Pacific Journal of Mathematics

Volume 330 No. 1 May 2024

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

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

The subscription price for 2024 is US $645/year for the electronic version, and $875/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber 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 1945-5844 electronic, 0030-8730 printed) 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. 330, No. 1, 2024
https://doi.org/10.2140/pjm.2024.330.1

MONOTONE TWIST MAPS AND DOWKER-TYPE THEOREMS

PETER ALBERS AND SERGE TABACHNIKOV

Given a planar oval, consider the maximal area of inscribed n-gons resp. the minimal area of circumscribed n-gons. One obtains two sequences indexed by n, and one of Dowker's theorems states that the first sequence is concave and the second is convex. In total, there are four such classic results, concerning areas resp. perimeters of inscribed resp. circumscribed polygons, due to Dowker, Molnár, and Eggleston. We show that these four results are all incarnations of the convexity property of Mather's β-function (the minimal average action function) of the respective billiard-type systems. We then derive new geometric inequalities of similar type for various other billiard systems. Some of these billiards have been thoroughly studied, and some are novel. Moreover, we derive new inequalities (even for conventional billiards) for higher rotation numbers.

1. Introduction

The classic Dowker theorem [16] concerns extremal polygons inscribed and circumscribed about an oval[1]. Here is its formulation.

Let γ be a smooth strictly convex closed plane curve (an oval). Denote by P_n the maximal area of n-gons inscribed in γ and by Q_n the minimal area of n-gons circumscribed about γ. Assume that $n \geq 4$. Then

$$P_{n-1} + P_{n+1} \leq 2P_n \quad \text{and} \quad Q_{n-1} + Q_{n+1} \geq 2Q_n, \tag{1}$$

see Figure 1. An analog of this result in the spherical geometry is due to L. Fejes Tóth [20].

A similar result holds for perimeters. Let R_n be the maximal perimeter of the n-gons inscribed in γ and S_n be the minimal perimeter of the n-gons circumscribed about γ. A theorem, due to Molnár [35] and Eggleston [18], states that

$$R_{n-1} + R_{n+1} \leq 2R_n \quad \text{and} \quad S_{n-1} + S_{n+1} \geq 2S_n. \tag{2}$$

See [19; 25; 29; 36] for surveys and ramifications of these results.

MSC2020: 37C83, 52-XX.
Keywords: monotone twist maps, Mather beta function, geometric approximation,

[1]Not to be confused with another classic Dowker theorem [17].

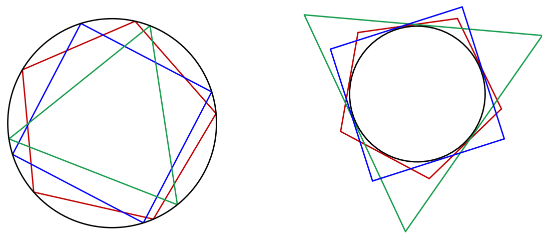

Figure 1. Left: $P_3 + P_5 \le 2P_4$. Right: $Q_3 + Q_5 \ge 2Q_4$.

In this article we show that these four results are particular cases of the convexity of the minimal average action function (Mather's β-function) of monotone twist maps, a result from the Aubry–Mather theory. The maps in question are various kinds of billiards: conventional billiards in γ for the perimeter of inscribed polygons, outer billiards about γ for the area of circumscribed polygons, symplectic billiards in γ for the area of inscribed polygons, and outer length billiards about γ for the perimeter of circumscribed polygons.

The first two billiard systems have been thoroughly studied for a long time; we refer to [42] and [43] for surveys. Symplectic billiards were introduced only recently [1]. As to outer length billiards (the "fourth billiards"), to the best of our knowledge, they were not studied before. We define them here and we plan to provide more details in the upcoming article [3].

We also apply the convexity of the minimal average action function to other billiard-like systems: wire billiards, wire symplectic billiards, magnetic billiards, and outer magnetic billiards.

Wire billiards were introduced and studied in [10]; this is a dynamical system on the set of chords of a closed curve (satisfying certain conditions) in $\mathbb{R}^n$. The resulting Dowker-type inequality concerns the maximal perimeters of inscribed n-gons, that is, n-gons whose vertices lie on the curve.

Similarly, wire symplectic billiards is a dynamical system on the set of chords of a closed curve (also satisfying certain conditions) in linear symplectic space $\mathbb{R}^{2n}$. The resulting inequality concerns the maximal symplectic areas of inscribed n-gons. Wire symplectic billiards are introduced here for the first time.

Magnetic billiards describe the motion of a charge in a magnetic field subject to the elastic reflections off the boundary of a plane domain. We consider the case of a weak constant magnetic field when the trajectory of a charge comprises circular arcs of a fixed radius that is greater than the greatest radius of curvature of the boundary of the domain, the oval γ. The resulting Dowker-style geometric inequality combines the perimeter of a trajectory with the area bounded by it.

Outer magnetic billiards are similar to outer billiards, but instead of tangent lines one considers tangent arcs of a sufficiently great fixed radius. The resulting inequality concerns the minimal areas of circumscribed curvilinear n-gons. Outer magnetic billiards were introduced and studied in [9].

Let us emphasize that the classic Dowker-type theorems concern simple polygons, whereas our inequalities include star-shaped polygons as well: the former have rotation numbers $1/q$, and the latter have the more general rotation numbers p/q. The corresponding inequalities are new even in the area/perimeter and inscribed/circumscribed cases.

We hope that the topic of this article is of interest to two research communities, the Hamiltonian dynamical and the convex geometrical ones. They use different methods to study closely related problems, and an interaction of these two communities would be beneficial. We make some comments on applications of the theory of interpolating Hamiltonians in convex geometry at the end of the article.

