for Anosov foliations, Theorem D of [Fenley 1998] contains results related to this theorem.

Let $\pi : \widetilde{M} \to M$ be the covering projection. For a leaf λ of $\widetilde{\mathcal{F}}$, we denote by $\underline{\lambda}$ the projected leaf $\pi(\lambda)$ of $\mathcal{F}$.

Theorem 5.3. *Let B be a branch locus of L. Then,*

(1) *if* $\mathrm{Stab}(B)$ *is trivial, $\underline{\lambda}$ is diffeomorphic to a plane, and*

(2) *if B is finite and* $\mathrm{Stab}(B)$ *is nontrivial, $\underline{\lambda}$ is diffeomorphic to a cylinder*

for any $\lambda \in B$.

Theorem 5.6. *Let B be a finite branch locus of L with a nontrivial stabilizer. Then the stabilizer* $\mathrm{Stab}(B)$ *is isomorphic to $\mathbb{Z}$.*

We say that $\alpha \in \pi_1(M)$ is *divisible* if there is some $\beta \in \pi_1(M)$ and an integer $k \geq 2$ such that $\alpha = \beta^k$. Otherwise we say α is *indivisible*.

Theorem 5.7. *Let B be a finite branch locus of L such that* $\mathrm{Stab}(B)$ *acts on B nontrivially. Then a generator of* $\mathrm{Stab}(B)$ $(\cong \mathbb{Z})$ *is indivisible.*

For an oriented loop γ in M, we say that γ is *tangentiable* if γ is freely homotopic to a leaf loop (a loop contained in a single leaf) of $\mathscr{F}$, and that γ is positively (resp. negatively) *transversable* if γ is freely homotopic to a loop positively (resp. negatively) transverse to $\mathscr{F}$. As a final topic of this article, we study relations between the infiniteness of branch loci and the existence of a nontransversable leaf loop in M (see Section 6). One of the results we obtain is the following:

Theorem 6.5. *Suppose $\mathscr{F}$ has branching. If there is a noncontractible leaf loop in M which is not freely homotopic to a loop transverse to $\mathscr{F}$, then $\mathscr{F}$ has an infinite branch locus.*

This article is organized as follows. In Section 2, we briefly review the Calegari–Dunfield construction of a universal circle. Using their construction, we prove the faithfulness of the natural action of $\pi_1(M)$ on L in Section 3. In Section 4, we introduce a notion of comparable sets and give several basic properties of such sets, which are applied in Section 5 to the investigation of the structure of finite branch loci and their stabilizers. In Section 6, we study how the nontransversability of leaf loops in M is related to the infiniteness of branch loci in L.

2. Universal circles

The theory of universal circles was originally developed in [Thurston 1998], and was written up carefully in [Calegari and Dunfield 2003]. In this section we briefly recall the definition of a universal circle after the latter reference.

Let M, $\mathscr{F}$ and L be as in the introduction. Here the topology of L is the quotient topology from $\widetilde{M}$; that is, there is a canonical projection map $q : \widetilde{M} \to L$ sending a point to the leaf containing it. The topology of L is the quotient topology from the map q.

For $\lambda, \mu \in L$ we write $\lambda < \mu$ if there is an oriented path in $\widetilde{M}$ from λ to μ which is positively transverse to $\widetilde{\mathscr{F}}$. We say that λ and μ are *comparable* if either $\lambda \le \mu$ or $\lambda \ge \mu$. For a leaf λ of $\widetilde{\mathscr{F}}$, the *endpoint map* $e : T_p\lambda - \{0\} \to S^1_\infty(\lambda)$ from the tangent space of λ at p to the ideal boundary of λ takes a vector v to the endpoint at infinity of the geodesic ray γ with $\gamma(0) = p$ and $\gamma'(0) = v$. The *circle bundle at infinity* is the disjoint union $E_\infty = \bigcup_{\lambda \in L} S^1_\infty(\lambda)$ with the finest topology such that the endpoint map $e : T\widetilde{\mathscr{F}} \setminus (\text{zero section}) \to E_\infty$ is continuous. A continuous map $\phi : X \to Y$ between oriented 1-manifolds homeomorphic to S^1 is *monotone* if it is of mapping degree one and if the preimage of any point of Y is contractible. A *gap* of ϕ is the interior in X of such a preimage. The *core* of ϕ is the complement of the union of gaps.

Definition 2.1. A *universal circle* S^1_{univ} for $\mathscr{F}$ is a circle together with a homomorphism $\rho_{\text{univ}} : \pi_1(M) \to \text{Homeo}^+(S^1_{\text{univ}})$ and a family of monotone maps $\phi_\lambda : S^1_{\text{univ}} \to S^1_\infty(\lambda)$, $\lambda \in \widetilde{\mathscr{F}}$, satisfying the following conditions:

1. For every $\alpha \in \pi_1(M)$, the following diagram commutes:
$$\begin{tikzcd} S^1_{\mathrm{univ}} \arrow[r, "\rho_{\mathrm{univ}}(\alpha)"] \arrow[d, "\phi_\lambda"'] & S^1_{\mathrm{univ}} \arrow[d, "\phi_{\alpha(\lambda)}"'] \\ S^1_\infty(\lambda) \arrow[r, "\alpha"] & S^1_\infty(\alpha(\lambda)). \end{tikzcd}$$
2. If λ and μ are incomparable, then the core of ϕ_λ is contained in the closure of a single gap of ϕ_μ and vice versa.

Calegari and Dunfield's construction for a universal circle is as follows. Let $I = [0, 1]$ be the unit interval. A *marker* for $\mathcal{F}$ is a continuous map $m : I \times \mathbb{R}^+ \to \widetilde{M}$ with the following properties:

- For each $s \in I$, the image $m(s \times \mathbb{R}^+)$ is a geodesic ray in a leaf of $\widetilde{\mathcal{F}}$. We call these the *horizontal rays* of m.
- For each $t \in \mathbb{R}^+$, the image $m(I \times t)$ is transverse to $\widetilde{\mathcal{F}}$ and of length smaller than some constant depending only on $\widetilde{\mathcal{F}}$.

We use the interval notation $[\lambda, \mu]$ to represent the oriented image of an injective continuous map $c : I \to L$ such that $c(0) = \lambda$ and $c(1) = \mu$. We call this the *interval* from λ to μ. Here, notice that the orientation of such an interval is induced from that of I (not from that of L).

Let $J = [\lambda, \mu]$ be an interval in L and let m be a marker which intersects only leaves of $\widetilde{\mathcal{F}}|_J$. Then the endpoints of the horizontal rays of m form an interval in $E_\infty|_J$ which is transverse to the circle fibers. By abuse of notation we refer to such an interval as a marker.

For each $\nu \in J$, the intersection of $S^1_\infty(\nu)$ with the union of all markers is dense in $S^1_\infty(\nu)$. If two markers m_1, m_2 in $E_\infty|_J$ are not disjoint, their union $m_1 \cup m_2$ is also an interval transverse to the circle fibers. It follows that a maximal such union of markers is still an interval. Again by abuse of notation we call such an interval a marker.

A continuous section $\tau : J \to E_\infty|_J$ is *admissible* if the image of τ does not cross (but might run into) any marker. The *leftmost section* $\tau(p, J) : J \to E_\infty|_J$ starting at $p \in S^1_\infty(\lambda)$ is an admissible section which is clockwisemost among all such sections if the order of J is compatible with that of L, and anticlockwisemost otherwise. Here, the meaning of "(anti-)clockwisemost" is the following: Consider the universal cover $\widetilde{E_\infty|_J} \cong \mathbb{R} \times J$ of $E_\infty|_J$ and take a lift $\tilde{p} \in \mathbb{R} \times J$ of p. Then, we say that τ is clockwisemost (resp. anticlockwisemost) if for any admissible section τ' the lifts $\tilde{\tau}, \tilde{\tau}'$ of τ, τ' to $\mathbb{R} \times J$ based at $\tilde{p}$ satisfy $\tilde{\tau}(\nu) \leq \tilde{\tau}'(\nu)$ (resp. $\tilde{\tau}'(\nu) \leq \tilde{\tau}(\nu)$) for any $\nu \in J$. For any p the leftmost section starting at p exists.

Let $B = \lim_{t \to 0} \nu_t$ be a branch locus and let $\mu_1, \mu_2 \in B$. For each $t > 0$, let $\alpha_t = [\mu_1, \nu_t]$ and $\beta_t = [\nu_t, \mu_2]$. Then, we can define a map $r_t : S^1_\infty(\mu_1) \to S^1_\infty(\mu_2)$

by $r_t(p) = \tau(\tau(p, \alpha_t)(\nu_t), \beta_t)(\mu_2)$. As t tends to 0, r_t converges to a constant map. We denote the image of the constant map by $r(\mu_1, \mu_2) \in S^1_\infty(\mu_2)$.

Definition 2.2. We call $r(\mu_1, \mu_2)$ the *turning point from μ_1 to μ_2*.

Given a pair $\lambda, \mu \in L$, we define a *geodesic spine* from λ to μ to be a disjoint union of finitely many intervals $[\hat{\nu}_{i-1}, \check{\nu}_i]$, $1 \le i \le n$, in L (some of them may degenerate to singletons), with the following properties:

(1) $\hat{\nu}_0 = \lambda$ and $\check{\nu}_n = \mu$,

(2) $\check{\nu}_i$ and $\hat{\nu}_i$ belong to a common branch locus for each $1 \le i \le n-1$, and

(3) n is minimal under the conditions (1) and (2).

Note that a geodesic spine connecting any two points in L exists and is unique. Geodesic spines have been extensively used in [Barbot 1996; 1998; Fenley 2003; Roberts et al. 2003].

For a point p in $S^1_\infty(\lambda)$, the *special section* $\sigma_p : L \to E_\infty$ at p is defined as follows. First, set $\sigma_p(\lambda) = p$. Next, pick any point $\mu \in L$. We define $\sigma_p(\mu)$ as follows: When μ is comparable with λ, then σ_p is defined on $[\lambda, \mu]$ to be the leftmost section starting at p. When μ is incomparable with λ, let $\coprod_{i=1}^n [\hat{\nu}_{i-1}, \check{\nu}_i]$ $(n > 1)$ be the geodesic spine from λ to μ. We then put $r = r(\check{\nu}_{n-1}, \hat{\nu}_{n-1}) \in S^1_\infty(\hat{\nu}_{n-1})$ and define σ_p on the interval $[\hat{\nu}_{n-1}, \check{\nu}_n]$ by $\sigma_p = \sigma_r$. This completes the definition of σ_p.

Let $\mathfrak{S}$ be the union of the special sections σ_p as p varies over all points in all circles $S^1_\infty(\lambda)$ of points λ in L. By [Calegari and Dunfield 2003, Lemma 6.25], the set $\mathfrak{S}$ admits a natural circular order. The universal circle S^1_{univ} will be derived from $\mathfrak{S}$ as a quotient of the order completion of $\mathfrak{S}$ with respect to the circular order. Remark that limits of special sections are also sections, hence that any element of S^1_{univ} is represented by a section $L \to E_\infty$.

3. Faithfulness of the action

We now show that if $\mathscr{F}$ has branching, the natural action of $\pi_1(M)$ on the leaf space L is faithful.

As explained in Section 2, every element σ of S^1_{univ} can be described as a section $\sigma : L \to E_\infty = \bigcup_{\lambda \in L} S^1_\infty(\lambda)$ and that the maps $\phi_\lambda : S^1_{\text{univ}} \to S^1_\infty(\lambda)$ are defined by $\phi_\lambda(\sigma) = \sigma(\lambda)$. For a point x in $S^1_\infty(\lambda)$, we define a (possibly degenerate) closed interval I_x in S^1_{univ} by $I_x = \{\sigma \in S^1_{\text{univ}} \mid \sigma(\lambda) = x\}$. Then, for any x the interval I_x is nonempty because the special section σ_x at x belongs to I_x. From the definition of a turning point, we have the following fact: If μ_1, μ_2 are in a branch locus and if z is in $S^1_\infty(\mu_2)$, then $\phi_{\mu_1}(\sigma_z) = r(\mu_2, \mu_1)$; that is, $\sigma_z \in I_{r(\mu_2, \mu_1)}$.

Let $\lambda \in L$ and $\alpha \in \pi_1(M)$ be such that $\alpha(\lambda) = \lambda$. Then α, as the restriction of a covering transformation of $\widetilde{M}$ to λ, induces an isometry of the hyperbolic plane λ, (hence also a projective transformation of $S^1_\infty(\lambda)$). We notice that this isometry is a hyperbolic element (meaning that its trace is greater than 2). In fact, since it has no fixed points in λ, it is not elliptic. If it were parabolic, then it would yield in M a noncontractible loop whose length can be made arbitrarily small, contradicting the compactness of M.

The following is a key lemma.

Lemma 3.1. *Let $B = \lim_{t\to 0} \nu_t$ be a branch locus of L. If $\alpha \in \pi_1(M)$ fixes two distinct points μ_1 and μ_2 in B and also fixes the interval $\{\nu_t \mid 0 < t < \epsilon\}$ pointwise, then α is trivial in $\pi_1(M)$.*

Proof. Suppose α is nontrivial. Let $p_1, q_1 \in S^1_\infty(\mu_1)$ and $p_2, q_2 \in S^1_\infty(\mu_2)$ be the fixed points of α, and let $r_1 \in S^1_\infty(\mu_1)$ be the turning point from μ_2 to μ_1. Without loss of generality, we assume that $p_1 \neq r_1$. Note that by construction of the universal circle, the special sections σ_{p_i} and σ_{q_i} in S^1_{univ} are fixed by $\rho_{\text{univ}}(\alpha)$ for $i = 1, 2$; therefore the images $\phi_{\nu_t}(\sigma_{p_i})$ and $\phi_{\nu_t}(\sigma_{q_i})$ are fixed by α for any $t \in (0, \epsilon)$.

We claim that if t is sufficiently close to 0, then $\phi_{\nu_t}(\sigma_{p_1})$ and $\phi_{\nu_t}(I_{r_1})$ are disjoint in $S^1_\infty(\nu_t)$. Take two distinct points x and y in $S^1_\infty(\mu_1) - \{p_1, r_1\}$ so that the 4-tuple (p_1, x, r_1, y) lies in circular order. Because of the density of markers, for sufficiently small $t > 0$ the 4-tuple $(\sigma_{p_1}(\nu_t), \sigma_x(\nu_t), \sigma_{r_1}(\nu_t), \sigma_y(\nu_t))$ lies in $S^1_\infty(\nu_t)$ also in circular order. Let K_t be the closed interval in $S^1_\infty(\nu_t)$ with boundary points $\sigma_x(\nu_t)$ and $\sigma_y(\nu_t)$ and containing $\sigma_{r_1}(\nu_t)$. Since I_{r_1} contains σ_{r_1} but not σ_{p_1}, σ_x and σ_y, and since special sections cannot cross, $\phi_{\nu_t}(I_{r_1})$ is contained in K_t. In particular, $\phi_{\nu_t}(\sigma_{p_1})$ and $\phi_{\nu_t}(I_{r_1})$ are disjoint. This shows the claim.

For t sufficiently close to 0, the two points $\sigma_{p_2}(\nu_t)$ and $\sigma_{q_2}(\nu_t)$ are distinct. Since both σ_{p_2} and σ_{q_2} pass through the turning point r_1 from μ_2 to μ_1, it follows that $\phi_{\mu_1}(\sigma_{p_2}) = \phi_{\mu_1}(\sigma_{q_2}) = r_1$; that is, σ_{p_2} and σ_{q_2} are contained in I_{r_1}. Therefore the 3 points $\sigma_{p_1}(\nu_t)$, $\sigma_{p_2}(\nu_t)$ and $\sigma_{q_2}(\nu_t)$ are also mutually distinct. Thus, we find at least 3 fixed points of α in $S^1_\infty(\nu_t)$, contradicting the fact that α is a nontrivial orientation preserving isometry of the hyperbolic plane ν_t. □

Now, the first main result of this article is the following:

Theorem 3.2. *Let M be a closed oriented 3-manifold, and $\mathscr{F}$ a transversely oriented leafwise hyperbolic taut foliation of M. If $\mathscr{F}$ has branching, then the natural action of $\pi_1(M)$ on the leaf space of $\widetilde{\mathscr{F}}$ is faithful.*

Proof. This is a direct consequence of Lemma 3.1. □

4. Comparable sets

In this section we do not assume leafwise hyperbolicity of $\mathscr{F}$. For $\alpha \in \pi_1(M)$, we define the *comparable set* C_α for α to be the subset of L consisting of points λ such that λ and $\alpha(\lambda)$ are comparable. Below we collect some basic properties of comparable sets.

Obviously, $\alpha(C_\alpha) = C_\alpha$, $C_\alpha = C_{\alpha^{-1}}$ and $C_\alpha \subset C_{\alpha^k}$ for every $k > 0$.

We say that $\mathcal{F}$ has *one-sided branching in the positive* (resp. *negative*) *direction* if L has positive (resp. negative) branch loci but has no negative (resp. positive) ones. If L has both positive loci and negative loci, then we say $\mathcal{F}$ has *two-sided branching*.

Lemma 4.1. *Let $\mathcal{F}$ have one-sided branching in the positive direction, and let $\alpha \in \pi_1(M)$. Suppose λ and μ are points in L such that λ is a common lower bound of μ and $\alpha(\mu)$, meaning that $\lambda \le \mu$ and $\lambda \le \alpha(\mu)$. Then $\lambda \in C_\alpha$.*

Proof. Since the natural action preserves the order of L, the inequality $\lambda \le \mu$ implies $\alpha(\lambda) \le \alpha(\mu)$. Thus, by the hypothesis, $\alpha(\mu)$ is a common upper bound of λ and $\alpha(\lambda)$. Since $\mathcal{F}$ has no branching in the negative direction, it follows that λ and $\alpha(\lambda)$ are comparable. □

From this lemma we see the following fact: Let $\mathcal{F}$ and α be as above. Then, there is $\lambda \in L$ such that $\{\mu \in L \mid \mu < \lambda\} \subset C_\alpha$.

Lemma 4.2. *Let $\alpha \in \pi_1(M)$ and let $\lambda, \mu \in C_\alpha$. Then the geodesic spine γ from λ to μ is entirely contained in C_α. Furthermore, if γ is written as $\gamma = \coprod_{i=1}^n [\hat{\nu}_{i-1}, \check{\nu}_i]$ ($\hat{\nu}_0 = \lambda$, $\check{\nu}_n = \mu$) by using a union of intervals, then $\check{\nu}_i$, $\hat{\nu}_i$ are fixed by α for each $1 \le i \le n-1$.*

Proof. Without loss of generality we may assume that $\lambda \le \check{\nu}_1$. We may also assume that $\alpha(\lambda) \le \lambda$, because if $\alpha^{-1}(\lambda) \le \lambda$ we may just consider α^{-1} instead of α.

We first treat the case when $n = 1$ (that is, the case when λ and μ are comparable). Suppose $\nu \notin C_\alpha$ for some $\nu \in [\lambda, \mu]$. Then we have $\nu \in [\alpha(\lambda), \mu]$ and $\alpha(\nu) \in [\alpha(\lambda), \alpha(\mu)]$. Since ν and $\alpha(\nu)$ are incomparable, it follows that μ and $\alpha(\mu)$ are also incomparable, which is a contradiction. Therefore, we have $[\lambda, \mu] \subset C_\alpha$.

Next, we assume $n \ge 2$. We claim that $\alpha(\check{\nu}_1) = \check{\nu}_1$ and $\alpha(\hat{\nu}_1) = \hat{\nu}_1$. Note that

$$[\alpha(\lambda), \lambda] \cup \gamma = [\alpha(\lambda), \check{\nu}_1] \cup \left(\coprod_{i=2}^{n} [\hat{\nu}_{i-1}, \check{\nu}_i] \right)$$

is the geodesic spine from $\alpha(\lambda)$ to μ, and that

$$\alpha(\gamma) = [\alpha(\lambda), \alpha(\check{\nu}_1)] \cup \left(\coprod_{i=2}^{n} [\alpha(\hat{\nu}_{i-1}), \alpha(\check{\nu}_i)] \right)$$

is the geodesic spine from $\alpha(\lambda)$ to $\alpha(\mu)$. Then the reader can work through the several possibilities ($\alpha(\check{\nu}_1) < \check{\nu}_1$, $\alpha(\check{\nu}_1) > \check{\nu}_1$, or $\alpha(\check{\nu}_1)$ and $\check{\nu}_1$ are incomparable) to deduce that any point $\nu \in \coprod_{i=2}^n [\hat{\nu}_{i-1}, \check{\nu}_i]$ is incomparable with $\alpha(\nu)$, contrary to the hypothesis that $\mu \in C_\alpha$. Similarly, if $\alpha(\hat{\nu}_1) \neq \hat{\nu}_1$, we also obtain that $\coprod_{i=2}^n [\hat{\nu}_{i-1}, \check{\nu}_i] \cap C_\alpha = \varnothing$, and therefore $\mu \notin C_\alpha$, which is a contradiction. The claim is proven. Since $\lambda, \check{\nu}_1 \in C_\alpha$, by arguing just as in the case of $n = 1$ we have that $[\lambda, \check{\nu}_1] \subset C_\alpha$. Now, since $\hat{\nu}_1 \in C_\alpha$, the induction on n proves the lemma. □

Lemma 4.3. *Let $\alpha \in \pi_1(M)$ and let B be an α-invariant branch locus. If $\{\nu_t\}_{0<t<\epsilon}$ is an embedded interval such that $B = \lim_{t\to 0} \nu_t$, then there exists $0 < \epsilon' < \epsilon$ such that ν_t is in C_α for any $t \in (0, \epsilon')$.*

Proof. Let $\{\nu_t\}_{0<t<\epsilon}$ be an embedded interval as described above. Then we have $\alpha(B) = \lim_{t\to 0} \alpha(\nu_t)$. Since $B = \alpha(B)$, the two intervals $\{\nu_t\}_{0<t<\epsilon}$ and $\{\alpha(\nu_t)\}_{0<t<\epsilon}$ are both asymptotic to B from the same direction as t tends to 0. This with the fact that L is a 1-manifold implies that the two intervals coincide near B. Thus, the conclusion of the lemma follows. □

Proposition 4.4. *For any $\alpha \in \pi_1(M)$, C_α is connected and open.*

Proof. First, we will show connectedness. Let λ and μ be any points in C_α, and $\gamma = \coprod_{i=1}^n [\hat{\nu}_{i-1}, \check{\nu}_i]$ ($\hat{\nu}_0 = \lambda$, $\check{\nu}_n = \mu$) the geodesic spine from λ to μ. By Lemma 4.2, we have that $\gamma \subset C_\alpha$, and that $\check{\nu}_i$ and $\hat{\nu}_i$ are fixed by α for each $1 \le i \le n-1$. Now let B_i $(1 \le i \le n-1)$ denote the branch locus which contains both $\check{\nu}_i$ and $\hat{\nu}_i$. Then B_i is α-invariant. Therefore, by Lemma 4.3, there is an interval $\{\nu_t^i\}_{0<t<\epsilon} \subset C_\alpha$ such that $B_i = \lim_{t\to 0} \nu_t^i$. It follows that $\check{\nu}_i$ and $\hat{\nu}_i$ can be joined by a path in $\{\nu_t^i\}_{0<t<\epsilon} \subset C_\alpha$, hence that λ and μ can be joined by some path.

Next, we will prove openness. Let λ be any point in C_α. If $\alpha(\lambda) \neq \lambda$ then the open interval bounded by $\alpha^{-1}(\lambda)$ and $\alpha(\lambda)$ is contained in C_α and contains λ. Thus, λ is an interior point of C_α. If $\alpha(\lambda) = \lambda$, take any point $\mu \in L$ with $\lambda < \mu$. Then the interval $[\lambda, \mu]$ is mapped by α orientation preservingly onto the interval $[\lambda, \alpha(\mu)]$. Since L is an oriented 1-manifold, there must exist $\nu \in (\lambda, \mu]$ such that $[\lambda, \nu)$ is contained in $[\lambda, \mu] \cap [\lambda, \alpha(\mu)]$. This implies that $[\lambda, \nu)$ is contained in C_α. Similarly, we can find $\eta < \lambda$ such that $(\eta, \lambda]$ is contained in C_α. Consequently, we have $\lambda \in (\eta, \nu) \subset C_\alpha$, which means λ is an interior point of C_α. This proves the proposition. □

Here we give some definitions. For a geodesic spine $\gamma = \coprod_{i=1}^n [\hat{\nu}_{i-1}, \check{\nu}_i]$, we call n the *length* of γ and denote it by $l(\gamma)$. Let $\lambda, \mu \in L$. As in [Barbot 1998], we set $d(\lambda, \mu) = l(\gamma) - 1$, where γ is the geodesic spine from λ to μ. Moreover, we define the *fundamental axis* A_α of α by $A_\alpha = \{\lambda \in L \mid d(\lambda, \alpha(\lambda)) \text{ is even}\}$. Notice that $C_\alpha = \{\lambda \in L \mid d(\lambda, \alpha(\lambda)) = 0\}$, and therefore, $C_\alpha \subset A_\alpha$.

Proposition 4.5. *Let $\alpha \in \pi_1(M)$. Suppose there is $\lambda \in L$ such that $d(\lambda, \alpha(\lambda))$ is nonzero and even. Then $C_{\alpha^k} = \varnothing$ for any $k > 0$.*

Proof. Let γ be the geodesic spine joining λ to $\alpha(\lambda)$. Since $d(\lambda, \alpha(\lambda))$ is even and since α preserves the orientation on L, there are no nontrivial overlappings in composing k geodesic spines $\gamma, \alpha(\gamma), \dots, \alpha^{k-1}(\gamma)$ successively, and the result $\gamma \cup \alpha(\gamma) \cup \cdots \cup \alpha^{k-1}(\gamma)$ is the geodesic spine from λ to $\alpha^k(\lambda)$. Then we have

$d(\lambda, \alpha^k(\lambda)) = kd(\lambda, \alpha(\lambda))$, and therefore $d(\lambda, \alpha^k(\lambda))$ is nonzero and even. By Corollary 2.20 of [Barbot 1998], α^k fixes no points, and stabilizes no branch loci.

By Proposition 2.10 of the same reference, we have that $A = \bigcup_{i\in\mathbb{Z}} \alpha^i(\gamma)$ is the fundamental axis of α^k. Then A can be expressed as a union of intervals $A = \coprod_{i\in\mathbb{Z}} [\mu_i, \nu_i]$ where ν_i and μ_{i+1} belong to a common branch locus. By [Barbot 1998, Corollary 2.11], there is an integer $m \neq 0$ such that $\alpha^k([\mu_i, \nu_i]) = [\mu_{i+m}, \nu_{i+m}]$. Since $d(\mu, \alpha^k(\mu)) = m \neq 0$ for any $\mu \in A$, it follows that $\mu \notin C_{\alpha^k}$. Therefore, we have $C_{\alpha^k} = \varnothing$, because $C_{\alpha^k} \subset A$. □

Lemma 4.6. *Let $\alpha \in \pi_1(M)$ and $\lambda \in L$ be such that $\lambda \notin C_\alpha$ and that $\lambda \in C_{\alpha^k}$ for some $k > 1$. Let $\gamma = \coprod_{i=1}^n [\hat{\nu}_{i-1}, \check{\nu}_i]$ ($\hat{\nu}_0 = \lambda$, $\check{\nu}_n = \alpha(\lambda)$) be the geodesic spine from λ to $\alpha(\lambda)$. Then $\alpha(\check{\nu}_m) = \hat{\nu}_m$ and $\alpha^k(\check{\nu}_m) = \check{\nu}_m$ where $m = l(\gamma)/2$ (which is an integer by the above proposition).*

Proof. Let γ_j be the geodesic spine from λ to $\alpha^j(\lambda)$, and let δ_0 and δ_1 be the geodesic spines from λ to $\check{\nu}_m$, and from $\hat{\nu}_m$ to $\alpha(\lambda)$, respectively. By reversing the transverse orientation of $\mathscr{F}$ if necessary, we can assume that $\check{\nu}_m$ and $\hat{\nu}_m$ belong to a common positive branch locus.

First, we show that $\check{\nu}_m \notin C_\alpha$. Suppose on the contrary that $\check{\nu}_m \in C_\alpha$. Note that the length of the geodesic spine $\alpha(\delta_0)$ joining $\alpha(\lambda)$ to $\alpha(\check{\nu}_m)$ is $l(\gamma)/2$. So if $\check{\nu}_m$ and $\alpha(\check{\nu}_m)$ are comparable, the intersection $\gamma \cap \alpha(\delta_0)$ must coincide with δ_1 as a set. In particular, $\alpha(\delta_0)$ cannot contain $\check{\nu}_m$. Therefore $\check{\nu}_m > \alpha(\check{\nu}_m)$. See Figure 1. Then $\check{\nu}_m > \alpha^{k-1}(\check{\nu}_m)$, and we have

$$\gamma_k = \delta_0 \cup [\check{\nu}_m, \alpha^{k-1}(\check{\nu}_m)] \cup \alpha^{k-1}(\delta_1).$$

Since γ_k passes through $\alpha^{k-1}(\check{\nu}_m)$ and $\alpha^{k-1}(\hat{\nu}_m)$, it follows that λ and $\alpha^k(\lambda)$ are incomparable, which contradicts the choice of λ.

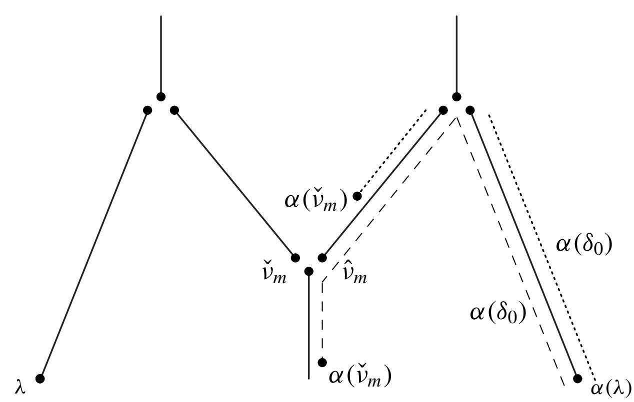


Figure 1. $\alpha(\delta_0)$ is shown as a broken line in the case $\check{\nu}_m \in C_\alpha$, and as a dotted line in the case $\alpha(\check{\nu}_m) \in (\hat{\nu}_m, \check{\nu}_{m+1}]$.

Next, we show that $\alpha(\check{\nu}_m) \notin (\hat{\nu}_m, \check{\nu}_{m+1}]$. Suppose not. Then $\alpha(\check{\nu}_m)$ is in the interval $(\hat{\nu}_m, \check{\nu}_{m+1}]$; that is, the branch locus obtained from the embedded interval $(\hat{\nu}_m, \alpha(\check{\nu}_m))$ contains $\alpha(\hat{\nu}_m)$. It follows that $\hat{\nu}_m$ and $\alpha(\hat{\nu}_m)$ are comparable. See Figure 1. Since we are assuming that $\check{\nu}_m$ and $\hat{\nu}_m$ belong to a common positive branch locus, we have $\hat{\nu}_m < \alpha(\hat{\nu}_m)$. Then $\hat{\nu}_m < \alpha^{k-1}(\hat{\nu}_m)$, and therefore

$$\gamma_k = \delta_0 \cup [\hat{\nu}_m, \alpha^{k-1}(\hat{\nu}_m)] \cup \alpha^{k-1}(\delta_1).$$

Since γ_k passes through $\check{\nu}_m$ and $\hat{\nu}_m$, it follows that λ and $\alpha^k(\lambda)$ are incomparable, which is a contradiction.

Finally, we consider other cases. If $\alpha(\check{\nu}_m) \neq \hat{\nu}_m$, we have

$$l(\alpha^{j+1}(\gamma) - \alpha^j(\gamma)) > l(\gamma)/2 \quad \text{for all } 0 \leq j < k.$$

Therefore, we have

$$1 < l(\gamma_1) < l(\gamma_2) < \cdots < l(\gamma_k) = 1.$$

This contradiction shows that $\alpha(\check{\nu}_m) = \hat{\nu}_m$. In particular, $\alpha(\check{\nu}_m)$ is nonseparated from $\check{\nu}_m$ on the negative side. So $\alpha^k(\check{\nu}_m)$ is also nonseparated from $\check{\nu}_m$ on the negative side. We also have that $\alpha^k(\check{\nu}_m) = \check{\nu}_m$. Otherwise, $\gamma_k = \delta_0 \cup \alpha^k(\delta_0)$, and therefore γ_k passes through $\check{\nu}_m$ and $\alpha^k(\check{\nu}_m)$, which belong the common branch locus. It follows that $\lambda \notin C_{\alpha^k}$, which is a contradiction. □

5. Branch loci and their stabilizers

In this section we focus on a branch locus of the leaf space L. We consider the case where a branch locus is a finite set and clarify the structure of the stabilizer of such a locus.

Lemma 5.1. *Let B be a finite branch locus and let $\alpha \in \mathrm{Stab}(B)$. If $\rho_{\text{univ}}(\alpha)$ has a fixed point in S^1_{univ}, then α fixes B pointwise.*

Proof. Let $\alpha \in \mathrm{Stab}(B)$ be a nontrivial element satisfying the hypothesis of the lemma, and let λ be any point of B. Then, since B is finite, there exists some $k \in \mathbb{N}$ such that $\alpha^k(\lambda) = \lambda$. Notice here that α^k is nontrivial in $\pi_1(M)$, because by tautness of $\mathscr{F}$ and by Novikov's theorem [1965] our manifold M is aspherical and hence has no torsion in $\pi_1(M)$ (see [Hempel 1976, Corollary 9.9]).

Now, let us suppose by contradiction that $\alpha(\lambda) \neq \lambda$. Let $r \in S^1_\infty(\lambda)$ be the turning point from $\alpha(\lambda)$ to λ and let $p \in S^1_\infty(\lambda)$ be one of the two fixed points of α^k which is different from r. Then the special section σ_p in S^1_{univ} is fixed by $\rho_{\text{univ}}(\alpha^k)$. This with the hypothesis that $\rho_{\text{univ}}(\alpha)$ has a fixed point implies that σ_p must be fixed by $\rho_{\text{univ}}(\alpha)$ itself. So we have $\rho_{\text{univ}}(\alpha)(\sigma_p) \in I_p$. On the other hand, since $\alpha(p) \in S^1_\infty(\alpha(\lambda))$, it follows from the definition of turning point that $\rho_{\text{univ}}(\alpha)(\sigma_p) = \sigma_{\alpha(p)} \in I_r$. This is a contradiction because I_p and I_r are disjoint. □

Theorem 5.2. *Let M be a closed oriented 3-manifold, and $\mathcal{F}$ a transversely oriented leafwise hyperbolic taut foliation of M. Suppose $\mathcal{F}$ has a finite branch locus B. If an element of* $\mathrm{Stab}(B)$ *fixes some point of B then it fixes all the points of B.*

Proof. Let λ be the α-fixed point in B, and let $p, q \in S^1_\infty(\lambda)$ be the fixed points of α. Then $\sigma_p, \sigma_q \in S^1_{\mathrm{univ}}$ are fixed by $\rho_{\mathrm{univ}}(\alpha)$. The result follows from Lemma 5.1. □

The next result gives information on topological types of leaves in a finite branch locus.

Theorem 5.3. *Let M be a closed oriented 3-manifold, and $\mathcal{F}$ a transversely oriented leafwise hyperbolic taut foliation of M. Let B be a branch locus of L. Then,*

(1) *if* $\mathrm{Stab}(B)$ *is trivial, $\underline{\lambda}$ is diffeomorphic to a plane, and*

(2) *if B is finite and if* $\mathrm{Stab}(B)$ *is nontrivial, $\underline{\lambda}$ is diffeomorphic to a cylinder,*

for any $\lambda \in B$.

Proof. Let $\lambda \in B$. Since $\mathcal{F}$ is taut, by Novikov's theorem the inclusion map of each leaf of $\mathcal{F}$ into M is π_1-injective. So, if $\underline{\lambda}$ is not a plane, there exists a nontrivial element $\alpha \in \pi_1(M)$ such that $\alpha(\lambda) = \lambda$. This α must belong to $\mathrm{Stab}(B)$, showing the first statement of the theorem.

To prove the second statement, suppose that B is finite and that $\mathrm{Stab}(B)$ is nontrivial. Then, we can first observe that $\underline{\lambda}$ is not a plane. In fact, let γ be any nontrivial element of $\mathrm{Stab}(B)$. Since B is finite, $\gamma^n(\lambda) = \lambda$ for some $n \in \mathbb{N}$. By the same argument as in the proof of Lemma 5.1 we see that γ^n nontrivial in $\pi_1(M)$. This shows the observation.

Now, by way of contradiction, let us assume $\underline{\lambda}$ is not a cylinder, either. Then, again by π_1-injectivity of the inclusion $\underline{\lambda} \to M$, we can find elements $\alpha, \beta \in \pi_1(M)$ generating a free subgroup of rank 2 such that $\alpha(\lambda) = \beta(\lambda) = \lambda$. These two elements are hyperbolic as isometries of λ and having no common fixed point on $S^1_\infty(\lambda)$. Let μ be another leaf in B, and let $r \in S^1_\infty(\lambda)$ be the turning point from μ to λ. By exchanging α and β if necessary, we may assume $\alpha(r) \neq r$. Then, $\alpha^k(r) \neq \alpha^l(r)$ for any $k \neq l \in \mathbb{Z}$. Pick a point $s \in S^1_\infty(\mu)$ and consider the special section σ_s at s. Then, $\rho_{\mathrm{univ}}(\alpha^k)(\sigma_s) = \sigma_{\alpha^k(s)}$ is the special section at $\alpha^k(s)$. Since $\sigma_{\alpha^k(s)}(\lambda) = \phi_\lambda \circ \rho_{\mathrm{univ}}(\alpha^k)(\sigma_s) = \alpha^k \circ \phi_\lambda(\sigma_s) = \alpha^k(r)$, it follows that $\alpha^k(r)$ is the turning point from $\alpha^k(\mu)$ to λ. In particular, $\alpha^k(\mu) \neq \alpha^l(\mu)$ for $k \neq l$; hence, B contains infinitely many elements $\alpha^k(\mu)$, $k \in \mathbb{Z}$, contradicting the finiteness of B. □

Remark 5.4. The author does not know whether or not there exists a branch locus which has a trivial stabilizer.

Proposition 5.5. *Let $B = \{\lambda_1, \dots, \lambda_n\}$ be a finite branch locus which has a nontrivial stabilizer and let $r_i^j \in S^1_\infty(\lambda_i)$ be the turning point from λ_j to λ_i. Then there exists $1 \leq k \leq n$ such that the set of turning points $\{r_k^j \mid j \neq k\}$ is a single point in $S^1_\infty(\lambda_k)$.*

Proof. By Theorem 5.3, each λ_i is a cylindrical leaf. Let γ be a generator of $\mathrm{Stab}(\lambda_1) = \{\alpha \in \pi_1(M) \mid \alpha(\lambda_1) = \lambda_1\}$. By Theorem 5.2, γ fixes all points in B. Let $p_i, q_i \in S^1_\infty(\lambda_i)$ be the fixed points of γ acting on $S^1_\infty(\lambda_i)$. Note that $r_i^j \in \{p_i, q_i\}$ for any i, j. Otherwise, B cannot be finite by the same argument as in the proof of Theorem 5.3.

We suppose that $\{r_1^j \mid j \neq 1\} = \{p_1, q_1\}$. After renumbering the indices if necessary, we can assume that $r_1^j = p_1$ for $2 \leq j < n_1$ and $r_1^j = q_1$ for $n_1 \leq j \leq n$, where $3 \leq n_1 \leq n$. Then, we claim that $r_{n_1}^j = r_{n_1}^1$ for $1 \leq j < n_1$. In fact, let $2 \leq j < n_1$, and take 4 points x, y, z, w as follows: x, y are in $S^1_\infty(\lambda_1) - \{p_1, q_1\}$ such that the 4-tuple (p_1, x, q_1, y) is circularly ordered, $z \in S^1_\infty(\lambda_j)$ and $w \in S^1_\infty(\lambda_{n_1}) - \{r_{n_1}^1\}$. Then, $\sigma_z \in I_{p_1}$, $\sigma_w \in I_{q_1}$ and the 4-tuple $(I_{p_1}, \sigma_x, I_{q_1}, \sigma_y)$ is circularly ordered in S^1_{univ}. Furthermore, $\sigma_x, \sigma_y \in I_{r_{n_1}^1}$ and $\sigma_w \notin I_{r_{n_1}^1}$. It follows that $\sigma_z \in I_{r_{n_1}^1}$; that is, $r_{n_1}^1$ is the turning point from λ_j to λ_{n_1}. This proves the claim.

Now, if $\{r_{n_1}^j \mid j \neq n_1\} = \{r_{n_1}^1\}$ we can put $k = n_1$. Otherwise, by renumbering the indices again, we can assume that $r_{n_1}^j = r_{n_1}^1 = p_{n_1}$ for $1 \leq j < n_2$ $(j \neq n_1)$, and $r_{n_1}^j = q_{n_1}$ for $n_2 \leq j \leq n$, where $n_1 < n_2 \leq n$. Similarly, we have $r_{n_2}^j = r_{n_2}^1$ for $1 \leq j < n_2$. Since B is finite, we can find a desired k after repeating this process finitely many times. □

Theorem 5.6. *Let M be a closed oriented 3-manifold, and $\mathcal{F}$ a transversely oriented leafwise hyperbolic taut foliation of M. Let B be a finite branch locus of L with a nontrivial stabilizer. Then the stabilizer* $\mathrm{Stab}(B)$ *is isomorphic to $\mathbb{Z}$.*

Proof. Let $B = \{\lambda_1, \ldots, \lambda_n\}$, and let $r_i^j \in S^1_\infty(\lambda_i)$ be the turning point from λ_j to λ_i for $i \neq j$. By Proposition 5.5, without loss of generality we can assume that $\{r_1^j \mid j \neq 1\}$ is a single point. Let γ be a generator of $\mathrm{Stab}(\lambda_1)$.

Now, if $\mathrm{Stab}(B)$ acts on B trivially, then each $\alpha \in \mathrm{Stab}(B)$ fixes λ_1. It follows that there exists an integer k such that $\alpha = \gamma^k$; that is, γ is a generator of $\mathrm{Stab}(B)$.

So we assume that $\mathrm{Stab}(B)$ acts on B nontrivially. By Theorem 5.2, γ fixes every point λ_i in B. Let $p_i, q_i \in S^1_\infty(\lambda_i)$ be the fixed points of γ acting on $S^1_\infty(\lambda_i)$. Put $\mathrm{Stab}(B)(\lambda_1) = \{\alpha(\lambda_1) \mid \alpha \in \mathrm{Stab}(B)\} = \{\lambda_1, \ldots, \lambda_m\}$ where $1 < m \leq n$. Since the natural action preserves the set of turning points, $\{r_i^j \mid 1 \leq j \leq n, j \neq i\}$ is also a single point for any $i \leq m$. Let us denote this single point by p_i. It follows that the subset $\{\sigma_{p_i} \mid 1 \leq i \leq m\}$ of S^1_{univ} is kept invariant by homeomorphisms $\rho_{\mathrm{univ}}(\alpha)$ for $\alpha \in \mathrm{Stab}(B)$. After renumbering indices if necessary, we can assume that the m-tuple $(\sigma_{p_1}, \ldots, \sigma_{p_m})$ is circularly ordered in S^1_{univ}. Let $\beta \in \mathrm{Stab}(B)$ be such that $\rho_{\mathrm{univ}}(\beta)(\sigma_{p_1}) = \sigma_{p_2}$; that is, $\beta(\lambda_1) = \lambda_2$. Since $\rho_{\mathrm{univ}}(\beta)$ preserves the circular order on S^1_{univ}, we have $\beta(\lambda_i) = \lambda_{i+1}$ where the indices i are taken modulo m.

Now, since $\beta\gamma\beta^{-1}(\lambda_1) = \lambda_1$, it follows that $\beta\gamma\beta^{-1} = \gamma^k$ for some $k \neq 0$. Moreover, there is $l \neq 0$ such that $\beta^m = \gamma^l$. It follows that $\beta^{km} = \gamma^{kl} = \beta\gamma^l\beta^{-1} = \beta^m$; that is, $\beta^{(k-1)m}$ is trivial. If $k \neq 1$, β is a torsion element in $\pi_1(M)$, which is a contradiction. Therefore $k = 1$ and we have that γ and β commute. Since $\pi_1(M)$ is torsion-free, the subgroup $\langle \gamma, \beta \mid \gamma^l \beta^{-m} \rangle$ must be isomorphic to $\mathbb{Z}$. It follows that there is $\delta \in \pi_1(M)$ such that $\gamma = \delta^i$ and $\beta = \delta^j$ where $i \neq 0$ and $j \neq 0$. Let α be any element in $\mathrm{Stab}(B)$. Then $\alpha(\lambda_1) = \lambda_i$ for some $1 \leq i \leq m$. By the choice of γ and β, we have that α can be represented as a word in γ and β, and hence in δ. It follows that $\mathrm{Stab}(B)$ is isomorphic to $\mathbb{Z}$. □

We say that $\alpha \in \pi_1(M)$ is *infinitely divisible* if for any integer ℓ, there are $k > \ell$ and $\beta \in \pi_1(M)$ such that $\alpha = \beta^k$.