2. Monotone twist maps and minimal action

We recall basic facts about monotone twist maps; we refer to [14; 24; 27; 33; 39].

We consider a cylinder $S^1 \times (a, b)$ where $-\infty \leq a < b \leq +\infty$ and an area preserving diffeomorphism $f : S^1 \times (a, b) \to S^1 \times (a, b)$, isotopic to the identity. Let $F = (F_1, F_2) : \mathbb{R} \times (a, b) \to \mathbb{R} \times (a, b)$ be a lift to the universal cover. We denote by $(x, y) \in \mathbb{R} \times (a, b)$ the coordinates on the strip $\mathbb{R} \times (a, b)$. In these coordinates the area form is $dy \wedge dx$. The monotone twist condition is

$$\frac{\partial F_1(x, y)}{\partial y} > 0.$$

If a, resp. b, is finite, we assume that F extends to the boundary of the strip as a rigid shift, that is $F(x, a) = (x + \omega_-, a)$, resp. $F(x, b) = (x + \omega_+, b)$. Otherwise, we set $\omega_- := -\infty$, resp. $\omega_+ := +\infty$. We call f a monotone twist map. The interval $(\omega_-, \omega_+) \subset \mathbb{R}$ is called the twist interval of the map f. The twist condition and the twist interval do not depend on the choice of the lift F.

In the following sections $S^1 \times (a, b)$ will appear as (diffeomorphic copy of) the phase space of various billiard systems. The foliation of the phase space corresponding to $\{\text{pt}\} \times (a, b)$ will be referred to as vertical foliation. Moreover, for now, $S^1 = \mathbb{R}/\mathbb{Z}$ for simplicity, but in the billiard situations S^1 has varying length, e.g., for conventional billiards the length is the perimeter of the billiard table.

Monotone twist maps can be defined via generating functions. A function

$$H : \{(x, x') \in \mathbb{R} \times \mathbb{R} \mid \omega_- < x' - x < \omega_+\} \longrightarrow \mathbb{R}$$

is a generating function for f if the following holds:

$$F(x, y) = (x', y') \quad \text{if and only if} \quad \frac{\partial H(x, x')}{\partial x} = -y, \ \frac{\partial H(x, x')}{\partial x'} = y'. \tag{3}$$

The variables are related via the diffeomorphism $(x, y) \mapsto (x, x') = (x, F_1(x, y))$. The function H is periodic in the diagonal direction: $H(x+k, x'+k) = H(x, x')$ for $k \in \mathbb{Z}$. The twist condition becomes the inequality

$$\frac{\partial^2 H}{\partial x \partial x'}(x, x') < 0. \tag{4}$$

In the coordinates (x, x') condition (3) becomes

$$F(x, x') = (x', x'') \quad \text{if and only if} \quad \frac{\partial}{\partial x'}\big(H(x, x') + H(x', x'')\big) = 0. \tag{5}$$

As a consequence, the differential 2-form

$$\frac{\partial^2 H}{\partial x \partial x'}(x, x')\, dx \wedge dx'$$

is invariant under the map.

In terms of the coordinates (x, x'), a generating function $H(x, x')$ is not uniquely defined, for instance it can be changed to $H(x, x') + h(x') - h(x)$ by any function $h(x)$ such that $h(x+1) - h(x)$ is constant. This changes the coordinates y and y', but does not change the periodicity in the diagonal directions of the function $H(x, x')$, the twist condition (4), or the variational characterization (5). In terms of (x, y)-coordinates the map F is conjugated by the area-preserving diffeomorphism $(x, y) \mapsto (x, y + h'(x))$. This explains for instance the different conventions for generating functions in the literature, e.g., for conventional billiards.

Birkhoff periodic orbits for f of type $(p, q) \in \mathbb{Z} \times \mathbb{Z}_+$ are bi-infinite sequences $(x_n, y_n)_{n\in\mathbb{Z}}$ in $\mathbb{R} \times (a, b)$ such that for all $n \in \mathbb{Z}$ we have

1) $x_{n+1} > x_n$;
2) $(x_{n+q}, y_{n+q}) = (x_n + p, y_n)$;
3) $F(x_n, y_n) = (x_{n+1}, y_{n+1})$.

The Birkhoff theorem asserts that for every rational number $\frac{p}{q} \in (\omega_-, \omega_+)$ in lowest terms, the map f possesses at least two Birkhoff periodic orbits of type (p, q). One of these orbits (the "easier one") corresponds to the minimum of the function

$$H(x_0, x_1) + H(x_1, x_2) + \cdots + H(x_{q-1}, x_q) \tag{6}$$

on the space of bi-infinite sequences of real numbers $X = (x_n)_{n\in\mathbb{Z}}$ satisfying the monotonicity condition $x_{n+1} \geq x_n$ and the periodicity condition $x_{n+q} = x_n + p$. Setting $y_n := \frac{\partial H}{\partial x'}(x_{n-1}, x_n)$, it turns out that, due to the twist condition (4), the sequence (x_n, y_n) actually satisfies the above conditions 1) – 3), in particular the stronger monotonicity condition 1).

Let us denote the minimal value of the function (6) on this space of bi-infinite sequences by $T_{p,q}$. Then the minimal action of a (p, q)-periodic orbit is defined by

$$\beta\left(\frac{p}{q}\right) := \frac{1}{q}\, T_{p,q}. \tag{7}$$

This is the celebrated Mather β-function. The amazing fact is that Mather's β-function is well-defined, i.e., does not depend on the representation of the rational number $\frac{p}{q}$, for example, $\beta\left(\frac{2}{6}\right) = \frac{1}{6}T_{2,6} = \frac{1}{3}T_{1,3} = \beta\left(\frac{1}{3}\right)$.

We note that the above mentioned change of the generating function $\tilde{H}(x, x') = H(x, x') + h(x') - h(x)$ with $h(x + 1) - h(x) = c$ leads to the change of the β-function:

$$\tilde{\beta}\left(\frac{p}{q}\right) = \beta\left(\frac{p}{q}\right) + c\,\frac{p}{q}.$$

Mather's β-function is a strictly convex continuous function of the rotation number: if $p_1/q_1 \neq p_2/q_2$, then

$$t\beta\left(\frac{p_1}{q_1}\right) + (1-t)\beta\left(\frac{p_2}{q_2}\right) > \beta\left(t\,\frac{p_1}{q_1} + (1-t)\,\frac{p_2}{q_2}\right) \tag{8}$$

for all $t \in (0, 1)$; see [31; 39]. Although the minimal average action extends as a strictly convex function to irrational rotations numbers too, we only consider rational ones, therefore, in what follows, t is also a rational number. The following lemma deduces the general Dowker-style inequality from the convexity of the minimal action.

Lemma 2.1. *For all relatively prime $(p, q) \in \mathbb{Z} \times \mathbb{Z}_+$ with $\frac{p}{q} \in (\omega_-, \omega_+)$ and $q \neq 1$, we have*

$$T_{p,q-1} + T_{p,q+1} > 2T_{p,q}.$$

Proof. Consider the inequality (8) with the choices

$$p_1 = p_2 = p, \quad q_1 = q - 1, \quad q_2 = q + 1, \quad t = \frac{q-1}{2q},$$

i.e.,

$$\frac{q-1}{2q}\beta\left(\frac{p}{q-1}\right) + \left(1 - \frac{q-1}{2q}\right)\beta\left(\frac{p}{q+1}\right) > \beta\left(\frac{q-1}{2q}\,\frac{p}{q-1} + \left(1 - \frac{q-1}{2q}\right)\frac{p}{q+1}\right),$$

and simplify to

$$\begin{aligned}\frac{q-1}{2q}\beta\left(\frac{p}{q-1}\right) + \frac{q+1}{2q}\beta\left(\frac{p}{q+1}\right) &> \beta\left(\frac{q-1}{2q}\,\frac{p}{q-1} + \frac{q+1}{2q}\,\frac{p}{q+1}\right)\\ &= \beta\left(\frac{p}{2q} + \frac{p}{2q}\right) = \beta\left(\frac{p}{q}\right).\end{aligned}$$

Then (7) gives

$$\frac{q-1}{2q}\frac{1}{q-1}T_{p,q-1}+\frac{q+1}{2q}\frac{1}{q+1}T_{p,q+1}=\frac{q-1}{2q}\beta\Big(\frac{p}{q-1}\Big)+\frac{q+1}{2q}\beta\Big(\frac{p}{q+1}\Big)$$
$$>\beta\Big(\frac{p}{q}\Big)=\frac{1}{q}T_{p,q},$$

which simplifies to $\frac{1}{2q}T_{p,q-1}+\frac{1}{2q}T_{p,q+1}>\frac{1}{q}T_{p,q}$, i.e.,

$$T_{p,q-1}+T_{p,q+1}>2T_{p,q},$$

as claimed. □

Next we describe a small extension of the above discussion, well-known to experts. It is sometimes convenient to consider as phase space a set of the form

$$\{(\bar{x},\bar{y})\mid \bar{x}\in\mathbb{R},\ 0\le\bar{y}\le o(\bar{x})\}$$

for some function $o:\mathbb{R}\to(0,\infty)$ together with an area-preserving (with respect to $d\bar{y}\wedge d\bar{x}$) self-map $\bar{F}$ which is a monotone twist map, i.e., $\frac{\partial\bar{F}_1}{\partial\bar{y}}>0$. This set-up can be transformed to the above standard setting. First, we observe that the map

$$\begin{aligned}\mathbb{R}\times[0,1]&\to\big\{(\bar{x},\bar{y})\mid \bar{x}\in\mathbb{R},\ 0\le\bar{y}\le o(\bar{x})\big\},\\ (x,y)&\mapsto\big(O^{-1}(x),o(O^{-1}(x))y\big),\end{aligned}\tag{9}$$

is an area-preserving diffeomorphism if $O(\bar{x})$ is an antiderivative of $o(\bar{x})$. Since $O'(\bar{x})=o(\bar{x})>0$, the function O is strictly monotone and thus invertible.

For simplicity, set $\varphi(x):=O^{-1}(x)$. Then the above maps reads

$$\begin{aligned}\bar{x}&=\varphi(x),\\ \bar{y}&=o(\varphi(x))y=o(\bar{x})y.\end{aligned}$$

That is, $\mathbb{R}\times\{0\}$ is mapped to itself and $\mathbb{R}\times\{1\}$ to $\{\bar{y}=o(\bar{x})\}$. The map (9) is area-preserving since $d\bar{x}\wedge d\bar{y}=\varphi'(x)o(\varphi(x))\,dx\wedge dy$ and

$$\varphi'(x)o(\varphi(x))=\frac{d}{dx}O(\varphi(x))=1$$

since $\varphi(x)=O^{-1}(x)$.

We point out that the two vertical foliations (given by fixing x, resp. $\bar{x}$) are mapped to each other by the map (9). Moreover, if we have area-preserving maps $\bar{F}$ on $\big\{(\bar{x},\bar{y})\mid \bar{x}\in\mathbb{R},\ 0\le\bar{y}\le o(\bar{x})\big\}$ and F on $\mathbb{R}\times[0,1]$ which are conjugate to each other by (9), then one satisfies the twist condition if and only if the other does. This uses again $o(x)>0$. Finally, the twist condition and the variational description of $\bar{F}$ in terms of a generating function continues to hold.