Theorem 5.7. *Let M be a closed oriented 3-manifold, and $\mathcal{F}$ a transversely oriented leafwise hyperbolic taut foliation of M. Let B be a finite branch locus of L such that* $\mathrm{Stab}(B)$ *acts on B nontrivially. Then a generator of* $\mathrm{Stab}(B)$ $(\cong \mathbb{Z})$ *is indivisible.*

Proof. Let $B = \{\lambda_1, \ldots, \lambda_n\}$. By Theorem 5.6, $\mathrm{Stab}(B)$ is generated by some single element α. We assume by contradiction that α is divisible. Since M is aspherical (as was noted in the proof of Lemma 5.1), $\pi_1(M)$ has no infinitely divisible elements (see [Friedl 2011, Theorem 4.1]). Hence, there exists an indivisible element β in $\pi_1(M)$ such that $\alpha = \beta^k$ for some $k > 1$.

Note that since $\beta \notin \mathrm{Stab}(B)$, the points $\lambda_i \in B$ and $\beta(\lambda_i) \in \beta(B)$ are distinct for any i. Moreover, we see that they are incomparable for any i. In fact, if λ_i and $\beta(\lambda_i)$ were comparable, say, $\lambda_i < \beta(\lambda_i)$, then $\lambda_i < \beta^k(\lambda_i) = \alpha(\lambda_i)$, contradicting the assumption that $\alpha \in \mathrm{Stab}(B)$.

Let $\{\nu_t\}_{0<t<\epsilon}$ be an embedded interval such that $B = \lim_{t\to 0} \nu_t$. Since B is α-invariant, it follows from Lemma 4.3 that there is some $\nu \in \{\nu_t\}_{0<t<\epsilon}$ such that $\nu \in C_\alpha = C_{\beta^k}$. We can (and do) take such ν so that ν also satisfies that $\nu \notin C_\beta$. Let $\coprod_{i=1}^{l} [\hat{\nu}_{i-1}, \check{\nu}_i]$ $(l > 1)$ be the geodesic spine joining ν to $\beta(\nu)$. By the choice of ν and by Lemma 4.6, we have $\beta^k(\check{\nu}_m) = \check{\nu}_m$ where $m = l/2$. It follows that $\rho_{\mathrm{univ}}(\beta^k)$ has a fixed point in S^1_{univ}. By Lemma 5.1, $\beta^k = \alpha$ fixes all points in B, which is a contradiction to the hypothesis, as α generates $\mathrm{Stab}(B)$ and $\mathrm{Stab}(B)$ acts on B nontrivially. □

Remark 5.8. The author does not know whether or not there is a finite branch locus B such that $\mathrm{Stab}(B)$ acts on B trivially and is generated by a divisible element.

We will give an example of a tautly foliated compact 3-manifold admitting a finite branch locus whose stabilizer acts on the locus nontrivially. We remark that a recipe how to construct such a locus has already been provided in [Calegari and Dunfield 2003, Example 3.7], and our construction follows it.

Example 5.9. Let $P = D^2 \setminus (E_1 \cup E_2)$ be the unit disk in $\mathbb{C}$ with two open disks removed, where E_1, E_2 are disks centered in $-\frac{1}{2}, \frac{1}{2}$ with radius $\frac{1}{4}$ respectively. Put $S_0 = \partial D^2$, $S_1 = \partial E_1$ and $S_2 = \partial E_2$. On P we consider a standard singular foliation $\mathcal{G}$ (see Figure 2) satisfying the following properties:

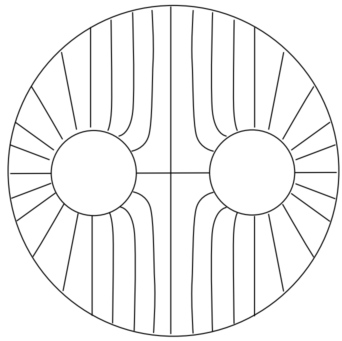


Figure 2. A singular foliation $\mathcal{G}$ of P.

(1) $\mathcal{G}$ has the origin as its unique singular point, which is of saddle type.

(2) $\mathcal{G}$ is transverse to ∂P.

(3) All leaves of $\mathcal{G}$ (except the 4 separatrices) are compact.

(4) $\mathcal{G}$ is symmetric with respect to both the x-axis and y-axis.

(5) The holonomy maps $h_1 : S_1 \setminus \left\{-\frac{1}{4}\right\} \to S_0$ and $h_2 : S_2 \setminus \left\{\frac{1}{4}\right\} \to S_0$ of $\mathcal{G}$ are given by

$$\begin{aligned} h_1\left(\tfrac{1}{4}e^{2\pi i\theta} - \tfrac{1}{2}\right) &= e^{\pi i(\theta+\frac{1}{2})} && \text{if } 0 < \theta < 1, \\ h_2\left(\tfrac{1}{4}e^{2\pi i\theta} + \tfrac{1}{2}\right) &= e^{\pi i\theta} && \text{if } -\tfrac{1}{2} < \theta < \tfrac{1}{2}. \end{aligned}$$

Let $(P', \mathcal{G}')$ be a copy of $(P, \mathcal{G})$, and let $c : P \to P'$ be the map induced by the identity. We construct a double $\Sigma = P \cup P'$ using diffeomorphisms $g_i : S_i \to c(S_i)$ (for $i = 0, 1, 2$) to glue S_i to $c(S_i)$, where $c^{-1} \circ g_0$ is given by

$$c^{-1} \circ g_0(e^{2\pi i\theta}) = e^{2\pi i(\theta+\alpha)}$$

for some $\alpha \in \mathbb{R} - \mathbb{Q}$, and $c^{-1} \circ g_i$ is the antipodal map of S_i for $i = 1, 2$. Since h_1, h_2 preserve rational (with respect to θ) points in S_1, S_2 and S_0, it follows that $\mathcal{G}$ and $\mathcal{G}'$ induce a singular foliation $\mathcal{G}''$ of Σ with two saddle singularities and without any saddle connection. By construction, the homeomorphism ρ of Σ which is defined to be the rotation by π in both P and P' preserves $\mathcal{G}''$.

Fix a hyperbolic structure on Σ. Then each leaf of $\mathcal{G}''$ except the singular points and the separatrices is isotopic to a unique embedded geodesic, and the closure of the union of these geodesics constitutes a geodesic lamination, say, λ, on Σ. Note that the two complementary regions Q_1 and Q_2 to λ are ideal open squares. There exists a λ-preserving homeomorphism ψ of Σ isotopic to ρ. Let M be the mapping torus of ψ, that is, $M = \Sigma \times [0, 1]/(s, 1) \sim (\psi(s), 0)$. Then λ induces a surface lamination Λ of M whose complementary regions R_i are Q_i-bundles over S^1 for $i = 1, 2$. Denote by $p_i : R_i \to S^1$ the bundle projection.

Now we extend Λ to a foliation $\mathscr{F}$ of M by filling R_i (for $i = 1, 2$) with leaves diffeomorphic to Q_i as follows. Denote the boundary components of R_i by C_{i1} and C_{i2}, which are open cylinders. Let γ_i be an oriented loop in R_i such that $p_i|\gamma_i$ is a diffeomorphism onto S^1. Then the composition $\gamma_i^2 = \gamma_i * \gamma_i$ is freely homotopic to a leaf loop γ_{ij} of C_{ij} which is a generator of $\pi_1(C_{ij})$. We foliate R_i as a product by leaves isotopic to the fibers Q_i so that the holonomy along γ_{i1} is contracting and the holonomy along γ_{i2} is expanding. Then the resulting foliation $\mathscr{F}$ is taut and has two-sided branching, and each end of a lift of γ_i to $\widetilde{M}$ gives a branch locus consisting of two points. Let α_i be an element in $\pi_1(M)$ whose conjugacy class corresponds with the free homotopy class of γ_i. Then α_i belongs to the stabilizer of some branch locus and acts on the locus nontrivially, as desired.

6. Loops and actions

Given a loop in a tautly foliated manifold $(M, \mathscr{F})$, it is natural to ask whether it is transversable, or tangentiable, to $\mathscr{F}$. In this section, we observe that these properties of loops are expressed completely in the language of the natural action. Furthermore, we consider relations between such properties and the branching phenomenon of $\widetilde{\mathscr{F}}$.

We do not need to assume leafwise hyperbolicity in the first two propositions below.

Proposition 6.1. *Let γ be a loop in M, and α an element in $\pi_1(M, p)$ whose conjugacy class corresponds with the free homotopy class of γ. Then, γ is tangentiable if and only if the action of α on L has a fixed point. Similarly, γ is positively (resp. negatively) transversable if and only if there is a point λ in L such that $\alpha(\lambda) > \lambda$ (resp. $\alpha(\lambda) < \lambda$).*

Proof. Let λ be a leaf of $\widetilde{\mathscr{F}}$ and suppose that the deck transformation α leaves λ invariant. Take any point x in λ and join x to $\alpha(x)$ by a path in λ. Then it projects down to a leaf loop in M freely homotopic to α. Conversely, suppose γ is a leaf loop in M. Join the base point p to a point of γ by a path c. Then, the loop $c * \gamma * c^{-1}$ represents an element of $\pi_1(M, p)$ conjugate to α. Obviously it has a fixed point, hence so does α. The claim on transversability is also shown easily. □

We remark here that $\pi_1(M)$ can have an element which is neither tangentiable nor transversable. Such an element exists if and only if $\mathscr{F}$ has two-sided branching. This fact is due to Barbot, and also follows from Lemma 4.1 and Proposition 4.5. (Notice that if $\mathscr{F}$ has two-sided branching, there are $\lambda, \mu \in L$ such that $d(\lambda, \mu)$ is nonzero and even. Then by the tautness of $\mathscr{F}$, we can find $\nu \in L$ which satisfies $d(\mu, \nu) = 0$, $d(\lambda, \nu) = d(\lambda, \mu)$, and $\alpha(\lambda) = \nu$ for some $\alpha \in \pi_1(M)$.)

Proposition 6.2. *Let $\alpha \in \pi_1(M)$. Suppose there are points $\lambda, \mu \in L$ such that $\alpha(\lambda) > \lambda$ and $\alpha(\mu) < \mu$. Then there exists a point $\nu \in L$ such that $\alpha(\nu) = \nu$.*

Moreover, if λ and μ are incomparable, then such ν can be found in some branch locus.

Proof. If λ and μ are comparable, then the conclusion follows immediately from the intermediate value theorem. If λ and μ are incomparable, then the conclusion follows from Lemma 4.2. □

This proposition means that if a loop in M is both positively and negatively transversable to $\mathcal{F}$, then it is tangentiable to $\mathcal{F}$.

In the following we assume leafwise hyperbolicity and observe that tangentiability and/or transversability of loops in M and the infiniteness of branch loci are closely related.

Theorem 6.3. *Let M be a closed oriented 3-manifold, and $\mathcal{F}$ a transversely oriented leafwise hyperbolic taut foliation of M with one-sided branching. Suppose that there is a noncontractible leaf loop γ in M which is not transversable. Then every branch locus of L is an infinite set.*

Proof. Suppose that there exists a finite (say, positive) branch locus $B = \{\lambda_1, \dots, \lambda_n\}$. Let α be an element in $\pi_1(M)$ whose conjugacy class corresponds with the free homotopy class of γ. By Proposition 6.1, α has a fixed point in L, and for each $\mu \in L$ if μ is not fixed by α then $\mu \notin C_\alpha$. Let ν be a fixed point of α. By Lemma 4.1, for every η with $\eta \leq \nu$, we have $\eta \in C_\alpha$, and therefore $\alpha(\eta) = \eta$. By replacing B with $\beta(B)$ for some $\beta \in \pi_1(M)$ if necessary, we can assume that $\lambda_1 \leq \nu$ and therefore $\alpha(\lambda_1) = \lambda_1$. This implies in particular that B is α-invariant. Since B is finite, by Theorem 5.2 we have $\alpha(\lambda_i) = \lambda_i$ for any $1 \leq i \leq n$. By Lemma 3.1, α must be trivial, which is a contradiction. □

Corollary 6.4. *Let M be a closed oriented 3-manifold, and $\mathcal{F}$ a transversely oriented leafwise hyperbolic taut foliation of M with every leaf dense. Suppose that there is noncontractible leaf loop γ in M which is not transversable. Then every branch locus of L is an infinite set.*

Proof. Suppose there is a finite branch locus B. Let $\alpha \in \pi_1(M)$ be as in the proof of the preceding theorem. By Proposition 4.4, there is an embedded open interval $I \subset L$ such that I is contained in C_α. Since every leaf of $\mathcal{F}$ is dense, there is $\beta \in \pi_1(M)$ such that $\beta(B) \cap I \neq \varnothing$. Then the same argument as in Theorem 6.3 shows the conclusion. □

Theorem 6.5. *Let M be a closed oriented 3-manifold, and $\mathcal{F}$ a transversely oriented leafwise hyperbolic taut foliation of M with branching. Suppose that there is a noncontractible leaf loop γ in M which is not transversable. Then L has an infinite branch locus.*

Proof. Let α be as in Theorem 6.3. Then α has a fixed point $\nu \in L$, and for each $\mu \in L$ if μ is not fixed by α then $\mu \notin C_\alpha$. Without loss of generality, we assume that $\mathcal{F}$ has a positive branch locus.

We claim that there exist some $\nu' > \nu$ such that ν' and $\alpha(\nu')$ are incomparable. Put $L' = \{\mu \mid \mu > \nu\}$. Notice that $\alpha(L') = L'$. Then we can observe that L' is a submanifold of L with one-sided branching in the positive direction and contains at least one branch locus. For, by the tautness of $\mathcal{F}$ we can find a positive branch locus B' in L and $\beta \in \pi_1(M)$ such that $\beta(\nu)$ is a common lower bound of all points in B'; that is, $\beta^{-1}(B') \subset L'$. If α fixes all leaves in L', then by applying Lemma 3.1 to a branch locus in L' we obtain that α is trivial in $\pi_1(M)$, which contradicts the hypothesis that α is represented by a noncontractible loop. Therefore, there exists some $\nu' \in L$ which is not fixed by α. Since such ν' does not belong to C_α, the claim is shown.

Since $\nu < \nu'$ and $\alpha(\nu) = \nu$, it follows that ν is a common lower bound for ν' and $\alpha(\nu')$. Thus, the fact that ν' and $\alpha(\nu')$ are incomparable implies that there is a unique $\lambda \in (\nu, \nu']$ such that $\mu \in [\nu, \nu']$ is fixed by α if and only if $\mu \in [\nu, \lambda)$. Evidently, λ belongs to some α-invariant branch locus, say, B. Also note that $\rho_{\text{univ}}(\alpha)$ has a fixed point because α fixes a point in L. We now show B is infinite. Suppose not. Then, by Lemma 5.1, all leaves in B are α-fixed, contradicting $\alpha(\lambda) \neq \lambda$. □

Acknowledgements

I thank T. Inaba for encouragement and conversations during the research. I am also grateful to the referee for careful reading of the manuscript and helpful suggestions.

References

[Barbot 1996] T. Barbot, "Flots d'Anosov sur les variétés graphées au sens de Waldhausen", *Ann. Inst. Fourier* (*Grenoble*) **46**:5 (1996), 1451–1517. MR 97j:57031 Zbl 0861.58028

[Barbot 1998] T. Barbot, "Actions de groupes sur les 1-variétés non séparées et feuilletages de codimension un", *Ann. Fac. Sci. Toulouse Math.* (6) **7**:4 (1998), 559–597. MR 2000e:57045 Zbl 0932.57027

[Calegari 2007] D. Calegari, *Foliations and the geometry of* 3*-manifolds*, Oxford University Press, 2007. MR 2008k:57048 Zbl 1118.57002

[Calegari and Dunfield 2003] D. Calegari and N. M. Dunfield, "Laminations and groups of homeomorphisms of the circle", *Invent. Math.* **152**:1 (2003), 149–204. MR 2005a:57013 Zbl 1025.57018

[Candel 1993] A. Candel, "Uniformization of surface laminations", *Ann. Sci. École Norm. Sup.* (4) **26**:4 (1993), 489–516. MR 94f:57025 Zbl 0785.57009

[Fenley 1998] S. R. Fenley, "The structure of branching in Anosov flows of 3-manifolds", *Comment. Math. Helv.* **73**:2 (1998), 259–297. MR 99a:58123 Zbl 0999.37008

[Fenley 2003] S. R. Fenley, "Pseudo-Anosov flows and incompressible tori", *Geom. Dedicata* **99** (2003), 61–102. MR 2004f:57035 Zbl 1047.37017

[Friedl 2011] S. Friedl, "Centralizers in 3-manifold groups", *RIMS Kōkyūroku* **1747** (2011), 23–34. arXiv 1205.2386

[Hempel 1976] J. Hempel, 3-*Manifolds*, Ann. of Math. Studies **86**, Princeton University Press, 1976. MR 54 #3702 Zbl 0345.57001

[Novikov 1965] S. P. Novikov, "Топология слоений", *Trudy Moskov. Mat. Obšč.* **14** (1965), 248–278. Translated as "Topology of foliations" in *Trans. Mosc. Math. Soc.* **14** (1965), 268–304. MR 34 #824 Zbl 0247.57006

[Roberts et al. 2003] R. Roberts, J. Shareshian, and M. Stein, "Infinitely many hyperbolic 3-manifolds which contain no Reebless foliation", *J. Amer. Math. Soc.* **16**:3 (2003), 639–679. MR 2004e:57023 Zbl 1012.57022

[Shields 2002] S. Shields, "Branching in the universal cover of taut foliations", *Pacific J. Math.* **203**:1 (2002), 235–256. MR 2003m:57061 Zbl 1063.57016

[Thurston 1998] W. P. Thurston, "Three-manifolds, foliations, and circles, II", Unfinished manuscript, 1998.

Received March 15, 2012. Revised July 16, 2013.

YOSUKE KANO
DEPARTMENT OF MATHEMATICS AND INFORMATICS
GRADUATE SCHOOL OF SCIENCE
CHIBA UNIVERSITY
JAPAN

ykano@g.math.s.chiba-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.417

A NEW MONOTONE QUANTITY ALONG THE INVERSE MEAN CURVATURE FLOW IN $\mathbb{R}^n$

KWOK-KUN KWONG AND PENGZI MIAO

We find a new monotone increasing quantity along smooth solutions to the inverse mean curvature flow in $\mathbb{R}^n$. As an application, we derive a sharp geometric inequality for mean convex, star-shaped hypersurfaces which relates the volume enclosed by a hypersurface to a weighted total mean curvature of the hypersurface.

1. Statement of the result

Monotone quantities along hypersurfaces evolving under the inverse mean flow have many applications in geometry and relativity. Huisken and Ilmanen [2001] applied the monotone increasing property of Hawking mass to give a proof of the Riemannian Penrose inequality. Brendle, Hung and Wang [Brendle et al. 2012] discovered a monotone decreasing quantity along the inverse mean curvature flow in anti-de Sitter–Schwarzschild manifolds and used it to establish a Minkowski-type inequality for star-shaped hypersurfaces.

In this note, we provide a new monotone increasing quantity along smooth solutions to the inverse mean curvature flow in $\mathbb{R}^n$:

Theorem 1. *Let Σ be a smooth, closed, embedded hypersurface with positive mean curvature in $\mathbb{R}^n$. Let I be an open interval and $X : \Sigma \times I \to \mathbb{R}^n$ be a smooth map satisfying*

$$\frac{\partial X}{\partial t} = \frac{1}{H}\nu, \tag{1-1}$$

where H is the mean curvature of the surface $\Sigma_t = X(\Sigma, t)$ and ν is the outward unit normal vector to Σ_t. Let Ω_t be the bounded region enclosed by Σ_t and $r = r(x)$ be the distance from x to a fixed point O. Then the function

$$Q(t) = e^{-\frac{n-2}{n-1}t}\left(n \operatorname{Vol}(\Omega_t) - \frac{1}{n-1}\int_{\Sigma_t} r^2 H \, d\mu\right) \tag{1-2}$$

MSC2010: primary 53C44; secondary 53A07.
Keywords: inverse mean curvature flow.

is monotone increasing and $Q(t)$ *is a constant function if and only if* Σ_t *is a round sphere for each* t*. Here* $\mathrm{Vol}(\Omega)$ *denotes the volume of a bounded region* Ω *and* $d\mu$ *denotes the volume form on a hypersurface.*

As an application, we derive a sharp inequality for star-shaped hypersurfaces in $\mathbb{R}^n$ which relates the volume enclosed by a hypersurface to an r^2-weighted total mean curvature of the hypersurface.

Theorem 2. *Let* Σ *be a smooth, star-shaped hypersurface with positive mean curvature in* $\mathbb{R}^n$*. Then*

$$n\,\mathrm{Vol}(\Omega) \le \frac{1}{n-1}\int_\Sigma r^2 H\, d\mu \tag{1-3}$$

where $\mathrm{Vol}(\Omega)$ *is the volume of the region* Ω *enclosed by* Σ, r *is the distance to a fixed point* O *and* H *is the mean curvature of* Σ*. Furthermore, equality in* (1-3) *holds if and only if* Σ *is a sphere centered at* O*.*

We give some remarks about Theorem 1 and Theorem 2. The discovery of the monotonicity of $Q(t)$ in Theorem 1 is motivated by [Brendle et al. 2012, Section 5]. To prove Theorem 1, we also need a result of Ros, proved using Reilly's formula. Once we know that $Q(t)$ is monotone increasing, to prove Theorem 2, it may be tempting to ask whether $\lim_{t\to\infty} Q(t) = 0$? We do not know if this is true because both $\mathrm{Vol}(\Omega_t)$ and $\int_{\Sigma_t} r^2 H\, d\mu$ grow like $\exp\bigl(\frac{n}{n-1}t\bigr)$ when $\{\Sigma_t\}$ are spheres, while there is only a factor of $\exp\bigl(-\frac{n-2}{n-1}t\bigr)$ in (1-2). Instead, we take an alternate approach by first proving Theorem 2 for a convex hypersurface Σ. The proof in that case again makes use of Reilly's formula. When Σ is merely assumed to be mean convex and star-shaped, we prove Theorem 2 by reducing it to the convex case using solutions to the inverse mean curvature flow provided by [Gerhardt 1990] and [Urbas 1990].

2. Proof of the theorems

Given a compact Riemannian manifold (Ω, g) with boundary Σ, Reilly's formula [1977] asserts that

$$\begin{aligned}\int_\Omega |\nabla^2 u|^2 + \langle \nabla(\Delta u), \nabla u\rangle + \mathrm{Ric}(\nabla u, \nabla u)\, dV \\ = \int_\Sigma (\Delta u)\frac{\partial u}{\partial \nu} - II\bigl(\nabla^\Sigma u, \nabla^\Sigma u\bigr) - 2(\Delta_\Sigma u)\frac{\partial u}{\partial \nu} - H\left(\frac{\partial u}{\partial \nu}\right)^2 d\mu.\end{aligned} \tag{2-1}$$

Here u is a smooth function on Ω; ∇^2, Δ and ∇ denote the Hessian, the Laplacian and the gradient on Ω; Δ_Σ and ∇^Σ denote the Laplacian and the gradient on Σ; ν is the unit outward normal vector to Σ; II and H are the second fundamental form and the mean curvature of Σ with respect to ν; and Ric is the Ricci curvature of g.

To prove Theorem 1, we need a result of [Ros 1987], which was proved by choosing $\Delta u = 1$ on Ω and $u = 0$ at Σ in the above Reilly's formula.

Theorem 3 [Ros 1987]. *Let (Ω, g) be an n-dimensional compact Riemannian manifold with nonnegative Ricci curvature with boundary Σ. Suppose Σ has positive mean curvature H; then*

$$n\,\mathrm{Vol}(\Omega) \le (n-1)\int_\Sigma \frac{1}{H}\,d\mu \tag{2-2}$$

and equality holds if and only if (Ω, g) is isometric to a round ball in $\mathbb{R}^n$.

Proof of Theorem 1. We use $'$ to denote differentiation with respect to t. Some basic formulas along the inverse mean curvature flow (1-1) in $\mathbb{R}^n$ are

$$H' = -\Delta_{\Sigma_t}\left(\frac{1}{H}\right) - \frac{|II|^2}{H}, \quad d\mu' = d\mu, \quad \mathrm{Vol}(\Omega_t)' = \int_{\Sigma_t}\frac{1}{H}\,d\mu. \tag{2-3}$$

Let $u = r^2$. Then u satisfies

$$\nabla^2 u = 2g \quad \text{and} \quad \Delta u = 2n, \tag{2-4}$$

where g is the Euclidean metric. Now

$$\left(\int_{\Sigma_t} uH\,d\mu\right)' = \int_{\Sigma_t}(u'H + uH' + uH)\,d\mu. \tag{2-5}$$

Let $\langle\cdot,\cdot\rangle$ be the Euclidean inner product. By (2-3), (2-4) and the divergence theorem, we have

$$\int_{\Sigma_t} u'H\,d\mu = \int_{\Sigma_t}\left\langle\nabla u, \frac{1}{H}\nu\right\rangle H\,d\mu = \int_{\Omega_t}\Delta u\,dV = 2n\,\mathrm{Vol}(\Omega_t). \tag{2-6}$$

By (2-4), we also have

$$\Delta_{\Sigma_t} u = \Delta u - H\frac{\partial u}{\partial\nu} - \nabla^2 u(\nu,\nu) = 2(n-1) - H\frac{\partial u}{\partial\nu},$$

which together with (2-3) and (2-4) implies

$$\begin{aligned}\int_{\Sigma_t} uH'\,d\mu &= \int_{\Sigma_t}\left(-\frac{\Delta_{\Sigma_t}u}{H} - \frac{u|II|^2}{H}\right)d\mu \\ &= \int_{\Sigma_t}\left(-\frac{2(n-1)}{H} + \frac{\partial u}{\partial\nu} - \frac{u|II|^2}{H}\right)d\mu \\ &= -\int_{\Sigma_t}\frac{2(n-1)}{H}\,d\mu + 2n\,\mathrm{Vol}(\Omega_t) - \int_{\Sigma_t}\frac{u|II|^2}{H}\,d\mu.\end{aligned} \tag{2-7}$$

Substituting (2-6) and (2-7) into (2-5) yields

$$\begin{aligned}
\text{(2-8)}\quad \Big(\int_{\Sigma_t} uH\,d\mu\Big)' &= 4n\,\mathrm{Vol}(\Omega_t) + \int_{\Sigma_t}\Big(-\frac{2(n-1)}{H} - \frac{u|II|^2}{H} + uH\Big)\,d\mu \\
&\le 4n\,\mathrm{Vol}(\Omega_t) + \int_{\Sigma_t}\Big(-\frac{2(n-1)}{H} - \frac{uH}{n-1} + uH\Big)\,d\mu \\
&= 4n\,\mathrm{Vol}(\Omega_t) + \int_{\Sigma_t}\Big(-\frac{2(n-1)}{H} + \frac{n-2}{n-1}uH\Big)\,d\mu \\
&\le 4n\,\mathrm{Vol}(\Omega_t) - 2n\,\mathrm{Vol}(\Omega_t) + \frac{n-2}{n-1}\int_{\Sigma_t} uH\,d\mu \\
&= 2n\,\mathrm{Vol}(\Omega_t) + \frac{n-2}{n-1}\int_{\Sigma_t} uH\,d\mu,
\end{aligned}$$

where we have used $|II|^2 \ge \frac{1}{n-1}H^2$ in the second line and Theorem 3 in the fourth. On the other hand, by Theorem 3 again, we have

$$\text{(2-9)}\qquad \mathrm{Vol}(\Omega_t)' = \int_{\Sigma_t}\frac{1}{H}\,d\mu \ge \frac{n}{n-1}\,\mathrm{Vol}(\Omega_t).$$

It follows from (2-8) and (2-9) that

$$\Big(n(n-1)\,\mathrm{Vol}(\Omega_t) - \int_{\Sigma_t} uH\,d\mu\Big)' \ge \frac{n-2}{n-1}\Big(n(n-1)\,\mathrm{Vol}(\Omega_t) - \int_{\Sigma_t} uH\,d\mu\Big)$$

or equivalently

$$\text{(2-10)}\qquad \Big[e^{-\frac{n-2}{n-1}t}\Big(n\,\mathrm{Vol}(\Omega_t) - \frac{1}{n-1}\int_{\Sigma_t} r^2H\,d\mu\Big)\Big]' \ge 0.$$

We conclude that $Q(t)$ is monotone increasing, moreover $Q(t)$ is a constant function if and only if equalities in (2-8) and (2-9) hold. By Theorem 3, we know these equalities hold if and only if Σ_t is a round sphere for all t. This completes the proof of Theorem 1. □

Next, we prove Theorem 2 in the case that Σ is a convex hypersurface.

Proposition 1. *Let Σ be a smooth, closed, convex hypersurface in $\mathbb{R}^n$. Then*

$$\text{(2-11)}\qquad n\,\mathrm{Vol}(\Omega) \le \frac{1}{n-1}\int_\Sigma r^2H\,d\mu,$$

where $\mathrm{Vol}(\Omega)$ *is the volume of the region Ω enclosed by Σ, r is the distance to a fixed point O and H is the mean curvature of Σ. Moreover, equality in* (2-11) *holds if and only if Σ is a sphere centered at O.*

Remark. Proposition 1 generalizes the first inequality in Theorem 3.2(1) of [Kwong 2012].

Proof. Apply Reilly's formula (2-1) to the Euclidean region Ω and choose $u = r^2$; we have

$$4n(n-1)\,\mathrm{Vol}(\Omega) = \int_\Sigma II\big(\nabla^\Sigma u, \nabla^\Sigma u\big) + 2(\Delta_\Sigma u)\frac{\partial u}{\partial \nu} + H\left(\frac{\partial u}{\partial \nu}\right)^2 d\mu$$

where $\Delta_\Sigma u = \Delta u - H\dfrac{\partial u}{\partial \nu} - \nabla^2 u(\nu,\nu) = 2(n-1) - H\dfrac{\partial u}{\partial \nu}$. Therefore,

$$\int_\Sigma H\left(\frac{\partial u}{\partial \nu}\right)^2 d\mu = \int_\Sigma II\big(\nabla^\Sigma u, \nabla^\Sigma u\big)\, d\mu + 4n(n-1)\,\mathrm{Vol}(\Omega). \tag{2-12}$$

Since Σ is convex, $II(\,\cdot\,,\cdot\,)$ is positive definite. Hence, (2-12) implies

$$n(n-1)\,\mathrm{Vol}(\Omega) \le \frac{1}{4}\int_\Sigma H\big\langle\nabla(r^2),\nu\big\rangle^2\, d\mu \le \int_\Sigma Hr^2\, d\mu. \tag{2-13}$$

When $n(n-1)\,\mathrm{Vol}(\Omega) = \int_\Sigma Hr^2\, d\mu$, we must have $II\big(\nabla^\Sigma u, \nabla^\Sigma u\big) = 0$, hence $\nabla^\Sigma u = 0$. This implies that $u = r^2$ is a constant on Σ, which shows that Σ is a sphere centered at O. □

To deform a star-shaped hypersurface to a convex hypersurface through the inverse mean curvature flow, we make use of a special case of a general result of Gerhardt and Urbas:

Theorem 4 [Gerhardt 1990; Urbas 1990]. *Let Σ be a smooth, closed hypersurface in $\mathbb{R}^n$ with positive mean curvature, given by a smooth embedding $X_0 : \mathbb{S}^{n-1} \to \mathbb{R}^n$. Suppose Σ is star-shaped with respect to a point P. Then the initial value problem*

$$\begin{cases} \dfrac{\partial X}{\partial t} = \dfrac{1}{H}\nu, \\ X(\,\cdot\,,0) = X_0(\,\cdot\,), \end{cases} \tag{2-14}$$

has a unique smooth solution $X : \mathbb{S}^{n-1} \times [0,\infty) \to \mathbb{R}^n$, where ν is the unit outer normal vector to $\Sigma_t = X(\mathbb{S}^{n-1}, t)$ and H is the mean curvature of Σ_t. Moreover, Σ_t is star-shaped with respect to P and the rescaled hypersurface $\widetilde{\Sigma}_t$, parametrized by $\widetilde{X}(\,\cdot\,,t) = e^{-t/(n-1)}X(\,\cdot\,,t)$, converges to a sphere centered at P in the $\mathscr{C}^\infty$ topology as $t \to \infty$.

Proof of Theorem 2. By Theorem 4, there exists a smooth solution $\{\Sigma_t\}$ to the inverse mean curvature flow with initial condition Σ. Moreover, the rescaled hypersurface $\widetilde{\Sigma}_t = \{e^{-t/(n-1)}x \mid x \in \Sigma_t\}$ converges exponentially fast in the C^∞ topology to a sphere. In particular, $\widetilde{\Sigma}_t$ and hence Σ_t, must be convex for large t.

Let T be a time when Σ_T becomes convex. By Proposition 1, we have

$$n\,\mathrm{Vol}(\Omega_T) \le \frac{1}{n-1}\int_{\Sigma_T} r^2 H\, d\mu;$$

thus $Q(T) \leq 0$. By Theorem 1, we know that $Q(t)$ is monotone increasing. Hence $Q(0) \leq Q(T) \leq 0$, which proves (1-3).

If the equality in (1-3) holds, then $Q(0) = 0$. It follows from the monotonicity of $Q(t)$ and the fact $Q(t) \leq 0$ for large t that $Q(t) = 0$ for all t. By Theorem 1, this implies that Σ_t is a sphere for each t. By Theorem 1, Σ_t is a sphere centered at O for large t. Therefore, we conclude that the initial hypersurface Σ is a sphere centered at O. $\square$

Remark. It can be shown that Theorem 2 is still true if the mean curvature is only assumed to be nonnegative. Please refer to the arXiv version of this paper (1212.1906) for details.

References

[Brendle et al. 2012] S. Brendle, P.-K. Hung, and M.-T. Wang, "A Minkowski-type inequality for hypersurfaces in the Anti-deSitter–Schwarzschild manifold", preprint, 2012. arXiv 1209.0669

[Gerhardt 1990] C. Gerhardt, "Flow of nonconvex hypersurfaces into spheres", *J. Differential Geom.* **32**:1 (1990), 299–314. MR 91k:53016 Zbl 0708.53045

[Huisken and Ilmanen 2001] G. Huisken and T. Ilmanen, "The inverse mean curvature flow and the Riemannian Penrose inequality", *J. Differential Geom.* **59**:3 (2001), 353–437. MR 2003h:53091 Zbl 1055.53052

[Kwong 2012] K.-K. Kwong, "On convex hypersurfaces in space forms and eigenvalues estimates for differential forms", preprint, 2012. arXiv 1207.3999

[Reilly 1977] R. C. Reilly, "Applications of the Hessian operator in a Riemannian manifold", *Indiana Univ. Math. J.* **26**:3 (1977), 459–472. MR 57 #13799 Zbl 0391.53019

[Ros 1987] A. Ros, "Compact hypersurfaces with constant higher order mean curvatures", *Rev. Mat. Iberoamericana* **3**:3-4 (1987), 447–453. MR 90c:53160 Zbl 0673.53003

[Urbas 1990] J. I. E. Urbas, "On the expansion of starshaped hypersurfaces by symmetric functions of their principal curvatures", *Math. Z.* **205**:3 (1990), 355–372. MR 92c:53037 Zbl 0691.35048

Received November 12, 2012.

KWOK-KUN KWONG
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MIAMI
CORAL GABLES, FL 33146
UNITED STATES
kwong@math.miami.edu

PENGZI MIAO
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MIAMI
CORAL GABLES, FL 33146
UNITED STATES
pengzim@math.miami.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.423

NONFIBERED L-SPACE KNOTS

TYE LIDMAN AND LIAM WATSON

We construct an infinite family of knots in rational homology spheres with irreducible, nonfibered complements, for which every nonlongitudinal filling is an L-space.

The Heegaard Floer homology of a rational homology three-sphere Y is an abelian group $\widehat{HF}(Y)$ satisfying rk $\widehat{HF}(Y) \geq |H_1(Y;\mathbb{Z})|$ [Ozsváth and Szabó 2004]. When equality is realized in this bound Y is called an L-space and any knot in Y admitting a nontrivial L-space surgery is called an L-space knot [Ozsváth and Szabó 2005]. A result of Ghiggini [2008] and Ni [2007] shows that L-space knots in the three-sphere must be fibered. Since manifolds with finite fundamental group provide examples of L-spaces,[1] this result implies that a knot K in S^3 admitting a finite filling must be fibered. This observation should be compared with other restrictions related to finite fillings such as the Cyclic Surgery Theorem [Culler et al. 1987] and its extensions [Boyer and Zhang 2001].

The restriction to knots in S^3 is not necessary. It is shown in [Boileau et al. 2012] that a primitive knot[2] in an irreducible L-space admitting a nontrivial L-space surgery must be fibered. Irreducibility of the complement is required: removing an unknot from an embedded three-ball in any L-space produces a nonfibered manifold with nontrivial L-space fillings. Even in the general setting of knots in rational homology spheres with irreducible complements fibered is not a necessary condition:

Theorem 1. *There exist infinitely many irreducible, nonfibered knot complements such that all nonlongitudinal Dehn fillings are L-spaces. Moreover, these examples arise as knots in manifolds with finite fundamental group.*

Lidman was supported by a UCLA Dissertation Year Fellowship. Watson was partially supported by an NSERC Postdoctoral Fellowship.

MSC2010: 57M27.

Keywords: Heegaard Floer homology, L-space, fibration.

[1] Ozsváth and Szabó [2005] have shown that manifolds admitting elliptic geometry are L-spaces; Perelman's Geometrization Theorem (see [Kleiner and Lott 2008], for example) implies that three-manifolds with finite fundamental group admit elliptic geometry.

[2] Recall that a knot K is primitive in Y if $[K] \in H_1(Y;\mathbb{Z})$ is a generator.

In particular, our examples are nonprimitive knots in L-spaces.

Before turning to the construction, we fix some terminology. Fibrations will always be locally trivial surface bundles over a circle and we say the total space fibers. To avoid confusion, we will refer to Seifert fibrations as Seifert structures; these are foliations of a manifold by circles. The base orbifold is the leaf space of such a foliation, where the (possibly empty) collection of cone points records the multiplicities of the exceptional fibers in the Seifert structure. A circle bundle is a Seifert structure for which there are no exceptional fibers.

Given a three-manifold M with torus boundary, a slope α is a primitive class in $H_1(\partial M; \mathbb{Z})/\{\pm 1\}$. We use $M(\alpha)$ to denote Dehn filling along α. If $\partial M = T_1 \cup T_2$, for tori T_i, then we denote α-filling on T_1 (respectively T_2) by $M(\alpha, -)$ (respectively $M(-, \alpha)$). When M admits a Seifert structure, the slope given by a regular fiber in the boundary is called the fiber slope. For background on Seifert structures and Dehn filling we refer the reader to [Boyer 2002]. A key fact is that Dehn filling a Seifert manifold with torus boundary along any slope α other than the fiber slope results in a Seifert manifold with a possible additional singular fiber. The multiplicity of this new fiber is $\Delta(\alpha, \varphi)$, the distance between the slopes α and ϕ [Heil 1974].

Finally, for knots in rational homology three-spheres recall that there is a preferred slope given by the rational longitude. This slope is characterized by the property that some number of like-oriented parallel copies in the boundary of the knot complement bounds a properly embedded surface. We will refer to this slope as the longitude. Note that an oriented three-manifold M with torus boundary for which $H_1(M; \mathbb{Q}) \cong \mathbb{Q}$ always arises (nonuniquely) as the complement of a knot in a rational homology three-sphere.

1. The twisted I-bundle over the Klein bottle

Let N denote the twisted I-bundle over the Klein bottle. As this orientable three-manifold with torus boundary plays a central role in our construction we will study its construction in depth.

First consider the group G generated by $f, g\colon \mathbb{R}^3 \to \mathbb{R}^3$, where

$$\begin{aligned} f(x, y, z) &= (x+1, y, z), \\ g(x, y, z) &= (-x, y+1, -z), \end{aligned}$$

and consider the noncompact, orientable three-manifold $N^\circ = \mathbb{R}^3/G$. Note that the z-component of $\mathbb{R}^3$ gives N° the structure of a line bundle, the zero-section of which is a Klein bottle; this is the unique line bundle over the Klein bottle with orientable total space. By restricting the action of G to $\widetilde{N} = \mathbb{R}^2 \times \left[-\frac{1}{2}, \frac{1}{2}\right] \subset \mathbb{R}^3$ we obtain the twisted I-bundle over the Klein bottle $N = \widetilde{N}/G$.

From this description two Seifert structures on N become apparent: the x- and y-components of $\widetilde{N}$ both determine foliations of N by circles. (This is essentially the observation that the Klein bottle is foliated by circles in two ways.) The leaf space of the foliation described by the x-components is a Möbius strip without cone points. Denote a regular fiber in this Seifert structure by ϕ_0. The base orbifold of the foliation determined by y-components is $D^2(2,2)$, with regular fiber denoted ϕ_1; this follows readily from a natural Heegaard decomposition which we now describe.

From the preceding construction, a fundamental domain for N is obtained by taking $\left[-\frac{1}{2},\frac{1}{2}\right)^2\times\left[-\frac{1}{2},\frac{1}{2}\right]\subset\mathbb{R}^3$. Then given a disk D^2 of radius less than $\frac{1}{2}$ (and centered at the origin of the xy-plane), the result of removing $D^2\times\left[-\frac{1}{2},\frac{1}{2}\right]$ is a genus-two handlebody. This gives rise to a Heegaard decomposition for N; a Heegaard diagram corresponding to this decomposition is described in Figure 1, from which the fundamental group $\pi_1(N)=\langle a,b\mid a^2b^2\rangle$ may be calculated. Note that since $fgfg^{-1}$ is trivial in the group G, the homomorphism determined by $a\mapsto fgf^{-1}$ and $b\mapsto fg^{-1}$ is well-defined and gives an isomorphism $G\cong\langle a,b\mid a^2b^2\rangle$. Further, by considering a separating disk decomposing the handlebody into solid

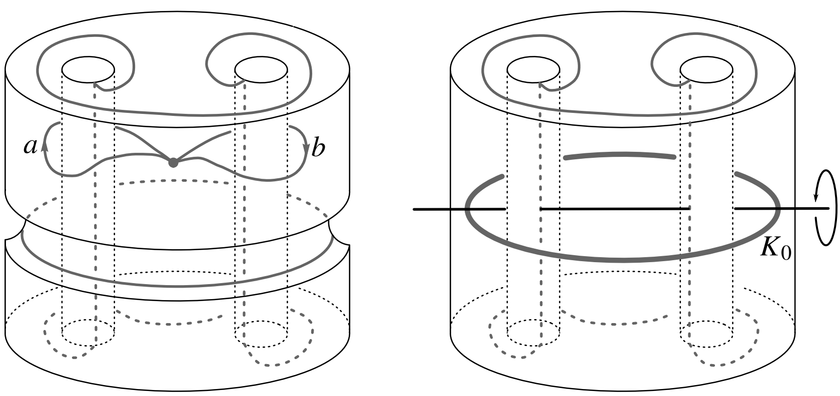


Figure 1. Two views of the Heegaard diagram for the twisted I-bundle over the Klein bottle N. With a and b generating the fundamental group of the genus-two handlebody, N is obtained by attaching a handle along a curve in the boundary representing a^2b^2 so that $\phi_0\simeq ab$ and $\phi_1\simeq b^2$. On the left, an annulus in the boundary with core representing the element $\phi_0\simeq ab$ may be used to find the fundamental group of M, the complement of a regular fiber in the interior of N, via HNN extension. On the right, the axis of rotational symmetry shows that the hyperelliptic involution on the handlebody induces a strong inversion on the pair (N,K_0) where K_0 is a knot in N isotopic to a regular fiber ϕ_0 in the interior of N.

tori, it is immediate that N is the union of two solid tori along essential annuli in the boundary. By fixing Seifert structures on each of these solid tori with base orbifolds $D^2(2)$, these annuli are foliated by regular fibers. The identification along these essential annuli therefore extends to a Seifert structure on N with base orbifold $D^2(2,2)$ as claimed.

Both Seifert structures induce foliations on the torus ∂N. Let ϕ_0 and ϕ_1 be regular fibers in ∂N, and notice that $\Delta(\phi_0, \phi_1) = 1$. (These conventions are consistent with [Boyer et al. 2013, Section 3].) The longitude of N is homotopic to the element ab (this element has order two in the abelianization of $\pi_1(N)$). That is, ϕ_0 represents the longitude of N. Any filling $N(\alpha)$ for which $\alpha \neq \phi_0, \phi_1$ admits a pair of Seifert structures with base orbifolds $\mathbb{R}P^2(\Delta(\alpha, \phi_0))$ and $S^2(2, 2, \Delta(\alpha, \phi_1))$. We point out that these manifolds always admit elliptic geometry [Scott 1983].

Now consider a knot K_0 in N that is isotopic to a regular fiber ϕ_0 in the interior of N. Define M by removing a neighborhood of K_0 from N; by construction M inherits a Seifert structure (the base orbifold is a punctured Möbius band). Now $\partial M = T_1 \cup T_2$ where T_2 denotes the boundary of a regular neighborhood of K_0.