3. The classic Dowker-type theorems revisited

In this section we consider four billiard-like systems: two inner, two outer, two with length, and two with area, as their generating functions. Lemma 2.1 will imply the four Dowker-style theorems mentioned in the introduction.

3.1. ***Conventional billiards.*** Let $\gamma : S^1 \to \mathbb{R}^2$ be a closed smooth strictly convex planar curve (an oval), oriented counterclockwise and parameterized by arc length. This is the boundary of the billiard table. We assume that it has unit length. Let x be the respective coordinate on $\mathbb{R}$, the universal cover of S^1.

The phase space of the billiard ball map F is the space of oriented chords of γ, the vertical foliation consists of the chords with a fixed initial point. The generating function is given by the formula[2]

$$H(x, x') = x' - x - |\gamma(x)\gamma(x')|,$$

where $|\gamma(x)\gamma(x')|$ denotes the chord length, i.e., the Euclidean length of the segment between $\gamma(x)$ and $\gamma(x')$.

One can calculate (see, e.g., [42]) that

$$\frac{\partial|\gamma(x)\gamma(x')|}{\partial x} = -\cos\alpha, \quad \frac{\partial|\gamma(x)\gamma(x')|}{\partial x'} = \cos\alpha', \quad \frac{\partial^2|\gamma(x)\gamma(x')|}{\partial x \partial x'} = \frac{\sin\alpha \sin\alpha'}{|\gamma(x)\gamma(x')|},$$

where $\alpha, \alpha' \in (0, \pi)$ are the angles made by the chord $\gamma(x)\gamma(x')$ with the curve γ. Therefore, as coordinates we obtain

$$y = 1 - \cos\alpha, \ \ y' = 1 - \cos\alpha' \in (0, 2),$$

and

$$\frac{\partial^2 H(x, x')}{\partial x \partial x'} = -\frac{\sin\alpha \sin\alpha'}{|\gamma(x)\gamma(x')|} < 0.$$

It follows that the quantity $T_{p,q}$ from Section 2 is p minus the greatest perimeter of the q-gons with the winding number p inscribed in γ. Denoting this perimeter by $R_{p,q}$, Lemma 2.1 implies that $R_{p,q-1} + R_{p,q+1} < 2R_{p,q}$ which, for $p = 1$, reduces to the statement of the Molnár–Eggleston theorem (2).

As explained above, even for conventional billiards the inequalities $R_{p,q-1} + R_{p,q+1} < 2R_{p,q}$ are new for $p > 1$. Figure 2 illustrates the inequality $R_{2,4} + R_{2,6} < 2R_{2,5}$. The extreme quadrilateral with $p = 2$ is the diameter of the curve, traversed four times, i.e., $R_{2,4} = 2R_{1,2}$, and the extreme hexagon with $p = 2$ is the extreme triangle, traversed twice: $R_{2,6} = 2R_{1,3}$. Hence the inequality can be rewritten as $R_{1,2} + R_{1,3} < R_{2,5}$, which has a different form from the Dowker-style inequalities.

[2]There are various conventions in the literature. Another common choice of generating function is $-|\gamma(x)\gamma(x')|$. As explained in the previous section, adding $x' - x$ does not change the twist condition, etc.

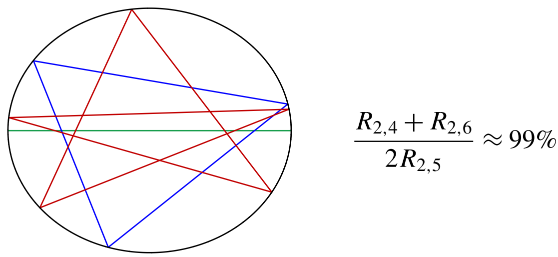

Figure 2. An ellipse with eccentricity of about 0.5.

3.2. ***Outer (area) billiards.*** The outer billiard map F about a smooth closed oriented strictly convex curve γ is depicted in Figure 3: one has $F(A) = A'$ if the orientation of the segment AA' coincides with the orientation of γ, and $|A\gamma(x')| = |\gamma(x')A'|$; see [43] for a survey. In this article we sometimes call this dynamical system outer area billiards to distinguish it from the upcoming outer length billiard in Section 3.4.

The map F is an area preserving map of the exterior of γ with respect to the standard area of the plane. It extends as the identity to γ. The phase space is foliated by the positive tangent rays to γ, and F is a twist map. Let $x \in \mathbb{R}$ be the (lifted to $\mathbb{R}$) angular coordinate on γ, that is, the direction of the oriented tangent line of γ. Moreover, if we write a point A in the exterior of γ as $\gamma(x) + r\gamma'(x)$, $r > 0$, then the map

$$\{\text{exterior of } \gamma\} \ni A \mapsto (x, r) \in S^1 \times (0, \infty)$$

is a symplectomorphism between the standard area form and $r\,dr \wedge dx = dy \wedge dx$ with $y = r^2/2$.