The fundamental group of M is presented by

$$\pi_1(M) = \langle a, b, t \mid a^2b^2, [t, ab]\rangle.$$

To see this, consult Figure 1 and notice that M may be constructed by identifying (disjoint neighborhoods of) each boundary component of the annulus with core ab in ∂N. This gives rise to the HNN extension presented above. Notice that $M(-, \mu) \cong N$ for any slope on T_2 satisfying $\Delta(\mu, \phi_0) = 1$. A preferred choice for μ is given by a representative of the homotopy class of t in the above presentation.

A final observation pertains to a natural strong inversion on (N, K_0) that descends to an involution on M with one-dimensional fixed point set. Recall that a strong inversion on (N, K_0) is an orientation preserving involution on N that reverses orientation on K_0; such a symmetry is illustrated in Figure 1. The involution on N is induced by the hyperelliptic involution on the genus-two handlebody since the attaching curve is fixed (as a set) by this involution. A fundamental domain for this involution is a three-ball, with one dimensional fixed point set. That is, N is

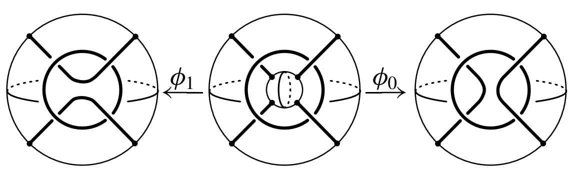


Figure 2. The branch set for the manifold $M = M(-,-)$ with branch sets for the fillings $M(\phi_1, -) = N$ and $M(\phi_0, -)$. Notice that $M(\phi_0, -)$ is reducible, containing an $S^2 \times S^1$ summand.

the twofold branched cover of a two-tangle; this is the leftmost tangle in Figure 2. We leave the following step to the reader: the genus-two handlebody is the twofold branched cover of a three-tangle, and attaching the handle closes one of the arcs (the arc meeting the attaching curve) to an unknotted curve in the branch set. The same construction may be applied to the complement of K_0 in N, to see that M is the twofold branched cover of a tangle in $S^2 \times I$. This tangle is shown in Figure 2.

Towards a proof of Theorem 1, our interest is in the family of manifolds

$$\{M(-,\alpha) \mid \text{for any slope } \alpha \text{ with } \Delta(\alpha,\phi_0) > 1\}.$$

Notice that each manifold in this set admits a Seifert structure with base orbifold a Möbius band with a single cone point of order $\Delta(\alpha,\phi_0)$. Since $M(\phi_1,\alpha)$ admits a Seifert structure with base orbifold $S^2(2,2,n)$ it follows that $M(-,\alpha)$ is the complement of a knot in an elliptic manifold for all α.

2. The proof of Theorem 1

Let M be the complement of K_0 in the twisted I-bundle over the Klein bottle N. We assume all of the notation introduced in the previous section.

Lemma 2. *Fix a slope α on T_2 with $\Delta(\alpha,\phi_0) = p$. Then*

$$M(\phi_0,\alpha) = \begin{cases} S^2\times S^1 \# S^2\times S^1 & \text{if } p=0,\\ S^2\times S^1 \# L(p,q) & \text{if } p>1,\\ S^2\times S^1 & \text{if } p=1.\end{cases}$$

Proof. Since $\pi_1(M) \cong \langle a,b,t \mid a^2b^2, [t,ab]\rangle$ and $\phi_0 \simeq ab$, we have that

$$\pi_1(M(\phi_0,-)) \cong \langle a,b,t \mid a^2b^2, [t,ab]\rangle/\langle\langle ab\rangle\rangle \cong \langle a,b,t \mid ab\rangle.$$

In other words, $\pi_1(M(\phi_0,-)) \cong \mathbb{Z} * \mathbb{Z}$. If $\alpha = p\mu + q\phi_0$, then

$$\pi_1(M(\phi_0,\alpha)) \cong \langle a,b,t \mid ab\rangle/\langle\langle t^p(ab)^q\rangle\rangle \cong \mathbb{Z} * \mathbb{Z}/p.$$

By Whitehead's proof of Kneser's conjecture [Whitehead 1958], $M(\phi_0,\alpha)$ is a connect-sum of closed manifolds Y_1 and Y_2 with $\pi_1(Y_1) \cong \mathbb{Z}$ and $\pi_1(Y_2) \cong \mathbb{Z}/p$. Geometrization now establishes the lemma. □

Remark 3. Alternatively, Lemma 2 follows from considering $M(\phi_0,-)$ as the double branched cover of a tangle as in Figure 2. The unknotted component gives rise to the $S^2 \times S^1$ summand. Dehn filling corresponds to attaching a rational tangle, which (ignoring the unknotted component) produces a two-bridge link and exhibits the lens space connect-summand.

Proposition 4. *For any α on T_2 with $\Delta(\alpha,\phi_0) > 1$, the manifold $M(-,\alpha)$ does not fiber over the circle.*

Proof. Suppose that $M(-,\alpha)$ fibers. Since ϕ_0 is the longitude, this is the only filling that extends the fibration on $M(-,\alpha)$ as any other filling of $M(-,\alpha)$ results in a rational homology sphere. By Lemma 2, $M(\phi_0,\alpha) \cong S^2 \times S^1 \# L(p,q)$ for $p = \Delta(\phi_0,\alpha) \geq 2$. Since $M(\phi_0,\alpha)$ is fibered and $\pi_2(M(\phi_0,\alpha)) \neq 0$, the fiber surface F must also have $\pi_2(F) \neq 0$ by the long exact sequence for a fibration. Hence F must be S^2 or $\mathbb{R}P^2$. However, $\pi_1(M(\phi_0,\alpha))$ is not the fundamental group of such a fibration, since it does not admit a surjective homomorphism onto $\mathbb{Z}$ with finite kernel. □

Proof of Theorem 1. Fix α with $\Delta(\alpha,\phi_0) \geq 2$. As the fiber slope of the Seifert structure on $M(-,\alpha)$ is the longitude, all nonlongitudinal fillings will extend the Seifert structure, yielding a base orbifold $\mathbb{R}P^2$ with two cone points. By [Boyer et al. 2013, Proposition 5], such manifolds are always L-spaces. Proposition 4 shows that $M(-,\alpha)$ is not fibered. Furthermore, $M(-,\alpha)$ is irreducible, since the only orientable, reducible Seifert manifolds are $S^2 \times S^1$ and $\mathbb{R}P^3 \# \mathbb{R}P^3$ (and in particular, are closed). Finally, $M(-,\alpha)$ is the complement of a knot in an elliptic manifold as observed in Section 1. □

Remark 5. Further examples may be constructed in an analogous way by removing a regular fiber from any manifold which has a Seifert structure with base orbifold $\mathbb{R}P^2$ with any positive number of singular fibers. It is also possible to construct examples, in a similar manner, admitting Sol geometry. The main observation is that every Sol rational homology sphere is an L-space [Boyer et al. 2013, Theorem 2]. Since every such L-space arises by identifying two twisted I-bundles along the boundary tori, one may consider the complement of the knot K_0 in one of the twisted I-bundles. In this setting, our construction goes through almost verbatim, having noticed that the obvious essential torus must be horizontal to the purported fibration of the exterior of K_0.

Question 6. All of our examples relied on the presence of an essential annulus, and have nonhyperbolic exterior. Do there exist examples of hyperbolic, nonfibered knots for which every nonlongitudinal surgery is an L-space?

Acknowledgments

The authors thank Ciprian Manolescu and Yi Ni for their comments on and interest in this problem. This paper owes its existence to the Workshop on Topics in Dehn Surgery, held at UT Austin in April 2012. The authors thank the organizers for putting together a great conference.

References

[Boileau et al. 2012] M. Boileau, S. Boyer, R. Cebanu, and G. S. Walsh, "Knot commensurability and the Berge conjecture", *Geom. Topol.* **16**:2 (2012), 625–664. MR 2928979 Zbl 1258.57001

[Boyer 2002] S. Boyer, "Dehn surgery on knots", pp. 165–218 in *Handbook of geometric topology*, edited by R. J. Daverman and R. B. Sher, North-Holland, Amsterdam, 2002. MR 2003f:57030 Zbl 1058.57004

[Boyer and Zhang 2001] S. Boyer and X. Zhang, "A proof of the finite filling conjecture", *J. Differential Geom.* **59**:1 (2001), 87–176. MR 2003k:57007 Zbl 1030.57024

[Boyer et al. 2013] S. Boyer, C. M. Gordon, and L. Watson, "On L-spaces and left-orderable fundamental groups", *Math. Ann.* **356**:4 (2013), 1213–1245. MR 3072799 Zbl 06194415

[Culler et al. 1987] M. Culler, C. M. Gordon, J. Luecke, and P. B. Shalen, "Dehn surgery on knots", *Ann. of Math.* (2) **125**:2 (1987), 237–300. MR 88a:57026 Zbl 0633.57006

[Ghiggini 2008] P. Ghiggini, "Knot Floer homology detects genus-one fibred knots", *Amer. J. Math.* **130**:5 (2008), 1151–1169. MR 2010f:57013 Zbl 1149.57019

[Heil 1974] W. Heil, "Elementary surgery on Seifert fiber spaces", *Yokohama Math. J.* **22** (1974), 135–139. MR 51 #11515 Zbl 0297.57006

[Kleiner and Lott 2008] B. Kleiner and J. Lott, "Notes on Perelman's papers", *Geom. Topol.* **12**:5 (2008), 2587–2855. MR 2010h:53098 Zbl 1204.53033

[Ni 2007] Y. Ni, "Knot Floer homology detects fibred knots", *Invent. Math.* **170**:3 (2007), 577–608. MR 2008j:57053 Zbl 1138.57031

[Ozsváth and Szabó 2004] P. Ozsváth and Z. Szabó, "Holomorphic disks and topological invariants for closed three-manifolds", *Ann. of Math.* (2) **159**:3 (2004), 1027–1158. MR 2006b:57016 Zbl 1073.57009

[Ozsváth and Szabó 2005] P. Ozsváth and Z. Szabó, "On knot Floer homology and lens space surgeries", *Topology* **44**:6 (2005), 1281–1300. MR 2006f:57034 Zbl 1077.57012

[Scott 1983] P. Scott, "The geometries of 3-manifolds", *Bull. London Math. Soc.* **15**:5 (1983), 401–487. MR 84m:57009 Zbl 0561.57001

[Whitehead 1958] J. H. C. Whitehead, "On finite cocycles and the sphere theorem", *Colloq. Math.* **6** (1958), 271–281. MR 20 #7271 Zbl 0119.38605

Received September 30, 2012. Revised September 16, 2013.

TYE LIDMAN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS AT AUSTIN
1 UNIVERSITY STATION
AUSTIN, TX 78712
UNITED STATES

tlid@math.utexas.edu

LIAM WATSON
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF GLASGOW
15 UNIVERSITY GARDENS
GLASGOW, G12 8QW
UNITED KINGDOM

liam.watson@glasgow.ac.uk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.431

FAMILIES AND SPRINGER'S CORRESPONDENCE

GEORGE LUSZTIG

We establish a relationship between the known parametrization of a family of irreducible representations of a Weyl group and Springer's correspondence.

Introduction

0.1. Let G be a connected reductive algebraic group over an algebraically closed field $\Bbbk$ of characteristic p. Let W be the Weyl group of G; let $\mathrm{Irr}W$ be a set of representatives for the isomorphism classes of irreducible representations of W over $\bar{\mathbb{Q}}_l$, an algebraic closure of the field of l-adic numbers (l is a fixed prime number other than p).

Now $\mathrm{Irr}W$ is partitioned into subsets called *families* as in [Lusztig 1979b, § 9; 1984a, 4.2]. Moreover to each family $\mathscr{F}$ in $\mathrm{Irr}W$, a certain set $X_{\mathscr{F}}$, a pairing $\{\ ,\ \}: X_{\mathscr{F}} \times X_{\mathscr{F}} \to \bar{\mathbb{Q}}_l$, and an imbedding $\mathscr{F} \to X_{\mathscr{F}}$ was canonically attached in [Lusztig 1979b; 1984a, Chapter 4]. (The set $X_{\mathscr{F}}$ with the pairing $\{\ ,\ \}$, which can be viewed as a nonabelian analogue of a symplectic vector space, plays a key role in the classification of unipotent representations of a finite Chevalley group [Lusztig 1984a] and in that of unipotent character sheaves on G.) In [Lusztig 1979b; 1984a] it is shown that $X_{\mathscr{F}} = M(\mathscr{G}_{\mathscr{F}})$ where $\mathscr{G}_{\mathscr{F}}$ is a certain finite group associated to $\mathscr{F}$ and, for any finite group Γ, $M(\Gamma)$ is the set of all pairs (g, ρ) where g is an element of Γ defined up to conjugacy and ρ is an irreducible representation over $\bar{\mathbb{Q}}_l$ (up to isomorphism) of the centralizer of g in Γ; moreover $\{\ ,\ \}$ is given by the "nonabelian Fourier transform matrix" of [Lusztig 1979b, § 4] for $\mathscr{G}_{\mathscr{F}}$.

In the remainder of this paper we assume that p is not a bad prime for G. In this case a uniform definition of the group $\mathscr{G}_{\mathscr{F}}$ was proposed in [Lusztig 1984a, 13.1] in terms of special unipotent classes in G and the Springer correspondence, but the fact that this leads to a group isomorphic to $\mathscr{G}_{\mathscr{F}}$ as defined in [Lusztig 1984a, Chapter 4] was stated in [Lusztig 1984a, (13.1.3)] without proof. One of the aims of this paper is to supply the missing proof.

Supported in part by National Science Foundation grant DMS-0758262.
MSC2010: 20G99.
Keywords: Weyl group, unipotent class, Springer correspondence.

To state the results of this paper we need some definitions. For $E \in \mathrm{Irr}W$ let $a_E \in \mathbb{N}$, $b_E \in \mathbb{N}$ be as in [Lusztig 1984a, 4.1]. As noted in [Lusztig 1979a], for $E \in \mathrm{Irr}W$ we have

(a) $$a_E \le b_E;$$

we say that E is *special* if $a_E = b_E$.

For $g \in G$ let $Z_G(g)$ or $Z(g)$ be the centralizer of g in G and let $A_G(g)$ or $A(g)$ be the group of connected components of $Z(g)$. Let C be a unipotent conjugacy class in G and let $u \in C$. Let $\mathcal{B}_u$ be the variety of Borel subgroups of G that contain u; this is a nonempty variety of dimension, say, e_C. The conjugation action of $Z(u)$ on $\mathcal{B}_u$ induces an action of $A(u)$ on $\boldsymbol{S}_u := H^{2e_C}(\mathcal{B}_u, \bar{\mathbb{Q}}_l)$. Now W acts on $\boldsymbol{S}_u$ by Springer's representation [Springer 1976]; however here we adopt the definition of the W-action on $\boldsymbol{S}_u$ given in [Lusztig 1984b] which differs from Springer's original definition by tensoring by sign. The W-action on $\boldsymbol{S}_u$ commutes with the $A(u)$-action. Hence we have canonically $\boldsymbol{S}_u = \oplus_{E \in \mathrm{Irr}W} E \otimes \mathcal{V}_E$ (as $W \times A(u)$-modules) where $\mathcal{V}_E$ are finite dimensional $\bar{\mathbb{Q}}_l$-vector spaces with $A(u)$-action. Let $\mathrm{Irr}_C W = \{E \in \mathrm{Irr}W; \mathcal{V}_E \ne 0\}$; this set does not depend on the choice of u in C. By [Springer 1976], the sets $\mathrm{Irr}_C W$ (for C variable) form a partition of $\mathrm{Irr}W$; also, if $E \in \mathrm{Irr}_C W$ then $\mathcal{V}_E$ is an irreducible $A(u)$-module and, if $E \ne E'$ in $\mathrm{Irr}_C W$, then the $A(u)$-modules $\mathcal{V}_E, \mathcal{V}_{E'}$ are not isomorphic. By [Borho and MacPherson 1981] we have

(b) $$e_C \le b_E \quad \text{for any } E \in \mathrm{Irr}_C W,$$

and the equality $b_E = e_C$ holds for exactly one $E \in \mathrm{Irr}_C W$ which we denote by E_C (for this E, $\mathcal{V}_E$ is the unit representation of $A(u)$).

Following [Lusztig 1984a, (13.1.1)] we say that C is *special* if E_C is special. (This concept was introduced in [Lusztig 1979a, § 9] although the word "special" was not used there.) From (b) we see that C is special if and only if $a_{E_C} = e_C$.

Now assume that C is special. We denote by $\mathcal{F} \subset \mathrm{Irr}W$ the family that contains E_C. (Note that $C \mapsto \mathcal{F}$ is a bijection from the set of special unipotent classes in G to the set of families in $\mathrm{Irr}W$.) We set $\mathrm{Irr}^*_C W = \{E \in \mathrm{Irr}_C W; E \in \mathcal{F}\}$ and

$$\mathcal{K}(u) = \{a \in A(u); a \text{ acts trivially on } \mathcal{V}_E \text{ for any } E \in \mathrm{Irr}^*_C W\}.$$

This is a normal subgroup of $A(u)$. We set $\bar{A}(u) = A(u)/\mathcal{K}(u)$, a quotient group of $A(u)$. Now, for any $E \in \mathrm{Irr}^*_C W$, $\mathcal{V}_E$ is naturally an (irreducible) $\bar{A}(u)$-module. Another definition of $\bar{A}(u)$ is given in [Lusztig 1984a, (13.1.1)]. In that definition $\mathrm{Irr}^*_C W$ is replaced by $\{E \in \mathrm{Irr}_C W; a_E = e_C\}$ and $\mathcal{K}(u)$, $\bar{A}(u)$ are defined as above but in terms of this modified $\mathrm{Irr}^*_C W$. However the two definitions are equivalent in view of the following result.

Proposition 0.2. *Assume that C is special. Let $E \in \mathrm{Irr}_C W$.*

(a) *We have $a_E \leq e_C$.*

(b) *We have $a_E = e_C$ if and only if $E \in \mathcal{F}$.*

This follows from [Lusztig 1992, 10.9]. Note that (a) was stated without proof in [Lusztig 1984a, (13.1.2)] (the proof I had in mind at the time of [Lusztig 1984a] was combinatorial).

0.3. The following result is equivalent to a result stated without proof in [Lusztig 1984a, (13.1.3)].

Theorem 0.4. *Let C be a special unipotent class of G, let $u \in C$ and let $\mathcal{F}$ be the family that contains E_C. Then we have canonically $\boldsymbol{X}_{\mathcal{F}} = M(\bar{A}(u))$ so that the pairing $\{\ ,\ \}$ on $\boldsymbol{X}_{\mathcal{F}}$ coincides with the pairing $\{\ ,\ \}$ on $M(\bar{A}(u))$. Hence $\mathcal{G}_{\mathcal{F}}$ can be taken to be $\bar{A}(u)$.*

This is equivalent to the corresponding statement in the case where G is adjoint, which reduces immediately to the case where G is adjoint simple. It is then enough to prove the theorem for one G in each isogeny class of semisimple, almost simple algebraic groups; this will be done in Section 3 after some combinatorial preliminaries in Sections 1 and 2. The proof uses the explicit description of the Springer correspondence: for type A_n, G_2 in [Springer 1976]; for type B_n, C_n, D_n in [Shoji 1979a; 1979b] (as an algorithm) and in [Lusztig 1984b] (by a closed formula); for type F_4 in [Shoji 1980]; for type E_n in [Alvis and Lusztig 1982; Spaltenstein 1982].

An immediate consequence of (the proof of) Theorem 0.4 is the following result which answers a question of R. Bezrukavnikov and which plays a role in [Losev and Ostrik 2012].

Corollary 0.5. *In the setup of Theorem 0.4 let $E \in \mathrm{Irr}^*_C W$ and let $\mathcal{V}_E$ be the corresponding $A(u)$-module viewed as an (irreducible) $\bar{A}(u)$-module. The image of E under the canonical imbedding $\mathcal{F} \to \boldsymbol{X}_{\mathcal{F}} = M(\bar{A}(u))$ is represented by the pair $(1, \mathcal{V}_E) \in M(\bar{A}(u))$. Conversely, if $E \in \mathcal{F}$ and the image of E under $\mathcal{F} \to \boldsymbol{X}_{\mathcal{F}} = M(\bar{A}(u))$ is represented by the pair $(1, \rho) \in M(\bar{A}(u))$ where ρ is an irreducible representation of $\bar{A}(u)$, then $E \in \mathrm{Irr}^*_C W$ and $\rho \cong \mathcal{V}_E$.*

0.6. Corollary 0.5 has the following interpretation. Let Y be a (unipotent) character sheaf on G whose restriction to the regular semisimple elements is $\neq 0$; assume that in the usual parametrization of unipotent character sheaves by $\bigsqcup_{\mathcal{F}'} \boldsymbol{X}_{\mathcal{F}'}$, Y corresponds to $(1, \rho) \in M(\bar{A}(u))$ where C is the special unipotent class corresponding to a family $\mathcal{F}$, $u \in C$ and ρ is an irreducible representation of $\bar{A}(u)$. Then $Y|_C$ is (up to shift) the irreducible local system on C defined by ρ.

0.7. ***Notation.*** If A, B are subsets of $\mathbb{N}$ we denote by $A \dot{\cup} B$ the union of A and B regarded as a multiset (each element of $A \cap B$ appears twice). For any set $\mathcal{X}$, we denote by $\mathcal{P}(\mathcal{X})$ the set of subsets of $\mathcal{X}$ viewed as an F_2-vector space with sum given by the symmetric difference. If $\mathcal{X} \neq \varnothing$ we note that $\{\varnothing, \mathcal{X}\}$ is a line in $\mathcal{P}(\mathcal{X})$ and we set $\bar{\mathcal{P}}(\mathcal{X}) = \mathcal{P}(\mathcal{X})/\{\varnothing, \mathcal{X}\}$, $\mathcal{P}_{\text{ev}}(\mathcal{X}) = \{L \in \mathcal{P}(\mathcal{X}); |L| = 0 \bmod 2\}$; let $\bar{\mathcal{P}}_{\text{ev}}(\mathcal{X})$ be the image of $\mathcal{P}_{\text{ev}}(\mathcal{X})$ under the obvious map $\mathcal{P}(\mathcal{X}) \to \bar{\mathcal{P}}(\mathcal{X})$ (thus $\bar{\mathcal{P}}_{\text{ev}}(\mathcal{X}) = \bar{\mathcal{P}}(\mathcal{X})$ if $|\mathcal{X}|$ is odd and $\bar{\mathcal{P}}_{\text{ev}}(\mathcal{X})$ is a hyperplane in $\bar{\mathcal{P}}(\mathcal{X})$ if $|\mathcal{X}|$ is even). Now if $\mathcal{X} \neq \varnothing$, the assignment $L, L' \mapsto |L \cap L'| \bmod 2$ defines a symplectic form on $\mathcal{P}_{\text{ev}}(\mathcal{X})$ which induces a nondegenerate symplectic form $(\ ,\ )$ on $\bar{\mathcal{P}}_{\text{ev}}(\mathcal{X})$ via the obvious linear map $\mathcal{P}_{\text{ev}}(\mathcal{X}) \to \bar{\mathcal{P}}_{\text{ev}}(\mathcal{X})$.

For $g \in G$ let g_s and g_ω be the semisimple and unipotent parts of g.

For $z \in \frac{1}{2}\mathbb{Z}$ we set $\lfloor z \rfloor = z$ if $z \in \mathbb{Z}$ and $\lfloor z \rfloor = z - \frac{1}{2}$ if $z \in \mathbb{Z} + \frac{1}{2}$.

Errata to [Lusztig 1984a]. On page 86, on line -6 delete "$b' < b$" and on line -4 before "In the language..." insert "The array above is regarded as identical to the array obtained by interchanging its two rows."

On page 343, line -5, after "respect to M" insert "and where the group $\mathcal{G}_{\mathcal{F}}$ defined in terms of (u', M) is isomorphic to the group $\mathcal{G}_{\mathcal{F}}$ defined in terms of (u, G)".

Erratum to [Lusztig 1984b]. In the definition of A_α, B_α in [Lusztig 1984b, 11.5], the condition $I \in \alpha$ should be replaced by $I \in \alpha'$ and the condition $I \in \alpha'$ should be replaced by $I \in \alpha$.

1. Combinatorics

1.1. Let N be an even integer ≥ 0. Let $a := (a_0, a_1, a_2, \dots, a_N) \in \mathbb{N}^{N+1}$ be such that $a_0 \leq a_1 \leq a_2 \leq \cdots \leq a_N$, $a_0 < a_2 < a_4 < \cdots$, $a_1 < a_3 < a_5 < \cdots$. Let $\mathcal{J} = \{i \in [0, N]; a_i \text{ appears exactly once in } a\}$. We have $\mathcal{J} = \{i_0, i_1, \dots, i_{2M}\}$ where $M \in \mathbb{N}$ and $i_0 < i_1 < \cdots < i_{2M}$ satisfy $i_s = s \bmod 2$ for $s \in [0, 2M]$. Hence for any $s \in [0, 2M-1]$ we have $i_{s+1} = i_s + 2m_s + 1$ for some $m_s \in \mathbb{N}$. Let $\mathcal{E}$ be the set of $b := (b_0, b_1, b_2, \dots, b_N) \in \mathbb{N}^{N+1}$ such that $b_0 < b_2 < b_4 < \cdots$, $b_1 < b_3 < b_5 < \cdots$ and such that $[b] = [a]$ (we denote by $[b], [a]$ the multisets $\{b_0, b_1, \dots, b_N\}$, $\{a_0, a_1, \dots, a_N\}$). We have $a \in \mathcal{E}$. For $b \in \mathcal{E}$ we set

$$\begin{aligned}\hat{b} &= (\hat{b}_0, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_N)\\ &= (b_0, b_1 + 1, b_2 + 1, b_3 + 2, b_4 + 2, \dots, b_{N-1} + N/2, b_N + N/2).\end{aligned}$$

Let $[\hat{b}]$ be the multiset $\{\hat{b}_0, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_N\}$. For $s \in \{1, 3, \dots, 2M-1\}$ we define $a^{\{s\}} = (a_0^{\{s\}}, a_1^{\{s\}}, a_2^{\{s\}}, \dots, a_N^{\{s\}}) \in \mathcal{E}$ by

$$\begin{aligned}&(a_{i_s}^{\{s\}}, a_{i_s+1}^{\{s\}}, a_{i_s+2}^{\{s\}}, a_{i_s+3}^{\{s\}}, \dots, a_{i_s+2m_s}^{\{s\}}, a_{i_s+2m_s+1}^{\{s\}})\\ &\qquad = (a_{i_s+1}, a_{i_s}, a_{i_s+3}, a_{i_s+2}, \dots, a_{i_s+2m_s+1}, a_{i_s+2m_s})\end{aligned}$$

and $a_i^{\{s\}} = a_i$ if $i \in [0, N] - [i_s, i_{s+1}]$. More generally, for $X \subset \{1, 3, \dots, 2M-1\}$ we define $a^X = (a_0^X, a_1^X, a_2^X, \dots, a_N^X) \in \mathscr{E}$ by $a_i^X = a_i^{\{s\}}$ if $s \in X$, $i \in [i_s, i_{s+1}]$, and $a_i^X = a_i$ for all other $i \in [0, N]$. Note that $[\widehat{a^X}] = [\hat{a}]$. Conversely, we have the following result.

Lemma 1.2. *Let $b \in \mathscr{E}$ be such that $[\hat{b}] = [\hat{a}]$. There exists $X \subset \{1, 3, \dots, 2M-1\}$ such that $b = a^X$.*

The proof is given in 1.3–1.5.

1.3. We argue by induction on M. We have

$$a = (y_1 = y_1 < y_2 = y_2 < \dots < y_r = y_r < a_{i_0} < \dots)$$

for some r. Since $[b] = [a]$, we must have

$$(b_0, b_2, b_4, \dots) = (y_1, y_2, \dots, y_r, \dots), (b_1, b_3, b_5, \dots) = (y_1, y_2, \dots, y_r, \dots).$$

Thus,

(a) $$b_i = a_i \quad \text{for } i < i_0.$$

We have $a = (\dots < a_{i_{2M}} < y'_1 = y'_1 < y'_2 = y'_2 < \dots < y'_{r'} = y'_{r'})$ for some r'. Since $[b] = [a]$, we must have

$$(b_0, b_2, b_4, \dots) = (\dots, y'_1, y'_2, \dots, y'_{r'}), (b_1, b_3, b_5, \dots) = (\dots, y'_1, y'_2, \dots, y'_{r'}).$$

Thus,

(b) $$b_i = a_i \quad \text{for } i > i_{2M}.$$

If $M = 0$ we see that $b = a$ and there is nothing further to prove. In the rest of the proof we assume that $M \geq 1$.

1.4. From 1.3 we see that

$$(a_0, a_1, a_2, \dots, a_{i_{2M}}) = (\dots, a_{i_{2M-1}} < x_1 = x_1 < x_2 = x_2 < \dots < x_q = x_q < a_{i_{2M}})$$

(for some q) has the same entries as $(b_0, b_1, b_2, \dots, b_{i_{2M}})$ (in some order). Hence the pair

$$(\dots, b_{i_{2M}-5}, b_{i_{2M}-3}, b_{i_{2M-1}}), (\dots, b_{i_{2M}-4}, b_{i_{2M}-2}, b_{i_{2M}})$$

must have one of the following four forms.

$$(\dots, a_{i_{2M-1}}, x_1, x_2, \dots, x_q), (\dots, x_1, x_2, \dots, x_q, a_{i_{2M}}),$$
$$(\dots, x_1, x_2, \dots, x_q, a_{i_{2M}}), (\dots, a_{i_{2M-1}}, x_1, x_2, \dots, x_q),$$
$$(\dots, x_1, x_2, \dots, x_q), (\dots, a_{i_{2M-1}}, x_1, x_2, \dots, x_q, a_{i_{2M}}),$$
$$(\dots, a_{i_{2M-1}}, x_1, x_2, \dots, x_q, a_{i_{2M}}), (\dots, x_1, x_2, \dots, x_q).$$

Hence $(\dots, b_{i_{2M}-2}, b_{i_{2M}-1}, b_{i_{2M}})$ must have one of the following four forms.

(I) $(\dots, a_{i_{2M-1}}, x_1, x_1, x_2, x_2, \dots, x_q, x_q, a_{i_{2M}})$,

(II) $(\dots, x_1, a_{i_{2M-1}}, x_2, x_1, x_3, x_2, \dots, x_q, x_{q-1}, a_{i_{2M}}, x_q)$,

(III) $(\dots, a_{i_{2M-1}}, z, x_1, x_1, x_2, x_2, \dots, x_q, x_q, a_{i_{2M}})$,

(IV) $(\dots, a_{i_{2M-1}}, z', x_1, z'', x_2, x_1, x_3, x_2, \dots, x_q, x_{q-1}, a_{i_{2M}}, x_q)$,

where $a_{i_{2M-1}} > z$, $a_{i_{2M-1}} > z'' \geq z'$ and all entries in $\dots$ are $< a_{i_{2M-1}}$. Correspondingly, $(\dots, \hat{b}_{i_{2M}-2}, \hat{b}_{i_{2M}-1}, \hat{b}_{i_{2M}})$ must have one of the following four forms.

(I) $(\dots, a_{i_{2M-1}}+h-q, x_1+h-q, x_1+h-q+1, x_2+h-q+1, x_2+h-q+2, \dots, x_q+h-1, x_q+h, a_{i_{2M}}+h)$,

(II) $(\dots, x_1+h-q, a_{i_{2M-1}}+h-q, x_2+h-q+1, x_1+h-q+1, x_3+h-q+2, x_2+h-q+1, \dots, x_q+h-1, x_{q-1}+h-1, a_{i_{2M}}+h, x_q+h)$,

(III) $(\dots, a_{i_{2M-1}}+h-q-1, z+h-q, x_1+h-q, x_1+h-q+1, x_2+h-q+1, x_2+h-q+2, \dots, x_q+h-1, x_q+h, a_{i_{2M}}+h)$,

(IV) $(\dots, a_{i_{2M-1}}+h-q-1, z'+h-q-1, x_1+h-q, z''+h-q, x_2+h-q+1, x_1+h-q+1, x_3+h-q+2, x_2+h-q+1, \dots, x_q+h-1, x_{q-1}+h-1, a_{i_{2M}}+h, x_q+h)$,

where $h = i_{2M}/2$ and in cases (III) and (IV), $a_{i_{2M-1}}+h-q$ is not an entry of $(\dots, \hat{b}_{i_{2M}-2}, \hat{b}_{i_{2M}-1}, \hat{b}_{i_{2M}})$.

Since $(\dots, \hat{a}_{i_{2M}-2}, \hat{a}_{i_{2M}-1}, \hat{a}_{i_{2M}})$ is given by (I) we see that $a_{i_{2M-1}}+h-q$ is an entry of $(\dots, \hat{a}_{i_{2M}-2}, \hat{a}_{i_{2M}-1}, \hat{a}_{i_{2M}})$. Using (b) in 1.3 we see that

$$\{\dots, \hat{a}_{i_{2M}-2}, \hat{a}_{i_{2M}-1}, \hat{a}_{i_{2M}}\} = (\dots, b_{i_{2M}-2}, b_{i_{2M}-1}, b_{i_{2M}})$$

as multisets. We see that cases (III) and (IV) cannot arise. Hence we must be in case (I) or (II). Thus we have *either*

(a) $$(b_{i_{2M-1}}, b_{i_{2M-1}+1}, \dots, b_{i_{2M}-2}, b_{i_{2M}-1}, b_{i_{2M}}) = (a_{i_{2M-1}}, a_{i_{2M-1}+1}, \dots, a_{i_{2M}-2}, a_{i_{2M}-1}, a_{i_{2M}})$$

or

(b) $$(b_{i_{2M-1}}, b_{i_{2M-1}+1}, \dots, b_{i_{2M}-2}, b_{i_{2M}-1}, b_{i_{2M}}) = (a_{i_{2M-1}+1}, a_{i_{2M-1}}, a_{i_{2M-1}+3}, a_{i_{2M-1}+2}, \dots, a_{i_{2M}}, a_{i_{2M}-1}).$$

1.5. Let $a' = (a_0, a_1, a_2, \dots, a_{i_{2M-1}-1})$, $b' = (b_0, b_1, b_2, \dots, b_{i_{2M-1}-1})$,

$$\hat{a}' = (a_0, a_1+1, a_2+1, a_3+2, a_4+2, \dots, a_{i_{2M-1}-1} + (i_{2M-1}-1)/2),$$
$$\hat{b}' = (b_0, b_1+1, b_2+1, b_3+2, b_4+2, \dots, b_{i_{2M-1}-1} + (i_{2M-1}-1)/2),$$

From $[\hat{b}] = [\hat{a}]$, (b) in 1.3 and (a)+(b) in 1.4 we see that the multiset formed by the entries of $\hat{a}'$ coincides with the multiset formed by the entries of $\hat{b}'$. Using the induction hypothesis we see that there exists $X' \subset \{1, 3, \dots, 2M-3\}$ such that $b' = a'^{X'}$ where $a'^{X'}$ is defined in terms of a', X' in the same way as a^X was defined (see 1.1) in terms of a, X. We set $X = X'$ if we are in case (a) of 1.4 and $X = X' \cup \{2M-1\}$ if we are in case (b). Then we have $b = a^X$ (see again (a) and (b) in 1.4), as required. This completes the proof of 1.2.

1.6. We shall use the notation of 1.1. Let $\mathfrak{T}$ be the set of all unordered pairs $(\mathfrak{A}, \mathfrak{B})$ where $\mathfrak{A}$, $\mathfrak{B}$ are subsets of $\{0, 1, 2, \dots\}$ and $\mathfrak{A} \dot\cup \mathfrak{B} = (a_0, a_1, a_2, \dots, a_N)$ as multisets. For example, setting $\mathfrak{A}_\varnothing = (a_0, a_2, a_4, \dots, a_N)$ and $\mathfrak{B}_\varnothing = (a_1, a_3, \dots, a_{N-1})$, we have $(\mathfrak{A}_\varnothing, \mathfrak{B}_\varnothing) \in \mathfrak{T}$. For any subset $\mathfrak{a}$ of $\mathcal{J}$ we consider

$$\mathfrak{A}_\mathfrak{a} = ((\mathcal{J} - \mathfrak{a}) \cap \mathfrak{A}_\varnothing) \cup (\mathfrak{a} \cap \mathfrak{B}_\varnothing) \cup (\mathfrak{A}_\varnothing \cap \mathfrak{B}_\varnothing),$$
$$\mathfrak{B}_\mathfrak{a} = ((\mathcal{J} - \mathfrak{a}) \cap \mathfrak{B}_\varnothing) \cup (\mathfrak{a} \cap \mathfrak{A}_\varnothing) \cup (\mathfrak{A}_\varnothing \cap \mathfrak{B}_\varnothing).$$

Then $(\mathfrak{A}_\mathfrak{a}, \mathfrak{B}_\mathfrak{a}) \in \mathfrak{T}$ and the map $\mathfrak{a} \mapsto (\mathfrak{A}_\mathfrak{a}, \mathfrak{B}_\mathfrak{a})$ induces a bijection $\bar{\mathcal{P}}(\mathcal{J}) \leftrightarrow \mathfrak{T}$. (Note that if $\mathfrak{a} = \varnothing$ then $(\mathfrak{A}_\mathfrak{a}, \mathfrak{B}_\mathfrak{a})$ agrees with the earlier definition of $(\mathfrak{A}_\varnothing, \mathfrak{B}_\varnothing)$.)

Let $\mathfrak{T}'$ be the set of all $(\mathfrak{A}, \mathfrak{B}) \in \mathfrak{T}$ such that $|\mathfrak{A}| = |\mathfrak{A}_\varnothing|$ and $|\mathfrak{B}| = |\mathfrak{B}_\varnothing|$.

Let $\mathcal{P}(\mathcal{J})_0$ be the subspace of $\mathcal{P}_{\mathrm{ev}}(\mathcal{J})$ spanned by the 2-element subsets

$$\{a_{i_0}, a_{i_1}\}, \{a_{i_2}, a_{i_3}\}, \dots, \{a_{i_{2M-2}}, a_{i_{2M-1}}\}$$

of $\mathcal{J}$. Let $\mathcal{P}(\mathcal{J})_1$ be the subspace of $\mathcal{P}_{\mathrm{ev}}(\mathcal{J})$ spanned by the 2-element subsets

$$\{a_{i_1}, a_{i_2}\}, \{a_{i_3}, a_{i_4}\}, \dots, \{a_{i_{2M-1}}, a_{i_{2M}}\}$$

of $\mathcal{J}$.

Let $\bar{\mathcal{P}}(\mathcal{J})_0$ and $\bar{\mathcal{P}}(\mathcal{J})_1$ be the images of $\mathcal{P}(\mathcal{J})_0$ and $\mathcal{P}(\mathcal{J})_1$ under the obvious map $\mathcal{P}(\mathcal{J}) \to \bar{\mathcal{P}}(\mathcal{J})$. Then:

(a) $\bar{\mathcal{P}}(\mathcal{J})_0$ *and* $\bar{\mathcal{P}}(\mathcal{J})_1$ *are opposed Lagrangian subspaces of the symplectic vector space* $\bar{\mathcal{P}}(\mathcal{J})$, (,) (see 0.7); *hence* (,) *defines an identification*

$$\bar{\mathcal{P}}(\mathcal{J})_0 = \bar{\mathcal{P}}(\mathcal{J})_1^*,$$

where $\bar{\mathcal{P}}(\mathcal{J})_1^*$ *is the vector space dual to* $\bar{\mathcal{P}}(\mathcal{J})_1$.

Let $\mathfrak{T}_0$ and $\mathfrak{T}_1$ be the subsets of $\mathfrak{T}$ corresponding to $\bar{\mathcal{P}}(\mathcal{J})_0$ and $\bar{\mathcal{P}}(\mathcal{J})_1$, respectively, under the bijection $\bar{\mathcal{P}}(\mathcal{J}) \leftrightarrow \mathfrak{T}$. Note that $\mathfrak{T}_0 \subset \mathfrak{T}'$, $\mathfrak{T}_1 \subset \mathfrak{T}'$, and $|\mathfrak{T}_0| = |\mathfrak{T}_1| = 2^M$.

For any $X \subset \{1, 3, \dots, 2M-1\}$ we set $\mathfrak{a}_X = \bigcup_{s \in X} \{a_{i_s}, a_{i_{s+1}}\} \in \mathcal{P}(\mathcal{J})$. Then $(\mathfrak{A}_{\mathfrak{a}_X}, \mathfrak{B}_{\mathfrak{a}_X}) \in \mathfrak{T}_1$ is related to a^X in 1.1 as follows:

$$\mathfrak{A}_{\mathfrak{a}_X} = \{a_0^X, a_2^X, a_4^X, \dots, a_N^X\}, \qquad \mathfrak{B}_{\mathfrak{a}_X} = \{a_1^X, a_3^X, \dots, a_{N-1}^X\}.$$

1.7. We shall use the notation of 1.1. Let T be the set of all ordered pairs (A, B) where A is a subset of $\{0, 1, 2, \dots\}$, B is a subset of $\{1, 2, 3, \dots\}$, A contains no consecutive integers, B contains no consecutive integers, and $A \dot{\cup} B = (\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_N)$ as multisets. For example, setting $A_\varnothing = (\hat{a}_0, \hat{a}_2, \hat{a}_4, \dots, \hat{a}_N)$ and $B_\varnothing = (\hat{a}_1, \hat{a}_3, \dots, \hat{a}_{N-1})$, we have $(A_\varnothing, B_\varnothing) \in T$.

For any $(A, B) \in T$ we define (A^-, B^-) as follows: A^- consists of $x_0 < x_1 - 1 < x_2 - 2 < \dots < x_p - p$ where $x_0 < x_1 < \dots < x_p$ are the elements of A; B^- consists of $y_1 - 1 < y_2 - 2 < \dots < y_q - q$ where $y_1 < y_2 < \dots < y_q$ are the elements of B.

We can enumerate the elements of T as in [Lusztig 1984b, 11.5]. Let J be the set of all $c \in \mathbb{N}$ such that c appears exactly once in the sequence

$$(\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_N) = (a_0, a_1+1, a_2+1, a_3+2, a_4+2, \dots, a_{N-1}+N/2, a_N+N/2).$$

A nonempty subset I of J is said to be an interval if it is of the form $\{i, i+1, i+2, \dots, j\}$ with $i-1 \notin J$, $j+1 \notin J$ and with $i \neq 0$. Let $\mathcal{I}$ be the set of intervals of J. For any $s \in \{1, 3, \dots, 2M-1\}$, the set $I_s := \{\hat{a}_{i_s}, \hat{a}_{i_s+1}, \hat{a}_{i_s+2}, \dots, \hat{a}_{i_s+2m_s+1}\}$ is either a single interval or a union of intervals $I_s^1 \sqcup I_s^2 \sqcup \dots \sqcup I_s^{t_s}$ ($t_s \geq 2$) where $\hat{a}_{i_s} \in I_s^1$, $\hat{a}_{i_s+2m_s+1} \in I_s^{t_s}$, $|I_s^1|$, $|I_s^{t_s}|$ are odd, $|I_s^h|$ are even for $h \in [2, t_s-1]$ and any element in I_s^e is $<$ than any element in $I_s^{e'}$ for $e < e'$. Let $\mathcal{I}_s$ be the set of all $I \in \mathcal{I}$ such that $I \subset I_s$. Let H be the set of all $c \in J$ such that c does not belong to any interval. For any subset $\alpha \subset \mathcal{I}$ we consider

$$A_\alpha = \bigcup_{I \in \mathcal{I}-\alpha} (I \cap A_\varnothing) \cup \bigcup_{I \in \alpha} (I \cap B_\varnothing) \cup (H \cap A_\varnothing) \cup (A_\varnothing \cap B_\varnothing),$$
$$B_\alpha = \bigcup_{I \in \mathcal{I}-\alpha} (I \cap B_\varnothing) \cup \bigcup_{I \in \alpha} (I \cap A_\varnothing) \cup (H \cap B_\varnothing) \cup (A_\varnothing \cap B_\varnothing).$$

Then $(A_\alpha, B_\alpha) \in T$ and the map $\alpha \mapsto (A_\alpha, B_\alpha)$ is a bijection $\mathcal{P}(\mathcal{I}) \leftrightarrow T$. (Note that if $\alpha = \varnothing$ then (A_α, B_α) agrees with the earlier definition of $(A_\varnothing, B_\varnothing)$.)

Let $T' = \{(A, B) \in T; |A| = |A_\varnothing|, |B| = |B_\varnothing|\}$, $T_1 = \{(A, B) \in T'; A^- \dot{\cup} B^- = A_\varnothing^- \dot{\cup} B_\varnothing^-\}$. Let $\mathcal{P}(\mathcal{I})'$ and $\mathcal{P}(\mathcal{I})_1$ be the subsets of $\mathcal{P}(\mathcal{I})$ corresponding to T' and T_1 under the bijection $\mathcal{P}(\mathcal{I}) \leftrightarrow T$.

Now let X be a subset of $\{1, 3, \dots, 2M-1\}$. Let $\alpha_X = \bigcup_{s \in X} \mathcal{I}_s \in \mathcal{P}(\mathcal{I})$. From the definitions we see that

(a) $$A^-_{\alpha_X} = \mathfrak{A}_{\mathfrak{a}_X}, \quad B^-_{\alpha_X} = \mathfrak{B}_{\mathfrak{a}_X}$$

(in the notation of 1.6). In particular we have $(A_{\alpha_X}, B_{\alpha_X}) \in T_1$. Thus $|T_1| \geq 2^M$. Using Lemma 1.2 we see that

(b) $$|T_1| = 2^M \quad \textit{and} \quad T_1 = \big\{(A_{\alpha_X}, B_{\alpha_X}); X \subset \{1, 3, \dots, 2M-1\}\big\}.$$

Using (a) and (b) we deduce:

(c) *The map $T_1 \to \mathfrak{T}_1$ given by $(A, B) \mapsto (A^-, B^-)$ is a bijection.*

2. Combinatorics (continued)

2.1. Let $N \in \mathbb{N}$. Let

$$a := (a_0, a_1, a_2, \dots, a_N) \in \mathbb{N}^{N+1}$$

be such that $a_0 \le a_1 \le a_2 \le \cdots \le a_N$, $a_0 < a_2 < a_4 < \cdots$, $a_1 < a_3 < a_5 < \cdots$ and such that the set $\mathcal{J} := \{i \in [0, N]; a_i$ appears exactly once in $a\}$ is nonempty. Now $\mathcal{J}$ consists of $\mu+1$ elements $i_0 < i_1 < \cdots < i_\mu$, where $\mu \in \mathbb{N}$, $\mu = N \bmod 2$. We have $i_s = s \bmod 2$ for $s \in [0, \mu]$. Hence for any $s \in [0, \mu-1]$ we have $i_{s+1} = i_s + 2m_s + 1$ for some $m_s \in \mathbb{N}$. Let $\mathcal{E}$ be the set of $b := (b_0, b_1, b_2, \dots, b_N) \in \mathbb{N}^{N+1}$ such that $b_0 < b_2 < b_4 < \cdots$, $b_1 < b_3 < b_5 < \cdots$ and such that $[b] = [a]$ (we denote by $[b]$, $[a]$ the multisets $\{b_0, b_1, \dots, b_N\}$, $\{a_0, a_1, \dots, a_N\}$). We have $a \in \mathcal{E}$. For $b \in \mathcal{E}$ we set

$$\mathring{b} = (\mathring{b}_0, \mathring{b}_1, \mathring{b}_2, \dots, \mathring{b}_N) = (b_0, b_1, b_2+1, b_3+1, b_4+2, b_5+2, \dots) \in \mathbb{N}^{N+1}.$$

Let $[\mathring{b}]$ be the multiset $\{\mathring{b}_0, \mathring{b}_1, \mathring{b}_2, \dots, \mathring{b}_N\}$. For any $s \in [0, \mu-1] \in 2\mathbb{N}$ we define $a^{\{s\}} = (a_0^{\{s\}}, a_1^{\{s\}}, a_2^{\{s\}}, \dots, a_N^{\{s\}}) \in \mathcal{E}$ by

$$\begin{aligned}(a_{i_s}^{\{s\}}, a_{i_s+1}^{\{s\}}, a_{i_s+2}^{\{s\}}, a_{i_s+3}^{\{s\}}, \dots, a_{i_s+2m_s}^{\{s\}}, a_{i_s+2m_s+1}^{\{s\}}) \\ = (a_{i_s+1}, a_{i_s}, a_{i_s+3}, a_{i_s+2}, \dots, a_{i_s+2m_s+1}, a_{i_s+2m_s})\end{aligned}$$

and $a_i^{\{s\}} = a_i$ if $i \in [0, N] - [i_s, i_{s+1}]$. More generally, for a subset X of $[0, \mu-1] \cap 2\mathbb{N}$ we define $a^X = (a_0^X, a_1^X, a_2^X, \dots, a_N^X) \in \mathcal{E}$ by $a_i^X = a_i^{\{s\}}$ if $s \in X$, $i \in [i_s, i_{s+1}]$, and $a_i^X = a_i$ for all other $i \in [0, N]$. Note that $[\mathring{a}^X] = [\mathring{a}]$. Conversely:

Lemma 2.2. *Let $b \in \mathcal{E}$ be such that $[\mathring{b}] = [\mathring{a}]$. Then there exists $X \subset [0, \mu-1] \cap 2\mathbb{N}$ such that $b = a^X$.*

The proof is given in 2.3–2.5.

2.3. We argue by induction on μ. By the argument in 1.3 we have

(a) $$b_i = a_i \quad \text{for } i < i_0,$$

(b) $$b_i = a_i \quad \text{for } i > i_\mu.$$

If $\mu = 0$ we see that $b = a$ and there is nothing further to prove. In the rest of the proof we assume that $\mu \ge 1$.

2.4. From 2.3 we see that $(a_{i_0}, a_{i_0+1}, \dots, a_N) = (a_{i_0} < x_1 = x_1 < x_2 = x_2 < \cdots < x_p = x_p < a_{i_1} < \dots)$ (for some p) has the same entries as $(b_{i_0}, b_{i_0+1}, \dots, b_N)$ (in some order). Hence the pair $(b_{i_0}, b_{i_0+2}, b_{i_0+4}, \dots)$, $(b_{i_0+1}, b_{i_0+3}, b_{i_0+5}, \dots)$ must have one of the following four forms.

$$(a_{i_0}, x_1, x_2, \dots, x_p, \dots), (x_1, x_2, \dots, x_p, a_{i_1}, \dots),$$

$$(x_1, x_2, \dots, x_p, a_{i_1}, \dots), (a_{i_0}, x_1, x_2, \dots, x_p, \dots),$$
$$(a_{i_0}, x_1, x_2, \dots, x_p, a_{i_1}, \dots), (x_1, x_2, \dots, x_p, \dots),$$
$$(x_1, x_2, \dots, x_p, \dots), (a_{i_0}, x_1, x_2, \dots, x_p, a_{i_1}, \dots).$$

Hence $(b_{i_0}, b_{i_0+1}, b_{i_0+2}, \dots, b_N)$ must have one of the following four forms.

(I) $(a_{i_0}, x_1, x_1, x_2, x_2, \dots, x_p, x_p, a_{i_1}, \dots)$,

(II) $(x_1, a_{i_0}, x_2, x_1, x_3, x_2, \dots, x_p, x_{p-1}, a_{i_1}, x_p, \dots)$,

(III) $(a_{i_0}, x_1, x_1, x_2, x_2, \dots, x_p, x_p, z, a_{i_1}, \dots)$,

(IV) $(x_1, a_{i_0}, x_2, x_1, x_3, x_2, \dots, x_p, x_{p-1}, z', x_p, z'', a_{i_1}, \dots)$,

where $a_{i_1} < z$, $a_{i_1} < z' \le z''$ and all entries in $\dots$ are $> a_{i_1}$. Correspondingly, $(\mathring{b}_{i_0}, \mathring{b}_{i_0+1}, \mathring{b}_{i_0+2}, \dots, \mathring{b}_N)$ must have one of the following four forms.

(I) $(a_{i_0} + h, x_1 + h, x_1 + h + 1, x_2 + h + 1, x_2 + h + 2, \dots, x_p + h + p - 1, x_p + h + p, a_{i_1} + h + p, \dots)$,

(II) $(x_1 + h, a_{i_0} + h, x_2 + h + 1, x_1 + h + 1, x_3 + h + 2, x_2 + h + 2, \dots, x_p + h + p - 1, x_{p-1} + h + p - 1, a_{i_1} + h + p, x_p + h + p, \dots)$,

(III) $(a_{i_0} + h, x_1 + h, x_1 + h + 1, x_2 + h + 1, x_2 + h + 2, \dots, x_p + h + p - 1, x_p + h + p, z + p, a_{i_1} + h + p + 1, \dots)$,

(IV) $(x_1 + h, a_{i_0} + h, x_2 + h + 1, x_1 + h + 1, x_3 + h + 2, x_2 + h + 2, \dots, x_p + h + p - 1, x_{p-1} + h + p - 1, z' + h + p, x_p + h + p, z'' + h + p + 1, a_{i_1} + h + p + 1, \dots)$,

where $h = i_0/2$ and in cases (III) and (IV) $a_{i_1} + h + p$ is not an entry of $(\mathring{b}_{i_0}, \mathring{b}_{i_0+1}, \mathring{b}_{i_0+2}, \dots)$.

Since $(\mathring{a}_{i_0}, \mathring{a}_{i_0+1}, \mathring{a}_{i_0+2}, \dots)$ is given by (I) we see that $a_{i_1} + h + p$ is an entry of $(\mathring{a}_{i_0}, \mathring{a}_{i_0+1}, \mathring{a}_{i_0+2}, \dots)$. Using 2.3 we see that

$$\{\mathring{a}_{i_0}, \mathring{a}_{i_0+1}, \mathring{a}_{i_0+2}, \dots\} = \{\mathring{b}_{i_0}, \mathring{b}_{i_0+1}, \mathring{b}_{i_0+2}, \dots\}$$

as multisets. We see that cases (III) and (IV) cannot arise. Hence we must be in case (I) or (II). Thus we have *either*

(a) $$(b_{i_0}, b_{i_0+1}, b_{i_0+2}, \dots, b_{i_1}) = (a_{i_0}, a_{i_0+1}, a_{i_0+2}, \dots, a_{i_1})$$

or

(b) $$(b_{i_0}, b_{i_0+1}, b_{i_0+2}, \dots, b_{i_1}) = (a_{i_0+1}, a_{i_0}, a_{i_0+3}, a_{i_0+2}, \dots, a_{i_1}, a_{i_1-1}).$$

From 2.3 and (a)+(b) we see that if $\mu = 1$ then Lemma 2.2 holds. Thus in the rest of the proof we can assume that $\mu \ge 2$.

2.5. Let $a' = (a_{i_1+1}, a_{i_1+2}, \dots, a_N)$, $b' = (b_{i_1+1}, b_{i_1+2}, \dots, b_N)$,

$$\mathring{a}' = (a_{i_1+1}, a_{i_1+2}, a_{i_1+3}+1, a_{i_1+4}+1, a_{i_1+5}+2, a_{i_1+6}+2, \dots),$$
$$\mathring{b}' = (b_{i_1+1}, b_{i_1+2}, b_{i_1+3}+1, b_{i_1+4}+1, b_{i_1+5}+2, b_{i_1+6}+2, \dots).$$

From $[\mathring{b}] = [\mathring{a}]$, (a) in 2.3 and (a)+(b) in 2.4 we see that the multiset formed by the entries of $\mathring{a}'$ coincides with the multiset formed by the entries of $\mathring{b}'$. Using the induction hypothesis we see that there exists $X' \subset [2, \mu-1] \cap 2\mathbb{N}$ such that $b' = a'^{X'}$ where $a'^{X'}$ is defined in terms of a', X' in the same way as a^X (see 2.1) was defined in terms of a, X. We set $X = X'$ if we are in case (a) of 2.4 and $X = \{0\} \cup X'$ if we are in case (b). Then we have $b = a^X$ (see again (a) and (b) of 2.4), as required. This completes the proof of Lemma 2.2.

2.6. We shall use the notation of 2.1. Let $\mathfrak{T}$ be the set of all unordered pairs $(\mathfrak{A}, \mathfrak{B})$ where $\mathfrak{A}$, $\mathfrak{B}$ are subsets of $\{0, 1, 2, \dots\}$ and $\mathfrak{A} \mathbin{\dot\cup} \mathfrak{B} = (a_0, a_1, a_2, \dots, a_N)$ as multisets. For example, setting

$$\mathfrak{A}_\varnothing = \{a_i; i \in [0, N] \cap 2\mathbb{N}\} \quad \text{and} \quad \mathfrak{B}_\varnothing = \{a_i; i \in [0, N] \cap (2\mathbb{N}+1)\},$$

we have $(\mathfrak{A}_\varnothing, \mathfrak{B}_\varnothing) \in \mathfrak{T}$. For any subset $\mathfrak{a}$ of $\mathcal{J}$ we consider

$$\mathfrak{A}_\mathfrak{a} = ((\mathcal{J} - \mathfrak{a}) \cap \mathfrak{A}_\varnothing) \cup (\mathfrak{a} \cap \mathfrak{B}_\varnothing) \cup (\mathfrak{A}_\varnothing \cap \mathfrak{B}_\varnothing),$$
$$\mathfrak{B}_\mathfrak{a} = ((\mathcal{J} - \mathfrak{a}) \cap \mathfrak{B}_\varnothing) \cup (\mathfrak{a} \cap \mathfrak{A}_\varnothing) \cup (\mathfrak{A}_\varnothing \cap \mathfrak{B}_\varnothing).$$

Then $(\mathfrak{A}_\mathfrak{a}, \mathfrak{B}_\mathfrak{a}) = (\mathfrak{A}_{\mathcal{J}-\mathfrak{a}}, \mathfrak{A}_{\mathcal{J}-\mathfrak{a}}) \in \mathfrak{T}$ and the map $\mathfrak{a} \mapsto (\mathfrak{A}_\mathfrak{a}, \mathfrak{B}_\mathfrak{a})$ induces a bijection $\bar{\mathcal{P}}(\mathcal{J}) \leftrightarrow \mathfrak{T}$. (Note that if $\mathfrak{a} = \varnothing$ then $(\mathfrak{A}_\mathfrak{a}, \mathfrak{B}_\mathfrak{a})$ agrees with the earlier definition of $(\mathfrak{A}_\varnothing, \mathfrak{B}_\varnothing)$.)

Let $\mathfrak{T}'$ be the set of all $(\mathfrak{A}, \mathfrak{B}) \in \mathfrak{T}$ such that $|\mathfrak{A}| = |\mathfrak{A}_\varnothing|$ and $|\mathfrak{B}| = |\mathfrak{B}_\varnothing|$. Let $\mathcal{P}(\mathcal{J})_1$ be the subspace of $\mathcal{P}(\mathcal{J})$ spanned by the following 2-element subsets of $\mathcal{J}$:

$$\{a_{i_1}, a_{i_2}\}, \{a_{i_3}, a_{i_4}\}, \dots, \{a_{i_{\mu-2}}, a_{i_{\mu-1}}\} \quad \text{if } N \text{ is odd},$$
$$\{a_{i_1}, a_{i_2}\}, \{a_{i_3}, a_{i_4}\}, \dots, \{a_{i_{\mu-1}}, a_{i_\mu}\} \quad \text{if } N \text{ is even}.$$

Let $\mathcal{P}(\mathcal{J})_0$ be the subspace of $\mathcal{P}(\mathcal{J})$ spanned by the following 2-element subsets of $\mathcal{J}$:

$$\{a_{i_0}, a_{i_1}\}, \{a_{i_2}, a_{i_3}\}, \dots, \{a_{i_{\mu-1}}, a_{i_\mu}\} \quad \text{if } N \text{ is odd},$$
$$\{a_{i_0}, a_{i_1}\}, \{a_{i_2}, a_{i_3}\}, \dots, \{a_{i_{\mu-2}}, a_{i_{\mu-1}}\} \quad \text{if } N \text{ is even}.$$

Let $\bar{\mathcal{P}}(\mathcal{J})_0$ and $\bar{\mathcal{P}}(\mathcal{J})_1$ be the images of $\mathcal{P}(\mathcal{J})_0$ and $\mathcal{P}(\mathcal{J})_1$ under the obvious map $\mathcal{P}(\mathcal{J}) \to \bar{\mathcal{P}}(\mathcal{J})$. Then:

(a) *$\bar{\mathcal{P}}(\mathcal{J})_0$ and $\bar{\mathcal{P}}(\mathcal{J})_1$ are opposed Lagrangian subspaces of the symplectic vector space $\bar{\mathcal{P}}_{\mathrm{ev}}(\mathcal{J})$*, (,) (see 0.7); *hence* (,) *defines an identification* $\bar{\mathcal{P}}(\mathcal{J})_1 = \bar{\mathcal{P}}(\mathcal{J})_0^*$, where $\bar{\mathcal{P}}(\mathcal{J})_0^*$ is the vector space dual to $\bar{\mathcal{P}}(\mathcal{J})_0$.

Let $\mathfrak{T}_0$ and $\mathfrak{T}_1$ be the subsets of $\mathfrak{T}$ corresponding to $\bar{\mathcal{P}}(\mathcal{J})_0$ and $\bar{\mathcal{P}}(\mathcal{J})_1$ under the bijection $\bar{\mathcal{P}}(\mathcal{J}) \leftrightarrow \mathfrak{T}$. Note that $\mathfrak{T}_0 \subset \mathfrak{T}'$, $\mathfrak{T}_1 \subset \mathfrak{T}'$, $|\mathfrak{T}_0| = |\mathfrak{T}_1| = 2^{\lfloor \mu/2 \rfloor}$.

For any $X \subset [0, \mu - 1] \cap 2\mathbb{N}$ we set $\mathfrak{a}_X = \bigcup_{s \in X} \{a_{i_s}, a_{i_{s+1}}\} \in \mathcal{P}(\mathcal{J})$. Then $(\mathfrak{A}_{\mathfrak{a}_X}, \mathfrak{B}_{\mathfrak{a}_X})$ is related to a^X in 2.1 as follows:

$$\mathfrak{A}_{\mathfrak{a}_X} = \{a_i^X; i \in [0, N] \cap 2\mathbb{N}\}, \mathfrak{B}_{\mathfrak{a}_X} = \{a_i^X; i \in [0, N] \cap (2\mathbb{N} + 1)\}.$$

2.7. We shall use the notation of 2.1. Let T be the set of all unordered pairs (A, B) where A is a subset of $\{0, 1, 2, \dots\}$, B is a subset of $\{1, 2, 3, \dots\}$, A contains no consecutive integers, B contains no consecutive integers, and $A \dot{\cup} B = (\mathring{a}_0, \mathring{a}_1, \mathring{a}_2, \dots, \mathring{a}_N)$ as multisets. For example, setting

$$A_\varnothing = \{\mathring{a}_i; i \in [0, N] \cap 2\mathbb{N}\} \quad \text{and} \quad B_\varnothing = (\mathring{a}_i; i \in [0, N] \cap (2\mathbb{N} + 1)\},$$

we have $(A_\varnothing, B_\varnothing) \in T$.

For any $(A, B) \in T$ we define (A^-, B^-) as follows: A^- consists of $x_1 < x_2 - 1 < x_3 - 2 < \cdots < x_p - p + 1$ where $x_1 < x_2 < \cdots < x_p$ are the elements of A; B^- consists of $y_1 < y_2 - 1 < \cdots < y_q - q + 1$ where $y_1 < y_2 < \cdots < y_q$ are the elements of B.

We can enumerate the elements of T as in [Lusztig 1984b, 11.5]. Let J be the set of all $c \in \mathbb{N}$ such that c appears exactly once in the sequence

$$(\mathring{a}_0, \mathring{a}_1, \mathring{a}_2, \dots, \mathring{a}_N) = (a_0, a_1, a_2 + 1, a_3 + 1, a_4 + 2, a_5 + 2, \dots).$$

A nonempty subset I of J is said to be an interval if it is of the form $\{i, i + 1, i + 2, \dots, j\}$ with $i - 1 \notin J$, $j + 1 \notin J$. Let $\mathcal{I}$ be the set of intervals of J. For any $s \in [0, \mu - 1] \cap 2\mathbb{N}$, the set $I_s := \{\mathring{a}_{i_s}, \mathring{a}_{i_s+1}, \mathring{a}_{i_s+2}, \dots, \mathring{a}_{i_s+2m_s+1}\}$ is either a single interval or a union of intervals $I_s^1 \sqcup I_s^2 \sqcup \dots \sqcup I_s^{t_s}$ $(t_s \geq 2)$ where $\mathring{a}_{i_s} \in I_s^1$, $\mathring{a}_{i_s+2m_s+1} \in I_s^{t_s}$, $|I_s^1|$, $|I_s^{t_s}|$ are odd, $|I_s^h|$ are even for $h \in [2, t_s - 1]$ and any element in I_s^e is $<$ than any element in $I_s^{e'}$ for $e < e'$. Let $\mathcal{I}_s$ be the set of all $I \in \mathcal{I}$ such that $I \subset I_s$. For any subset $\alpha \subset \mathcal{I}$ we consider

$$A_\alpha = \bigcup_{I \in \mathcal{I} - \alpha} (I \cap A_\varnothing) \cup \bigcup_{I \in \alpha} (I \cap B_\varnothing) \cup (A_\varnothing \cap B_\varnothing),$$
$$B_\alpha = \bigcup_{I \in \mathcal{I} - \alpha} (I \cap B_\varnothing) \cup \bigcup_{I \in \alpha} (I \cap A_\varnothing) \cup (A_\varnothing \cap B_\varnothing).$$

Then $(A_\alpha, B_\alpha) \in T$ and the map $\alpha \mapsto (A_\alpha, B_\alpha)$ is a bijection $\bar{\mathcal{P}}(\mathcal{I}) \leftrightarrow T$. (Note that if $\alpha = \varnothing$ then (A_α, B_α) agrees with the earlier definition of $(A_\varnothing, B_\varnothing)$.)

Let

$$T' = \{(A, B) \in T; |A| = |A_\varnothing|, |B| = |B_\varnothing|\},$$
$$T_1 = \{(A, B) \in T'; A^- \dot{\cup} B^- = A_\varnothing^- \dot{\cup} B_\varnothing^-\}.$$

Let $\bar{\mathcal{P}}(\mathcal{I})'$ and $\bar{\mathcal{P}}(\mathcal{I})_1$ be the subsets of $\bar{\mathcal{P}}(\mathcal{I})$ corresponding to T' and T_1 under the bijection $\bar{\mathcal{P}}(\mathcal{I}) \leftrightarrow T$.

Now let X be a subset of $[0, \mu-1]\cap 2\mathbb{N}$. Let $\alpha_X = \bigcup_{s\in X} \mathcal{I}_s \in \mathcal{P}(\mathcal{I})$. From the definitions we see that

(a) $$A^-_{\alpha_X} = \mathfrak{A}_{\mathfrak{a}_X}, \quad B^-_{\alpha_X} = \mathfrak{B}_{\mathfrak{a}_X}$$

(in the notation of 2.6). In particular we have $(A_{\alpha_X}, B_{\alpha_X}) \in T_1$. Thus $|T_1| \geq 2^{\lfloor \mu/2 \rfloor}$. Using Lemma 2.2 we see that

(b) $$|T_1| = 2^{\lfloor \mu/2 \rfloor} \quad \textit{and} \quad T_1 = \big\{(A_{\alpha_X}, B_{\alpha_X});\, X \subset [0, \mu-1]\cap 2\mathbb{N}\big\}.$$

Using (a) and (b) we deduce:

(c) *The map $T_1 \to \mathfrak{T}_1$ given by $(A, B) \mapsto (A^-, B^-)$ is a bijection.*

3. Proof of Theorem 0.4 and of Corollary 0.5

3.1. If G is simple adjoint of type A_n, $n \geq 1$, then Theorem 0.4 and Corollary 0.5 are obvious: we have $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.

3.2. Assume that $G = Sp_{2n}(\Bbbk)$ where $n \geq 2$. Let N be a sufficiently large even integer. Now $u : \Bbbk^{2n} \to \Bbbk^{2n}$ has i_e Jordan blocks of size e ($e = 1, 2, 3, \dots$). Here $i_1, i_3, i_5, \dots$ are even. Let $\Delta = \{e \in \{2, 4, 6, \dots\}; i_e \geq 1\}$. Then $A(u)$ can be identified in the standard way with $\mathcal{P}(\Delta)$. Hence the group of characters $\hat{A}(u)$ of $A(u)$ (which may be canonically identified with the F_2-vector space dual to $\mathcal{P}(\Delta)$) may be also canonically identified with $\mathcal{P}(\Delta)$ itself (so that the basis given by the one-element subsets of Δ is self-dual).

To the partition $1i_1+2i_2+3i_3+\cdots$ of $2n$ we associate a pair (A, B) as in [Lusztig 1984b, 11.6] (with N, $2m$ replaced by $2n$, N). We have $A = (\hat{a}_0, \hat{a}_2, \hat{a}_4, \dots, \hat{a}_N)$, $B = (\hat{a}_1, \hat{a}_3, \dots, \hat{a}_{N-1})$, where $\hat{a}_0 \leq \hat{a}_1 \leq \hat{a}_2 \leq \cdots \leq \hat{a}_N$ is obtained from a sequence $a_0 \leq a_1 \leq a_2 \leq \cdots \leq a_N$ as in 1.1. (Here we use that C is special.) Now the definitions and results in Section 1 are applicable. As in [Lusztig 1984a, 4.5] the family $\mathcal{F}$ is in canonical bijection with $\mathfrak{T}'$ in 1.6.

We arrange the intervals in $\mathcal{I}$ in increasing order $I_{(1)}, I_{(2)}, \dots, I_{(f)}$ (any element in $I_{(1)}$ is smaller than any element in $I_{(2)}$, etc.). We arrange the elements of Δ in increasing order $e_1 < e_2 < \cdots < e_{f'}$; then $f = f'$ and we have a bijection $\mathcal{I} \leftrightarrow \Delta$, $I_{(h)} \leftrightarrow e_h$; moreover we have $|I_{(h)}| = i_{e_h}$ for $h \in [1, f]$; see [Lusztig 1984b, 11.6]. Using this bijection we see that $A(u)$ and $\hat{A}(u)$ are identified with the F_2-vector space $\mathcal{P}(\mathcal{I})$ with basis given by the one-element subsets of $\mathcal{I}$. Let $\pi : \mathcal{P}(\mathcal{I}) \to \mathcal{P}(\mathcal{I})_1^*$ (the dual of $\mathcal{P}(\mathcal{I})_1$ in 1.7) be the (surjective) F_2-linear map which to $X \subset \mathcal{I}$ associates the linear form $L \mapsto |X \cap L| \bmod 2$ on $\mathcal{P}(\mathcal{I})_1$. We will show that

(a) $$\ker \pi = \mathcal{K}(u), \quad \text{with } \mathcal{K}(u) \text{ as in 0.1.}$$

We identify $\mathrm{Irr}_C W$ with T' (see 1.7) via the restriction of the bijection in [Lusztig 1984b, (12.2.4)] (we also use the description of the Springer correspondence in [Lusztig 1984b, 12.3]). Under this identification the subset $\mathrm{Irr}_C^* W$ of $\mathrm{Irr}_C W$ becomes the subset T_1 (see 1.7) of T'. Via the identification $\mathcal{P}(\mathcal{I})' \leftrightarrow T'$ in 1.7 and $\hat{A}(u) \leftrightarrow \mathcal{P}(\mathcal{I})$ (see above), the map $E \mapsto \mathcal{V}_E$ from T' to $\hat{A}(u)$ becomes the obvious imbedding $\mathcal{P}(\mathcal{I})' \to \mathcal{P}(\mathcal{I})$ (we use again [Lusztig 1984b, 12.3]). By definition, $\mathcal{K}(u)$ is the set of all $X \in \mathcal{P}(\mathcal{I})$ such that for any $L \in \mathcal{P}(\mathcal{I})_1$ we have $|X \cap L| = 0 \bmod 2$. Thus, (a) holds.

Using (a) we have canonically $\bar{A}(u) = \mathcal{P}(\mathcal{I})_1^*$ via π. We define an F_2-linear map $\mathcal{P}(\mathcal{I})_1 \to \bar{\mathcal{P}}(\mathcal{J})_1$ (see 1.6) by $I_s \mapsto \{a_{i_s}, a_{i_{s+1}}\}$ for $s \in \{1, 3, \dots, 2M-1\}$ (I_s as in 1.7). This is an isomorphism; it corresponds to the bijection 1.7(c) under the identification $T_1 \leftrightarrow \mathcal{P}(\mathcal{I})_1$ in 1.7 and the identification $\mathfrak{T}_1 \leftrightarrow \bar{\mathcal{P}}(\mathcal{J})_1$ in 1.6. Hence we can identify $\mathcal{P}(\mathcal{I})_1^*$ with $\bar{\mathcal{P}}(\mathcal{J})_1^*$ and with $\bar{\mathcal{P}}(\mathcal{J})_0$ (see 1.6(a)). We obtain an identification $\bar{A}(u) = \bar{\mathcal{P}}(\mathcal{J})_0$.

By [Lusztig 1984a, 4.5] we have $X_{\mathcal{F}} = \bar{\mathcal{P}}(\mathcal{J})$. Using 1.6(a) we see that $\bar{\mathcal{P}}(\mathcal{J}) = M(\bar{\mathcal{P}}(\mathcal{J})_0) = M(\bar{A}(u))$ canonically so that Theorem 0.4 holds in our case. From the arguments above we see that in our case Corollary 0.5 follows from 1.7(c).

3.3. Assume that $G = \mathrm{SO}_n(\Bbbk)$ where $n \ge 7$. Let N be a sufficiently large integer such that $N = n \bmod 2$. Now $u : \Bbbk^n \to \Bbbk^n$ has i_e Jordan blocks of size e ($e = 1, 2, 3, \dots$). Here $i_2, i_4, i_6, \dots$ are even. Let $\Delta = \{e \in \{1, 3, 5, \dots\}; i_e \ge 1\}$. If $\Delta = \varnothing$ then $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$ and $\mathcal{G}_{\mathcal{F}} = \{1\}$ so that the result is trivial.

In the remainder of this subsection we assume that $\Delta \neq \varnothing$. Then $A(u)$ can be identified in the standard way with the F_2-subspace $\mathcal{P}_{\mathrm{ev}}(\Delta)$ of $\mathcal{P}(\Delta)$ and the group of characters $\hat{A}(u)$ of $A(u)$ (which may be canonically identified with the F_2-vector space dual to $A(u)$) becomes $\bar{\mathcal{P}}(\Delta)$; the obvious pairing $A(u) \times \hat{A}(u) \to F_2$ is induced by the inner product $L, L' \mapsto |L \cap L'| \bmod 2$ on $\mathcal{P}(\Delta)$.

To the partition $1i_1 + 2i_2 + 3i_3 + \cdots$ of n we associate a pair (A, B) as in [Lusztig 1984b, 11.7] (with N, M replaced by n, N). We have $A = \{\mathring{a}_i; i \in [0, N] \cap 2\mathbb{N}\}$, $B = (\mathring{a}_i; i \in [0, N] \cap (2\mathbb{N}+1)\}$ where $\mathring{a}_0 \le \mathring{a}_1 \le \mathring{a}_2 \le \cdots \le \mathring{a}_N$ is obtained from a sequence $a_0 \le a_1 \le a_2 \le \cdots \le a_N$ as in 2.1. (Here we use that C is special.) Now the definitions and results in §2 are applicable. As in [Lusztig 1984a, 4.5] (if N is even) or [Lusztig 1984a, 4.6] (if N is odd) the family $\mathcal{F}$ is in canonical bijection with $\mathfrak{T}'$ in 2.6.

We arrange the intervals in $\mathcal{I}$ in increasing order $I_{(1)}, I_{(2)}, \dots, I_{(f)}$ (any element in $I_{(1)}$ is smaller than any element in $I_{(2)}$, etc.). We arrange the elements of Δ in increasing order $e_1 < e_2 < \cdots < e_{f'}$; then $f = f'$ and we have a bijection $\mathcal{I} \leftrightarrow \Delta$, $I_{(h)} \leftrightarrow e_h$; moreover we have $|I_{(h)}| = i_{e_h}$ for $h \in [1, f]$; see [Lusztig 1984b, 11.7]. Using this bijection we see that $A(u)$ is identified with $\mathcal{P}_{\mathrm{ev}}(\mathcal{I})$ and $\hat{A}(u)$ is identified with $\bar{\mathcal{P}}(\mathcal{I})$. For any $X \in \mathcal{P}_{\mathrm{ev}}(\mathcal{I})$, the assignment $L \mapsto |X \cap L| \bmod 2$ can be viewed as an element of $\bar{\mathcal{P}}(\mathcal{I})_1^*$ (the dual space of $\bar{\mathcal{P}}(\mathcal{I})_1$ in 2.7 which by 2.7(b) is an F_2-vector space of dimension $2^{\lfloor \mu/2 \rfloor}$). This induces a (surjective) F_2-linear map $\pi : \mathcal{P}_{\mathrm{ev}}(\mathcal{I}) \to \bar{\mathcal{P}}(\mathcal{I})_1^*$. We will show that

(a) $$\ker \pi = \mathcal{K}(u), \quad \text{with } \mathcal{K}(u) \text{ as in 0.1.}$$

We identify $\mathrm{Irr}_C W$ with T' (see 2.7) via the restriction of the bijection in [Lusztig 1984b, (13.2.5)] if N is odd or [ibid., (13.2.6)] if N is even (we also use the description of the Springer correspondence in [Lusztig 1984b, 13.3]). Under this identification the subset $\mathrm{Irr}_C^* W$ of $\mathrm{Irr}_C W$ becomes the subset T_1 (see 2.7) of T'. Via the identification $\bar{\mathcal{P}}(\mathcal{I})' \leftrightarrow T'$ in 2.7 and $\hat{A}(u) \leftrightarrow \bar{\mathcal{P}}(\mathcal{I})$ (see above), the map $E \mapsto \mathcal{V}_E$ from T' to $\hat{A}(u)$ becomes the obvious imbedding $\bar{\mathcal{P}}(\mathcal{I})_0 \to \bar{\mathcal{P}}(\mathcal{I})$ (we use again [ibid., 13.3]). By definition, $\mathcal{K}(u)$ is the set of all $X \in \mathcal{P}_{\mathrm{ev}}(\mathcal{I})$ such that for any $L \in \mathcal{P}(\mathcal{I})$ representing a vector in $\bar{\mathcal{P}}(\mathcal{I})_1$ we have $|X \cap L| = 0 \bmod 2$. Thus, (a) holds.

Using (a) we have canonically $\bar{A}(u) = \bar{\mathcal{P}}(\mathcal{I})_1^*$ via π. We have an F_2-linear map $\bar{\mathcal{P}}(\mathcal{I})_1 \to \bar{\mathcal{P}}(\mathcal{J})_0$ (see 2.6) induced by $I_s \mapsto \{a_{i_s}, a_{i_{s+1}}\}$ for $s \in [0, \mu-1] \cap 2\mathbb{N}$ (I_s as in 2.7). This is an isomorphism; it corresponds to the bijection 2.7(c) under the identification $T_1 \leftrightarrow \bar{\mathcal{P}}(\mathcal{I})_1$ in 2.7 and the identification $\mathfrak{T}_1 \leftrightarrow \bar{\mathcal{P}}(\mathcal{J})_0$ in 2.6. Hence we can identify $\bar{\mathcal{P}}(\mathcal{I})_1^*$ with $\bar{\mathcal{P}}(\mathcal{J})_0^*$ and with $\bar{\mathcal{P}}(\mathcal{J})_1$ (see 2.6(a)). We obtain an identification $\bar{A}(u) = \bar{\mathcal{P}}(\mathcal{J})_1$.

By [Lusztig 1984a, 4.6] we have $X_{\mathcal{F}} = \bar{\mathcal{P}}_{\mathrm{ev}}(\mathcal{J})$. Using 2.6(a) we see that $\bar{\mathcal{P}}(\mathcal{J}) = M(\bar{\mathcal{P}}(\mathcal{J})_1) = M(\bar{A}(u))$ canonically so that Theorem 0.4 holds in our case. From the arguments above we see that in our case Corollary 0.5 follows from 2.7(c).

3.4. In 3.5–3.9 we consider the case where G is simple adjoint of exceptional type. In each case we list the elements of the set $\mathrm{Irr}_C W$ for each special unipotent class C of G; an element e of $\mathrm{Irr}_C W - \mathrm{Irr}_C^* W$ is listed as $[e]$. (The notation for the various C is as in [Spaltenstein 1985]; the notation for the objects of $\mathrm{Irr} W$ is as in [Spaltenstein 1985] (for type E_n) and as in [Lusztig 1984a, 4.10] for type F_4.) In each case the structure of $A(u)$, $\bar{A}(u)$ (for $u \in C$) is indicated; here S_n denotes the symmetric group in n letters. The order in which we list the objects in $\mathrm{Irr}_C W$ corresponds to the following order of the irreducible representations of $A(u) = S_n$:

$$\begin{aligned}
&1, \epsilon \ (n=2),\\
&1, r, \epsilon \ (n=3,\ G \neq G_2),\\
&1, r \ (n=3,\ G = G_2),\\
&1, \lambda^1, \lambda^2, \sigma \ (n=4),\\
&1, \nu, \lambda^1, \nu', \lambda^2, \lambda^3 \ (n=5),
\end{aligned}$$

in the notation of [Lusztig 1984a, 4.3]. Now Theorem 0.4 and Corollary 0.5 follow in our case from the tables in 3.5–3.9 and the definitions in [Lusztig 1984a, 4.8–4.13]. (In those tables S_n is the symmetric group in n letters.)

3.5. Assume that G is of type E_8.

$\mathrm{Irr}_{E_8} W = \{1_0\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{E_8(a_1)} W = \{8_1\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{E_8(a_2)} W = \{35_2\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{E_7A_1} W = \{112_3, 28_8\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_8} W = \{210_4, 160_7\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{E_7(a_1)A_1} W = \{560_5, [50_8]\}$; $A(u) = S_2$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{E_7(a_1)} W = \{567_6\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_8(a_1)} W = \{700_6, 300_8\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{E_7(a_2)A_1} W = \{1400_7, 1008_9, 56_{19}\}$; $A(u) = S_3$, $\bar{A}(u) = S_3$.
$\mathrm{Irr}_{A_8} W = \{1400_8, 1575_{10}, 350_{14}\}$; $A(u) = S_3$, $\bar{A}(u) = S_3$.
$\mathrm{Irr}_{D_7(a_1)} W = \{3240_9, [1050_{10}]\}$; $A(u) = S_2$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_8(a_3)} W = \{2240_{10}, [175_{12}], 840_{13}\}$; $A(u) = S_3$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_6A_1} W = \{2268_{10}, 1296_{13}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{E_6(a_1)A_1} W = \{4096_{11}, 4096_{12}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{E_6} W = \{525_{12}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_7(a_2)} W = \{4200_{12}, 3360_{13}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{E_6(a_1)} W = \{2800_{13}, 2100_{16}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_5A_2} W = \{4536_{13}, [840_{14}]\}$; $A(u) = S_2$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_6(a_1)A_1} W = \{6075_{14}, [700_{16}]\}$; $A(u) = S_2$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_6A_1} W = \{2835_{14}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_6} W = \{4200_{15}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_6(a_1)} W = \{5600_{15}, 2400_{17}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{2A_4} W = \{4480_{16}, 4536_{18}, 5670_{18}, 1400_{20}, 1680_{22}, 70_{32}\}$; $A(u) = S_5$, $\bar{A}(u) = S_5$.
$\mathrm{Irr}_{D_5} W = \{2100_{20}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{(A_5A_1)''} W = \{5600_{21}, 2400_{23}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_4A_2} W = \{4200_{15}, [168_{24}]\}$; $A(u) = S_2$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_4A_2A_1} W = \{2835_{22}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_4A_2} W = \{4536_{23}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_5(a_1)} W = \{2800_{25}, 2100_{28}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{A_42A_1} W = \{4200_{24}, 3360_{25}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_4} W = \{525_{36}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_4A_1} W = \{4096_{26}, 4096_{27}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{A_4} W = \{2268_{30}, 1296_{33}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_4(a_1)A_2} = \{2240_{28}, 840_{31}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{A_3A_2} W = \{3240_{31}, [972_{32}]\}$; $A(u) = S_2$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_4(a_1)A_1} W = \{1400_{32}, 1575_{34}, 350_{38}\}$; $A(u) = S_3$, $\bar{A}(u) = S_3$.
$\mathrm{Irr}_{D_4(a_1)} W = \{1400_{37}, 1008_{39}, 56_{49}\}$; $A(u) = S_3$, $\bar{A}(u) = S_3$.

$\mathrm{Irr}_{2A_2} W = \{700_{42}, 300_{44}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{A_3} W = \{567_{46}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_2 2A_1} W = \{560_{47}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_2 A_1} W = \{210_{52}, 160_{55}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{A_2} W = \{112_{63}, 28_{68}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{2A_1} W = \{35_{74}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_1} W = \{8_{91}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{\varnothing} W = \{1_{120}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.

3.6. Assume that G is adjoint of type E_7.
$\mathrm{Irr}_{E_7} W = \{1_0\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{E_7(a_1)} W = \{7_1\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{E_7(a_2)} W = \{27_2\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_6 A_1} W = \{56_3, 21_6\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{E_6} W = \{21_3\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{E_6(a_1)} W = \{120_4, 105_5\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_6(a_1) A_1} W = \{189_5, [15_7]\}$; $A(u) = S_2$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_6(a_1)} W = \{210_6\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_6} W = \{105_6\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_5 A_1} W = \{168_6\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_5} W = \{189_7\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_6(a_2) A_1} W = \{315_7, 280_9, 35_{13}\}$; $A(u) = S_3$, $\bar{A}(u) = S_3$.
$\mathrm{Irr}_{(A_5 A_1)'} = \{405_8, 189_{10}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_5(a_1) A_1} W = \{378_9\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_4 A_2} W = \{210_{10}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_5(a_1)} W = \{420_{10}, 336_{11}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{A_5''} W = \{105_{12}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_4 A_1} W = \{512_{11}, 512_{12}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_4} W = \{105_{15}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_4} W = \{420_{13}, 336_{14}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{A_3 A_2 A_1} W = \{210_{13}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_3 A_2} W = \{378_{14}, [84_{15}]\}$; $A(u) = S_2$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_4(a_1) A_1} W = \{405_{15}, 189_{17}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_4(a_1)} W = \{315_{16}, 280_{18}, 35_{22}\}$; $A(u) = S_3$, $\bar{A}(u) = S_3$.
$\mathrm{Irr}_{(A_3 A_1)''} W = \{189_{20}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{2A_2} W = \{168_{21}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_2 3A_1} W = \{105_{21}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_3} W = \{210_{21}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.

$\mathrm{Irr}_{A_2 2A_1} W = \{189_{22}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_2 A_1} W = \{120_{25}, 105_{26}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{3A_1''} W = \{21_{36}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_2} W = \{56_{30}, 21_{33}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{2A_1} W = \{27_{37}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_1} W = \{7_{46}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{\varnothing} W = \{1_{63}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.

3.7. Assume that G is adjoint of type E_6.
$\mathrm{Irr}_{E_6} W = \{1_0\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{E_6(a_1)} W = \{6_1\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_5} W = \{20_2\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_5 A_1} W = \{30_3, 15_5\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{D_5(a_1)} W = \{64_4\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_4 A_1} W = \{60_5\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_4} W = \{81_6\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_4} W = \{24_6\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{D_4(a_1)} W = \{80_7, 90_8, 20_{10}\}$; $A(u) = S_3$, $\bar{A}(u) = S_3$.
$\mathrm{Irr}_{2A_2} W = \{24_{12}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_3} W = \{81_{10}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_2 2A_1} W = \{60_{11}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_2 A_1} w = \{64_{13}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_2} W = \{30_{15}, 15_{17}\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{2A_1} W = \{20_{20}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_1} W = \{6_{25}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{\varnothing} W = \{1_{36}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.

3.8. Assume that G is of type F_4.
$\mathrm{Irr}_{F_4} W = \{1_1\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{F_4(a_1)} W = \{4_2, 2_3\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.
$\mathrm{Irr}_{F_4(a_2)} W = \{9_1\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{B_3} W = \{8_1\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{C_3} W = \{8_3\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{F_4(a_3)} W = \{12_1, 9_3, 6_2, 1_3\}$; $A(u) = S_4$, $\bar{A}(u) = S_4$.
$\mathrm{Irr}_{\tilde{A}_2} W = \{8_2\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_2} W = \{8_4, [1_2]\}$; $A(u) = S_2$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{A_1 \tilde{A}_1} W = \{9_4\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.
$\mathrm{Irr}_{\tilde{A}_1} W = \{4_5, 2_2\}$; $A(u) = S_2$, $\bar{A}(u) = S_2$.

$\mathrm{Irr}_{\varnothing} W = \{1_4\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.

3.9. Assume that G is of type G_2.

$\mathrm{Irr}_{G_2} W$ is the unit representation; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.

$\mathrm{Irr}_{G_2(a_1)} W$ consists of the reflection representation and the one dimensional representation on which the reflection with respect to a long (resp. short) simple coroot acts nontrivially (resp. trivially); $A(u) = S_3$, $\bar{A}(u) = S_3$.

$\mathrm{Irr}_{\varnothing} W = \{\mathrm{sgn}\}$; $A(u) = \{1\}$, $\bar{A}(u) = \{1\}$.

3.10. This completes the proof of Theorem 0.4 and that of Corollary 0.5.

We note that the definition of $\mathcal{G}_{\mathcal{F}}$ given in [Lusztig 1984a] (for type C_n, B_n) is $\bar{\mathcal{P}}(\mathcal{J})_1$ (in the setup of 3.2) and $\bar{\mathcal{P}}(\mathcal{J})_0$ (in the setup of 3.3) which is noncanonically isomorphic to $\bar{A}(u)$, unlike the definition adopted here that is, $\bar{\mathcal{P}}(\mathcal{J})_0$ (in the setup of 3.2) and $\bar{\mathcal{P}}(\mathcal{J})_1$ (in the setup of 3.3) which makes $\mathcal{G}_{\mathcal{F}}$ canonically isomorphic to $\bar{A}(u)$.