The generating function $H : \{(x, x') \mid 0 < x' - x < \pi\} \to \mathbb{R}$ of the map F is the area of the (oriented) curvilinear triangle obtained by first following the segments $\gamma(x)A$ and $A\gamma(x')$ and then the arc $\gamma(x')\gamma(x)$; see Figure 3. One has

$$\frac{\partial H(x,x')}{\partial x} = -\frac{r^2}{2}, \quad \frac{\partial H(x,x')}{\partial x'} = \frac{(r')^2}{2}, \quad \frac{\partial^2 H(x,x')}{\partial x \partial x'} = -r\frac{\partial r}{\partial x'} < 0.$$

It follows that the quantity $T_{p,q}$ is the minimal area of the circumscribed q-gon with the winding number p minus a constant (p times the area bounded by γ). Denoting

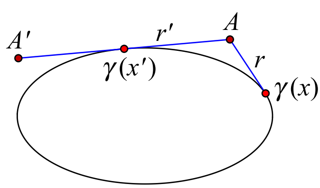

Figure 3. The outer billiard map.

this circumscribed area by $Q_{p,q}$, Lemma 2.1 implies that $Q_{p,q-1}+Q_{p,q+1} > 2Q_{p,q}$ which, for $p=1$, is a statement of the Dowker theorem (1).

3.3. ***Symplectic billiards.*** Symplectic billiards were introduced and studied in [1]; see also the recent papers [5; 6].

Let $\gamma(x)$ be a positively oriented parameterized smooth closed strictly convex planar curve. For a point $\gamma(x)$, let $\gamma(x^*)$ be the other point on γ where the tangent line is parallel to that in $\gamma(x)$. The phase space of symplectic billiard is then the set of the oriented chords $\gamma(x)\gamma(x')$ where $x < x' < x^*$ according to the orientation of γ. That is, the phase space is the set of pairs (x, x') such that $\omega(\gamma'(x), \gamma'(x')) > 0$. Here, ω is the standard area form in the plane, the determinant made by two vectors.

The vertical foliation consists of the chords with a fixed initial point. The symplectic billiard map F sends a chord $\gamma(x)\gamma(x')$ to $\gamma(x')\gamma(x'')$ if the tangent line $T_{\gamma(x')}\gamma$ is parallel to the line $\gamma(x)\gamma(x'')$; see Figure 4. Unlike the conventional billiards, this reflection law is not local. We note that if $\omega(\gamma'(x), \gamma'(x')) > 0$, then $\omega(\gamma'(x'), \gamma'(x'')) > 0$ as well (see [1]).

We extend the map F to the boundary of the phase space by continuity: $F(x, x) :=(x, x)$ and $F(x, x^*) := (x^*, x)$.

The generating function $H : \{(x, x') \mid x < x' < x^*\} \to \mathbb{R}$ is the area bounded by the oriented bigon formed by following first the arc $\gamma(x)\gamma(x')$ and then the segment $\gamma(x')\gamma(x)$. Note the similarity of this generating function with the one of the conventional billiard: the length is replaced by the area. One has

$$y = -\frac{\partial H(x, x')}{\partial x} = \frac{1}{2}\omega(\gamma'(x), \gamma(x') - \gamma(x))$$

and

$$\frac{\partial^2 H(x, x')}{\partial x \partial x'} = -\frac{1}{2}\omega(\gamma'(x), \gamma'(x')) < 0.$$

Thus, this is a situation where the phase space naturally is of the form

$$\big\{(\bar{x}, \bar{y}) \mid \bar{x} \in \mathbb{R},\ 0 \le \bar{y} \le o(\bar{x})\big\},$$

where $o(x) = \frac{1}{2}\omega(\gamma'(x), \gamma(x^*) - \gamma(x))$.

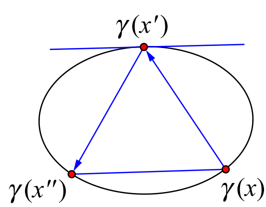

Figure 4. The symplectic billiard map.

Let $\gamma(x), \gamma(x'), \gamma(x'')$ be three consecutive vertices of an inscribed polygon of the maximal area. Then $T_{\gamma(x')}\gamma$ is parallel to the line $\gamma(x)\gamma(x'')$, hence $F(x,x') = (x',x'')$. Then either

$$\omega(\gamma'(x), \gamma'(x')) > 0, \quad \text{or} \quad \omega(\gamma'(x), \gamma'(x')) < 0, \quad \text{or} \quad \omega(\gamma'(x), \gamma'(x')) = 0.$$

One has the same inequality for $\omega(\gamma'(x'), \gamma'(x''))$. If this is negative, the total area is negative as well, and changing the orientation makes it greater. Likewise if $\omega(\gamma'(x), \gamma'(x')) = 0$, then the area is not maximal either. Hence the maximal area polygon corresponds to the Birkhoff minimal (p,q)-periodic orbit of the symplectic billiard map.

It follows that, up to an additive constant (p times the area bounded by γ), the quantity $T_{p,q}$ is minus the greatest area of the q-gon with the winding number p, inscribed in γ. Denoting this area by $P_{p,q}$, Lemma 2.1 implies that $P_{p,q-1} + P_{p,q+1} < 2P_{p,q}$ which, for $p = 1$, is a statement of the Dowker theorem (1).

3.4. ***Outer length billiards.*** As far as we know, the outer length billiard system, defined by extremizing the perimeter of a circumscribed polygons, has not been described in the literature yet. We provide necessary details here, and will return to a more detailed study of this dynamical system in [3]. See [15] for a study of the polygons circumscribed about a convex curve and having the minimal perimeter.

The map F acts on the exterior of an arc length parameterized oval $\gamma(x)$ and is given by the following geometrical construction; see Figure 5. Let A be a point outside of γ, and let $A\gamma(x')$ and $A\gamma(x)$ be the positive and negative tangent segments to γ (the sign given by the orientation of the oval). Consider the circle tangent to γ at the point $\gamma(x')$, tangent to the line $A\gamma(x)$, and lying on the opposite side of γ with respect to the line $A\gamma(x')$. Then $A' = F(A)$ is defined as the intersection point of the line $A\gamma(x')$ and the other common tangent line of the circle and γ (tangent to γ in $\gamma(x'')$ in Figure 5).

The map F extends as identity to γ. The vertical foliation in phase space, i.e., the exterior of γ, is the same as for the outer (area) billiard: it consists of the positive tangent rays to γ. Similarly to outer billiards, F is a twist map. We recall from symplectic billiards that the points $\gamma(x)$ and $\gamma(x^*)$ have parallel tangent lines.

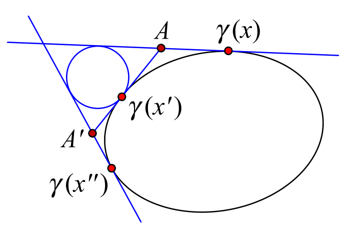

Figure 5. Outer length billiard.

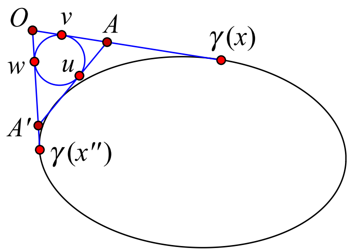

Figure 6. Construction of the extremizer.

Lemma 3.1. *The generating function* $H : \{(x, x') \mid x < x' < x^*\} \to \mathbb{R}$ *of the map* F *is given by the formula*

$$H(x, x') = |\gamma(x)A| + |A\gamma(x')| - x' + x.$$

Proof. Fix points $\gamma(x)$ and $\gamma(x'')$ and consider

$$\begin{aligned} H(x, x') + H(x', x'') &= |\gamma(x)A| + |A\gamma(x')| + |\gamma(x')A'| + |A'\gamma(x'')| - x'' + x' - x' + x \\ &= |\gamma(x)A| + |AA'| + |A\gamma(x'')| - x'' + x. \end{aligned}$$

We claim that the specific point $\gamma(x')$ described above (Figure 5) extremizes the length $|\gamma(x)A| + |AA'| + |A'\gamma(x'')|$.

To prove this claim, consider Figure 6. We use the fact that the two tangent segments to a circle through a common point have equal lengths, e.g., $|Au| = |Av|$ in Figure 6.

Therefore, we can rewrite as follows

$$|\gamma(x)A| + |AA'| + |A'\gamma(x'')| = |\gamma(x)v| + |\gamma(x'')w| = |\gamma(x)O| + |\gamma(x'')O| - 2|Ov|.$$

The left-hand side is minimal when $|Ov|$ is maximal, and this happens when the circle is greatest possible, i.e., if u lies on γ, that is, $u = \gamma(x')$ in Figure 5.

A similar argument applies if the point O is on the other side of the line AA', and if the tangent lines at points $\gamma(x)$ and $\gamma(x'')$ are parallel. □

The partial derivatives of the generating function are

$$y = -\frac{\partial H(x, x')}{\partial x} = k(x)|A\gamma(x)| \cot\frac{\varphi}{2} \in (0, \infty),$$

where k is the curvature function of γ, and φ is the angle between the tangent segments $A\gamma(x)$ and $A\gamma(x')$, and

$$\frac{\partial^2 H(x, x')}{\partial x \partial x'} = -\frac{k(x)k(x')(|A\gamma(x)| + |A\gamma(x')|)}{2\sin^2(\varphi/2)}.$$

It follows that, up to an additive constant (p times the perimeter of γ), the quantity $T_{p,q}$ is the minimal perimeter of the q-gons with the winding number p, circumscribed about γ. Denoting this perimeter by $S_{p,q}$, Lemma 2.1 implies that $S_{p,q-1} + S_{p,q+1} > 2S_{p,q}$ which, for $p = 1$, is a statement of the Molnár–Eggleston theorem (2).

Remark 3.2. The area form that is invariant under this billiard map is, in terms of the generating function,

$$-\frac{\partial^2 H(x,x')}{\partial x \partial x'}\, dx \wedge dx'.$$

This is a functional multiple of the standard area form ω in the exterior of the oval and, at the point A (in Figure 5), its value is, see [3],

$$\cot\frac{\varphi}{2}\left(\frac{1}{|A\gamma(x)|} + \frac{1}{|A\gamma(x')|}\right)\omega.$$

Remark 3.3. The quantity $H = |\gamma(x)A| + |A\gamma(x')| - x' + x$ is known in the study of (the conventional) billiards as the Lazutkin parameter. Given an oval γ, consider the locus of points A for which H has a constant value. This locus is a curve Γ, and the billiard inside Γ has the curve γ as a caustic: a billiard trajectory tangent to γ remains tangent to it after the reflection in Γ. This is known as the string construction of a billiard curve by its caustic (see, e.g., [42]).

A similar relation exists between the level curves of the generating function of the symplectic billiard and the invariant curves of the outer billiard. Consider the set of chords that cut off a fixed area from an oval γ, that is, a level curve of the generating function of the symplectic billiard in γ. The envelope of these chords is a curve Γ, and γ is an invariant curve of the outer billiard about Γ. This is the area construction of an outer billiard curve by its invariant curve.

The meaning of this relation between the level curves of a generating function of one billiard system and invariant curves of another one is not clear to us.

4. More examples

4.1. ***Inner and outer billiards in S^2 and H^2.*** Inner (conventional) billiards are defined in the same way in the spherical and hyperbolic geometries as in the Euclidean plane: the boundary of the billiard table γ has positive geodesic curvature and, in the case of S^2, this implies that γ is contained in an open hemisphere. The billiard ball travels along geodesics and reflects in γ subject to the law of equal angles.