References

[Alvis and Lusztig 1982] D. Alvis and G. Lusztig, "On Springer's correspondence for simple groups of type E_n ($n = 6, 7, 8$)", *Math. Proc. Cambridge Philos. Soc.* **92**:1 (1982), 65–72. MR 83k:20040 Zbl 0492.20028

[Borho and MacPherson 1981] W. Borho and R. MacPherson, "Représentations des groupes de Weyl et homologie d'intersection pour les variétés nilpotentes", *C. R. Acad. Sci. Paris Sér. I Math.* **292**:15 (1981), 707–710. MR 82f:14002 Zbl 0467.20036

[Losev and Ostrik 2012] I. Losev and V. Ostrik, "Classification of finite dimensional irreducible modules over W-algebras", preprint, 2012. arXiv 1202.6097

[Lusztig 1979a] G. Lusztig, "A class of irreducible representations of a Weyl group", *Indag. Math. (Proc.)* **82**:1 (1979), 323–335. MR 81a:20052 Zbl 0435.20021

[Lusztig 1979b] G. Lusztig, "Unipotent representations of a finite Chevalley group of type E_8", *Quart. J. Math. Oxford Ser.* (2) **30**:119 (1979), 315–338. MR 80j:20041 Zbl 0418.20038

[Lusztig 1984a] G. Lusztig, *Characters of reductive groups over a finite field*, Annals of Mathematics Studies **107**, Princeton University Press, Princeton, NJ, 1984. MR 86j:20038 Zbl 0556.20033

[Lusztig 1984b] G. Lusztig, "Intersection cohomology complexes on a reductive group", *Invent. Math.* **75**:2 (1984), 205–272. MR 86d:20050 Zbl 0547.20032

[Lusztig 1992] G. Lusztig, "A unipotent support for irreducible representations", *Adv. Math.* **94**:2 (1992), 139–179. MR 94a:20073 Zbl 0789.20042

[Shoji 1979a] T. Shoji, "On the Springer representations of the Weyl groups of classical algebraic groups", *Comm. Algebra* **7**:16 (1979), 1713–1745. MR 81h:20007a Zbl 0423.20042

[Shoji 1979b] T. Shoji, "Correction to 'On the Springer representations of the Weyl groups of classical algebraic groups'", *Comm. Algebra* **7**:18 (1979), 2027–2033. MR 81h:20007b Zbl 0426.20028

[Shoji 1980] T. Shoji, "On the Springer representations of Chevalley groups of type F_4", *Comm. Algebra* **8**:5 (1980), 409–440. MR 81b:20013 Zbl 0434.20026

[Spaltenstein 1982] N. Spaltenstein, "Appendix", *Math. Proc. Cambridge Philos. Soc.* **92**:1 (1982), 73–78. MR 83k:20040 Zbl 0492.20028

[Spaltenstein 1985] N. Spaltenstein, "On the generalized Springer correspondence for exceptional groups", pp. 317–338 in *Algebraic groups and related topics* (Kyoto/Nagoya, 1983), edited by R. Hotta, Adv. Stud. Pure Math. **6**, North-Holland, Amsterdam, 1985. MR 87g:20072b Zbl 0574.20029

[Springer 1976] T. A. Springer, "Trigonometric sums, Green functions of finite groups and representations of Weyl groups", *Invent. Math.* **36** (1976), 173–207. MR 56 #491 Zbl 0374.20054

Received March 5, 2012. Revised May 11, 2012.

GEORGE LUSZTIG
DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ROOM 2-276
CAMBRIDGE, MA 02139-4307
UNITED STATES
gyuri@math.mit.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.451

REFLEXIVE OPERATOR ALGEBRAS ON BANACH SPACES

FLORENCE MERLEVÈDE, COSTEL PELIGRAD AND MAGDA PELIGRAD

In this paper we study the reflexivity of a unital strongly closed algebra of operators with complemented invariant subspace lattice on a Banach space. We prove that if such an algebra contains a complete Boolean algebra of projections of finite uniform multiplicity and with the direct sum property, then it is reflexive, i.e., it contains every operator that leaves invariant every closed subspace in the invariant subspace lattice of the algebra. In particular, such algebras coincide with their bicommutant.

1. Introduction

Let $A \subset B(X)$ denote a strongly closed algebra of operators on the Banach space X. Suppose that A has the property that each of its invariant subspaces has an invariant complement. If A contains a complete Boolean algebra of projections of finite uniform multiplicity and with the direct sum property as defined below, we prove that A is reflexive in the sense that it contains all the operators which leave its closed invariant subspaces invariant (Theorem 15). In particular such an algebra is equal to its bicommutant A'' (Corollary 22). The problem of whether a strongly closed algebra of operators with complemented invariant subspace lattice is reflexive started to be studied in the sixties. This problem is a generalization of the invariant subspace problem in operator theory. Arveson [1967] introduced a technique for studying the particular case of transitive algebras on Hilbert spaces, namely the strongly closed algebras of operators on Hilbert spaces that have no nontrivial closed invariant subspaces. He proved that every transitive algebra that contains a maximal abelian von Neumann algebra coincides with the full algebra $B(X)$ if X is a complex Hilbert space. Douglas and Pearcy [1972] extended the result of Arveson to the case of transitive operator algebras containing an abelian von Neumann algebra of finite multiplicity. Hoover [1973] extended the result of Douglas and Pearcy to the case of reductive operator algebras on Hilbert spaces that contain

C. Peligrad and M. Peligrad were supported in part by a Charles Phelps Taft Memorial Fund grant. M. Peligrad was also supported by NSF grant DMS-1208237.
MSC2010: primary 47A15, 47B48; secondary 47C05.
Keywords: operator algebras, invariant subspace lattice, Boolean algebra of projections, spectral operator.

abelian von Neumann algebras of finite multiplicity. Hoover proved that every reductive operator algebra (that is a strongly closed subalgebra for which every closed invariant subspace is reducing) which contains an abelian von Neumann algebra of finite multiplicity is self-adjoint. The transitive algebra result of Douglas and Pearcy was generalized in [Önder and Orhon 1989] to the case of transitive algebras on Banach spaces that contain a n-fold direct sum of a cyclic complete Boolean algebra of projections. The case of operator algebras on Banach spaces with complemented invariant subspace lattice was considered by Rosenthal and Sourour [1977]. They proved that every strongly closed algebra of operators with complemented invariant subspace lattice containing a complete Boolean algebra of projections of uniform multiplicity one is reflexive.

In this paper we build upon the techniques introduced by Arveson and developed in [Douglas and Pearcy 1972; Radjavi and Rosenthal 1973] for invariant subspaces of operator algebras as well as Bade's multiplicity theory of Boolean algebras of projections [Bade 1955; 1959]. We also use the results of [Foguel 1959] and [Tzafriri 1967] about the commutant of Boolean algebras of projections of finite multiplicity.

2. Notation and preliminary results

2.1. *Invariant subspaces of operator algebras.* Let X be a complex Banach space and $B(X)$ the algebra of all bounded linear operators on X. We will denote by $X^{(n)}$ the direct sum of n copies of X and, if $S \subset B(X)$, we set

$$S^{(n)} = \{a \oplus a \oplus \cdots \oplus a \in B(X^{(n)}) \mid a \in S\}.$$

If $S \subset B(Y)$, where Y is a Banach space, we denote by $\operatorname{Lat} S$ the collection of all closed linear subspaces of Y that are invariant under every element of S. If L is a collection of closed linear subspaces of Y, we denote by $\operatorname{alg} L$ the (strongly closed) algebra of operators on Y that leave every element of L invariant. An algebra $A \subset B(X)$ is called *reflexive* if $\operatorname{algLat} A = A$.

In what follows all the subalgebras $A \subset B(X)$ will be assumed to be strongly closed and containing the identity operator $I \in B(X)$.

Remark 1. Let $A \subset B(X)$ be a strongly closed algebra with $I \in A$ and $b \in B(X)$. If $\operatorname{Lat} A^{(n)} \subset \operatorname{Lat} b^{(n)}$ for every $n \in \mathbb{N}$, then $b \in A$.

Proof. Indeed, then for every finite set of elements $\{x_1, x_2, \dots, x_n\} \subset X$ we have that $K = \overline{\{ax_1 \oplus ax_2 \oplus \cdots \oplus ax_n \mid a \in A\}} \in \operatorname{Lat} A^{(n)}$ and therefore $K \in \operatorname{Lat} b^{(n)}$. This means that $b \in A$, since A is strongly closed. □

Proposition 2. *Let $A \subset B(X)$ be a strongly closed algebra with complemented invariant subspace lattice and with $I \in A$. Let $q \in B(X)$ be a projection.*

(i) *If $q \in A$, the algebra $qAq \subset B(qX)$ has complemented invariant subspace lattice and* $\operatorname{algLat}(qAq) = q(\operatorname{algLat} A)q$.

(ii) *If $q \in A'$, where A' denotes the commutant of A, the strong operator closure $\overline{qAq}^{\mathrm{so}} \subset B(qX)$ is an algebra with complemented invariant subspace lattice.*

Proof. We prove first (i). Clearly, qAq is a strongly closed subalgebra of $B(qX)$ whose unit is q. Let $L \subset qX$, $L \in \operatorname{Lat}(qAq)$. We define $\tilde{L} = \overline{AqL}$, the closure being taken in X. Then, obviously, $\tilde{L} \in \operatorname{Lat} A$ and, therefore $\tilde{L}$ has a complement $\tilde{L}^c$ in $\operatorname{Lat} A$. Since $q \in A$, we have $q\tilde{L} \subset \tilde{L}$, also $q\tilde{L}^c \subset \tilde{L}^c$ and $q\tilde{L}^c \in \operatorname{Lat}(qAq)$. Moreover, it is immediate that $q\tilde{L}$ and $q\tilde{L}^c$ are closed linear subspaces of qX such that

$$q\tilde{L} \oplus q\tilde{L}^c = qX.$$

On the other hand we have $L \subset q\tilde{L} = q\overline{AqL} \subset \overline{qAqL} \subset \bar{L} = L$. Hence

$$q\tilde{L} = L.$$

It follows that L is complemented in $\operatorname{Lat}(qAq)$ and so qAq has complemented invariant subspace lattice. Let now $b \in \operatorname{algLat} A$ and $L \in \operatorname{Lat}(qAq)$. By the above argument, there exists $\widetilde{L} \in \operatorname{Lat} A$ such that $L = q\tilde{L}$. Hence $b\widetilde{L} \subset \tilde{L}$. Therefore, since $q\tilde{L} = L \subset \tilde{L}$ it follows that $qbqL \subset L$, so $qbq \in \operatorname{algLat}(qAq)$. Conversely, let $c \in \operatorname{algLat}(qAq)$ and let $\tilde{c} \in B(X)$ be the extension of c to X that equals 0 on $(I-q)X$. Then, it is straightforward to show that $\tilde{c} \in \operatorname{algLat} A$ and $c = q\tilde{c}q$ and so the proof is completed.

To establish (ii), let $K \in \operatorname{Lat}(qAq)$. Since $q \in A'$, it follows that $K \in \operatorname{Lat} A$ and therefore K has a complement $K^c \in \operatorname{Lat} A$. Then, clearly $K^c \cap qX \in \operatorname{Lat}(qAq)$ and $K + K^c \cap qX = qX$. $\square$

We will also need the following:

Lemma 3. *Let $A \subset B(X)$ be an algebra with complemented invariant subspace lattice and let $K \in \operatorname{Lat} A$. If $p \in A'$ is the projection on K and $t_1, t_2, \ldots, t_n \in (pAp)'$, for some $n \in \mathbb{N}$, then the subspace*

$$\Gamma_{\{t_1,t_2,\ldots,t_n;p\}} = \left\{x \oplus t_1x \oplus t_2x \oplus \cdots \oplus t_nx \;\middle|\; x \in pX\right\} \in \operatorname{Lat} A^{(n+1)}$$

is complemented in $\operatorname{Lat} A^{(n+1)}$.

Proof. Since A has complemented invariant subspace lattice and $pX = K \in \operatorname{Lat} A$, it follows that the subspace $(1-p)X = (pX)^c = K^c$ belongs to $\operatorname{Lat} A$. It is then clear that $(pX)^c \oplus X^{(n)}$ is a complement of $\Gamma_{\{t_1,t_2,\ldots,t_n;p\}}$ in $\operatorname{Lat} A^{(n+1)}$. $\square$

Remark 4. Let $I \in A \subset B(X)$ be a strongly closed subalgebra with complemented invariant subspace lattice. If A is reflexive, then $A'' = A$ where A'' denotes the bicommutant of A.

Proof. If $a \in A''$, then, in particular, a commutes with every projection on an invariant subspace of A. Therefore $a \in \operatorname{algLat} A = A$. □

The following concept is defined in [Radjavi and Rosenthal 1973, §8.2], for instance.

Definition 5. Let $A \subset B(X)$ be a subalgebra. A linear operator T defined on a not necessarily closed linear subspace $P \subset X$ is called a graph transformation for A if there exist finitely many linear operators $T_1, T_2, \dots, T_l$, all defined on P, such that

$$\{x \oplus Tx \oplus T_1 x \oplus T_2 x \oplus \cdots \oplus T_l x \mid x \in P\} \in \operatorname{Lat} A^{(l+2)}.$$

Remark 6. Let $K \in \operatorname{Lat} A^{(n)}$, $n \in \mathbb{N}$. Define

$$K_0 = \{\boldsymbol{x} \in X^{(n-1)} \mid 0 \oplus \boldsymbol{x} \in K\} \in \operatorname{Lat} A^{(n-1)}.$$

Then, if K_0 is complemented in $\operatorname{Lat} A^{(n-1)}$ with complement K_0^c it follows that there exist graph transformations for A: $T_1, T_2, \dots, T_{n-1}$, defined on a linear subspace, $P \subset X$, such that

$$(X \oplus K_0^c) \cap K = \{x \oplus T_1 x \oplus T_2 x \oplus \cdots \oplus T_{n-1} x \mid x \in P\}.$$

Proof. Straightforward. □

2.2. *Boolean algebras of projections in Banach spaces and spectral operators.* Let $\mathcal{B}$ be a complete Boolean algebra of projections in a (complex) Banach space X (as defined for instance in [Bade 1955] or in [Dunford and Schwartz 1988, Chapter XVII]). It is known [Stone 1949] that there exists an extremally disconnected compact Hausdorff topological space Ω (that is a compact Hausdorff space in which the closure of every open set in it is open), such that $\mathcal{B}$ is equivalent as a Boolean algebra with the Boolean algebra of open and closed subsets of Ω. We will denote by Σ the collection of Borel sets of Ω. Such a compact Hausdorff space is called a Stonean space.

The following remark collects some results about the complete Boolean algebras of projections in Banach spaces that will be used in this paper.

Remark 7. (i) If $\mathcal{B}$ is a complete Boolean algebra of projections, then there is a regular countably additive spectral measure E in X defined on the family of Borel sets in Ω such that the mapping

$$S(f) = \int_\Omega f(w)\, E(dw)$$

is a continuous isomorphism of the algebra $C(\Omega)$ of continuous functions on Ω onto the uniformly closed algebra of operators, B, generated by $\mathcal{B}$.

(ii) The algebra B coincides with the strongly closed algebra generated by $\mathcal{B}$ and consists of spectral operators of scalar type.

(iii) The range of E is precisely the Boolean algebra $\mathcal{B}$.

(iv) $\mathcal{B}$ is norm bounded.

Proof. (i) and (iii) follow from [Dunford and Schwartz 1988, Lemma XVII.3.9]. Point (ii) is Corollary XVII.3.17 of the same reference, and (iv) follows from [Bade 1955, Theorem 2.2]. □

Let $\mathcal{B} \subset B(X)$ be a complete Boolean algebra of projections that contains the identity projection $I \in B(X)$. We say that $I \in \mathcal{B}$ has multiplicity k, $k \in \mathbb{N}$, if there are $x_1, x_2, \ldots, x_k \in X$ such that $\overline{\text{lin}}\,\{ex_i \mid e \in \mathcal{B}, 1 \le i \le k\} = X$ and no subset of X of cardinality less than k has this property [Bade 1959, Definition 3.2]. The Boolean algebra $\mathcal{B}$ is said to be of uniform multiplicity k if every projection $e \in \mathcal{B}$, $e \neq 0$ has multiplicity k. For each i, $1 \le i \le k$, define $\mathfrak{M}(x_i) = \overline{\text{lin}}\{ex_i \mid e \in \mathcal{B}\}$. Here, $\overline{\text{lin}}\{ex_i \mid e \in \mathcal{B}\}$ denotes the closed linear subspace of X spanned by $\{ex_i \mid e \in \mathcal{B}\}$.

The next remark collects some known results from [Bade 1959] (see also [Dunford and Schwartz 1988]).

Remark 8. Let $\mathcal{B}$ be a complete Boolean algebra of finite uniform multiplicity n, $n \in \mathbb{N}$, and let $\{x_1, \ldots, x_n\}$ be a set of vectors such that

$$\overline{\text{lin}}\{ex_i \mid e \in \mathcal{B}, 1 \le i \le n\} = X.$$

(i) There are $x_i^* \in X^*$, $i = 1, 2, \ldots, n$, where X^* is the dual Banach space of X, such that each of the measures $\mu_i(\delta) = x_i^* E(\delta) x_i$, $i \in \{1, 2, \ldots, n\}$, $\delta \in \Sigma$ vanishes on sets of first category of Ω and $\mu_i(\sigma) \neq 0$ if σ has nonempty interior. The measures μ_i are equivalent and $x_i^*(\mathfrak{M}(x_j)) = \{0\}$ for $i \neq j$.

(ii) There exists a continuous injective linear map V of X onto a dense linear subspace $L \subset \sum_{i=1}^n L^1(\Omega, \Sigma, \mu_i)$ such that if $V(x) = \boldsymbol{f} = \sum f_i$, then:

(a)
$$x_i^* E(\delta) x = \int_\delta f_i(\omega)\, \mu_i(d\omega),$$

for $\delta \in \Sigma$. In particular, $V(x_i) = 0 \oplus \cdots \oplus \chi_\Omega \oplus \cdots \oplus 0$, where $\chi_\Omega = 1$ is in the i-th place in the direct sum.

(b)
$$x = \lim_{m \to \infty} \sum_{i=1}^n S(f_i \chi_{\delta_m}) x_i,$$

where χ_{δ_m} is the characteristic function of

$$\delta_m = \big\{\omega \bigm| |f_i(\omega)| \le m,\ i = 1, 2, \ldots, n\big\}.$$

(iii) The linear space L is a Banach space when endowed with the norm

$$\|\boldsymbol{f}\|_0 = \max_{1 \le i \le n} \|f_i\|_1 + \|V^{-1}(\boldsymbol{f})\|,$$

and V is a Banach space isomorphism between X and $(L, \|\cdot\|_0)$.

Proof. Points (i) and (ii) follow from [Bade 1959, Lemma 5.1 and Theorem 5.2] (see also [Dunford and Schwartz 1988, Theorem XVIII.3.19]). The proof of (iii) is immediate. □

A function f is called E-essentially bounded if

$$\inf_{E(\delta)=1} \sup_{\omega\in\delta} |f(\omega)|$$

is finite [Dunford and Schwartz 1988, Definition 7].

Denote by $EB(\Omega, \Sigma)$ the set of all E-essentially bounded Σ-measurable functions.

Lemma 9. *With the notations in Remark 8, if $\varphi \in EB(\Omega, \Sigma)$, then the operator $M_\varphi(\boldsymbol{f}) = \varphi\boldsymbol{f}$ is a well defined, bounded operator on $(L, \|\cdot\|_0)$ and $M_\varphi = VS(\varphi)V^{-1}$. Here $\varphi\boldsymbol{f} = \varphi f_1 \oplus \varphi f_2 \oplus \cdots \oplus \varphi f_n$. Thus*

$$VBV^{-1} = \{M_\varphi \mid \varphi \in EB(\Omega, \Sigma)\}.$$

Proof. Let $\boldsymbol{f} \in L$ and $x = S(\varphi)V^{-1}(\boldsymbol{f})$. Then, according to point (a) in Remark 8 (ii), if $\boldsymbol{g} = V(x)$, we have $x_i^* S(\chi_\delta)x = \int_\delta g_i(w)\,\mu_i(dw)$ for every Borel set $\delta \in \Sigma$. On the other hand,

$$x_i^* S(\chi_\delta)x = x_i^* S(\chi_\delta)S(\varphi)V^{-1}(\boldsymbol{f}) = x_i^* S(\chi_\delta\varphi)V^{-1}(\boldsymbol{f}) = \int_\delta \varphi(w) f_i(w)\,\mu_i(dw).$$

Hence $g_i = \varphi f_i$ μ_i-a.e., so $\boldsymbol{g} = \varphi\boldsymbol{f}$ a.e. and the proof is completed. □

In [Dieudonné 1956] is presented an example of a Boolean algebra of projections, $\mathcal{B}$, such that every nonzero projection $e \in \mathcal{B}$ has multiplicity 2. However, for no choice of $x_1, x_2 \in X$ or $e \in \mathcal{B}$, $e \neq 0$ is eX the algebraic sum of $\mathfrak{M}(ex_1)$ and $\mathfrak{M}(ex_2)$. In the rest of this paper we will consider only Boolean algebras of finite uniform multiplicity with the direct sum property:

Definition 10. We say that the complete Boolean algebra $\mathcal{B}$ of uniform multiplicity k has the direct sum property if X is the algebraic (and therefore, Banach) direct sum of $\mathfrak{M}(x_i)$, $1 \le i \le k$.

A particular case of a Boolean algebra of uniform multiplicity k with the direct sum property is the k-fold direct sum of k copies of a cyclic Boolean algebra of projections. Other examples are presented in [Foguel 1959].

Lemma 11. *Suppose that $\mathcal{B}$ is a complete Boolean algebra of projections of uniform multiplicity k with the direct sum property. Then, for every $\epsilon > 0$ there exist $e \in \mathcal{B}$, $e = E(\rho)$, and $\rho \in \Sigma$ with $\mu_l(\rho^c) < \epsilon$ for every $1 \le l \le k$ (where ρ^c is the complement of ρ) such that for every $\{\varphi_{ij} \mid 1 \le i, j \le k\} \subset EB(\Omega, \Sigma)$, the matrix $[\varphi_{ij}\chi_\rho]$ is a bounded linear operator on $(L, \|\cdot\|_0)$ and $[\varphi_{ij}\chi_\rho]$ belongs to the commutant $\mathcal{B}'$ of $\mathcal{B}$.*

Proof. Since the measures μ_l, $1 \le l \le k$ are equivalent, let $h_{ml} = d\mu_m/d\mu_l$, $1 \le m, l \le k$ be the corresponding Radon Nikodym derivative. Let $\epsilon > 0$ be arbitrary. Fix $1 \le m, l \le k$. Then, since $\bigcup_{n=1}^{\infty} \{1/n \le h_{ml} \le n\} = \Omega$, there is a $n \in \mathbb{N}$ such that $\mu_l(\{1/n \le h_{ml} \le n\}^c) < \epsilon/k^2$. Therefore there is a $n \in \mathbb{N}$ such that $\mu_l(\{1/n \le h_{ml} \le n\}^c) < \epsilon$ for every $1 \le m, l \le k$. Let $\rho = \{1/n \le h_{ml} \le n\} \in \Sigma$. It is easy to see that for every Borel subset $\sigma \subset \rho$ we have $\mu_i(\sigma)/n \le \mu_j(\sigma) \le n\mu_i(\sigma)$ for all $1 \le i, j \le k$. Hence all the spaces $M_{\chi_\rho} L^1(\mu_i) = \chi_\rho L^1(\mu_i)$, $1 \le i \le k$, are equal as sets and mutually isomorphic as Banach spaces. Then, clearly,

$$\chi_\rho L = \chi_\rho L^1(\mu_1) \oplus \chi_\rho L^1(\mu_2) \oplus \cdots \oplus \chi_\rho L^1(\mu_k).$$

Since $\mathcal{B}$ has the direct sum property, we also have

$$E(\rho)X = E(\rho)\mathfrak{M}(x_1) \oplus \cdots \oplus E(\rho)\mathfrak{M}(x_k)$$

and the lemma follows. □

For the definition and basic facts about spectral operators on Banach spaces we refer to [Dunford and Schwartz 1988]. We will need the following result, which follows from [Tzafriri 1967, Theorem 2] and [Foguel 1959, Lemma 2.1 and Theorem 2.3].

Remark 12. Let $T \in B(X)$ and let $\mathcal{B}$ be a complete Boolean algebra of projections in X, of uniform multiplicity k, $k \in \mathbb{N}$. If T commutes with the strongly closed algebra B generated by $\mathcal{B}$, then there exists an increasing sequence of projections $\{e_m = E(\chi_{\delta_m}) \mid m \in \mathbb{N}\} \subset \mathcal{B}$ such that $\{e_m\}$ converges strongly to the identity $I \in B(X)$ and Te_m is a spectral operator of finite type for every m. Moreover, if $T \in B'$ is a spectral operator then T is the sum of a spectral operator R of scalar type in B' and a nilpotent operator Q of order k, $Q \in B'$.

Next we will study the dense linear subspaces of X that are invariant under every element of B, where B is the strongly closed algebra generated by $\mathcal{B}$, the complete Boolean algebra of projections of uniform multiplicity k with the direct sum property. The following lemma is an extension to the case of Banach spaces and an improvement on [Douglas and Pearcy 1972, Lemma 3.3]. Using Remark 8 and Lemma 9, we will identify X with L and B with $\{VS(\varphi)V^{-1} \mid S(\varphi) \in B\}$.

Lemma 13. *Let $k \in \mathbb{N}$ and B the strongly closed algebra generated by the Boolean algebra of projections of uniform multiplicity k, $\mathcal{B} \subset B(X)$ and with the direct sum property. With the above notations, suppose that $\mathfrak{D} \subset X$ is a dense linear subspace which is invariant under all operators in B. Then, for every $\epsilon > 0$, there exists an open and closed set $\lambda_\epsilon \subset \Omega$ such that*

(i) $\mu_i(\lambda_\epsilon^c) < \epsilon$, $i = 1, 2, \ldots, k$, *where λ_ϵ^c is the complement of λ_ϵ in Ω, and*

(ii) $\chi_{\lambda_\epsilon} \boldsymbol{e}_j \in \mathfrak{D}$ *for all* $j \in \{1, 2, \ldots, k\}$, *where* $\{\boldsymbol{e}_j \mid j = 1, 2, \ldots, k\}$ *is the standard basis of* $\mathbb{C}^{(k)}$.

Proof. If $\boldsymbol{z} = (z^1, z^2, \ldots, z^k) \in \mathbb{C}^{(k)}$, consider the norm

$$\|\boldsymbol{z}\| = \max\bigl\{|z^p| \bigm| 1 \le p \le k\bigr\}.$$

It is easy to see that there exists $\alpha > 0$ such that if the set $\{h_1, h_2, \ldots, h_k\} \subset \mathbb{C}^{(k)}$ satisfies $\|h_i - \boldsymbol{e}_i\| < \alpha$, $i = 1, 2, \ldots, k$, then the set $\{h_1, h_2, \ldots, h_k\}$ is linearly independent. Let now $\epsilon > 0$ be arbitrary. We can choose $\alpha < \epsilon^2/2$. Let $\rho \in \Sigma$, $\mu_l(\rho) < \epsilon/2$, $1 \le l \le k$, be as in Lemma 11. Since Ω is extremally disconnected, we can assume that ρ is an open and closed set. For every j, $1 \le j \le k$ let $\boldsymbol{g}_j(w) = \boldsymbol{e}_j$, if $w \in \rho$ and $\boldsymbol{g}_j(w) = 0$ if $\omega \in \rho^c$. Since by point (a) of Remark 8 (ii) we have that $\boldsymbol{g}_j \in \chi_\rho L$ for every j, $1 \le j \le k$ and $\chi_\rho \mathfrak{D}$ is dense in $\chi_\rho L$, it follows that there exists a set of elements $\{\boldsymbol{l}_i \mid 1 \le i \le k\} \subset \chi_\rho \mathfrak{D}$, $\boldsymbol{l}_i = l_i^1 \oplus l_i^2 \oplus \cdots \oplus l_i^k$ such that

$$\|\boldsymbol{l}_i - \boldsymbol{g}_i\|_0 = \max_{1 \le p \le k} \bigl\{\|l_i^p - g_i^p\|_0 = \|l_i^p - g_i^p\|_1 + \|T^{-1}(l_i^p - g_i)^p\|\bigr\} < \alpha < \epsilon^2.$$

Let $\delta_\epsilon = \bigcap_{i=1}^{k} \bigl\{\omega \in \rho \bigm| |l_i^p(w) - g_i^p(w)| \ge \epsilon \text{ and } 1 \le p \le k\bigr\}$. Then we have

$$\epsilon^2/2 > \alpha > \max\bigl\{\|l_i^p - g_i^p\|_1 \bigm| 1 \le i, p \le k\bigr\} \ge \epsilon \mu_m(\delta_\epsilon) \quad \text{for } 1 \le m \le k.$$

Hence $\mu_m(\delta_\epsilon) < \epsilon/2$ for $m = 1, 2, \ldots, k$. Assuming that $\epsilon < 2$, it follows that $\mu_m(\delta_\epsilon^c) \ne 0$ and since Ω is a Stonean space, and μ_m a normal measure, $\mu_m(\delta_\epsilon^c) = \mu_m((\delta_\epsilon^c)^\circ)$ where $(\delta_\epsilon^c)^\circ$ is the interior of δ_ϵ^c. The same argument as the preceding one shows that there exists an open and closed subset $\sigma_\epsilon \subset (\delta_\epsilon^c)^\circ$ with $\mu_m(\sigma_\epsilon^c) < \epsilon/2$. Let $\lambda_\epsilon = \rho \cap \sigma_\epsilon$. Then, $\mu_m(\lambda_\epsilon) < \epsilon$ for all $1 \le m \le k$. It follows that all the components of the vectors $l_i^\epsilon = l_i \chi_{\lambda_\epsilon} \in L$ are in $EB(\Omega, \Sigma)$. Let M be the matrix whose i-th column is l_i^ϵ. Then, using Lemma 11, it follows that M is a bounded linear operator that commutes with every element in B, so $M \in B'$. The choice of α implies that $M(w)$ is nonsingular for every $\omega \in \lambda_\epsilon$. Consider the matrix N defined as follows:

$$N(w) = \begin{cases} M(w)^{-1} & \text{if } w \in \lambda_\epsilon, \\ 0 & \text{if } w \in \lambda_\epsilon^c. \end{cases}$$

By restricting N to an open and closed subset of λ_ϵ, if necessary, we can apply Lemma 11 again and get $N \in B'$. It follows that the columns of the product MN are linear combinations of vectors in $\mathfrak{D}$ with coefficients in B. Since $\mathfrak{D}$ is invariant under B we have that these columns belong to $\mathfrak{D}$. Since $M(w)N(w) = I$ for $w \in \lambda_\epsilon$ the proof is completed. $\square$

We will use next the following results about spectral operators and their resolutions of the identity from [Dunford and Schwartz 1988].

Remark 14. If the operator M commutes with the spectral operator T, then M commutes with every resolution of the identity of T.

Proof. This is [Dunford and Schwartz 1988, Corollary XV.3.7]. □

3. Algebras with complemented invariant subspace lattices

In this section we will prove our main result:

Theorem 15. *Let B be the strongly closed subalgebra of $B(X)$ generated by a complete Boolean algebra of projections $\mathcal{B} \subset B(X)$ of finite uniform multiplicity, k, with the direct sum property. If $A \subset B(X)$ is a strongly closed algebra with complemented invariant subspace lattice that contains B, then A is reflexive.*

The proof of this theorem will be given after a series of auxiliary results. In the rest of this section $\mathcal{B}$ and B will be as in Theorem 15. We will identify X with $(L, \|\cdot\|_0)$ as in Remark 8.

Proposition 16. *Let B as in Theorem 15 and let T be a densely defined closed operator on X which commutes with B. There exists an increasing sequence of projections $\{q_p\}_{p=1}^{\infty} \subset \mathcal{B}$ that converges strongly to I such that Tq_p is a spectral operator of finite type for every $p \in \mathbb{N}$.*

Proof. Let $\mathcal{D} \subset X$ be the (dense) domain of T. Since T commutes with B it follows that $\mathcal{D}$ is invariant under B. By Lemma 13 it follows that for every $p \in \mathbb{N}$ there is an open and closed subset $\sigma_p \subset \Omega$ such that $\chi_{\sigma_p} \oplus \chi_{\sigma_p} \oplus \cdots \oplus \chi_{\sigma_p} \in \mathcal{D}$ and $\mu_l(\sigma_p^c) < 1/2p$ for every $1 \leq l \leq k$. Define $r_p = S(\chi_{\sigma_p}) \in \mathcal{B}$. Obviously, we can take $r_p \leq r_{p+1}$ (in the sense that $r_pX \subset r_{p+1}X$) for every $p \in \mathbb{N}$. Therefore Tr_p $(p \in \mathbb{N})$ is a bounded operator and $r_p \nearrow I$. On the other hand, by Remark 12, since $Tr_p \in B^{'}$, for every $p \in \mathbb{N}$, there exists a Borel set $\delta_p \in \Sigma$ such that, for all $1 \leq l \leq p$, we have $\mu_l(\delta_p^c) < 1/2p$. Furthermore, if $q_p = S(\chi_{\delta p \cap \sigma_p})$, then Tq_p is a spectral operator of finite type. Clearly $\{q_p\}$ is an increasing sequence of projections in $\mathcal{B}$ that converges strongly to I and the proof is completed. □

Proposition 17. *Assume that B is as in the statement of Theorem 15. Let T be a densely defined graph transformation for $B \subset B(X)$. Then there exists an increasing sequence of projections $\{q_p\}_{p=1}^{\infty} \subset \mathcal{B}$ that converges strongly to I such that Tq_p is a spectral operator of finite type for every $p \in \mathbb{N}$. In particular every such transformation is closable and its closure commutes with B.*

Proof. Let T be a densely defined graph transformation for B with domain $\mathcal{D}_T$. Since T is a graph transformation for B, there exists $l \in \mathbb{N}$ and operators $T_1, T_2, \ldots, T_{l-2}$ such that the subspace

$$Z = \left\{x \oplus Tx \oplus T_1x \oplus T_2x \oplus \cdots \oplus T_{l-2}x \;\middle|\; x \in \mathcal{D}_T\right\}$$

belongs to Lat $B^{(l)}$. Define $\Delta_{l-1} = \{x \oplus x \oplus \cdots \oplus \mid x \in X\} \subset X^{(l-1)}$. Then it can be easily seen that the subspace

$$\Delta_{l-1}^c = \left\{ x_1 \oplus x_2 \oplus \cdots \oplus x_{l-1} \;\middle|\; x_i \in X \text{ with } \sum_{i=1}^{l-1} x_i = 0 \right\}$$

is a Banach subspace complement of Δ_{l-1} which is invariant under every element of $B^{(l-1)}$. The operator $\widetilde{T}$ defined by

$$\widetilde{T}(x \oplus x \oplus \cdots \oplus x) = Tx \oplus T_1 x \oplus \cdots \oplus T_{l-2}x \quad \text{if } x \in \mathcal{D}_T$$

and

$$\widetilde{T}(x_1 \oplus x_2 \oplus \cdots \oplus x_{l-1}) = 0 \quad \text{if } x_1 \oplus x_2 \oplus \cdots \oplus x_{l-1} \in \Delta_{l-1}^c$$

is a closed, densely defined operator which commutes with $B^{(l-1)}$. An application of Proposition 16 with k replaced by $k(l-1)$ completes the proof. □

Remark 18. Let $A \subset B(X)$ be a strongly closed algebra with complemented invariant subspace lattice and $I \in A$. Then, if $Q \in A'$ is such that $Q^2 = 0$ it follows that $Q \in (\text{algLat}\, A)'$.

Proof. The proof of [Feintuch and Rosenthal 1973, Lemma 3] for the particular case of Hilbert spaces can be extended to the case of Banach spaces. Indeed, let $Q \in A'$ be such that $Q^2 = 0$. Then, if $Y = \ker Q$ is the null space of Q, Y is in Lat A and since A has a complemented invariant subspace lattice, Y has a complement, Y^c in Lat A. Therefore Q can be written as a matrix

$$Q = \begin{bmatrix} 0 & c \\ 0 & 0 \end{bmatrix},$$

and every $a \in A$ can be written as the matrix

$$a = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}.$$

Moreover, every $b \in \text{algLat}\, A$, can be written as a matrix

$$b = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix}.$$

Since $aQ = Qa$ it follows that $ca_2 = a_1 c$. Hence the subspace $\{cx \oplus x \mid x \in Y^c\}$ belongs to Lat A and is therefore invariant for algLat A. It follows that $cb_2 = b_1 c$, so $Qb = bQ$. □

Part (i) of the next result is a generalization of Remark 18.

Proposition 19. *Let $A \subset B(X)$ be an algebra with complemented invariant subspace lattice.*

(i) *If $Q \in A'$ is a nilpotent operator, then $Q \in (\operatorname{algLat} A)'$.*

(ii) *If $T = R + Q$ is a spectral operator of finite type (where R is spectral of scalar type and Q is nilpotent) and $T \in A'$, then $R \in (\operatorname{algLat} A)'$ and $N \in (\operatorname{algLat} A)'$.*

Proof. We will prove point (i) of this proposition by induction. By Remark 18, if $Q \in A'$ and $Q^2 = 0$, then $Q \in (\operatorname{algLat} A)'$. Suppose that for every operator $Q \in A'$ with $Q^n = 0$ it follows that $Q \in (\operatorname{algLat} A)'$ and let $Q \in A'$ with $Q^{n+1} = 0$. Let p_0 denote a projection on $\ker Q$ such that $p_0 \in A'$. Since $Qp_0 = 0$ it follows that

$$(1 - p_0)Q = (1 - p_0)Q(1 - p_0)$$

and therefore

$$(1 - p_0)Q^k = ((1 - p_0)Q(1 - p_0))^k, \quad k \in \mathbb{N}.$$

Since $Q^{n+1} = 0$ we have $Q^n(X) \subset \ker Q$ and therefore

$$0 = (1 - p_0)Q^n = ((1 - p_0)Q(1 - p_0))^n.$$

By hypothesis, $(1 - p_0)Q = (1 - p_0)Q(1 - p_0) \in (\operatorname{algLat} A)'$. On the other hand, since $Q \in A'$ and $p_0 \in A'$ we have $p_0 Q \in A'$. Since obviously $(p_0 Q)^2 = 0$, by Remark 18, it follows that $p_0 Q \in (\operatorname{algLat} A)'$. Therefore

$$Q = p_0 Q + (1 - p_0)Q \in (\operatorname{algLat} A)'$$

and the proof of (i) is completed.

We turn now to prove point (ii). By Remark 14, every resolution of the identity of T, $\boldsymbol{E}(\delta)$, where δ is a Borel subset of the spectrum of T, $\delta \subset \operatorname{sp}(T)$, is in A'. Therefore, since A has complemented invariant subspace lattice, it follows that $\boldsymbol{E}(\delta) \in (\operatorname{algLat} A)'$ for every Borel set $\delta \subset \operatorname{sp}(T)$. Hence $R = \int \lambda\, \boldsymbol{E}(d\lambda) \in (\operatorname{algLat} A)'$. Since $T \in A'$ and $R \in A'$ it follows that $Q \in A'$. By part (i) it follows that $Q \in (\operatorname{algLat} A)'$. □

Lemma 20. *Let A be a strongly closed algebra with complemented invariant subspace lattice that contains a complete Boolean algebra of projections of finite uniform multiplicity k with the direct sum property. Then, if $K \in \operatorname{Lat} A^{(n)}$ for some $n \in \mathbb{N}$, then, there exists an increasing sequence of projections $\{p_m\} \subset \mathcal{B}$, $p_m \nearrow I$ such that $p_m^{(n)} K$ is complemented in $\operatorname{Lat}(p_m A p_m)^{(n)}$ for every $m \in \mathbb{N}$.*

Proof. We will prove the lemma by induction on n. For $n = 1$ the statement is obvious with $p_m = I$ for every m. Let $K \in \operatorname{Lat} A^{(n)}$. Define

$$K_0 = \{\boldsymbol{x} \in X^{(n-1)} \mid 0 \oplus \boldsymbol{x} \in K\}.$$

Obviously, $K_0 \in \operatorname{Lat} A^{(n-1)}$, so there exists an increasing sequence of projections $\{r_m\} \subset \mathcal{B}$, $r_m \nearrow I$ such that $r_m^{(n-1)} K_0$ is complemented in $\operatorname{Lat}(r_m A r_m)^{(n-1)}$. Let

$(r_m^{(n-1)}K_0)^c$ be the complement of $r_m^{(n-1)}K_0$ in $\mathrm{Lat}(r_m A r_m)^{(n-1)}$. Then,

$$\big(r_m X \oplus (r_m^{(n-1)}K_0)^c\big) \cap K \in \mathrm{Lat}\, A^{(n)}$$

and

$$r_m^{(n)}K = \big(0 \oplus r_m^{(n-1)}K_0\big) + \big(r_m X \oplus (r_m^{(n-1)}K_0)^c\big) \cap r_m^{(n)}K.$$

Since $\big(r_m X \oplus (r_m^{(n-1)}K_0)^c\big) \cap r_m^{(n)}K$ is the complement of $0 \oplus r_m^{(n-1)}K_0$ in $r_m^{(n)}K$, there exist graph transformations $T_1, T_2, \dots, T_{n-1}$ such that

$$\big(r_m X \oplus (r_m^{(n-1)}K_0)^c\big) \cap r_m^{(n)}K = \big\{x \oplus T_1 x \oplus T_2 x \oplus \cdots \oplus T_{n-1}x \;\big|\; x \in P\big\},$$

where P is a linear subspace of $r_m X$ invariant under every element of $r_m A r_m$. The closure of P in $r_m X$, $\bar{P}$, belongs to $\mathrm{Lat}(r_m A r_m)$ and hence has a complement $\bar{P}^c$ in $\mathrm{Lat}(r_m A r_m)$. For $1 \le i \le n-1$, consider the following densely defined, graph transformation on $r_m X$:

$$\widetilde{T}_i x = \begin{cases} T_i x & \text{if } x \in P, \\ 0 & \text{if } x \in \bar{P}^c. \end{cases}$$

Then $\widetilde{T}_i$ commutes with A. By Proposition 17, there exists an increasing sequence of projections $\{q_p\} \subset \mathcal{B}$, $q_p \nearrow I$ such that $\widetilde{T}_i q_p$ are bounded spectral operators of finite type. From Lemma 3 it follows that the subspace

$$\big\{q_p x \oplus \widetilde{T}_1 q_p x \oplus \widetilde{T}_2 q_p x \oplus \cdots \oplus \widetilde{T}_{n-1} q_p x \;\big|\; x \in q_p P \oplus q_p \bar{P}^c\big\}$$

is complemented in $\mathrm{Lat}(q_p A q_p)^{(n)}$. By the definition of the transformations $\widetilde{T}_i$ it follows immediately that the subspace

$$\big\{q_p x \oplus T_1 q_p x \oplus T_2 q_p x \oplus \cdots \oplus T_{n-1} q_p x \;\big|\; x \in P\big\}$$

is complemented in $\mathrm{Lat}(q_p A q_p)^{(n)}$. If we set $p_m = r_m q_m \in \mathcal{B}$ we have that $p_m \nearrow I$, $p_m^{(n)}K$ is complemented in $\mathrm{Lat}(p_m A p_m)^{(n)}$:

$$p_m^{(n)}X = \big(0 \oplus p_m^{(n-1)}K_0\big) + \big\{p_m x \oplus T_1 p_m x \oplus \cdots \oplus T_{n-1} p_m x \;\big|\; x \in P\big\} + \big((p_m P)^c \oplus p_m^{(n-1)}K_0^c\big).$$

Hence

$$p_m^{(n)}X = \big(p_m^{(n)}K\big) + \big((p_m P)^c \oplus p_m^{(n-1)}K_0^c\big),$$

and the proof of the lemma is completed. □

The following statement follows from the proof of Lemma 20.

Remark 21. If A is as in the statement of Lemma 20 and $K \in \mathrm{Lat}\, A^{(n)}$ for some $n \in \mathbb{N}$, then there exists an increasing sequence of projections $\{p_m\} \subset \mathcal{B}$, $p_m \nearrow I$ such that $p_m^{(n)}K = (0 \oplus p_m^{(n-1)}K_0) + \{p_m x \oplus T_1 p_m x \oplus \cdots \oplus T_{n-1} p_m x \;|\; x \in P\}$, where $K_0 = \{x \in X^{(n-1)} \mid 0 \oplus x \in K\}$ and $T_i p_m$, $1 \le i \le n-1$, $m \in \mathbb{N}$, are bounded spectral operators of finite type on the closed A-invariant subspace $p_m P$ that commute with $p_m A p_m$.

Proof of Theorem 15. Let $b \in \operatorname{algLat} A$ and $K \in \operatorname{Lat} A^{(n)}$. We will prove by induction on n that there exists an increasing sequence of projections $\{p_m\} \subset \mathcal{B}$ such that $p_m \nearrow I$ and $p_m^{(n)} K \in \operatorname{Lat}(p_m b p_m)^{(n)}$ for every $m \in \mathbb{N}$ and therefore $K \in \operatorname{Lat} b^{(n)}$; then apply Remark 1 to conclude that $b \in A$. By Remark 21, there exists an increasing sequence of projections $\{p_m\} \subset \mathcal{B}$, $p_m \nearrow I$ such that $p_m^{(n)} K = (0 \oplus p_m^{(n-1)} K_0) + \{p_m x \oplus T_1 p_m x \oplus T_2 p_m x \oplus \cdots \oplus T_{n-1} p_m x \mid x \in P\}$ where $K_0 = \{x \in X^{(n-1)} \mid 0 \oplus x \in K\}$ and $T_i p_m$, $1 \le i \le n-1$, $m \in \mathbb{N}$, are bounded spectral operators of finite type on the closed A-invariant subspace $p_m P$ that commute with $p_m A p_m$. The induction hypothesis and Proposition 2 (i) imply that $0 \oplus p_m^{(n-1)} K_0 \in \operatorname{Lat}(p_m b p_m)^{(n)}$. By Proposition 19 (ii) it follows that the bounded spectral operators of finite type $T_i p_m$, $1 \le i \le n-1$, $m \in \mathbb{N}$ commute with $p_m b p_m$. Hence $p_m^{(n)} K \in \operatorname{Lat}(p_m b p_m)^{(n)}$. Since $p_m \nearrow I$ and, by Remark 7 (iv), $\mathcal{B}$ is norm bounded, it follows that $K \in \operatorname{Lat} b^{(n)}$ and the result follows. □

Corollary 22. *Let $A \subset B(X)$ be a strongly closed algebra that contains a complete Boolean algebra of projections $\mathcal{B}$ of finite uniform multiplicity with the direct sum property. If A has complemented invariant subspace lattice, then $A = A''$ where A'' is the bicommutant of A.*

Proof. Follows from Theorem 15 and Remark 4. □

References

[Arveson 1967] W. B. Arveson, "A density theorem for operator algebras", *Duke Math. J.* **34** (1967), 635–647. MR 36 #4345 Zbl 0183.42403

[Bade 1955] W. G. Bade, "On Boolean algebras of projections and algebras of operators", *Trans. Amer. Math. Soc.* **80** (1955), 345–360. MR 17,513d Zbl 0066.36202

[Bade 1959] W. G. Bade, "A multiplicity theory for Boolean algebras of projections in Banach spaces", *Trans. Amer. Math. Soc.* **92** (1959), 508–530. MR 21 #7443 Zbl 0086.31501

[Dieudonné 1956] J. Dieudonné, "Champs de vecteurs non localement triviaux", *Arch. Math.* (*Basel*) **7** (1956), 6–10. MR 17,1112c Zbl 0074.10001

[Douglas and Pearcy 1972] R. G. Douglas and C. Pearcy, "Hyperinvariant subspaces and transitive algebras", *Michigan Math. J.* **19** (1972), 1–12. MR 45 #4186 Zbl 0233.47005

[Dunford and Schwartz 1988] N. Dunford and J. T. Schwartz, *Linear operators, III: Spectral operators*, Wiley, New York, 1988. MR 90g:47001c Zbl 0635.47003

[Feintuch and Rosenthal 1973] A. Feintuch and P. Rosenthal, "Remarks on reductive operator algebras", *Israel J. Math.* **15** (1973), 130–136. MR 48 #9410 Zbl 0264.46069

[Foguel 1959] S. R. Foguel, "Boolean algebras of projections of finite multiplicity", *Pacific J. Math.* **9** (1959), 681–693. MR 21 #7451 Zbl 0087.31701

[Hoover 1973] T. B. Hoover, "Operator algebras with complemented invariant subspace lattices", *Indiana Univ. Math. J.* **22** (1973), 1029–1035. MR 49 #1163 Zbl 0259.46054

[Önder and Orhon 1989] H. Önder and M. Orhon, “Transitive operator algebras on the n-fold direct sum of a cyclic Banach space”, *J. Operator Theory* **22**:1 (1989), 99–107. MR 90k:47098 Zbl 0701.47024

[Radjavi and Rosenthal 1973] H. Radjavi and P. Rosenthal, *Invariant subspaces*, Ergebnisse der Math. **77**, Springer, New York, 1973. Reprinted Dover, Mineola, 2003. MR 51 #3924 Zbl 0269.47003

[Rosenthal and Sourour 1977] P. Rosenthal and A. R. Sourour, “On operator algebras containing cyclic Boolean algebras, II”, *J. London Math. Soc.* (2) **16**:3 (1977), 501–506. MR 58 #17876b Zbl 0391.47028

[Stone 1949] M. H. Stone, “Boundedness properties in function-lattices”, *Canadian J. Math.* **1** (1949), 176–186. MR 10,546a Zbl 0032.16901

[Tzafriri 1967] L. Tzafriri, “Operators commuting with Boolean algebras of projections of finite multiplicity”, *Pacific J. Math.* **20** (1967), 571–587. MR 34 #8197 Zbl 0147.12004

Received April 3, 2012.

FLORENCE MERLEVÈDE
UPEM, LAMA, UMR 8050 CNRS
UNIVERSITÉ PARIS EST
BÂTIMENT COPERNIC
5 BOULEVARD DESCARTES
77435 CHAMPS-SUR-MARNE
FRANCE
florence.merlevede@u-pem.fr

COSTEL PELIGRAD
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF CINCINNATI
PO BOX 210025
CINCINNATI, OH 45221-0025
UNITED STATES
peligrc@ucmail.uc.edu

MAGDA PELIGRAD
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF CINCINNATI
PO BOX 210025
CINCINNATI, OH 45221-0025
UNITED STATES
peligrm@ucmail.uc.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.465

HARER STABILITY AND ORBIFOLD COHOMOLOGY

NICOLA PAGANI

In this paper we review the combinatorics of the twisted sectors of $\mathcal{M}_{g,n}$, and we exhibit a formula for the age of each of them in terms of the combinatorial data. Then we show that orbifold cohomology of $\mathcal{M}_{g,n}$ when $g \to \infty$ reduces to its ordinary cohomology. We do this by showing that the twisted sector of minimum age is always the hyperelliptic twisted sector with all markings in the Weierstrass points; the age of the latter moduli space is just half its codimension in $\mathcal{M}_{g,n}$.

1. Introduction

In recent years there have been lots of new results on geometrical and topological properties of the moduli space $\mathcal{M}_{g,n}$ parametrizing smooth curves of genus g with n distinct marked points on it. When $2g - 2 + n > 0$, this moduli space is a smooth Deligne–Mumford stack, or an orbifold, and its coarse moduli space is a quasiprojective variety of dimension $3g - 3 + n$. When $n > 2g + 2$, every marked curve is rigid, therefore the moduli space is actually a smooth quasiprojective variety.

A celebrated result states that there are isomorphisms

$$H^k(\mathcal{M}_{g,n}, \mathbb{Q}) \cong H^k(\mathcal{M}_{g+1,n}, \mathbb{Q}) \quad \text{when } 3k + 2 \leq 2g. \tag{1}$$

These isomorphisms were introduced in [Harer 1985], but the ranges of their validity have been gradually improved over time by the efforts of different authors. This allows the definition of the *stable cohomology*, denoted $H^*(\mathcal{M}_{\infty,n}, \mathbb{Q})$. The tautological classes κ and ψ are preserved by the above isomorphisms when g is sufficiently large.

A recent result, whose proof was completed in [Madsen and Weiss 2007], asserts that the resulting maps

$$\mathbb{Q}[\kappa_1, \kappa_2, \dots] \otimes \mathbb{Q}[\psi_1, \dots, \psi_n] \to H^*(\mathcal{M}_{\infty,n}). \tag{2}$$

are also isomorphisms. (More precisely, the paper cited shows the result in the case $n = 0$; the extension to $n > 0$ follows from [Looijenga 1996, Proposition 2.1].)

MSC2010: primary 14H10, 32G15, 55N32; secondary 14N35, 14D23, 14H37, 55P50.
Keywords: Harer stability, orbifold cohomology, moduli of curves, Chen–Ruan cohomology, automorphisms, age, twisted sector, inertia stack.

We refer the reader to [Kirwan 2002; Wahl 2012] for a survey of these topological results.

In the latest years, building on earlier results in topology [Kawasaki 1979] and theoretical physics [Dixon et al. 1985; 1986], it has become clearer that when studying the geometry and topology of orbifolds, one should include in the study the *twisted sectors* of the orbifold itself. We refer to [Adem et al. 2007] for an introduction to this emerging new subject. In particular, the cohomology theory of an orbifold is enriched by the so-called *orbifold cohomology*, introduced by Chen and Ruan in [2004]. As a graded vector space, the orbifold cohomology is the direct sum of the cohomology of the original orbifold and of the cohomology of the twisted sectors; the degree of the cohomology classes of each twisted sector is shifted in orbifold cohomology by (twice) a rational number called *age*. This number is not of topological nature, in fact it depends on the complex structure. Its geometric significance appears in [Jarvis et al. 2007] as the virtual rank of an element in the rational K-theory of the twisted sector also known as "half of the normal bundle", this element plays a key role in orbifold intersection theory.

In this note, we introduce the twisted sectors of $\mathcal{M}_{g,n}$ in the combinatorial description of [Pagani 2012; Pagani and Tommasi 2013], we write a closed formula for the age of the twisted sectors of $\mathcal{M}_{g,n}$. (The two papers just cited contain the special cases of this formula for $\mathcal{M}_{2,n}$ and $\mathcal{M}_g$, respectively.) Our main result is Theorem 1, which states that for fixed (g, n), the twisted sector of minimum age is the hyperelliptic twisted sector with marked Weierstrass points. It is a well-known and classical fact, which we review in Proposition 1, that the twisted sectors of $\mathcal{M}_{g,n}$ have codimension higher than $g-2+n$, with equality only for the hyperelliptic locus. Our novel contribution here is that the virtual rank of "half of the normal bundle" (see above) is strictly greater than $\frac{1}{2}(g-2+n)$, with equality only for the hyperelliptic twisted sector. This inequality might have further geometric consequences, besides the implications in orbifold cohomology investigated in this note. (The study of the age of the twisted sectors of various types of moduli spaces of curves, has also recently played a significant role in the investigation of the singularities of the coarse moduli space.)

Combining Theorem 1 with Harer stability, we obtain that the orbifold cohomology of $\mathcal{M}_{g,n}$ stabilizes. Combining further our main result with the theorem of Madsen–Weiss, we can explicitly compute the orbifold cohomology of $\mathcal{M}_{g,n}$ in low degrees. Indeed, from Theorem 1, we deduce

$$H^k_{\mathrm{orb}}(\mathcal{M}_{g,n}, \mathbb{Q}) = H^k(\mathcal{M}_{g,n}, \mathbb{Q}) \quad \text{if } k < g-2+n \text{ or } n > 2g+2. \tag{3}$$

(There are no twisted sectors of $\mathcal{M}_{g,n}$ if and only if $n > 2g+2$).

The stabilization of orbifold cohomology was conjectured by Fantechi in the discussion following her talk [Fantechi 2009] at MSRI. We acknowledge her for

the insight in this topic. We also thank Stefano Maggiolo for having significantly improved the computer program that plays a role at the end of the proof of our main result. The author was supported by DFG project Hu 337/6-2.

2. The twisted sectors of $\mathcal{M}_{g,n}$ and their age

In this section we review the combinatorics of the twisted sectors of $\mathcal{M}_{g,n}$. This description of the twisted sectors of $\mathcal{M}_{g,n}$ was obtained in [Pagani 2012] for $n \geq 1$ or $g = 2$, and in [Pagani and Tommasi 2013] for the remaining cases $\mathcal{M}_{g,0}$, $g \geq 3$.

Let us fix (g, n) with $2g-2+n > 0$. A (g, n)-admissible datum consists of non-negative integers $(g', N; d_1, \ldots, d_{N-1}, a_1, \ldots, a_{N-1})$ such that $N \geq 2$ and

$$2g-2 = N(2g'-2) + \sum_{i=1}^{N-1} (N - \gcd(i, N))\, d_i, \tag{4}$$

$$\sum_{i=1}^{N-1} i\, d_i \equiv 0 \pmod{N}, \tag{5}$$

$$\sum_{i=1}^{N-1} a_i = n, \quad a_i \leq d_i, \quad a_i = 0 \text{ if } \gcd(i, N) \neq 1, \tag{6}$$

$$n = g' = 0 \implies \text{the g.c.d. of } N \text{ and of the } i\text{'s such that } d_i \neq 0 \text{ is } 1. \tag{7}$$

Each (g, n)-admissible datum corresponds to $\binom{n}{a_1, \ldots, a_{N-1}}$ twisted sectors of $\mathcal{M}_{g,n}$ that are related each to the other by an $(a_1, \ldots, a_{N-1})$-permutation of the n marked points. Since we will only investigate properties of the twisted sectors of $\mathcal{M}_{g,n}$ that do not depend on this permutation, from now on we shall slightly abuse the notation and identify each twisted sector Y of $\mathcal{M}_{g,n}$ with its (g, n)-admissible datum

$$Y \sim (g', N; d_1, \ldots, d_{N-1}, a_1, \ldots, a_{N-1}).$$

These facts follow from [Pagani 2012, Proposition 2.13] for $n \geq 1$ and from [Pagani and Tommasi 2013, Corollary 2.16, Theorem 2.19] in the case $n = 0$.

We observe that, from condition (4), there are no (g, n)-admissible data when $n > 2g+2$; in particular, this is the case when g equals 0.

For completeness, we briefly recall our description of the twisted sectors of $\mathcal{M}_{g,n}$, from which the above correspondence follows. For more details, we refer to [Pagani 2012, Section 2.b] for the case $n \geq 1$ and to [Pagani and Tommasi 2013, Section 2.b] for the case $n = 0$.

Construction 1. A twisted sector of $\mathcal{M}_{g,n}$ parametrizes connected cyclic covers of order N of curves of genus g' with total space a curve of genus g, where the n marked points are chosen among the points of total ramification. The branch divisor of the cyclic cover splits into $N-1$ divisors, some of which are possibly empty. Indeed to any point p in the branch divisor D, let q be any point in the fiber of p under the cyclic cover map; we define H_p as the stabilizer of the action of $\mathbb{Z}/N\mathbb{Z}$ at q, and ψ_p as the character of the action of H_p on the cotangent space in q. Then, for $0 < i < N$, we define D_i as the subset of D of those p such that H_p equals the subgroup generated by i in $\mathbb{Z}/N\mathbb{Z}$, and such that $\psi_p(i)$ equals ω_N, a fixed generator for μ_N: the group of N-th roots of 1. In addition to (g, g', N), the admissible datum consists of $d_i := |D_i|$ and of a_i, the number of chosen marked points in the preimage of D_i under the cyclic cover map.

Given a (g, n)-admissible datum, we can construct a moduli space of cyclic covers as in the paragraph above. Condition (4) is the Riemann–Hurwitz formula, condition (5) is a compatibility condition that guarantees the existence of a (not necessarily connected) cyclic cover with the data d_i and N, condition (6) corresponds to the fact that the marked points must be points of total ramification for the cover. Now if $n \geq 1$, it is easy to see that the total space of the cover is forced to be connected and that the moduli space parametrizing such covers is also connected. If instead $n = 0$, it is shown in [Pagani and Tommasi 2013, Theorem 2.19] that there is always one connected component of the moduli space that parametrizes *connected* cyclic covers. This component may possibly be empty only when $g' = 0$, condition (7) rules out precisely these cases.

Let us fix a twisted sector $(g', N; d_1, \dots, d_{N-1}, a_1, \dots, a_{N-1})$. Since Y admits a finite map to $\mathcal{M}_{g', \sum d_i}$, its dimension is $3g' - 3 + \sum d_i$, its *codimension in* $\mathcal{M}_{g,n}$ is

$$\text{(8)} \qquad \operatorname{codim}(Y) := 3g - 3g' - \sum_{i=1}^{N-1} d_i + n,$$

and its *twin* is $(g', N; d_{N-1}, \dots, d_1, a_{N-1}, \dots, a_1)$. If (g, n) is fixed and n is at most $2g + 2$, the *hyperelliptic twisted sector with n marked Weierstrass points* is $(g' = 0, N = 2; d_1 = 2g + 2, a_1 = n)$. In short, we will also call it simply the *hyperelliptic twisted sector*; from (8) it has codimension $g - 2 + n$.

The next result is classical, but we review it for completeness.

Proposition 1. *The codimension of any twisted sector Y of $\mathcal{M}_{g,n}$ satisfies*

$$\operatorname{codim}(Y) \geq g - 2 + n,$$

with equality if and only if Y is the hyperelliptic twisted sector with n marked Weierstrass points.

Proof. Using (8), our statement is reduced to proving the inequality

$$\text{(9)} \qquad \sum d_i \leq 2g - 3g' + 2.$$

Using (4), we have

$$\frac{N}{2}\sum d_i \le \sum d_i(N-\gcd(i,N)) = 2g-2-N(2g'-2); \tag{10}$$

therefore, it is enough to show that

$$\frac{2}{N}(2g-2-N(2g'-2)) \le 2g-3g'+2.$$

Or, rearranging the terms, that

$$(2N-4)(g-1)+Ng' \ge 0. \tag{11}$$

This is clearly always true. Equality holds if and only if $g'=0$ and $N=2$. □

Every twisted sector Y is assigned a rational number, first defined by Chen and Ruan in [2004], which is called the *degree shifting number*, *age*, or *fermionic shift*. Orbifold cohomology is then the direct sum of the ordinary cohomology and of the cohomology of all the twisted sectors, where the latter is shifted in degree by twice the age. For completeness, we briefly review the Chen–Ruan definition of the degree shifting number, building on Construction 1.

Construction 2. Let $f\colon Y \to \mathcal{M}_{g,n}$ be the natural map from the twisted sector to the moduli stack of curves. The group μ_N of N-th roots of 1 acts on $f^*(T_{\mathcal{M}_{g,n}})$, the action can be diagonalized, and each eigenvalue at a point of Y has the form $\lambda_k = e^{2\pi i \alpha_k}$, where the $\alpha_k \in [0,1)\cap\mathbb{Q}$ are the "logarithms" of the eigenvalues. It is not difficult to see that the function $\sum_k \alpha_k$ is well-defined and constant on Y, thus the *age* of Y is defined as

$$a(Y) := \sum_k \alpha_k \in \mathbb{Q}. \tag{12}$$

Moreover, by the very definition of twisted sector, the action of μ_N on T_Y is trivial, thus in the definition (12) it is equivalent to sum the "logarithms" of the eigenvalues of the normal bundle $N_Y\mathcal{M}_{g,n}$, where the latter is defined by the exact sequence of vector bundles

$$0 \to T_Y \to f^*(T_{\mathcal{M}_{g,n}}) \to N_Y\mathcal{M}_{g,n} \to 0.$$

The age of a twisted sector can be interpreted as the virtual rank of an element in the rational K-theory of Y that plays an important role in orbifold intersection theory, see [Jarvis et al. 2007, Definition 1.3, Sections 1.3 and 4].

The age of a twisted sector of $\mathcal{M}_{g,n}$ can explicitly be determined in terms of its admissible datum. From [Pagani and Tommasi 2013, Proposition 5.6] and [Pagani

2012, Lemma 4.6], we have the following formula for the age:

$$a(Y) = \frac{(3g'-3)(N-1)}{2} + \frac{1}{N} \sum_{\gcd(i,N)=1} a_i \sum_{k=1}^{N-1} k\sigma(k,i) + \frac{1}{N} \sum_{i=1}^{N-1} d_i \sum_{k=1}^{N-1} k\left(\left\{\frac{ki}{N}\right\} + \sigma(k,i)\right), \tag{13}$$

where $\{x\} := x - \lfloor x \rfloor$ denotes the fractional part of $x \in \mathbb{Q}^+$, and

$$\sigma(k,i) := \begin{cases} 0 & \text{if } ki + \gcd(i,N) \equiv 0 \pmod{N}, \\ 1 & \text{if } ki + \gcd(i,N) \not\equiv 0 \pmod{N}. \end{cases}$$

Using only (13), it is an easy exercise to check that, if Y and Y' are twins, then

$$a(Y) + a(Y') = \operatorname{codim}(Y) = \operatorname{codim}(Y'). \tag{14}$$

For example, when a twisted sector Y is twin to itself (this happens always, for example, when $N = 2$), its age is half its codimension.

3. The twisted sectors of minimum age

Using only the combinatorial description of the previous section, and in analogy with Proposition 1, we can prove the main result of this note. From now on, we assume $2g - 2 + n > 0$.

Theorem 1. *The age of any twisted sector Y satisfies $2a(Y) \geq g - 2 + n$, with equality if and only if Y is the hyperelliptic twisted sector with n marked Weierstrass points.*

The marked hyperelliptic twisted sector is, using the terminology established in the previous section, twin to itself. Therefore its age is half its codimension: $\frac{1}{2}(g - 2 + n)$.

This implies the following corollary, relevant for orbifold cohomology:

$$H^k(\mathcal{M}_{g,n}, \mathbb{Q}) = H^k_{\mathrm{orb}}(\mathcal{M}_{g,n}, \mathbb{Q}) \quad \text{if } k < g - 2 + n \text{ or } n > 2g + 2. \tag{15}$$

There are no twisted sectors of $\mathcal{M}_{g,n}$ if and only if $n > 2g + 2$; otherwise our bound on the cohomological degree k is sharp.

Using the stability results for ordinary cohomology, we deduce:

Corollary 1. *The isomorphisms* (1) *are*, *in fact*, *isomorphisms*

$$H^k_{\mathrm{orb}}(\mathcal{M}_{g,n}, \mathbb{Q}) \cong H^k_{\mathrm{orb}}(\mathcal{M}_{g+1,n}, \mathbb{Q})$$

when $k \leq \min(g - 3 + n, 2g/3 - 2/3)$.

In particular, we can interpret this by saying that orbifold cohomology of $\mathcal{M}_{g,n}$ "trivially stabilizes" when $g \to \infty$, and the stable orbifold cohomology of $\mathcal{M}_{g,n}$ coincides with its ordinary stable cohomology. The only pairs (g, n) for which $2g/3 - 2/3 > g - 3 + n$ occur are

$$(1, 1), (2, 0), (3, 0), (4, 0).$$

In these special cases, the ranges for k in Corollary 1 are optimal, whereas in all other cases our ranges coincide with the ranges of stability for ordinary cohomology: $k < 2g/3 - 2/3$. The latter ranges are known to be optimal when $g \equiv 2 \pmod 3$. More details on the sharpness of the ranges for cohomological stability can be found in [Wahl 2012, p. 2].

Combining (15) with the isomorphisms (2), we see how the orbifold cohomology of $\mathcal{M}_{g,n}$ is explicitly computable in low degrees.

We now move to the proof of Theorem 1. Thanks to Proposition 1 and to (14), what we have to prove is in fact

$$\begin{aligned} |a(Y) - a(Y')| &\leq \operatorname{codim}(Y) - (g - 2 + n) \\ &= 2g + 2 - 3g' - \textstyle\sum d_i, \end{aligned} \tag{16}$$

with equality only when Y is the hyperelliptic twisted sector with n marked Weierstrass points.

We introduce some notation. Let

$$\Sigma := \{d \in \mathbb{N} \mid d \text{ divides } N, d \neq N\}$$

be the set of proper divisors of N, and let

$$a(Y)_{\text{mark}} := \frac{1}{N} \sum_{\gcd(i,N)=1} a_i \sum_{k=1}^{N-1} k\, \sigma(k, i), \tag{17}$$

$$a(Y)_\sigma := \frac{1}{N} \sum_{\gcd(i,N)=\sigma} d_i \sum_{k=1}^{N-1} k \left(\left\{ \frac{ki}{N} \right\} + \sigma(k, i) \right). \tag{18}$$

We can rewrite formula (13) for the age of a twisted sector Y as:

$$a(Y) = \frac{(3g' - 3)(N - 1)}{2} + a(Y)_{\text{mark}} + \sum_{\sigma \in \Sigma} a(Y)_\sigma .$$

The term $a(Y)_{\text{mark}}$ is the contribution to the age of Y coming from the marked points, and as such it is zero when $n = 0$. Of course now we have the estimate

$$|a(Y) - a(Y')| \leq \left|a(Y)_{\text{mark}} - a(Y')_{\text{mark}}\right| + \sum_{\sigma \in \Sigma} \left|a(Y)_\sigma - a(Y')_\sigma\right| . \tag{19}$$

We can give estimates for each term in the right hand side of (19).

Lemma 1. *The following inequalities hold*:

$$|a(Y)_{\text{mark}} - a(Y')_{\text{mark}}| \leq \frac{N-2}{N} \sum_{\gcd(i,N)=1} a_i, \tag{20}$$

$$|a(Y)_\sigma - a(Y')_\sigma| \leq \frac{(N-2\sigma)(N/\sigma+5)}{6N} \sum_{\gcd(i,N)=\sigma} d_i. \tag{21}$$

Proof. Let us begin with the contribution coming from the marked points. The left hand side of (20) is equal to

$$\frac{1}{N}\left|\sum a_i(\lambda(i)-\lambda(N-i))\right|, \tag{22}$$

where $\lambda(s)$ is the multiplicative inverse of s modulo N. The maximum of the absolute value of

$$\lambda(i)-\lambda(N-i) = 2\lambda(i) - N$$

is obtained when i is either 1 or $N-1$.

As for the second inequality, we separate the two summands in the right hand side of (18). For the first term, consider the function of i

$$g_N^\sigma(i) := \left|\sum_{k=1}^{N-1} k\left(\left\{\frac{ik}{N}\right\} - \left\{\frac{(N-i)k}{N}\right\}\right)\right|.$$

Its maximum among the values of i such that $\gcd(i, N) = \sigma$ is obtained for $i = \sigma$ or for $i = N-\sigma$. From this, we obtain

$$\left|\sum_{k=1}^{N-1} k\left\{\frac{ik}{N}\right\} - \left\{\frac{(N-i)k}{N}\right\}\right| \leq g_N^\sigma(\sigma) = \frac{1}{6}\left(\frac{N}{\sigma}-1\right)(N-2\sigma). \tag{23}$$

The second term is treated similarly to the contribution coming from the marked points. The maximum of the absolute value of

$$\sum_k (\sigma(k,i) - \sigma(k, N-i)) = 2i - N$$

is obtained when i is either σ or $N-\sigma$. Combining this fact with (23), we get the desired inequality. □

Proof of Theorem 1. As we have already observed, it suffices to prove (16). By using the Riemann–Hurwitz formula (4) to eliminate the variable g, the right hand side of (16) can be rearranged to

$$(2N-3)g' - 2(N-2) + \sum_{\sigma\in\Sigma}(N-\sigma-1)\sum_{\gcd(i,N)=\sigma} d_i.$$

Now let us define for convenience the function

$$f_N(\sigma) := (N-\sigma-1) - \frac{(N-2\sigma)\left(\frac{N}{\sigma}+5\right)}{6N} = \frac{\left(6-\frac{1}{\sigma}\right)N^2-(6\sigma+9)N+10\sigma}{6N}$$

for any integer $N \geq 2$ and any σ a real number between 1 and $N/2$. By using (20) and (21), in order to prove (16) it is enough to prove

$$-\frac{N-2}{N}\sum_{\gcd(i,N)=1} a_i + \sum_{\sigma\in\Sigma} f_N(\sigma) \sum_{\gcd(i,N)=\sigma} d_i \geq (2-2g')(N-2)-g', \tag{24}$$

with equality only in the case of the hyperelliptic twisted sector. Note that $a_i \leq d_i$ from inequality (6).

The left hand side of (24) is always nonnegative for any integer $N \geq 2$, because

$$\tilde{f_N}(1) := f_N(1) - \frac{N-2}{N} = \frac{(N-2)(5N-11)}{6N} \geq 0.$$

Therefore when $g' > 0$, the strict inequality (24) holds evidently, as the right hand side is strictly smaller than 0. Thus all we have to prove is (24) when $g'=0$, a case in which we always have that $\sum d_i \geq 3$ (this follows from condition (4) with $g'=0$ and $g>0$).

We start with the case $g'=0$ and $\sum d_i \geq 4$. The function f_N is concave (just look at its second derivative with respect to σ), thus it has its minimum either in 1 or in $N/2$, and we have $f_N(N/2)=(N-2)/2$. The following two inequalities hold in this case:

$$\tilde{f_N}(1)\sum d_i \leq 2(N-2), \tag{25}$$

$$f_N(N/2)\sum d_i \leq 2(N-2), \tag{26}$$

and they suffice to prove (24). If (24) is an equality, then either (25) or (26) must be an equality. If $N=2$, we are precisely in the case of the hyperelliptic twisted sector. If $N>2$, the inequality (25) is strict, so (26) must be an equality and therefore $\sum d_i = 4$. So if (24) is an equality, with $g'=0$, $N>2$ and $\sum d_i = 4$, then $d_{N/2}=4$, but this implies $g=1$ by (4), hence $n \geq 1$, and this case does not exist because of (6).

So we are left with the case $g'=0$ and $\sum d_i = 3$. A large number of twisted sectors still falls into this last category, but not the hyperelliptic twisted sector. We set the three nonzero d_i's to 1, and denote them by $d_{\sigma_1}=d_{\sigma_2}=d_{\sigma_3}=1$. Then it suffices to prove the strict inequality

$$\left(6-\sum_{i=1}^{3}\frac{1}{\sigma_i}\right)N^2-\left(3+6\sum_{i=1}^{3}\sigma_i\right)N+10\sum_{i=1}^{3}\sigma_i > 6n(N-2). \tag{27}$$

If N is fixed, there are only finitely many possibilities for the variables involved in (27). The constraints are

$$\begin{cases} \sigma_1+\sigma_2+\sigma_3 < N, \\ a\sigma_1+b\sigma_2+c\sigma_3 = N \quad \text{for some } a,b,c \in \mathbb{N}^+, \\ n \leq \big|\{i \mid \sigma_i = 1\}\big| \leq 3, \\ \sigma_i \text{ divides } N, \sigma_i \neq N, \end{cases} \tag{28}$$

where all the quantities involved are integers. The first is a consequence of Riemann–Hurwitz (4) (assuming $g > 1$), the second follows from (5) and the third from (6). From now on, we aim at proving (27) for N greater than a certain explicit constant. We will repeatedly use that the left hand side of (27), for fixed n, N, is a concave function in the domain of definition (28). We can also assume for convenience that $\sigma_1 \leq \sigma_2 \leq \sigma_3$.

- If $n = 3$, then from (28) we deduce $\sigma_1 = \sigma_2 = \sigma_3 = 1$. The inequality (27) is satisfied when $N > 11$.
- If $n = 2$, from (28), we have that $\sigma_1 = \sigma_2 = 1$. It is enough to check (27) for the extreme values $\sigma_3 = 1$ and $\sigma_3 = N/2$. The first follows from the case $n = 3$, by checking the case of $\sigma_3 = N/2$ we see that (27) is valid when $N > 22$.
- If $n = 1$, from (28) $\sigma_1 = 1$, so we have $1 \leq \sigma_2 \leq \sigma_3 \leq N/2$ and $\sigma_2+\sigma_3 < N-1$. It is enough to check the extremal values. The case when $\sigma_2 = 1$ follows from the case $n = 2$. From the second point in (28), if $\sigma_3 = N/2$, then σ_2 is either 1 or 2; in the latter case (27) is valid when $N > 14$. Finally, when $\sigma_2 = \sigma_3 = N/3$, (27) is always valid.
- If $n = 0$, we can assume $\sigma_i \geq 2$, since the other cases fall in the above paragraph. Moreover, there are six extremal cases that fulfill the first and the last of (28):

$$(2,2,2), \left(2,2,\frac{N}{2}\right), \left(2,\frac{N}{3},\frac{N}{2}\right), \left(\frac{N}{7},\frac{N}{3},\frac{N}{2}\right), \left(\frac{N}{5},\frac{N}{4},\frac{N}{2}\right), \left(\frac{N}{4},\frac{N}{3},\frac{N}{3}\right).$$

  We check that (27) for the extremal cases is satisfied when $N > 36$ (the inequality is sharp in the case of the fourth triple).

To conclude the proof, we have to check that (16) holds in the cases when $g' = 0$, $\sum d_i = 3$ and $N < 37$, which imply $g \leq 17$. These cases are only finitely many, and can be handled with the help of a computer program.[1] $\square$

Let us conclude with some remarks.

---

[1] The source code of a C++ program that lists all twisted sectors of $\mathcal{M}_g$, each one with its age, is available at http://pcwww.liv.ac.uk/~pagani/twisted.cpp.

Remark 1. We list the number of twisted sectors of $\mathcal{M}_g$ for $1 \leq g \leq 17$, to give an idea of its rapid growth:

$$(7, 17, 47, 72, 76, 203, 196, 225, 415, 537, 427, 1040, 811, 779, 1750, 1860, 1371).$$

Then the number of twisted sectors of $\mathcal{M}_g$ with $g' = 0$:

$$(7, 16, 43, 65, 64, 193, 163, 207, 372, 485, 359, 983, 657, 866, 1592, 1636, 1115).$$

And, finally, the number of twisted sectors of $\mathcal{M}_g$ with $g' = 0$ and $\sum d_i = 3$:

$$(6, 12, 32, 38, 42, 108, 76, 100, 184, 190, 150, 352, 162, 286, 544, 382, 196).$$

These final 2860 twisted sectors are those for which we performed the computer assisted calculation mentioned in the last paragraph of the proof of Theorem 1.

Remark 2. One can also ask what are the twisted sectors of small age in $\mathcal{M}_{g,n}$, after the marked hyperelliptic one. Here we list, for fixed (g, n), the first twisted sectors in order of increasing age: marked hyperelliptic, marked bielliptic, ..., (marked) double covers of curves of genus $\lceil g/2 \rceil$. We remark that the ranges of existence of those twisted sectors, in terms of g and n, are, respectively,

$$n \leq 2g+2,\ n \leq 2g-2,\ n \leq 2g-6,\ \ldots,\ n \leq 1+(-1)^g;$$

their ages are, respectively,

$$\frac{g-2+n}{2},\ \frac{g-1+n}{2},\ \frac{g+n}{2},\ \ldots,\ \frac{3\lfloor g/2 \rfloor - 1 - (-1)^g + n}{2}.$$

After all these, there is one marked trigonal cyclic twisted sector (when $n \leq g+2$). Then the picture becomes more complicated, and we do not know the answer. For example, we have empirically observed that the minimum age among twisted sectors of codimension k can be bigger than the minimum age among twisted sectors of codimension $k+1$.

The validity of the statements that we made in this remark require a long combinatorial proof along the lines of the proof of Theorem 1, which we do not include in this note as it is not really relevant to our scope.

Remark 3. Condition (7) has not been used in any of the steps of the proof of Theorem 1, which could then have been stated slightly more generally for the twisted sectors of the moduli spaces of not necessarily connected smooth curves of genus g.

Remark 4. There is no such thing as a stable cohomology in low degrees for $\overline{\mathcal{M}}_{g,n}$; it is a classical fact for example that even the second Betti number (which equals the dimension of the Picard group in this case) grows exponentially in g. It still makes sense to ask for the twisted sector of minimum age of $\overline{\mathcal{M}}_{g,n}$, but the answer is much easier. To fix the ideas, we give the answer when $g+n>3$ and $g>0$: the cases in which the generalized hyperelliptic locus has codimension >1. Then the unique twisted sector of minimum age is the codimension-1 locus, consisting generically of a smooth elliptic curve glued at the origin to a smooth curve of genus $g-1$ carrying all the marked points, and with the automorphism induced by the pair (elliptic involution on the elliptic curve, identity). Its age is $\frac{1}{2}$: half its codimension.

References

[Adem et al. 2007] A. Adem, J. Leida, and Y. Ruan, *Orbifolds and stringy topology*, Cambridge Tracts in Mathematics **171**, Cambridge University Press, 2007. MR 2009a:57044 Zbl 1157.57001

[Chen and Ruan 2004] W. Chen and Y. Ruan, "A new cohomology theory of orbifold", *Comm. Math. Phys.* **248**:1 (2004), 1–31. MR 2005j:57036 Zbl 1063.53091

[Dixon et al. 1985] L. Dixon, J. A. Harvey, C. Vafa, and E. Witten, "Strings on orbifolds", *Nuclear Phys. B* **261**:4 (1985), 678–686. MR 87k:81104a

[Dixon et al. 1986] L. Dixon, J. A. Harvey, C. Vafa, and E. Witten, "Strings on orbifolds, II", *Nuclear Phys. B* **274**:2 (1986), 285–314. MR 87k:81104b

[Fantechi 2009] B. Fantechi, "Inertia stack of $\mathcal{M}_{g,n}$", February 23–27 2009, Available at http://www.msri.org/workshops/472/schedules/3657. Talk at the conference "Modern Moduli Theory", MSRI, Berkeley, CA.

[Harer 1985] J. L. Harer, "Stability of the homology of the mapping class groups of orientable surfaces", *Ann. of Math.* (2) **121**:2 (1985), 215–249. MR 87f:57009 Zbl 0579.57005

[Jarvis et al. 2007] T. J. Jarvis, R. Kaufmann, and T. Kimura, "Stringy K-theory and the Chern character", *Invent. Math.* **168**:1 (2007), 23–81. MR 2007i:14059 Zbl 1132.14047

[Kawasaki 1979] T. Kawasaki, "The Riemann–Roch theorem for complex V-manifolds", *Osaka J. Math.* **16**:1 (1979), 151–159. MR 80f:58042 Zbl 0405.32010

[Kirwan 2002] F. Kirwan, "Cohomology of moduli spaces", pp. 363–382 in *Proceedings of the International Congress of Mathematicians* (Beijing, 2002), vol. I, edited by T. T. Li et al., Higher Education Press, Beijing, 2002. MR 2004e:14044 Zbl 1049.14017

[Looijenga 1996] E. Looijenga, "Stable cohomology of the mapping class group with symplectic coefficients and of the universal Abel–Jacobi map", *J. Algebraic Geom.* **5**:1 (1996), 135–150. MR 97g:14026 Zbl 0860.57010

[Madsen and Weiss 2007] I. Madsen and M. Weiss, "The stable moduli space of Riemann surfaces: Mumford's conjecture", *Ann. of Math.* (2) **165**:3 (2007), 843–941. MR 2009b:14051 Zbl 1156.14021

[Pagani 2012] N. Pagani, "The Chen–Ruan cohomology of moduli of curves of genus 2 with marked points", *Adv. Math.* **229**:3 (2012), 1643–1687. MR 2871153 Zbl 1236.14032

[Pagani and Tommasi 2013] N. Pagani and O. Tommasi, "The orbifold cohomology of moduli of genus 3 curves", *Manuscripta Math.* **142**:3-4 (2013), 409–437. MR 3117170 Zbl 06228467

[Wahl 2012] N. Wahl, "Homological stability for mapping class groups of surfaces", pp. 547–583 in *Handbook of Moduli*, vol. III, edited by G. Farkas and I. Morrison, Advanced Lectures in Mathematics **26**, International Press, Somerville, MA, 2012.

Received August 16, 2012. Revised October 27, 2013.

NICOLA PAGANI
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF LIVERPOOL
LIVERPOOL
L697ZL
UNITED KINGDOM
pagani@liv.ac.uk
and
INSTITUT FÜR ALGEBRAISCHE GEOMETRIE
LEIBNIZ UNIVERSITÄT HANNOVER
WELFENGARTEN 1
30167 HANNOVER
GERMANY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.479

SPECTRA OF PRODUCT GRAPHS AND PERMANENTS OF MATRICES OVER FINITE RINGS

LE ANH VINH

We study the spectra of product graphs over the finite cyclic ring $\mathbb{Z}_m$. Using this spectra, we show that if $\mathscr{E}$ is a sufficiently large subset of $\mathbb{Z}_m^k$ then the set of permanents of $k \times k$ matrices with rows in $\mathscr{E}$ contains all nonunits of $\mathbb{Z}_m$.

1. Introduction

Let $\mathbb{F}_q$ be a finite field of q elements where q is an odd prime power. The prime base field $\mathbb{F}_p$ of $\mathbb{F}_q$ may then be naturally identified with $\mathbb{Z}_p$. Let M be an $k \times k$ matrix. Two basic parameters of M are its determinant

$$\mathrm{Det}(M) := \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma) \prod_{i=1}^{k} a_{i\sigma(i)},$$

and its permanent

$$\mathrm{Per}(M) := \sum_{\sigma \in S_k} \prod_{i=1}^{k} a_{i\sigma(i)}.$$

The distribution of the determinants of matrices with entries in a finite field $\mathbb{F}_q$ has been studied by various researchers. Suppose that the ground field $\mathbb{F}_q$ is fixed and $M = M_k$ is a random $k \times k$ matrix with entries chosen independently from $\mathbb{F}_q$. If the entries are chosen uniformly from $\mathbb{F}_q$, then it is well known that

$$\Pr(M_k \text{ is nonsingular}) \to \prod_{i \geqslant 1} (1 - q^{-i}) \quad \text{as } k \to \infty. \tag{1-1}$$

It is interesting that (1-1) is quite robust. Specifically, J. Kahn and J. Komlós [2001] proved a strong necessary and sufficient condition for (1-1).

Theorem 1.1 [Kahn and Komlós 2001]. *Let M_k be a random $k \times k$ matrix with entries chosen according to some fixed nondegenerate probability distribution μ on*

This research is supported by Vietnam National University, Hanoi, under project QG.12.43, “Some problems on matrices over finite fields”.
MSC2010: 05C50.
Keywords: permanent, finite ring, expander graph.

$\mathbb{F}_q$. *Then* (1-1) *holds if and only if the support of* μ *is not contained in any proper affine subfield of* $\mathbb{F}_q$.

An extension of the uniform limit to random matrices with μ depending on k was considered by Kovalenko, Levitskaya, and Savchuk [1986]. They proved the standard limit (1-1) under the condition that the entries m_{ij} of M are independent and $\Pr(m_{ij} = \alpha) > (\log k + \alpha(1))/n$ for all $\alpha \in \mathbb{F}_q$. The behavior of the nullity of M_k for $1 - \mu(0)$ close to $\log k/k$ and $\mu(\alpha) = (1 - \mu(0))/(q - 1)$ for $\alpha \neq 0$ was also studied by Blömer, Karp, and Welzl [1997].

Another direction is to fix the dimension k of matrices and view the size of the finite field as an asymptotic parameter. Note that the implied constants in the symbols O, o, $\lesssim$, and $\ll$ may depend on the integer parameter k. We recall that the notations $U = O(V)$ and $U \lesssim V$ are equivalent to the assertion that the inequality $|U| \leq c|V|$ holds for some constant $c > 0$. The notations $U = o(V)$ and $U \ll V$ are equivalent to the assertion that for any $\epsilon > 0$, the inequality $|U| \leq \epsilon|V|$ holds when the variables of U and V are sufficiently large. For an integer k and a subset $\mathscr{E} \subseteq \mathbb{F}_q^k$, let $M_k(\mathscr{E})$ denote the set of $k \times k$ matrices with rows in $\mathscr{E}$. For any $t \in \mathbb{F}_q$, let $D_k(\mathscr{E}; t)$ be the number of $k \times k$ matrices in $M_k(\mathscr{E})$ having determinant t. Ahmadi and Shparlinski [2007] studied some natural classes of matrices over finite fields $\mathbb{F}_p$ of p elements with components in a given subinterval $[-H, H] \subseteq [-(p-1)/2, (p-1)/2]$. They showed that

$$D_k([-H, H]^k; t) = (1 + o(1)) \frac{(2H + 1)^{k^2}}{p}, \tag{1-2}$$

if $t \in \mathbb{F}_p^*$ and $H \gtrsim p^{3/4+\varepsilon}$ for any constant $\varepsilon > 0$. In the case $k = 2$, the lower bound of the size of the interval can be improved to $H \gtrsim p^{1/2}$.

Using the geometry incidence machinery developed in [Covert et al. 2010], and some properties of nonsingular matrices, the author [Vinh 2009] obtained the following result for higher-dimensional cases ($k \geq 4$):

$$D_k(\mathscr{A}^k; t) = (1 + o(1)) \frac{|\mathscr{A}|^{k^2}}{q},$$

if $t \in \mathbb{F}_q^*$ and $\mathscr{A} \subseteq \mathbb{F}_q$ of cardinality $|\mathscr{A}| \gg q^{k/(2k-1)}$. Covert et al. [2010] studied this problem in a more general setting. A subset $\mathscr{E} \subseteq \mathbb{F}_q^k$ is called a product-like set if $|\mathscr{H}_l \cap \mathscr{E}| \lesssim |\mathscr{E}|^{l/k}$ for any l-dimensional subspace $\mathscr{H}_l \subset \mathbb{F}_q^k$. Covert et al. showed that

$$D_3(\mathscr{E}; t) = (1 + o(1)) \frac{|\mathscr{E}|^3}{q},$$

if $t \in \mathbb{F}_q^*$ and $\mathscr{E} \subset \mathbb{F}_q^3$ is a product-like set of cardinality $|\mathscr{E}| \gg q^{15/8}$. In the singular case, the author [Vinh 2012b] showed that for any subset $\mathscr{E} \subseteq \mathbb{F}_q^k$ with $|\mathscr{E}| \gg q^{k-1+2/k}$ then the number of singular matrices whose rows are in $\mathscr{E}$ is close to the expected number $(1+o(1))|\mathscr{E}|^k/q$. In the general case, the author [Vinh 2013a] showed that if $\mathscr{E}$ is a subset of the k-dimensional vector space over a finite field $\mathbb{F}_q$ ($k \geq 3$) of cardinality $|\mathscr{E}| \geq (k-1)q^{k-1}$, then the set of volumes of k-dimensional parallelepipeds determined by $\mathscr{E}$ covers $\mathbb{F}_q$. This bound is sharp up to a factor of $(k-1)$ as taking $\mathscr{E}$ to be a $(k-1)$-hyperplane through the origin shows.