Outer billiards are defined in H^2 similarly to the Euclidean case, but the case of S^2 is somewhat different.

Let γ be a closed smooth oriented geodesically convex spherical curve, and let $-\gamma$ be its antipodal curve. The curve γ lies in a hemisphere, and $-\gamma$ lies in the antipodal hemisphere. The spherical belt (topologically, a cylinder) bounded by γ and $-\gamma$ is the phase space of the outer billiard about γ; it is foliated by the arc of the positive tangent great circles to γ: these segments have the initial points on γ, and the terminal points on $-\gamma$. This is the vertical foliation that appears in the definition of twist maps.

Likewise, the phase space is foliated by the arc of the negative tangent great circles to γ. This makes it possible to define the outer billiard similarly to the planar case: given a point A, there is a unique point $x \in \gamma$ such that the arc of the negative tangent great circle at x contains A. The image point A' lies on the arc of the positive tangent great circle at x at the same spherical distance from x as point A.

The outer billiard map about $-\gamma$ is conjugated to that about γ by the antipodal involution of the sphere.

Inner and outer billiards in S^2 are conjugated by spherical duality; see Figure 7. Spherical duality interchanges oriented great circles with their poles, and the angle between two circles is equal to the spherical distance between their poles. The duality extends to convex smooth curves: the poles of the 1-parameter family of the tangent great circles of a curve γ comprise the dual curve γ^*. Equivalently, γ^* is the $\pi/2$-equidistant curve of γ, that is, γ^* is the locus of the endpoints of the arcs of the great circles, orthogonal to γ and having length $\pi/2$.

Let γ be a convex smooth curve and γ^* be its dual. Let L, A, L^*, A^* be the perimeters of these curves and the areas of the convex domains bounded by them. Then $L^* = 2\pi - A$, or equivalently $A^* = 2\pi - L$; see [4].

These relations are easy to see if P is a spherical convex n-gon and P^* is its dual. Then the angles and the sides lengths of these polygons are related by

$$\alpha_i^* = \pi - \ell_i, \quad \ell_i^* = \pi - \alpha_i.$$

The Gauss–Bonnet theorem implies

$$A^* = \sum \alpha_i^* - \pi(n-2) = 2\pi - \sum \ell_i = 2\pi - L,$$

as claimed. The other equality follows by interchanging P and P^*. For details see [23, Chapter 20], for example.

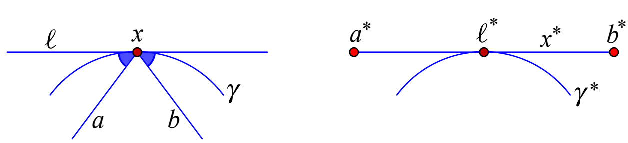

Figure 7. Spherical duality conjugates inner and outer billiards.

If γ is a closed smooth convex spherical curve and P is a circumscribed spherical polygon, then the dual polygon P^* is inscribed in the dual curve γ^*. It follows that P has the minimal area or perimeter if and only if P^* has the maximal perimeter or area, respectively.

See Section 2.1.3 of [1] for a discussion of symplectic billiards in the spherical and hyperbolic geometries. The convexity of Mather's β-function implies the spherical version of Dowker's inequalities due to L. Fejes Tóth [20].

4.2. ***Wire billiards.*** Wire billiards were introduced and studied in [10], see also [12; 13; 21].

Let $\gamma(x)$ be a smooth closed arc length parameterized curve in $\mathbb{R}^n$ (a wire). One defines the wire billiard relation in the same way as for the conventional billiards in the plane: chords $\gamma(x)\gamma(y)$ and $\gamma(y)\gamma(z)$ are in this relation if

$$\frac{\partial}{\partial y}(|\gamma(x)\gamma(y)|+|\gamma(y)\gamma(z)|)=0.$$

Thus, the phase space for wire billiards is given by oriented chords of γ and the vertical foliation is given by the chords with a fixed initial point.

There is a class of curves, including small C^2 perturbations of planar ovals, for which the wire billiard relation is a map which then is an area-preserving twist map, called the wire billiard map. This class of curves is given by the following three conditions.

(1) Any line in $\mathbb{R}^n$ intersects γ in at most two points, and if it intersects at two points, the intersections are with nonzero angles.

(2) The curvature of γ does not vanish.

(3) Let π_{xy} be the 2-plane spanned by the tangent vector $\gamma'(x)$ and the chord $\gamma(y)-\gamma(x)$. Then for every x, y the planes π_{xy} and π_{yx} are not orthogonal.

If these conditions are satisfied, the generating function of the wire billiard map is given by the same expression as for the conventional planar billiards, $H(x,x')=x'-x-|\gamma(x)\gamma(x')|$, and

$$y=-\frac{\partial H}{\partial x}=1-\cos\alpha\in(0,2),$$

where α is the angle between $\gamma'(x)$ and $\gamma(x')-\gamma(x)$. Moreover,

$$\frac{\partial^2 H(x,x')}{\partial x\partial x'}=-\frac{\cos\varphi\sin\alpha\sin\alpha'}{|\gamma(x)\gamma(x')|}<0,$$

where φ is the angle between the planes π_{xy} and π_{yx}.

Let $\gamma\subset\mathbb{R}^n$ be a curve satisfying the above conditions. Let $R_{p,q}$ be the greatest perimeter of a q-gon with the winding number p inscribed in γ, that is, whose vertices lie on γ. Then Lemma 2.1 implies that $R_{p,q-1} + R_{p,q+1} < 2R_{p,q}$, a generalization of the Molnár–Eggleston theorem (2) for nonplanar curves.