On the other hand, little is known about the permanent. The only known uniform limit similar to (1-1) for the permanent is due to Lyapkov and Sevast′yanov [Lyapkov and Sevast'yanov 1996]. They proved that the permanent of a random $k \times l$ matrix M_{kl} with elements from $\mathbb{F}_p$ and independent rows has the limit distribution of the form

$$\lim_{k\to\infty} \Pr(\mathrm{Per}(M_{kl}) = \lambda) = \rho_l \delta_{\lambda 0} + (1-\rho_l)/p, \quad \lambda \in \mathbb{F}_p,$$

where $\delta_{\lambda 0}$ is Kronecker's symbol. In [Vinh 2012a], the author studied the distribution of the permanent when the dimension of matrices is fixed. We are interested in the set of all permanents, $P_k(\mathscr{E}) = \{\mathrm{Per}(M) : M \in M_k(\mathscr{E})\}$. Using Fourier analytic methods, the author [Vinh 2012a] proved the following result.

Theorem 1.2 [Vinh 2012a]. *Suppose that q is an odd prime power and* $\gcd(q, k) = 1$. *If $\mathscr{E} \cap (\mathbb{F}_q^*)^k \neq \varnothing$, and $|\mathscr{E}| \gtrsim q^{(k+1/2)}$, then $\mathbb{F}_q^* \subseteq P_k(\mathscr{E})$.*

Note that if $|\mathscr{E}| > nq^{n-1}$ then $\mathscr{E} \cap (\mathbb{F}_q^*)^k \neq \varnothing$. Hence we have an immediate corollary of Theorem 1.2.

Corollary 1.3 [Vinh 2012a]. *Suppose that q is an odd prime power and* $\gcd(q, n) = 1$.

(a) *If $\mathscr{E} \subset \mathbb{F}_q^n$ of cardinality $|\mathscr{E}| > nq^{n-1}$, then $\mathbb{F}_q^* \subseteq P_n(\mathscr{E})$.*

(b) *If $\mathscr{A} \subset \mathbb{F}_q$ of cardinality $|\mathscr{A}| \gg q^{1/2+1/(2n)}$, then $\mathbb{F}_q^* \subseteq P_n(\mathscr{A}^n)$.*

The bound in the first part of Corollary 1.3 is tight up to a factor of n. For example, $|\{\boldsymbol{x} \in \mathbb{F}_q^n : x_1 = 0\}| = q^{n-1}$ and $P_n(\{\boldsymbol{x} \in \mathbb{F}_q^n : x_1 = 0\}) = 0$. However, we conjecture that the bound in the second part of Corollary 1.3 can be further improved to $|\mathscr{A}| \gg q^{1/2+\epsilon}$ (for any $\epsilon > 0$) when n is sufficiently large.

Let m be a large nonprime integer and $\mathbb{Z}_m$ be the ring of residues modulo m. Let $\gamma(m)$ be the smallest prime divisor of m, $\omega(m)$ the number of prime divisors of m, and $\tau(m)$ the number of divisors of m. We identify $\mathbb{Z}_m$ with $\{0, 1, \ldots, m-1\}$. Define the set of units and the set of nonunits in $\mathbb{Z}_m$ by $\mathbb{Z}_m^\times$ and $\mathbb{Z}_m^0$, respectively. The finite Euclidean space $\mathbb{Z}_m^k$ consists of column vectors $\boldsymbol{x}$, with j-th entries $x_j \in \mathbb{Z}_m$. The main purpose of this paper is to extend Theorem 1.2 to the setting of finite cyclic rings $\mathbb{Z}_m$. One reason for considering this situation is that if one is interested in answering similar questions in the setting of rational points, one can ask questions for such sets and see how they compare to the answers in $\mathbb{R}^k$. By scale invariance of these questions, the problem for a subset $\mathscr{E}$ of $\mathbb{Q}^k$ would be the same as for subsets of $\mathbb{Z}_m^k$. More precisely, we have the following analog of Theorem 1.2 over the finite cyclic rings.

Theorem 1.4. *Suppose that m is a large integer and* $\gcd(m,k)=1$. *If $\mathscr{E}\cap(\mathbb{Z}_m^\times)^k\neq\varnothing$, and*

$$|\mathscr{E}|\gtrsim\frac{\tau(m)m^k}{\gamma(m)^{(k-1)/2}},$$

then $\mathbb{Z}_m^\times\subseteq P_k(\mathscr{E})$.

Notice that if $|\mathscr{E}|>k(m-\phi(m))m^{k-1}$ then $\mathscr{E}\cap(\mathbb{Z}_m^\times)^k\neq\varnothing$. Hence, we have an immediate corollary of Theorem 1.4.

Corollary 1.5. *Suppose that m is a large integer and* $\gcd(m,k)=1$.

(a) *Suppose that*

$$(m-\phi(m))\gamma(m)^{(k-1)/2}\gtrsim\tau(m)m$$

and

$$|\mathscr{E}|\gtrsim(m-\phi(m))m^{k-1},$$

then $\mathbb{Z}_m^\times\subseteq P_k(\mathscr{E})$.

(b) *Suppose that $\mathscr{A}\subset\mathbb{Z}_m$ of cardinality*

$$|\mathscr{A}|\gtrsim\frac{\tau(m)m}{\gamma(m)^{(k-1/2k)}},$$

then $\mathbb{Z}_m^\times\subseteq P_k(\mathscr{A}^k)$.

Note that the bound in Corollary 1.5 is sharp. For example, if $\mathscr{E}=\mathbb{Z}_m^0\times\mathbb{Z}_m^{k-1}$ then $P_k(\mathscr{E})\subset\mathbb{Z}_m^0$. Theorem 1.4 and Corollary 1.5 are most effective when m has only a few prime divisors. For example, if $m=p^r$, we have the following result.

Theorem 1.6. *Suppose that p^r is a large prime power and* $\gcd(p,k)=1$. *If $\mathscr{E}\cap(\mathbb{Z}_{p^r}^\times)^k\neq\varnothing$, and*

$$|\mathscr{E}|\gtrsim(r+1)p^{rk-(k-1/2)},$$

then $\mathbb{Z}_{p^r}^\times\subseteq P_k(\mathscr{E})$.

In particular, suppose that $k\geq 3$, $p\gg r$, and $|\mathscr{E}|\gtrsim p^{kr-1}$, then $\mathbb{Z}_{p^r}^\times\subset P_k(\mathscr{E})$. The lower bound of $|\mathscr{E}|$ in this case is sharp, as taking $\mathscr{E}$ to be the set $\mathbb{Z}_{p^r}^0\times\mathbb{Z}_{p^r}^{k-1}$ shows.

Note that, the bounds in Corollary 1.5 and Theorem 1.6 are sharp in general cases. When $\mathscr{E}=\mathscr{A}^n$ is a product set, we conjecture that these bounds can be further improved when n is sufficiently large.

For any $t\in\mathbb{F}_q$ and $\mathscr{E}\subset\mathbb{F}_q^k$, let $P_k(\mathscr{E};t)$ be the number of $k\times k$ matrices with rows in $\mathscr{E}$ having permanent t. In [Vinh 2012a], the author studied the distribution of $P_n(\mathscr{E};t)$ when $\mathscr{E}=\mathscr{A}^k$ for a large subset $\mathscr{A}\subset\mathbb{F}_q$. It would be of interest to extend these results to the setting of finite rings.

2. Product graphs over rings

For a graph G, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of its adjacency matrix. The quantity $\lambda(G) = \max\{\lambda_2, -\lambda_n\}$ is called the second eigenvalue of G. A graph $G = (V, E)$ is called an (n, d, λ)-graph if it is d-regular, has n vertices, and the second eigenvalue of G is at most λ. It is well known (see [Ahmadi and Shparlinski 2007, Chapter 9] for more details) that if λ is much smaller than the degree d, then G has certain random-like properties. For two (not necessarily) disjoint subsets of vertices $U, W \subset V$, let $e(U, W)$ be the number of ordered pairs (u, w) such that $u \in U$, $w \in W$, and (u, w) is an edge of G. For a vertex v of G, let $N(v)$ denote the set of vertices of G adjacent to v and let $d(v)$ denote its degree. Similarly, for a subset U of the vertex set, let $N_U(v) = N(v) \cap U$ and $d_U(v) = |N_U(v)|$. We first recall the following well-known fact.

Theorem 2.1 [Ahmadi and Shparlinski 2007, Corollary 9.2.5]. *Let $G = (V, E)$ be an (n, d, λ)-graph. For any two sets $B, C \subset V$, we have*

$$\left| e(B, C) - \frac{d|B||C|}{n} \right| \leq \lambda\sqrt{|B||C|}.$$

For any $\lambda \in \mathbb{Z}_m$, the product graph $B_m(k, \lambda)$ is defined as follows. The vertex set of the product graph $B_m(k, \lambda)$ is the set $V(B_m(k, \lambda)) = \mathbb{Z}_m^k \backslash (\mathbb{Z}_m^0)^k$. Two vertices $\boldsymbol{a}$ and $\boldsymbol{b} \in V(B_m(k, \lambda))$ are connected by an edge, $(\boldsymbol{a}, \boldsymbol{b}) \in E(B_m(k, \lambda))$, if and only if $\boldsymbol{a} \cdot \boldsymbol{b} = \lambda$. When $\lambda = 0$, the graph is a variant of the Erdős–Rényi graph, which has several interesting applications. We will study this case in a separate paper. We now study the product graph when $\lambda \in \mathbb{Z}_m^\times$.

Theorem 2.2 [Vinh 2013b]. *For any $k \geq 2$ and $\lambda \in \mathbb{Z}_m^\times$, the product graph $B_m(k, \lambda)$ is an*

$$\left(m^k - (m - \phi(m))^k, m^{k-1}, \frac{\tau(m) m^{k-1}}{\gamma(m)^{(k-1)/2}} \right)\text{-}\mathit{graph}.$$

Proof. This proof follows from the proof of [Vinh 2013b, Theorem 3.1]. We include its proof here for completeness. It follows from the definition of the product graph $B_m(k, \lambda)$ that $B_m(k, \lambda)$ is a graph of order $m^k - (m - \phi(m))^k$. The valency of the graph is also easy to compute. Given a vertex $\boldsymbol{x} \in V(B_m(k, \lambda))$, there exists an index $x_i \in \mathbb{Z}_m^\times$. We can assume that $x_1 \in \mathbb{Z}_m^\times$. We can choose $y_2, \ldots, y_k \in \mathbb{Z}_m$ arbitrarily, then y_1 is determined uniquely such that $\boldsymbol{x} \cdot \boldsymbol{y} = \lambda$. Hence, $B_m(k, \lambda)$ is a regular graph of valency m^{d-1}. It remains to estimate the eigenvalues of this multigraph (that is, graph with loops). For any $\boldsymbol{a} \neq \boldsymbol{b} \in \mathbb{Z}_m^k \backslash (\mathbb{Z}_m^0)^k$, we count the number of solutions of the following system:

$$\boldsymbol{a} \cdot \boldsymbol{x} \equiv \boldsymbol{b} \cdot \boldsymbol{x} \equiv \lambda \bmod m, \quad \boldsymbol{x} \in \mathbb{Z}_m^k \backslash (\mathbb{Z}_m^0)^k. \tag{2-1}$$

There exist uniquely $n|m$ and $\boldsymbol{b}_1 \in (\mathbb{Z}_{m/n})^k \backslash (\mathbb{Z}^0_{m/n})^k$ such that $\boldsymbol{b} = \boldsymbol{a} + n\boldsymbol{b}_1$. The system (2-1) becomes

$$\boldsymbol{a} \cdot \boldsymbol{x} \equiv \lambda \bmod m, \quad n\boldsymbol{b}_1 \cdot \boldsymbol{x} \equiv 0 \bmod m, \quad \boldsymbol{x} \in (\mathbb{Z}_{m/n})^k \backslash (\mathbb{Z}^0_{m/n})^k. \tag{2-2}$$

Let $\boldsymbol{a}_n \in (\mathbb{Z}_{m/n})^k \backslash (\mathbb{Z}^0_{m/n})^k \equiv \boldsymbol{a} \bmod m/n$, $\boldsymbol{x}_n \in (\mathbb{Z}_{m/n})^k \backslash (\mathbb{Z}^0_{m/n})^k \equiv \boldsymbol{x} \bmod m/n$, and $\lambda_n \equiv \lambda \bmod m/n$. To solve (2-2), we first solve the following system:

$$\boldsymbol{a}_n \cdot \boldsymbol{x}_n \equiv \lambda_n \bmod m/n, \; \boldsymbol{b}_1 \cdot \boldsymbol{x}_n \equiv 0 \bmod m/n, \quad \boldsymbol{x}_n \in (\mathbb{Z}_{m/n})^k \backslash (\mathbb{Z}^0_{m/n})^k. \tag{2-3}$$

The system (2-3) has no solution when $\boldsymbol{a}_n \equiv t\boldsymbol{b}_1 \bmod p$ for some prime $p|(m/n)$ and $t \in \mathbb{Z}_m^\times$, and $(m/n)^{k-2}$ solutions otherwise. For each solution $\boldsymbol{x}_n$ of (2-3), putting back into the system

$$\boldsymbol{a} \cdot \boldsymbol{x} \equiv \lambda \bmod m, \quad \boldsymbol{x} \equiv \boldsymbol{x}_n \bmod m/n, \tag{2-4}$$

gives us n^{k-1} solutions of the system (2-2). Hence, the system (2-2) has $m^{k-2}n$ solutions when $\boldsymbol{a}_n \not\equiv t\boldsymbol{b}_1 \bmod p$ and no solution otherwise. Let A be the adjacency matrix of $B_m(k, \lambda)$. It follows that

$$A^2 = m^{k-2}J + (m^{k-1} - m^{k-2})I - m^{k-2} \sum_{\substack{n|m \\ 1 \le n < m}} E_n + \sum_{\substack{n|m \\ 1 < n < m}} (m^{k-2}n - m^{k-2})F_n, \tag{2-5}$$

where J is the all-ones matrix; I is the identity matrix; E_n is the adjacency matrix of the graph $B_{E,n}$, where for any two vertices $\boldsymbol{a}, \boldsymbol{b} \in V(B_m(k, \lambda))$, $(\boldsymbol{a}, \boldsymbol{b})$ is an edge of $B_{E,n}$ if and only if $\boldsymbol{b} = \boldsymbol{a} + n\boldsymbol{b}_1$, $\boldsymbol{b}_1 \in (\mathbb{Z}_{m/n})^k \backslash (\mathbb{Z}^0_{m/n})^k$ and $\boldsymbol{a}_n \equiv t\boldsymbol{b}_1 \bmod p$ for some prime $p|(m/n)$; and F_n is the adjacency matrix of the graph $B_{F,n}$, where for any two vertices $\boldsymbol{a}, \boldsymbol{b} \in V(B_m(k, \lambda))$, $(\boldsymbol{a}, \boldsymbol{b})$ is an edge of $B_{F,n}$ if and only if $\boldsymbol{b} = \boldsymbol{a} + n\boldsymbol{b}_1$, $\boldsymbol{b}_1 \in (\mathbb{Z}_{m/n})^k \backslash (\mathbb{Z}^0_{m/n})^k$, and $\boldsymbol{a}_n \not\equiv t\boldsymbol{b}_1 \bmod p$ for any prime $p|(m/n)$.

Therefore, $B_{E,n}$ is a regular graph of valency at most

$$\sum_{p|(m/n),\ p \in \mathcal{P}} (p-1)\Big(\frac{m}{np}\Big)^k < \omega(m)(m/n)^k \gamma(m)^{1-k}.$$

Hence all eigenvalues of E_n are at most $\omega(m)(m/n)^k\gamma(m)^{1-k}$. Besides, it is clear that all eigenvalues of F_n are at most $(m/n)^k$. Since $B_m(k, \lambda)$ is a m^{k-1}-regular graph, m^{k-1} is an eigenvalue of A with the all-one eigenvector $\mathbf{1}$. The graph $B_m(k, \lambda)$ is connected, therefore the eigenvalue m^{k-1} has multiplicity one. Since the graph $B_m(k, \lambda)$ contains (many) triangles, it is not bipartite. Hence, for any other eigenvalue θ, $|\theta| < m^{k-1}$. Let $\boldsymbol{v}_\theta$ denote the corresponding eigenvector of θ. Note that $\boldsymbol{v}_\theta \in \mathbf{1}^\perp$, so $J\boldsymbol{v}_\theta = 0$. It follows from (2-5) that

$$(\theta^2 - m^{k-1} + m^{k-2})\boldsymbol{v}_\theta = \Big(m^{k-2} \sum_{\substack{n|m \\ 1 \le n < m}} E_n - \sum_{\substack{n|m \\ 1 < n < m}} (m^{k-2}n - m^{k-2})F_n\Big)\boldsymbol{v}_\theta.$$

Hence, $\boldsymbol{v}_\theta$ is also an eigenvalue of

$$m^{k-2}\sum_{\substack{n|m\\1\le n<m}} E_n - \sum_{\substack{n|m\\1<n<m}} (m^{k-2}n - m^{k-2})F_n.$$

Since the absolute values of the eigenvalues of a sum of matrices are bounded by the sums of the largest absolute values of eigenvalues of the summands, we have

$$\begin{aligned}\theta^2 &\le m^{k-1}-m^{k-2}+m^{k-2}\sum_{\substack{n|m\\1\le n<m}}\omega(m)(m/n)^k\gamma(m)^{1-k}+\sum_{\substack{n|m\\1<n<m}}(m^{k-2}n-m^{k-2})(m/n)^k\\ &< m^{k-1}+\omega(m)(\tau(m)-1)m^{2k-2}\gamma(m)^{1-k}+\sum_{\substack{n|m\\1<n<m}} m^{2k-2}n^{1-k}\\ &< (\omega(m)+1)(\tau(m)-1)m^{2k-2}\gamma(m)^{1-k}\le \tau(m)^2m^{2k-2}\gamma(m)^{1-k}.\end{aligned}$$

The lemma follows. □

The following lemma is an immediate corollary of Theorems 2.1 and 2.2.

Lemma 2.3. *For any $\mathscr{E}, \mathscr{F} \subset \mathbb{Z}_m^k\backslash(\mathbb{Z}_m^0)^k$ and $\lambda \in \mathbb{Z}_m^\times$, let*

$$e_\lambda(\mathscr{E}, F) = |\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathscr{E}\times\mathscr{F} : \boldsymbol{x}\cdot\boldsymbol{y} = \lambda\}|.$$

Then

$$e_\lambda(\mathscr{E},\mathscr{F}) = \frac{(1+o(1))|\mathscr{E}||\mathscr{F}|}{m} + O\biggl(\frac{\tau(m)m^{k-1}}{\gamma(m)^{(k-1)/2}}\sqrt{|\mathscr{E}||\mathscr{F}|}\biggr).$$

See also [Covert et al. 2012, Theorem 1.3.2] for another proof using character sums over finite rings of Lemma 2.3 in the case of $m = p^r$.

3. Proof of Theorem 1.4

Fix an $\boldsymbol{a} = (a_1, \dots, a_n) \in \mathscr{E}\cap(\mathbb{Z}_m^\times)^k$. For any $\boldsymbol{x} = (x_1, \dots, x_k)$, and $\boldsymbol{y} = (y_1, \dots, y_k) \in \mathscr{E}$, let $M(\boldsymbol{a}; \boldsymbol{x}, \boldsymbol{y})$ denote the matrix whose rows are $\boldsymbol{x}$, $\boldsymbol{y}$, and $(k-2)$ $\boldsymbol{a}$'s. Let $\mathbf{1} := (1, \dots, 1)$, $\boldsymbol{x}/\boldsymbol{a} := (x_1/a_1, \dots, x_k/a_k)$, and $\boldsymbol{y}/\boldsymbol{a} := (y_1/a_1, \dots, y_k/a_k)$; we have

$$\mathrm{Per}(M(\boldsymbol{a}; \boldsymbol{x}, \boldsymbol{y})) = \prod_{i=1}^k a_i\,\mathrm{Per}(M(\mathbf{1}; \boldsymbol{x}/\boldsymbol{a}, \boldsymbol{y}/\boldsymbol{a})) = \Bigl(\prod_{i=1}^k a_i\Bigr)\sum_{i=1}^k \frac{x_i}{a_i}\sum_{j\ne i}\frac{y_j}{a_j}.$$

Set

$$\mathscr{E}_1 := \{(x_i/a_i)_{i=1}^k : (x_1, \dots, x_k) \in \mathscr{E}\}, \tag{3-1}$$

$$\mathscr{E}_2 := \biggl\{\Bigl(\sum_{j\ne i} y_i/a_i\Bigr)_{i=1}^k : (y_1, \dots, y_k) \in \mathscr{E}\biggr\}. \tag{3-2}$$

It is clear that $|\mathscr{E}_1| = |\mathscr{E}_2| = |\mathscr{E}|$ (as $\gcd(k, m) = 1$). For any $\lambda \in \mathbb{Z}_m^\times$, it follows from Lemma 2.3 that

$$
\begin{aligned}
e_\lambda(\mathscr{E}_1, \mathscr{E}_2) &= \frac{(1+o(1))|\mathscr{E}_1||\mathscr{E}_2|}{m} + O\left(\frac{\tau(m)m^{k-1}}{\gamma(m)^{(k-1)/2}}\sqrt{|\mathscr{E}_1||\mathscr{E}_2|}\right) \\
&= \frac{(1+o(1))|\mathscr{E}|^2}{m} + O\left(\frac{\tau(m)m^{k-1}}{\gamma(m)^{(k-1)/2}}|\mathscr{E}|\right).
\end{aligned}
\tag{3-3}
$$

Since

$$|\mathscr{E}| \gtrsim \frac{\tau(m)m^k}{\gamma(m)^{(k-1)/2}},$$

(3-3) implies that

$$\mathbb{Z}_m^\times \subset \{\mathrm{Per}(M(\boldsymbol{a}; \boldsymbol{x}, \boldsymbol{y})) : \boldsymbol{x}, \boldsymbol{y} \in \mathscr{E}\} \subset P_k(\mathscr{E}),$$

completing the proof of Theorem 1.4.

References

[Ahmadi and Shparlinski 2007] O. Ahmadi and I. E. Shparlinski, "Distribution of matrices with restricted entries over finite fields", *Indag. Math.* (*N.S.*) **18**:3 (2007), 327–337. MR 2008k:11127 Zbl 1181.11030

[Blömer et al. 1997] J. Blömer, R. Karp, and E. Welzl, "The rank of sparse random matrices over finite fields", *Random Structures Algorithms* **10**:4 (1997), 407–419. MR 99b:15028 Zbl 0877.15027

[Covert et al. 2010] D. Covert, D. Hart, A. Iosevich, D. Koh, and M. Rudnev, "Generalized incidence theorems, homogeneous forms and sum-product estimates in finite fields", *European J. Combin.* **31**:1 (2010), 306–319. MR 2010m:11014 Zbl 1243.11009

[Covert et al. 2012] D. Covert, A. Iosevich, and J. Pakianathan, "Geometric configurations in the ring of integers modulo p^ℓ", *Indiana Univ. Math. J.* **61**:5 (2012), 1949–1969. MR 3119606 Zbl 06236912

[Kahn and Komlós 2001] J. Kahn and J. Komlós, "Singularity probabilities for random matrices over finite fields", *Combin. Probab. Comput.* **10**:2 (2001), 137–157. MR 2002c:15043 Zbl 0979.15022

[Kovalenko et al. 1986] I. N. Kovalenko, A. A. Levitskaya, and M. N. Savchuk, *Izbrannye zadachi veroyatnostnoi kombinatoriki*, Naukova Dumka, Kiev, 1986. MR 88m:60022

[Lyapkov and Sevast'yanov 1996] L. A. Lyapkov and B. A. Sevast'yanov, "Limit distribution of the probabilities of the permanent of a random matrix in the field GF(p)", *Diskret. Mat.* **8**:2 (1996), 3–13. In Russian; translated in *Discrete Math. Appl.* **6**(2), 107–116 (1996). MR 97g:60017 Zbl 0869.15004

[Vinh 2009] L. A. Vinh, "Distribution of determinant of matrices with restricted entries over finite fields", *J. Comb. Number Theory* **1**:3 (2009), 203–212. MR 2011g:11056 Zbl 1234.11030

[Vinh 2012a] L. A. Vinh, "On the permanents of matrices with restricted entries over finite fields", *SIAM J. Discrete Math.* **26**:3 (2012), 997–1007. MR 3022119 Zbl 1260.15008

[Vinh 2012b] L. A. Vinh, "Singular matrices with restricted rows in vector spaces over finite fields", *Discrete Math.* **312**:2 (2012), 413–418. MR 2012h:15062 Zbl 1246.15035

[Vinh 2013a] L. A. Vinh, "On the volume set of point sets in vector spaces over finite fields", *Proc. Amer. Math. Soc.* **141**:9 (2013), 3067–3071. MR 3068960 Zbl 06203435

[Vinh 2013b] L. A. Vinh, "Product graphs, sum-product graphs and sum-product estimates over finite rings", *Forum Mathematicum* (2013).

Received December 18, 2011. Revised September 9, 2013.

LE ANH VINH
UNIVERSITY OF EDUCATION
VIETNAM NATIONAL UNIVERSITY, HANOI
144 XUAN THUY
CAU GIAY
HANOI 100000
VIETNAM
vinhla@vnu.edu.vn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 267, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.267.489

THE CONCAVITY OF THE GAUSSIAN CURVATURE OF THE CONVEX LEVEL SETS OF MINIMAL SURFACES WITH RESPECT TO THE HEIGHT

PEI-HE WANG

For the minimal graph with strictly convex level sets, we find an auxiliary function to study the Gaussian curvature of the level sets. We prove that this curvature function is a concave function with respect to the height of the minimal surface while this auxiliary function is almost sharp when the minimal surface is the catenoid.

1. Introduction

Consider a function whose graph is minimal and whose level sets are strictly convex. Extending work of Longinetti [1987], we explore the relation between the Gaussian curvature of the level sets and the height.

The nature of the level sets of the solutions of elliptic partial differential equations is a subject with a long history, going back to results of Shiffman in the 1950s for minimal surfaces. The curvature of such level sets has also been studied for several decades. Some key contributions to these problems are listed in the introduction of [Chen and Shi 2011]. Here we just mention some recent developments directly relevant to our problem.

Jost, Ma, and Ou [Jost et al. 2012] and Ma, Ye, and Ye [Ma et al. 2011] proved that the Gaussian and principal curvatures of convex level sets of three-dimensional harmonic functions attain their minima on the boundary. Ma, Ou, and Zhang [2010] gave estimates of the Gaussian curvature of convex level sets of higher-dimensional harmonic functions based on the Gaussian curvature of the boundary and the norm of the gradient on the boundary. Wang and Zhang [2012] have given estimates for the Gaussian curvature of convex level sets of minimal surfaces, Poisson equations, and a class of semilinear elliptic partial differential equations studied by Caffarelli and Spruck [1982].

Research was supported by STPF of University (number J11LA05), NSFC (number ZR2012AM010), the Postdoctoral fund (number 201203030) of Shandong Province and the Postdoctoral Fund (number 2012M521302) of China.
MSC2010: 35B45.
Keywords: concavity, minimal surface, Gaussian curvature, level sets.

In this paper we use the support function of strictly convex level sets and the maximum principle to obtain the concavity of the Gaussian curvature of convex level sets of minimal graphs with respect to the height:

Theorem 1.1. *Let Ω be a bounded smooth domain in $\mathbb{R}^n$, $n \geq 2$, and let*

$$u \in C^4(\Omega) \cap C^2(\overline{\Omega}), \quad t_0 \leq u(x) \leq t_1$$

be a minimal graph in Ω, that is, one such that

$$\operatorname{div} \frac{\nabla u}{\sqrt{1+|\nabla u|^2}} = 0 \quad \text{in } \Omega. \tag{1-1}$$

Assume $|\nabla u| \neq 0$ in $\overline{\Omega}$. Let

$$\Gamma_t = \{x \in \Omega : u(x) = t\} \quad \text{for } t_0 < t < t_1$$

be the level sets of u and let K be their Gaussian curvature function. For

$$f(t) = \min\left\{\left[\left(\frac{|\nabla u|^2}{1+|\nabla u|^2}\right)^{\frac{n-3}{2}} K\right]^{\frac{1}{n-1}}(x) : x \in \Gamma_t\right\},$$

if the level sets of u are strictly convex with respect to the normal ∇u, we have the differential inequality

$$D^2 f(t) \leq 0 \quad \text{in } (t_0, t_1).$$

Under the same assumption as in Theorem 1.1, Wang and Zhang [2012] proved the following statement: *for $n \geq 2$, the function $(|\nabla u|^2/(1+|\nabla u|^2))^\theta K$ attains its minimum on the boundary, where $\theta = -\frac{1}{2}$ or $\theta \geq \frac{1}{2}(n-3)$.* From this fact they got the lower bound estimates for the Gaussian curvature of the level sets.

Corollary 1.2. *Let u satisfy*

$$\begin{cases} \operatorname{div} \dfrac{\nabla u}{\sqrt{1+|\nabla u|^2}} = 0 & \text{in } \Omega = \Omega_0 \backslash \overline{\Omega}_1, \\ u = 0 & \text{on } \partial\Omega_0, \\ u = 1 & \text{on } \partial\Omega_1, \end{cases} \tag{1-2}$$

where Ω_0 and Ω_1 are bounded smooth convex domains in $\mathbb{R}^n$, $n \geq 2$, $\overline{\Omega}_1 \subset \Omega_0$. Assume $|\nabla u| \neq 0$ in $\overline{\Omega}$ and the level sets of u are strictly convex with respect to normal ∇u. Let K be the Gaussian curvature of the level sets. For any point $x \in \Gamma_t$, $0 < t < 1$, we have the following estimates.

- *For $n = 3$, we have*

$$K(x)^{1/2} \geq (1-t)(\min_{\partial\Omega_0} K)^{1/2} + t(\min_{\partial\Omega_1} K)^{1/2}. \tag{1-3}$$

- *For $n \neq 3$, we have*

$$\left[\left(\frac{|\nabla u|^2}{1+|\nabla u|^2}\right)^{\frac{n-3}{2}} K\right]^{\frac{1}{n-1}}(x) \geq (1-t)\min_{\partial\Omega_0}\left[\left(\frac{|\nabla u|^2}{1+|\nabla u|^2}\right)^{\frac{n-3}{2}} K\right]^{\frac{1}{n-1}} + t\min_{\partial\Omega_1}\left[\left(\frac{|\nabla u|^2}{1+|\nabla u|^2}\right)^{\frac{n-3}{2}} K\right]^{\frac{1}{n-1}}. \tag{1-4}$$

Remark 1.3. The following example shows that our estimates are almost sharp in a sense. Let $u(r,\theta)$, $r > 2$, be the n-dimensional catenoid:

$$u(r,\theta) = \int_{-r}^{-2} \frac{1}{\sqrt{s^{2(n-1)}-1}}\,ds. \tag{1-5}$$

Then

$$|\nabla u| = \frac{1}{\sqrt{r^{2(n-1)}-1}}, \tag{1-6}$$

and the Gaussian curvature of the level set at x is $K(x) = r^{1-n}$. Hence,

$$f(t) = \left[\left(\frac{|\nabla u|^2}{1+|\nabla u|^2}\right)^{\frac{n-3}{2}} K\right]^{\frac{1}{n-1}} = r^{2-n}. \tag{1-7}$$

For $n = 2$, $f(t)$ becomes a constant function, which shows that our estimate of its concavity is sharp. Now we turn to the case $n > 2$.

Set
$$R = \int_{-\infty}^{-2} \frac{1}{\sqrt{s^{2(n-1)}-1}}\,ds.$$

Then we have

$$\begin{aligned} -u + R &= \int_{-\infty}^{-r} \frac{1}{s^{n-1}}\,ds - \int_{-\infty}^{-r} \frac{1}{s^{n-1}}\left[1 - \frac{1}{\sqrt{1-s^{-2(n-1)}}}\right] ds \\ &= \frac{(-1)^n}{2-n} r^{2-n} + \mathbb{O}(r^{4-3n}). \end{aligned} \tag{1-8}$$

This means that

$$f(t) = (-1)^n (2-n)(R-t) + \mathbb{O}(r^{4-3n}), \tag{1-9}$$

which shows the "almost sharpness" of our estimate in higher dimensions.

To prove these theorems, let K be the Gaussian curvature of the convex level sets, and let $\varphi = \log K(x) + \rho(|\nabla u|^2)$. For suitable choices of ρ and β, we shall show the elliptic differential inequality

$$L(e^{\beta\varphi}) \leq 0 \quad \mathrm{mod}\ \nabla_\theta \varphi \quad \text{in } \Omega, \tag{1-10}$$

where L is the elliptic operator associated with the equation we discussed and here we have suppressed the terms involving $\nabla_\theta\varphi$ (see the notations below) with locally bounded coefficients. Then we apply the strong minimum principle to obtain the main results.

In Section 2, we first give brief definitions on the support function of the level sets, and then we obtain the equation of the minimal graph in terms of the support function. We prove Theorem 1.1 in Section 3 by formal calculations. The main technique in the proof consists of rearranging the second and third derivative terms using the equation and the first derivative condition for φ. The key idea is Pogorelov's method in a priori estimates for fully nonlinear elliptic equations.

2. Notations and preliminaries

Let Ω_0 and Ω_1 be bounded smooth open convex subsets of $\mathbb{R}^n$ such that $\overline{\Omega}_1 \subset \Omega_0$, and let $\Omega = \Omega_0 \backslash \overline{\Omega}_1$. Let $u : \overline{\Omega} \to \mathbb{R}$ be a smooth function with $|Du| > 0$ in Ω and let its level sets be strictly convex with respect to the normal direction Du.

For simplicity, we will assume that

$$\begin{aligned} u &= 0 \quad \text{on } \partial\Omega_0, \\ u &= 1 \quad \text{on } \partial\Omega_1, \end{aligned}$$

and we extend u to Ω_1 with the value 1. For $0 \le t \le 1$, we set

$$\overline{\Omega}_t = \{x \in \overline{\Omega}_0 : u \ge t\};$$

Then every $x \in \Omega$ belongs to the boundary of $\overline{\Omega}_{u(x)}$.

Next we define the *support function* of u, denoted by

$$H : \mathbb{R}^n \times [0, 1] \to \mathbb{R}$$

as follows: for each $t \in [0, 1]$, $H(\,\cdot\,, t)$ is the support function of the convex body $\overline{\Omega}_t$, that is,

$$H(X, t) = H_{\overline{\Omega}_t}(X) \quad \text{for all } X \in \mathbb{R}^n,\ t \in [0, 1].$$

For details, see [Colesanti and Salani 2003; Longinetti and Salani 2007].

The rest of this section is devoted to deriving the minimal graph by means of the support function. For this we need a reformulation of the first and second derivatives of u in terms of the support function h_{Ω_t}, which is the restriction of $H(\,\cdot\,, t)$ to the unit sphere $\mathbb{S}^{n-1}$; see [Chiti and Longinetti 1992; Longinetti and Salani 2007]. For the convenience of the reader, we report the main steps here.

Recall that h is the restriction of H to $\mathbb{S}^{n-1} \times [0, 1]$, so $h(\theta, t) = H(Y(\theta), t) = h_{\overline{\Omega}_t}(Y(\theta))$ where $t \in [0, 1]$ and $Y(\theta) \in \mathbb{S}^n$ is a unit vector with coordinate θ. Since the level sets of u are strictly convex and $h(\theta, t)$ is well defined, the map

$$x(X, t) = x_{\overline{\Omega}_t}(X),$$

which assigns to every $(X, t) \in \mathbb{R}^n \backslash \{0\} \times (0, 1)$ the unique point $x \in \Omega$ on the level surface $\{u = t\}$, where the gradient of u is parallel to X (and orientation reversed).

Let
$$T_i = \frac{\partial Y}{\partial \theta_i},$$
so that $\{T_1, \dots, T_{n-1}\}$ is a tangent frame field on $\mathbb{S}^{n-1}$, and let
$$x(\theta, t) = x_{\overline{\Omega}_t}(Y(\theta));$$
we denote its inverse map by
$$\nu : (x_1, \dots, x_n) \to (\theta_1, \dots, \theta_{n-1}, t).$$
Notice that all these maps (h, x, and ν) depend on the considered function u (like H), even if we do not adopt any explicit notation to stress this fact.

For $h(\theta, t) = \langle x(\theta, t), Y(\theta)\rangle$, since Y is orthogonal to $\partial\overline{\Omega}_t$ at $x(\theta, t)$, deriving the previous equation, we obtain
$$h_i = \langle x, T_i\rangle.$$
In order to simplify some computations, we can also assume that $\theta_1, \dots, \theta_{n-1}, Y$ is an orthonormal frame positively oriented. Hence, from the previous two equalities, we have
$$x = hY + \sum_i h_i T_i$$
and
$$\frac{\partial T_i}{\partial \theta_j} = -\delta_{ij} Y \quad \text{at } x,$$
where the summation index runs from 1 to $n-1$ if no extra explanation is given, and δ_{ij} is the standard Kronecker symbol. Following [Chiti and Longinetti 1992], we obtain, at the point x under consideration,
$$\begin{aligned} \frac{\partial x}{\partial t} &= h_t Y + \sum_i h_{ti} T_i, \\ \frac{\partial x}{\partial \theta_j} &= h T_j + \sum_i h_{ij} T_i, \qquad j = 1, \dots, n-1. \end{aligned}$$
The inverse of the above Jacobian matrix is
$$\begin{aligned} \frac{\partial t}{\partial x_\alpha} &= h_t^{-1}[Y]_\alpha, && \alpha = 1, \dots, n, \\ \frac{\partial \theta_i}{\partial x_\alpha} &= \sum_j b^{ij}[T_j - h_t^{-1} h_{tj} Y]_\alpha, && \alpha = 1, \dots, n, \end{aligned} \tag{2-1}$$

where $[\,\cdot\,]_i$ denotes the i-coordinate of the vector in the bracket and

$$b_{ij} = \left\langle \frac{\partial x}{\partial \theta_i}, \frac{\partial Y}{\partial \theta_j} \right\rangle = h\delta_{ij} + h_{ij} \tag{2-2}$$

denotes the inverse tensor of the second fundamental form of the level surface $\partial\overline{\Omega}_t$ at $x(\theta, t)$. The eigenvalues of the tensor b^{ij} are the principal curvatures $\kappa_1, \dots, \kappa_{n-1}$ of $\partial\overline{\Omega}_t$ at $x(\theta, t)$; see [Schneider 1993].

The first equation of (2-1) can be rewritten as

$$Du = \frac{Y}{h_t},$$

where the left hand side is computed at $x(\theta, t)$, while the right hand side is computed at (θ, t). It follows that

$$|Du| = -\frac{1}{h_t}.$$

By the chain rule and (2-1), the second derivatives of u in terms of h can be computed as

$$u_{\alpha\beta} = \sum_{i,j} [-h_t^{-2} h_{ti} Y + h_t^{-1} T_i]_\alpha b^{ij} [T_j - h_t^{-1} h_{tj} Y]_\beta - h_t^{-3} h_{tt} [Y]_\alpha [Y]_\beta \tag{2-3}$$

for $\alpha, \beta = 1, \dots, n$.

In these new coordinates, the minimal graph equation, $\operatorname{div} \dfrac{\nabla u}{\sqrt{1 + |\nabla u|^2}} = 0$, reads

$$h_{tt} = \sum_{i,j} [(1 + h_t^2)\delta_{ij} + h_{ti} h_{tj}] b^{ij}, \tag{2-4}$$

and the associated linear elliptic operator is

$$L = \sum_{i,j,p,q} [(1 + h_t^2)\delta_{pq} + h_{tp} h_{tq}] b^{ip} b^{jq} \frac{\partial^2}{\partial \theta_i\, \partial \theta_j} - 2 \sum_{i,j} h_{tj} b^{ij} \frac{\partial^2}{\partial \theta_i\, \partial t} + \frac{\partial^2}{\partial t^2}. \tag{2-5}$$

Now we recall the well-known commutation formulas for the covariant derivatives of a smooth function $u \in C^4(S^n)$.

$$u_{ijk} - u_{ikj} = -u_k \delta_{ij} + u_j \delta_{ik}, \tag{2-6}$$

$$u_{ijkl} - u_{ijlk} = u_{ik}\delta_{jl} - u_{il}\delta_{jk} + u_{kj}\delta_{il} - u_{lj}\delta_{ik}. \tag{2-7}$$

They will be used during the calculations in the next section. By the definition of b_{ij} and the above commutation formulas, we easily get the following Codazzi-type formula:

$$b_{ij,k} = b_{ik,j}. \tag{2-8}$$

3. Gauss curvature of the level sets of minimal graph

In this section we prove Theorem 1.1. We state a technical lemma.

Lemma 3.1 [Ma et al. 2010]. *Let $\lambda \geq 0$, $\mu \in \mathbb{R}$, $b_k > 0$, and $c_k \in \mathbb{R}$ for $2 \leq k \leq n-1$. Define the quadratic polynomial*

$$Q(X_2, \ldots, X_{n-1}) = -\sum_{2\leq k\leq n-1} b_k X_k^2 - \lambda\Bigl(\sum_{2\leq k\leq n-1} X_k\Bigr)^2 + 4\mu \sum_{2\leq k\leq n-1} c_k X_k.$$

Then we have

$$Q(X_2, \ldots, X_{n-1}) \leq 4\mu^2\Gamma,$$

where

$$\Gamma = \sum_{2\leq k\leq n-1} \frac{c_k^2}{b_k} - \lambda\Bigl(1 + \lambda \sum_{2\leq k\leq n-1} \frac{1}{b_k}\Bigr)^{-1}\Bigl(\sum_{2\leq k\leq n-1} \frac{c_k}{b_k}\Bigr)^2.$$

For a continuous function $f(t)$ on [0, 1], we define its *generalized second-order derivative* at any point t in (0, 1) as

$$D^2 f(t) = \limsup_{h\to 0} \frac{f(t+h) + f(t-h) - 2f(t)}{h^2}.$$

Let B be the quotient set $B \equiv \mathbb{R}^n/2\pi\mathbb{Z}^n$ and let $Q \equiv B \times (0, 1)$. Let $G(\theta, t)$ be a regular function in Q such that $\mathscr{L}(G(\theta, t)) \geq 0$ for $(\theta, t) \in Q$, where $\mathscr{L}$ is an elliptic operator of the form

$$\mathscr{L} = \sum_{i,j} a^{ij} \frac{\partial^2}{\partial\theta_i\, \partial\theta_j} + \sum_i b^i \frac{\partial^2}{\partial\theta_i\, \partial t} + \frac{\partial^2}{\partial t^2} + \sum_i c^i \frac{\partial}{\partial\theta_i}$$

with regular coefficients a^{ij}, b^i, c^i.

Lemma 3.2 [Longinetti 1987]. *The function $\phi(t) = \max\{G(\theta, t) : \theta \in B\}$ satisfies the differential inequality*

$$D^2\phi(t) \geq 0.$$

Moreover, $\phi(t)$ is a convex function with respect to t.

The lemma is proved only in dimension $n = 2$ in [Longinetti 1987], but it is easy to see that it is valid for the general case $n \geq 2$.

Since the level sets of u are strictly convex with respect to the normal Du, the matrix of second fundamental form (b_{ij}) is positive definite in Ω. Set

$$\varphi = \rho(h_t^2) - \log K(x),$$

where $K = \det(b^{ij})$ is the Gaussian curvature of the level sets and $\rho(t)$ is a smooth function defined on $(0, +\infty)$. For suitable choices of ρ and β, we will derive the differential inequality

$$L(e^{\beta\varphi}) \le 0 \quad \mathrm{mod}\ \nabla_\theta\varphi \quad \text{in } \Omega, \tag{3-1}$$

where the elliptic operator L is given in (2-5) and we have modified the terms involving $\nabla_\theta\varphi$ with locally bounded coefficients. Then, by applying a maximum principle argument in Lemma 3.2, we can obtain the desired result.

In order to prove (3-1) at an arbitrary point $x_0 \in \Omega$, we may assume the matrix $(b_{ij}(x_0))$ is diagonal by rotating the coordinate system suitably. From now on, all the calculations will be done at the fixed point x_0.

Proof of Theorem 1.1. We shall prove the theorem in three steps.

Step 1: *computation* $L(\varphi)$. Taking the first derivative of φ, we get

$$\frac{\partial\varphi}{\partial\theta_j} = 2\rho' h_t h_{tj} + \sum_{k,l} b^{kl} b_{kl,j}, \tag{3-2}$$

$$\frac{\partial\varphi}{\partial t} = 2\rho' h_t h_{tt} + \sum_{k,l} b^{kl} b_{kl,t}. \tag{3-3}$$

Taking the derivative of (3-2) and (3-3) once more, we have

$$\frac{\partial^2\varphi}{\partial\theta_i\,\partial\theta_j} = (2\rho'+4\rho'' h_t^2) h_{ti} h_{tj} + 2\rho' h_t h_{tji} - \sum_{k,l,r,s} b^{kr} b_{rs,i} b^{sl} b_{kl,j} + \sum_{k,l} b^{kl} b_{kl,ji},$$

$$\frac{\partial^2\varphi}{\partial\theta_i\,\partial t} = (2\rho'+4\rho'' h_t^2) h_{ti} h_{tt} + 2\rho' h_t h_{tti} - \sum_{k,l,r,s} b^{kr} b_{rs,i} b^{sl} b_{kl,t} + \sum_{k,l} b^{kl} b_{kl,ti},$$

$$\frac{\partial^2\varphi}{\partial t^2} = (2\rho'+4\rho'' h_t^2) h_{tt}^2 + 2\rho' h_t h_{ttt} - \sum_{k,l,r,s} b^{kr} b_{rs,t} b^{sl} b_{kl,t} + \sum_{k,l} b^{kl} b_{kl,tt}.$$

So we can wrtie

$$L(\varphi) = I_1 + I_2 + I_3 + I_4, \tag{3-4}$$

with

$$I_1 = (2\rho'+4\rho'' h_t^2)\Big[\sum_{i,j}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}] b^{ii} b^{jj} h_{ti} h_{tj} - 2\sum_i h_{ti}^2 b^{ii} h_{tt} + h_{tt}^2\Big],$$

$$I_2 = 2\rho' h_t\Big[\sum_{i,j}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}] b^{ii} b^{jj} h_{tji} - 2\sum_i h_{ti} b^{ii} h_{tti} + h_{ttt}\Big],$$

$$I_3 = -\sum_{k,l} b^{kk} b^{ll}\Big[\sum_{i,j}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}] b^{ii} b^{jj} b_{kl,i} b_{kl,j},\ -2\sum_i h_{ti} b^{ii} b_{kl,i} b_{kl,t} + b_{kl,t}^2\Big]$$

$$I_4 = \sum_k b^{kk} L(b_{kk}).$$

In the rest of this section, we will deal with the four terms above respectively. For the term I_1, by recalling our equation, that is,

$$h_{tt} = \sum_{i,j} [(1 + h_t^2)\delta_{ij} + h_{ti} h_{tj}] b^{ij}, \tag{3-5}$$

we have, by recalling that (b^{ij}) is diagonal at x_0,

$$\begin{aligned} I_1 &= (2\rho' + 4\rho'' h_t^2)\Bigl[\sum_{i,j} [(1 + h_t^2)\delta_{ij} + h_{ti} h_{tj}] b^{ii} b^{jj} h_{ti} h_{tj} - 2\sum_i h_{ti}^2 b^{ii} h_{tt} + h_{tt}^2\Bigr] \\ &= (2\rho' + 4\rho'' h_t^2)\Bigl[(1 + h_t^2)\sum_i (h_{ti} b^{ii})^2 + \Bigl(\sum_i h_{ti}^2 b^{ii} - h_{tt}\Bigr)^2\Bigr] \\ &= (2\rho' + 4\rho'' h_t^2)(1 + h_t^2)\sum_i (h_{ti} b^{ii})^2 + (2\rho' + 4\rho'' h_t^2)(1 + h_t^2)^2 \sigma_1^2, \end{aligned} \tag{3-6}$$

where $\sigma_1 = \sum_i b^{ii}$ is the mean curvature.

Now we treat the term I_2. Differentiating (3-5) with respect to t, we have

$$h_{ttt} = 2h_t h_{tt}\sigma_1 + 2\sum_{i,j} h_{tti} h_{tj} b^{ij} - \sum_{i,j} [(1 + h_t^2)\delta_{ij} + h_{ti} h_{tj}] b^{ii} b^{jj} b_{ij,t}. \tag{3-7}$$

By inserting (3-7) into I_2, we can get

$$\begin{aligned} I_2 &= 2\rho' h_t\Bigl[\sum_{i,j} [(1 + h_t^2)\delta_{ij} + h_{ti} h_{tj}] b^{ii} b^{jj} h_{tji} - 2\sum_i h_{ti} b^{ii} h_{tti} + h_{ttt}\Bigr] \\ &= 2\rho' h_t\Bigl[\sum_{i,j} [(1 + h_t^2)\delta_{ij} + h_{ti} h_{tj}] b^{ii} b^{jj} (h_{tji} - b_{ij,t}) + 2h_t h_{tt}\sigma_1\Bigr]. \end{aligned}$$

Recalling the definition of the second fundamental form, that is, (2-2), together with (3-5), we obtain

$$\begin{aligned} I_2 &= 2\rho' h_t\Bigl[\sum_{i,j} [(1 + h_t^2)\delta_{ij} + h_{ti} h_{tj}] b^{ii} b^{jj} (-h_t\delta_{ij}) + 2h_t h_{tt}\sigma_1\Bigr] \\ &= -2\rho' h_t^2 (1 + h_t^2)\sum_i (b^{ii})^2 - 2\rho' h_t^2 \sum_i (h_{ti} b^{ii})^2 + 4\rho' h_t^2 (1 + h_t^2)\sigma_1^2 \\ &\qquad + 4\rho' h_t^2 \sigma_1 \sum_i h_{ti}^2 b^{ii}. \end{aligned} \tag{3-8}$$

Combining (3-6) and (3-8),

$$\begin{aligned} I_1 + I_2 &= 4\rho' h_t^2 \sigma_1 \sum_i h_{ti}^2 b^{ii} + [4\rho' h_t^2 (1+h_t^2) + (2\rho' + 4\rho'' h_t^2)(1+h_t^2)^2]\sigma_1^2 \\ &\quad + [(2\rho' + 4\rho'' h_t^2)(1+h_t^2) - 2\rho' h_t^2]\sum_i (h_{ti} b^{ii})^2 - 2\rho' h_t^2 (1+h_t^2)\sum_i (b^{ii})^2. \end{aligned} \tag{3-9}$$

In order to deal with the last two terms, we shall compute $L(b_{kk})$ in advance. In this process, the index k is not summed. By differentiating (3-5) twice with respect to θ_k, we have

$$h_{ttkk} = J_1 + J_2 + J_3 + J_4, \tag{3-10}$$

with

$$\begin{aligned}
J_1 &= \sum_{i,j} [(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]_{kk} b^{ij},\\
J_2 &= 2\sum_{ij,p,q} [(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]_k (-b^{ip} b_{pq,k} b^{qj}),\\
J_3 &= \sum_{ij,p,q,r,s} [(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}](2b^{ir} b_{rs,k} b^{sp} b_{pq,k} b^{qj}),\\
J_4 &= \sum_{ij,p,q} [(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}](-b^{ip} b_{pq,kk} b^{qj}).
\end{aligned}$$

For the term J_1, we have

$$\begin{aligned}
J_1 &= \sum_{i,j} (2h_t h_{tk}\delta_{ij} + h_{tik}h_{tj} + h_{ti}h_{tjk})_k b^{ij}\\
&= 2h_{tk}^2\sigma_1 + 2h_t h_{tkk}\sigma_1 + 2\sum_i h_{tikk}h_{ti}b^{ii} + 2\sum_i h_{tik}^2 b^{ii}.
\end{aligned}$$

Noticing that

$$\begin{aligned}
h_{tik} &= h_{kit} = b_{ki,t} - h_t\delta_{ki},\\
h_{tikk} &= h_{ikkt} = b_{ik,kt} - h_{kt}\delta_{ik} = b_{kk,it} - h_{kt}\delta_{ik},
\end{aligned}$$

we obtain

$$\begin{aligned}
J_1 = 2h_{tk}^2\sigma_1 + 2h_t b_{kk,t}\sigma_1 - 2h_t^2\sigma_1 + 2\sum_i b_{kk,it}h_{ti}b^{ii}&\\
- 2h_{tk}^2 b^{kk} + 2\sum_l b_{kl,t}^2 b^{ll} - 4h_t b_{kk,t}b^{kk} + 2h_t^2 b^{kk}&.
\end{aligned} \tag{3-11}$$

For the term J_2, we have

$$\begin{aligned}
J_2 &= 2\sum_{i,j} (2h_t h_{tk}\delta_{ij} + h_{tik}h_{tj} + h_{ti}h_{tjk})(-b^{ii} b_{ij,k} b^{jj})\\
&= -4h_t h_{tk}\sum_i (b^{ii})^2 b_{ii,k} - 4\sum_{i,j} h_{tik}h_{tj}b^{ii}b^{jj}b_{ij,k}\\
&= -4h_t h_{tk}\sum_i (b^{ii})^2 b_{ii,k} - 4\sum_{i,l} h_{ti}b^{ii}b^{ll}b_{kl,i}b_{kl,t}\\
&\qquad + 4h_t\sum_j h_{tj}b^{kk}b^{jj}b_{kk,j}.
\end{aligned} \tag{3-12}$$

Note that we have changed the lower index during the above calculations and this will happen frequently in the following procedure.

Also we have

$$J_3 = 2\sum_{i,j,l}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]b^{ii}b^{jj}b^{ll}b_{kl,i}b_{kl,j}. \tag{3-13}$$

Applying the commutation rule $b_{ij,kl} - b_{ij,lk} = b_{jk}\delta_{il} - b_{jl}\delta_{ik} + b_{ik}\delta_{jl} - b_{il}\delta_{jk}$, for the term J_4, we have

$$\begin{aligned} J_4 &= -\sum_{i,j}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]b^{ii}b^{jj}b_{ij,kk} \\ &= -\sum_{i,j}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]b^{ii}b^{jj}(b_{kk,ij} + b_{ij} - b_{kk}\delta_{ij}). \end{aligned} \tag{3-14}$$

On the other hand,

$$h_{ttkk} = h_{kktt} = b_{kk,tt} - h_{tt} = b_{kk,tt} - \sum_{i,j}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]b^{ij}. \tag{3-15}$$

By putting (3-11)–(3-15) into (3-10), recalling the definition of the operator L, we obtain

$$\begin{aligned} L(b_{kk}) = &\sum_{i,j}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]b^{ij} + 2h_{tk}^2\sigma_1 + 2h_t b_{kk,t}\sigma_1 - 2h_t^2\sigma_1 \\ &- 2h_{tk}^2 b^{kk} + 2\sum_l b_{kl,t}^2 b^{ll} - 4h_t b^{kk}b_{kk,t} + 2h_t^2 b^{kk} - 4h_t h_{tk}\sum_i (b^{ii})^2 b_{ii,k} \\ &- 4\sum_{i,l} h_{ti}b^{ii}b^{ll}b_{kl,i}b_{kl,t} + 2\sum_{i,j,l}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]b^{ii}b^{jj}b^{ll}b_{kl,i}b_{kl,j} \\ &+ 4h_t\sum_i h_{ti}b^{kk}b^{ii}b_{kk,i} - \sum_{i,j}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]b^{ii}b^{jj}(b_{ij} - b_{kk}\delta_{ij}). \end{aligned}$$

Therefore,

$$\begin{aligned} I_4 = 2&\sum_{i,j,k,l}[(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}]b^{ii}b^{jj}b^{kk}b^{ll}b_{kl,i}b_{kl,j} - 4\sum_{i,k,l} h_{ti}b^{ii}b^{kk}b^{ll}b_{kl,i}b_{kl,t} \\ &+ 2h_t\sigma_1\sum_k b^{kk}b_{kk,t} - 4h_t\sum_k (b^{kk})^2 b_{kk,t} - 2h_t^2\sigma_1^2 + 2\sum_{k,l} b^{kk}b^{ll}b_{kl,t}^2 \\ &+ [(n-1)(1+h_t^2) + 2h_t^2]\sum_i (b^{ii})^2 + 2\sigma_1\sum_i h_{ti}^2 b^{ii} \\ &+ (n-3)\sum_i (h_{ti}b^{ii})^2. \end{aligned} \tag{3-16}$$

By substituting (3-9) and (3-16) in (3-4), we obtain

(3-17)
$$
\begin{aligned}
L(\varphi) = {} & \sum_{i,j,k,l} [(1+h_t^2)\delta_{ij} + h_{ti}h_{tj}] b^{ii} b^{jj} b^{kk} b^{ll} b_{kl,i} b_{kl,j} - 2\sum_{i,k,l} h_{ti} b^{ii} b^{kk} b^{ll} b_{kl,i} b_{kl,t} \\
& + \sum_{k,l} b^{kk} b^{ll} b_{kl,t}^2 + 2h_t\sigma_1 \sum_k b^{kk} b_{kk,t} - 4h_t \sum_k (b^{kk})^2 b_{kk,t} \\
& + (2+4\rho' h_t^2)\sigma_1 \sum_i h_{ti}^2 b^{ii} + [(n-1)(1+h_t^2) + 2h_t^2 - 2\rho' h_t^2 (1+h_t^2)] \sum_i (b^{ii})^2 \\
& + [4\rho' h_t^2 (1+h_t^2) + (2\rho' + 4\rho'' h_t^2)(1+h_t^2)^2 - 2h_t^2]\sigma_1^2 \\
& + [(2\rho' + 4\rho'' h_t^2)(1+h_t^2) - 2\rho' h_t^2 + (n-3)] \sum_i (h_{ti} b^{ii})^2 .
\end{aligned}
$$

Step 2: *calculation of* $L(e^{\beta\varphi})$ *and estimation of the third-order derivatives involving* $b_{kk,t}$. Notice that

$$
\begin{aligned}
L(e^{\beta\varphi}) = \beta e^{\beta\varphi}\{L(\varphi) + \beta\varphi_t^2\} + \beta^2 e^{\beta\varphi} \sum_{i,j,p,q} [(1+h_t^2)\delta_{pq} + h_{tp}h_{tq}] b^{ip} b^{jq} \frac{\partial\varphi}{\partial\theta_i}\frac{\partial\varphi}{\partial\theta_j} \\
- 2\beta^2 e^{\beta\varphi} \sum_{i,j} h_{tj} b^{ij} \frac{\partial\varphi}{\partial\theta_i}\frac{\partial\varphi}{\partial t}.
\end{aligned}
$$

To reach (3-1), we only need to prove that, for some constant $\beta < 0$,

$$L(\varphi) + \beta\varphi_t^2 \geq 0 \quad \operatorname{mod} \nabla_\theta \varphi.$$

We now compute $\beta\varphi_t^2$.

By (3-3), we have

$$
\begin{aligned}
\text{(3-18)} \quad \varphi_t^2 &= 4(\rho')^2 h_t^2 h_{tt}^2 + 4\rho' h_t h_{tt} \sum_k b^{kk} b_{kk,t} + \Big(\sum_k b^{kk} b_{kk,t}\Big)^2 \\
&= 4(\rho')^2 h_t^2 (1+h_t^2)^2 \sigma_1^2 + 8(\rho')^2 h_t^2 (1+h_t^2)\sigma_1 \sum_i h_{ti}^2 b^{ii} \\
&\quad + 4(\rho')^2 h_t^2 \Big(\sum_i h_{ti}^2 b^{ii}\Big)^2 + 4\rho' h_t (1+h_t^2)\sigma_1 \sum_k b^{kk} b_{kk,t} \\
&\quad + 4\rho' h_t \Big(\sum_i h_{ti}^2 b^{ii}\Big)\Big(\sum_k b^{kk} b_{kk,t}\Big) + \Big(\sum_k b^{kk} b_{kk,t}\Big)^2 .
\end{aligned}
$$

Joining (3-17) with (3-18), we regroup the terms in $L(\varphi) + \beta\varphi_t^2$ as follows:

$$L(\varphi) + \beta\varphi_t^2 = P_1 + P_2 + P_3,$$

where

$$P_1 = \sum_{k\neq l}\biggl(\sum_{i,j} h_{ti}h_{tj}b^{ii}b^{jj}b^{kk}b^{ll}b_{kl,i}b_{kl,j} - 2\sum_i h_{ti}b^{ii}b^{kk}b^{ll}b_{kl,i}b_{kl,t} + b^{kk}b^{ll}b^2_{kl,t}\biggr),$$

$$\begin{aligned} P_2 = {} & \sum_k (b^{kk}b_{kk,t})^2 + \beta\Bigl(\sum_k b^{kk}b_{kk,t}\Bigr)^2 \\ & + 2\sum_k\biggl[[1+2\beta\rho'(1+h_t^2)]h_t\sigma_1 + 2\beta\rho' h_t\Bigl(\sum_i h_{ti}^2 b^{ii}\Bigr) \\ & \qquad - \sum_i h_{ti}b^{ii}b^{kk}b_{kk,i} - 2h_t b^{kk}\biggr]\cdot(b^{kk}b_{kk,t}), \end{aligned}$$

$$\begin{aligned} P_3 = {} & (1+h_t^2)\sum_{i,k,l}(b^{ii})^2 b^{kk}b^{ll}b^2_{kl,i} + \sum_{i,j,k} h_{ti}h_{tj}b^{ii}b^{jj}b^{kk}b_{kk,i}b^{kk}b_{kk,j} \\ & + [2+4\rho' h_t^2 + 8\beta(\rho')^2 h_t^2(1+h_t^2)]\sigma_1\sum_i h_{ti}^2 b^{ii} \\ & + [(n-1)(1+h_t^2) + 2h_t^2 - 2\rho' h_t^2(1+h_t^2)]\sum_i (b^{ii})^2 \\ & + [4\rho' h_t^2(1+h_t^2) + (2\rho' + 4\rho'' h_t^2)(1+h_t^2)^2 - 2h_t^2 + 4\beta(\rho')^2 h_t^2(1+h_t^2)^2]\sigma_1^2 \\ & + [(2\rho' + 4\rho'' h_t^2)(1+h_t^2) - 2\rho' h_t^2 + (n-3)]\sum_i (h_{ti}b^{ii})^2 \\ & + 4\beta(\rho')^2 h_t^2\Bigl(\sum_i h_{ti}^2 b^{ii}\Bigr)^2. \end{aligned}$$

In the rest of this step, we will deal with the term P_2. Let $X_k = b^{kk}b_{kk,t}(k = 1, 2, \dots, n-1)$. Then P_2 can be rewritten as

$$P_2(X_1, X_2, \dots, X_{n-1}) = \sum_k X_k^2 + \beta\Bigl(\sum_k X_k\Bigr)^2 + 2\sum_k c_k X_k,$$

where

$$c_k = [1+2\beta\rho'(1+h_t^2)]h_t\sigma_1 + 2\beta\rho' h_t\Bigl(\sum_i h_{ti}^2 b^{ii}\Bigr) - \sum_i h_{ti}b^{ii}b^{kk}b_{kk,i} - 2h_t b^{kk}.$$

Denote by $\mathcal{P}_2$ the matrix

$$\begin{pmatrix} 1+\beta & \beta & \cdots & \beta \\ \beta & 1+\beta & \cdots & \beta \\ \vdots & \vdots & \ddots & \cdots \\ \beta & \beta & \cdots & 1+\beta \end{pmatrix}.$$

In a word, we want to bound $P_2(X_1, X_2, \ldots, X_{n-1})$ from below. Thus the nonnegativity of $\mathcal{P}_2$ is necessary, and this requires

$$\beta \geq -\frac{1}{n-1}.$$

For convenience, Let us choose the degenerate case, that is, $\beta = -1/(n-1)$. By setting $\tau = (1, 1, \ldots, 1)$, the null eigenvector of the matrix $\mathcal{P}_2$, we then have, by (3-2),

$$(\star) \qquad P_2(1, 1, \ldots, 1) = 2\sum_k c_k = 2[n-3-2\rho'(1+h_t^2)]h_t\sigma_1 - 2\sum_i h_{ti}b^{ii}\frac{\partial\varphi}{\partial\theta_i},$$

which suggests that the simplest selection should be $\rho(t) = ((n-3)/2)\log(1+t)$.

From now on, let us fix $\rho(t) = ((n-3)/2)\log(1+t)$ and $\beta = -1/(n-1)$. But, for simplicity, we do not always substitute for the values of ρ and β.

By straightforward computation and $(\star)$, we have

$$\sum_k \Big(X_k + \beta\sum_i X_i + c_k\Big)^2 = P_2(X_1, X_2, \ldots, X_{n-1}) + \sum_k c_k^2 + P_2(\nabla_\theta\varphi),$$

where

$$P_2(\nabla_\theta\varphi) = 2\beta\Big(\sum_i X_i\Big)\sum_k c_k = 2\beta\Big(\sum_j X_j\Big)\sum_i h_{ti}b^{ii}\frac{\partial\varphi}{\partial\theta_i}.$$

Putting ρ and β into some terms in c_k, we derive that

$$c_k = \frac{2}{n-1}h_t\sigma_1 - \frac{2}{n-1}\rho' h_t\Big(\sum_i h_{ti}^2 b^{ii}\Big) - \sum_i h_{ti}b^{ii}b^{kk}b_{kk,i} - 2h_t b^{kk}.$$

Therefore, together with (3-2), we get

$$\begin{aligned}
&P_2(X_1, X_2, \ldots, X_{n-1}) \\
&\quad \geq -\sum_k c_k^2 - P_2(\nabla_\theta\varphi) \\
&\quad = -\sum_{i,j,k} h_{ti}h_{tj}b^{ii}b^{jj}b^{kk}b_{kk,i}b^{kk}b_{kk,j} - 4h_t\sum_{i,k} h_{ti}b^{ii}(b^{kk})^2 b_{kk,i} \\
&\qquad\qquad - 4h_t^2\sum_k (b^{kk})^2 + \frac{4}{n-1}h_t^2\sigma_1^2 - \frac{8}{n-1}\rho' h_t^2\sigma_1\sum_i h_{ti}^2 b^{ii} \\
&\qquad\qquad\qquad + \frac{4}{n-1}h_t^2(\rho')^2\Big(\sum_i h_{ti}^2 b^{ii}\Big)^2 + \widetilde{P}_2(\nabla_\theta\varphi),
\end{aligned}$$

where

$$\widetilde{P}_2(\nabla_\theta\varphi) = -P_2(\nabla_\theta\varphi) - \frac{4}{n-1}h_t\Big[\sigma_1 - \rho'\sum_j h_{tj}^2 b^{jj}\Big]\sum_i h_{ti}b^{ii}\frac{\partial\varphi}{\partial\theta_i}.$$

Observing that $P_1 \geq 0$,

$$
\begin{aligned}
(3\text{-}19)\quad & L(\varphi)+\beta\varphi_t^2 \\
&\geq (1+h_t^2)\sum_{i,k,l}(b^{ii})^2 b^{kk} b^{ll} b_{kl,i}^2 - 4h_t \sum_{i,k} h_{ti} b^{ii}(b^{kk})^2 b_{kk,i} \\
&\qquad + \Big[2+4\rho' h_t^2 + 8\beta(\rho')^2 h_t^2(1+h_t^2) - \frac{8}{n-1}\rho' h_t^2\Big]\sigma_1 \sum_i h_{ti}^2 b^{ii} \\
&\qquad + [(n-1)(1+h_t^2) - 2h_t^2 - 2\rho' h_t^2(1+h_t^2)] \sum_i (b^{ii})^2 \\
&\quad + \Big[4\rho' h_t^2(1+h_t^2) + [(2\rho'+4\rho'' h_t^2) + 4\beta(\rho')^2 h_t^2](1+h_t^2)^2 - \frac{2n-6}{n-1}h_t^2\Big]\sigma_1^2 \\
&\quad + [(2\rho'+4\rho'' h_t^2)(1+h_t^2) - 2\rho' h_t^2 + (n-3)] \sum_i (h_{ti} b^{ii})^2 + \widetilde{P}_2(\nabla_\theta \varphi).
\end{aligned}
$$

In the next step we will concentrate on the following two terms:

$$
R = (1+h_t^2)\sum_{i,k,l}(b^{ii})^2 b^{kk} b^{ll} b_{kl,i}^2 - 4h_t \sum_{i,k} h_{ti} b^{ii}(b^{kk})^2 b_{kk,i}.
$$

Step 3: *conclusion of the proof of* (3-1). Recalling our first-order condition (3-2), we have

$$
(3\text{-}20)\qquad b^{11} b_{11,j} = \frac{\partial\varphi}{\partial\theta_j} - \sum_{k\geq 2} b^{kk} b_{kk,j} - 2\rho' h_t h_{tj} \quad \text{for } j=1,2,\ldots,n-1.
$$

For the term R, we have

$$
\begin{aligned}
R &= (1+h_t^2)\Big[\sum_i \sum_{k\neq l}(b^{ii})^2 b^{kk} b^{ll} b_{kl,i}^2 + \sum_{i,k}(b^{ii})^2 (b^{kk} b_{kk,i})^2\Big] \\
&\qquad\qquad - 4\sum_{i,k} h_t h_{ti} b^{ii}(b^{kk})^2 b_{kk,i} \\
&= (1+h_t^2)\Big[2\sum_{k\geq 2}(b^{11})^2 b^{kk} b^{11} b_{k1,1}^2 + 2\sum_{i,k\geq 2}(b^{ii})^2 b^{kk} b^{11} b_{k1,i}^2 \\
&\qquad + \sum_i \sum_{\substack{k,l\geq 2\\ k\neq l}} (b^{ii})^2 b^{kk} b^{ll} b_{kl,i}^2 + \sum_i (b^{ii})^2 (b^{11} b_{11,i})^2 \\
&\qquad + \sum_i \sum_{k\geq 2}(b^{ii})^2 (b^{kk} b_{kk,i})^2\Big] \\
&\qquad - 4\sum_i h_t h_{ti} b^{ii}(b^{11})^2 b_{11,i} - 4\sum_i \sum_{k\geq 2} h_t h_{ti} b^{ii}(b^{kk})^2 b_{kk,i} \\
&= R_1 + R_2 + R_3,
\end{aligned}
$$

where

$$R_1=(1+h_t^2)\Big[2\sum_{k\geq 2}(b^{11})^2b^{kk}b^{11}b_{k1,1}^2+\sum_i(b^{ii})^2(b^{11}b_{11,i})^2\Big]\\ -4\sum_i h_th_{ti}b^{ii}(b^{11})^2b_{11,i},$$

$$R_2=2\sum_{i,k\geq 2}(1+h_t^2)(b^{ii})^2b^{kk}b^{11}b_{k1,i}^2+\sum_i\sum_{\substack{k,l\geq 2\\ k\neq l}}(1+h_t^2)(b^{ii})^2b^{kk}b^{ll}b_{kl,i}^2,$$

$$R_3=\sum_i\sum_{k\geq 2}(1+h_t^2)(b^{ii})^2(b^{kk}b_{kk,i})^2-4\sum_i\sum_{k\geq 2}h_th_{ti}b^{ii}(b^{kk})^2b_{kk,i}.$$

By (3-20), one has

$$R_1=(1+h_t^2)\Big[2b^{11}\sum_{i,k,l\geq 2}b^{ii}b^{kk}b^{ll}b_{kk,i}b_{ll,i}+8\rho'h_tb^{11}\sum_{i,k\geq 2}h_{ti}b^{ii}b^{kk}b_{kk,i}\\ +8(\rho')^2h_t^2b^{11}\sum_{i\geq 2}h_{ti}^2b^{ii}+\sum_i\sum_{k,l\geq 2}(b^{ii})^2b^{kk}b^{ll}b_{kk,i}b_{ll,i}\\ +4\rho'h_t\sum_i\sum_{k\geq 2}h_{ti}(b^{ii})^2b^{kk}b_{kk,i}+4(\rho')^2h_t^2\sum_i(h_{ti}b^{ii})^2\Big]\\ +4h_t\sum_i\sum_{k\geq 2}h_{ti}b^{ii}b^{11}b^{kk}b_{kk,i}+8\rho'h_t^2b^{11}\sum_i h_{ti}^2b^{ii}+R(\nabla_\theta\varphi),$$

where

$$R(\nabla_\theta\varphi)=(1+h_t^2)\Big[2b^{11}\sum_{k\geq 2}b^{kk}\Big(\frac{\partial\varphi}{\partial\theta_k}\Big)^2-4b^{11}\sum_{k,l\geq 2}b^{kk}b^{ll}b_{ll,k}\frac{\partial\varphi}{\partial\theta_k}\\ -8\rho'h_tb^{11}\sum_{k\geq 2}b^{kk}h_{tk}\frac{\partial\varphi}{\partial\theta_k}+\sum_i(b^{ii})^2\Big(\frac{\partial\varphi}{\partial\theta_i}\Big)^2\\ -2\sum_i\sum_{k\geq 2}(b^{ii})^2b^{kk}b_{kk,i}\frac{\partial\varphi}{\partial\theta_i}-4\rho'h_t\sum_i(b^{ii})^2h_{ti}\frac{\partial\varphi}{\partial\theta_i}\Big]\\ -4h_tb^{11}\sum_i b^{ii}h_{ti}\frac{\partial\varphi}{\partial\theta_i}.$$

On the other hand,

$$R_2=(1+h_t^2)\Big[2b^{11}\sum_{k\geq 2}(b^{kk})^3b_{kk,1}^2+2\sum_{\substack{i,k\geq 2\\ i\neq k}}(b^{ii})^2b^{kk}b^{11}b_{k1,i}^2\\ +2\sum_{\substack{i,k\geq 2\\ i\neq k}}b^{ii}(b^{kk})^3b_{kk,i}^2+\sum_i\sum_{\substack{k,l\geq 2\\ k\neq l,k\neq i,l\neq i}}(b^{ii})^2b^{kk}b^{ll}b_{kl,i}^2\Big].$$

Recall that $2\rho'(1+h_t^2) = n-3$, which will be denoted by α for simplicity in the following calculations. Now we are at a stage where we can rewrite the terms in R in a natural way: we denote by T_1 the terms involving $b_{kk,1} (k \geq 2)$, by T_2 the terms involving $b_{kk,i} (k, i \geq 2)$, and by T_3 all of the rest of the terms. More precisely,

$$T_1 = \sum_{k\geq 2}(1+2b_{11}b^{kk})\cdot((1+h_t^2)^{1/2}b^{11}b^{kk}b_{kk,1})^2 + \left(\sum_{k\geq 2}(1+h_t^2)^{1/2}b^{11}b^{kk}b_{kk,1}\right)^2 + 4h_t h_{t1}b^{11}(1+h_t^2)^{-1/2}\sum_{k\geq 2}\left(1+\frac{\alpha}{2}-b_{11}b^{kk}\right)\cdot((1+h_t^2)^{1/2}b^{11}b^{kk}b_{kk,1})$$

and

$$T_2 = (1+h_t^2)\sum_{i\geq 2}\Bigg\{(1+2b_{ii}b^{11})\cdot\left(\sum_{k\geq 2}b^{ii}b^{kk}b_{kk,i}\right)^2 + \sum_{\substack{k\geq 2\\ k\neq i}}2b_{ii}b^{kk}\cdot(b^{ii}b^{kk}b_{kk,i})^2 + \sum_{k\geq 2}(b^{ii}b^{kk}b_{kk,i})^2 + 4h_t h_{ti}b^{ii}(1+h_t^2)^{-1} \times \sum_{k\geq 2}\left[-b_{ii}b^{kk}+\frac{\alpha}{2}+(1+\alpha)b_{ii}b^{11}\right]\cdot(b^{ii}b^{kk}b_{kk,i})\Bigg\};$$

the rest of the terms are

$$T_3 = h_t^2(1+h_t^2)^{-1}\left[2\alpha^2 b^{11}\sum_{i\geq 2}h_{ti}^2 b^{ii} + \alpha^2\sum_i (h_{ti}b^{ii})^2 + 4\alpha b^{11}\sum_i h_{ti}^2 b^{ii}\right] + (1+h_t^2)\left[2\sum_{\substack{i,k\geq 2\\ i\neq k}}(b^{ii})^2 b^{kk}b^{11}b_{k1,i}^2 + \sum_i \sum_{\substack{k,l\geq 2\\ k\neq l, k\neq i, l\neq i}}(b^{ii})^2 b^{kk}b^{ll}b_{kl,i}^2\right] + R(\nabla_\theta\varphi). \tag{3-21}$$

We shall minimize the terms T_1 and T_2 via Lemma 3.1 for different choices of parameters.

At first, let us examine the term T_1. set $X_k = (1+h_t^2)^{1/2}b^{11}b^{kk}b_{kk,1}$, $\lambda = 1$, $\mu = h_{t1}b^{11}h_t(1+h_t^2)^{-1/2}$, $b_k = 1+2b_{11}b^{kk}$, and $c_k = b_{11}b^{kk} - (1+\alpha/2)$, where $k \geq 2$. By Lemma 3.1, we have

$$T_1 \geq -4h_t^2(1+h_t^2)^{-1}(h_{t1}b^{11})^2\Gamma_1,$$

where

$$\Gamma_1 = \sum_{k\geq 2}\frac{c_k^2}{b_k} - \left(1+\sum_{k\geq 2}\frac{1}{b_k}\right)^{-1}\left(\sum_{k\geq 2}\frac{c_k}{b_k}\right)^2.$$

Next we shall simplify Γ_1. By denoting

$$\beta_k = \frac{1}{b_k},$$

we have

$$b_{11}b^{kk} = \frac{1}{2\beta_k} - \tfrac{1}{2}, \qquad c_k = \frac{1}{2\beta_k} - \frac{3+\alpha}{2}.$$

Hence

$$\begin{aligned}\Gamma_1 &= \sum_{k\geq 2}\beta_k\Big(\frac{1}{2\beta_k} - \frac{3+\alpha}{2}\Big)^2 - \Big(1+\sum_{k\geq 2}\beta_k\Big)^{-1}\Big[\sum_{k\geq 2}\beta_k\Big(\frac{1}{2\beta_k} - \frac{3+\alpha}{2}\Big)\Big]^2 \\ &= \frac{1}{4}\sum_{k\geq 2}\frac{1}{\beta_k} - \Big(1+\sum_{k\geq 2}\beta_k\Big)^{-1}\frac{(n+1+\alpha)^2}{4} + \frac{(3+\alpha)^2}{4}.\end{aligned}$$

Since

$$1 \leq 1 + \sum_{k\geq 2}\beta_k \leq n-1,$$

it follows that

$$\begin{aligned}\Gamma_1 &\leq \frac{1}{4}\sum_{k\geq 2}\frac{1}{\beta_k} - \frac{(n+1+\alpha)^2}{4(n-1)} + \frac{(3+\alpha)^2}{4} \\ &= \frac{n-2}{4(n-1)}(2+\alpha)^2 + \tfrac{1}{4}(2\sigma_1 b_{11} - 2).\end{aligned}$$

Therefore,

$$T_1 \geq -\Big[\frac{(n-2)}{n-1}(2+\alpha)^2 + 2\sigma_1 b_{11} - 2\Big]h_t^2(1+h_t^2)^{-1}(h_{t1}b^{11})^2. \tag{3-22}$$

Now we will deal with T_2. For every $i \geq 2$ fixed, set $X_k = b^{ii}b^{kk}b_{kk,i}$, $\lambda = 1 + 2b_{ii}b^{11}$, $\mu = -h_{ti}b^{ii}h_t(1+h_t^2)^{-1}$, $b_k = 1 + 2b_{ii}b^{kk} (k \neq i)$, $b_i = 1$, and $c_k = b_{ii}b^{kk} - \frac{1}{2}\alpha - (1+\alpha)b_{ii}b^{11}$. By Lemma 3.1, we have

$$T_2 \geq -4(1+h_t^2)\sum_{i\geq 2}(h_{ti}b^{ii})^2\Gamma_i,$$

where

$$\Gamma_i = c_i^2 + \sum_{\substack{k\geq 2\\ k\neq i}}\frac{c_k^2}{b_k} - \Big(\frac{1}{\lambda} + 1 + \sum_{\substack{k\geq 2\\ k\neq i}}\frac{1}{b_k}\Big)^{-1}\Big(c_i + \sum_{\substack{k\geq 2\\ k\neq i}}\frac{c_k}{b_k}\Big)^2.$$

For $k \neq i$, denoting

$$\beta_k = \frac{1}{b_k},$$

we have

$$b_{ii}b^{kk} = \frac{1}{2\beta_k} - \tfrac{1}{2}, \quad c_k = \frac{1}{2\beta_k} - \delta,$$

where

$$\delta = \frac{1+\alpha}{2} + (1+\alpha)b_{ii}b^{11}.$$

Noticing that

$$c_i = \tfrac{3}{2} - \delta, \quad \frac{\delta}{\lambda} = \frac{1+\alpha}{2},$$

we obtain

$$\begin{aligned}
\Gamma_i &= c_i^2 + \sum_{\substack{k\geq 2\\ k\neq i}} \beta_k \big(\frac{1}{2\beta_k} - \delta\big)^2 - \Big(\frac{1}{\lambda} + 1 + \sum_{\substack{k\geq 2\\ k\neq i}} \beta_k\Big)^{-1} \Big[c_i + \sum_{\substack{k\geq 2\\ k\neq i}} \beta_k \Big(\frac{1}{2\beta_k} - \delta\Big)\Big]^2 \\
&= \frac{1}{4} \sum_{\substack{k\geq 2\\ k\neq i}} \frac{1}{\beta_k} - \Big(\frac{1}{\lambda} + 1 + \sum_{\substack{k\geq 2\\ k\neq i}} \beta_k\Big)^{-1} \Big(\frac{n}{2} + \frac{\delta}{\lambda}\Big)^2 + \frac{9}{4} + \frac{\delta^2}{\lambda} \\
&= \frac{1}{4} \sum_{\substack{k\geq 2\\ k\neq i}} \frac{1}{\beta_k} - \Big(\frac{1}{\lambda} + 1 + \sum_{\substack{k\geq 2\\ k\neq i}} \beta_k\Big)^{-1} \frac{(n+1+\alpha)^2}{4} + \frac{9}{4} + \frac{1+\alpha}{2}\delta.
\end{aligned}$$

Obviously,

$$1 \leq \frac{1}{\lambda} + 1 + \sum_{\substack{k\geq 2\\ k\neq i}} \beta_k \leq n-1,$$

hence

$$\begin{aligned}
\Gamma_i &\leq \frac{1}{4} \sum_{\substack{k\geq 2\\ k\neq i}} \frac{1}{\beta_k} - \frac{(n+1+\alpha)^2}{4(n-1)} + \frac{9}{4} + \frac{1+\alpha}{2}\delta \\
&= \frac{n-2}{4(n-1)}\alpha^2 - \frac{1}{n-1}\alpha + \frac{n-3}{2(n-1)} + \tfrac{1}{2}\sigma_1 b_{ii} + \tfrac{1}{2}\alpha^2 b_{ii} b^{11} + \alpha b_{ii} b^{11}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
T_2 \geq -\frac{h_t^2}{1+h_t^2} \sum_{i\geq 2} \Big(&\frac{n-2}{n-1}\alpha^2 - \frac{4}{n-1}\alpha + \frac{2n-6}{n-1} \\
&+ 2\sigma_1 b_{ii} + 2\alpha^2 b_{ii} b^{11} + 4\alpha b_{ii} b^{11}\Big)(h_{ti} b^{ii})^2.
\end{aligned} \tag{3-23}$$

Now, combining (3-21) , (3-22), and (3-23), we obtain

$$R \geq \frac{h_t^2}{1+h_t^2} \sum_i \Big(\frac{1}{n-1}\alpha^2 + \frac{4}{n-1}\alpha - \frac{2n-6}{n-1} - 2\sigma_1 b_{ii}\Big)(h_{ti} b^{ii})^2 + R(\nabla_\theta \varphi). \tag{3-24}$$

For choices of ρ and β, by (3-19) and (3-24), we have, for $n \geq 2$,

$$\begin{aligned}
L(\varphi) - \frac{1}{n-1}\varphi_t^2 &\geq \frac{2\sigma_1}{1+h_t^2} \sum_i h_{ti}^2 b^{ii} + (n-1) \sum_i (b^{ii})^2 + (n-3)\sigma_1^2 \\
&\qquad + \frac{2(n-3)}{1+h_t^2} \sum_i (h_{ti} b^{ii})^2 + \widetilde{P}_2(\nabla_\theta \varphi) + R(\nabla_\theta \varphi) \\
&\geq 0 \mod \nabla_\theta \varphi.
\end{aligned}$$

The proof of (3-1) is completed. □

Now we give a remark on Theorem 1.1.

Remark 3.3. In the proof of Theorem 1.1, if we restrict to the case $n = 2$ and just set $\rho = 0$, then (3-2) shows that

$$b_{11,1} = 0 \mod \nabla_\theta \varphi.$$

Applying this to the expression of $L(\varphi)$ in (3-17) will give

$$\begin{aligned} L(\varphi) &= (b^{11} b_{11,t})^2 - 2h_t (b^{11})^2 b_{11,t} + (b^{11})^2 h_{t1}^2 + (1 + h_t^2)(b^{11})^2 \\ &= [b^{11} b_{11,t} - h_t b^{11}]^2 + (b^{11})^2 h_{t1}^2 + (b^{11})^2 \geq 0 \mod \nabla_\theta \varphi, \end{aligned}$$

and this means that, for any point $x \in \Gamma_t$, $0 < t < 1$,

$$\log K(x) \geq (1-t) \min_{\partial\Omega_0} \log K + t \min_{\partial\Omega_1} \log K,$$

which has already been proved by Longinetti [1987]. Also, by Remark 1.3 we know that this estimate is not sharp in the two-dimensional case.

Acknowledgments

The author thanks Professor X. Ma for many useful discussions on this subject, and the School of Mathematical Sciences of University of Sciences and Technology of China for hospitality.

The author also thanks the referees for their careful efforts to make the paper clearer.

Part of this work was done while the first author was staying at his postdoctoral mobile research station in QFNU.

References

[Caffarelli and Spruck 1982] L. A. Caffarelli and J. Spruck, "Convexity properties of solutions to some classical variational problems", *Comm. Partial Differential Equations* **7**:11 (1982), 1337–1379. MR 85f:49062 Zbl 0508.49013

[Chen and Shi 2011] C. Chen and S. Shi, "Curvature estimates for the level sets of spatial quasiconcave solutions to a class of parabolic equations", *Sci. China Math.* **54**:10 (2011), 2063–2080. MR 2838121

[Chiti and Longinetti 1992] G. Chiti and M. Longinetti, "Differential inequalities for Minkowski functionals of level sets", pp. 109–127 in *General inequalities, 6* (Oberwolfach, 1990), edited by W. Walter, Internat. Ser. Numer. Math. **103**, Birkhäuser, Basel, 1992. MR 94b:49073 Zbl 0764.53005

[Colesanti and Salani 2003] A. Colesanti and P. Salani, "The Brunn–Minkowski inequality for p-capacity of convex bodies", *Math. Ann.* **327**:3 (2003), 459–479. MR 2004j:31007 Zbl 1052.31005

[Jost et al. 2012] J. Jost, X.-N. Ma, and Q. Ou, "Curvature estimates in dimensions 2 and 3 for the level sets of p-harmonic functions in convex rings", *Trans. Amer. Math. Soc.* **364**:9 (2012), 4605–4627. MR 2922603 Zbl 06191423

[Longinetti 1987] M. Longinetti, "On minimal surfaces bounded by two convex curves in parallel planes", *J. Differential Equations* **67**:3 (1987), 344–358. MR 88m:58035 Zbl 0626.53002

[Longinetti and Salani 2007] M. Longinetti and P. Salani, "On the Hessian matrix and Minkowski addition of quasiconvex functions", *J. Math. Pures Appl.* (9) **88**:3 (2007), 276–292. MR 2008k:35163 Zbl 1144.26017

[Ma et al. 2010] X.-N. Ma, Q. Ou, and W. Zhang, "Gaussian curvature estimates for the convex level sets of p-harmonic functions", *Comm. Pure Appl. Math.* **63**:7 (2010), 935–971. MR 2011d:35168 Zbl 1193.35031

[Ma et al. 2011] X.-N. Ma, J. Ye, and Y.-H. Ye, "Principal curvature estimates for the level sets of harmonic functions and minimal graph in $\mathbb{R}^3$", *Comm. Pure Appl. Anal.* **10**:1 (2011), 225–243. MR 2012d:35134 Zbl 1235.35083

[Schneider 1993] R. Schneider, *Convex bodies: the Brunn–Minkowski theory*, Encyclopedia of Mathematics and its Applications **44**, Cambridge University Press, 1993. MR 94d:52007 Zbl 0798.52001

[Wang and Zhang 2012] P. Wang and W. Zhang, "Gaussian curvature estimates for the convex level sets of solutions for some nonlinear elliptic partial differential equations", *J. Partial Differ. Equ.* **25**:3 (2012), 239–275. MR 2987367 Zbl 06178153 arXiv 1003.2057

Received February 16, 2012. Revised June 10, 2013.

PEI-HE WANG
SCHOOL OF MATHEMATICAL SCIENCES
QUFU NORMAL UNIVERSITY
QUFU, 273165, SHANDONG PROVINCE
CHINA

peihewang@hotmail.com

CONTENTS

Volume 267, no. 1 and no. 2

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 267 No. 2 February 2014

0030-8730(2014)267:2;1-6