4.3. ***Minkowski plane.*** The Molnár–Eggleston theorem holds in Minkowski planes (2-dimensional Banach spaces), see Theorem 10 in [29]. It is geometrically clear that inner and outer length billiards are still twist maps if the Euclidean metric is replaced by a general norm. Therefore, the Molnár–Eggleston theorem for Minkowski planes should be deducible, as in the Euclidean case, from Lemma 2.1. We decided not to resolve the details here.

As for the outer area billiard, we recall that a point and its reflection under the outer area billiard map lie on the same tangent line to the table and having the same distance to the tangent point, see Figure 3 in Section 3.2. That is, the ratio of the distances is 1. Since the ratio of two distances measured in any norm in $\mathbb{R}^2$ is the same, we see that the reflection rule of outer area billiards is independent of the choice of a norm. This is, of course, no surprise since the generating function is the standard area and does not involve a choice of a norm.

Finally, the symplectic billiard in the plane simply does not involve a metric in its definition or the reflection rule.

4.4. ***Wire symplectic billiards.*** Let $\gamma(x)$ be a smooth parameterized closed curve in the linear symplectic space $(\mathbb{R}^{2n}, \omega)$. We define the symplectic billiard relation on the chords of γ that generalizes symplectic billiards in the plane.

Two chords $\gamma(x)\gamma(x')$ and $\gamma(x')\gamma(x'')$ of γ are said to be in symplectic billiard relation if $\gamma(x'')-\gamma(x) \in T^{\omega}_{x'}\gamma$. Here, $T^{\omega}_{x'}\gamma := (T_{x'}\gamma)^{\omega}$ is the symplectic orthogonal complement of the tangent line $T_{\gamma(x')}\gamma$. Therefore, as for wire billiards, the phase space for symplectic wire billiards is given by oriented chords of γ and the vertical foliation is given by the chords with fixed initial point.

If x and x'' are fixed, then $\gamma(x)\gamma(x')$ and $\gamma(x')\gamma(x'')$ are in symplectic billiard relation if and only if

$$\frac{\partial[\omega(\gamma(x), \gamma(x')) + \omega(\gamma(x'), \gamma(x''))]}{\partial x'} = \omega(\gamma(x) - \gamma(x''), \gamma'(x')) = 0,$$

since $\gamma'(x') - \gamma(x) \in T^{\omega}_{x'}\gamma$ is equivalent to $\omega(\gamma(x) - \gamma(x''), \gamma'(x')) = 0$.

As for wire billiards, this relation does not necessarily define a map. Furthermore, for symplectic wire billiard there is the additional complication that, even if it defines a map, this map need not be a twist map.

We describe a class of curves γ for which this relation is indeed an area preserving twist map. We recall from [2] that a curve γ is called symplectically convex if $\omega(\gamma'(x), \gamma''(x)) > 0$ for all x. Consider such a curve, and fix a value x_0 of the parameter. Then the function $F_{x_0}(x) := \omega(\gamma'(x_0), \gamma(x))$ has a critical point at x_0.

Moreover, this zero critical value is a local minimum since

$$F''_{x_0}(x_0) = \omega(\gamma'(x_0), \gamma''(x_0)) > 0.$$

The class of curves γ that we consider is given by the following properties:

(1) γ is symplectically convex.

(2) For every x_0, the function $F_{x_0} : S^1 \to \mathbb{R}$ is a perfect Morse function, that is, it has exactly two nondegenerate critical points, a maximum and a minimum.

We call any such curve *admissible*. This class of curves is open in the C^2-topology, and a sufficiently small perturbation of an oval that lies in a symplectic plane $\mathbb{R}^2 \subset \mathbb{R}^{2n}$ is an admissible curve.

For an admissible curve, we denote by x^* the maximum point of the function F_x, i.e., $\omega(\gamma'(x), \gamma'(x^*)) = 0$. We observe that, for an admissible curve γ and a chord $\gamma(x)\gamma(x')$, there exists a unique chord $\gamma(x')\gamma(x'')$ such that $\gamma(x'')-\gamma(x) \in T^{\omega}_{\gamma(x')}\gamma$. Indeed, the noncritical level sets of the function $F_{x'} : S^1 \to \mathbb{R}$ consist of two points. This makes it possible to extend what we said above about symplectic billiards in the plane to this setting.

Namely, the phase space of wire symplectic billiard is the set of oriented chords $\gamma(x)\gamma(x')$ of γ satisfying $x < x' < x^*$ or, equivalently, $\omega(\gamma(x), \gamma(x')) > 0$. The vertical foliation consists of the chords with a fixed initial point. As before, we extend the map to the boundary of the phase space by continuity as follows:

$$(xx) \mapsto (xx), \ \ (xx^*) \mapsto (x^*x).$$

The next lemma justifies this statement and repeats a result from [1].

Lemma 4.1. *Let the wire symplectic billiard map take xx' to $x'x''$. If $x < x' < x^*$, then $x' < x'' < (x')^*$.*

Proof. If x' is close to x, then $x' < x'' < (x')^*$. If $(x')^* < x''$, then, by continuity, we move point x' toward x until $x'' = (x')^*$. Then

$$(xx') \mapsto (x'(x')^*) \mapsto ((x')^*x').$$

But the map is reversible: if $(xy) \mapsto (yz)$, then $(zy) \mapsto (yx)$. Hence $x = x'$, which is a contradiction. □

Similarly to the plane case, as the generating function we take

$$H(x, x') = \int_C \lambda,$$

where the integral is over the closed curve C made of the arc $\gamma(x)\gamma(x')$ of γ and the chord $\gamma(x')\gamma(x)$. The curve C is oriented according to the orientation of γ

and λ is a differential 1-form such that $\omega = d\lambda$. The result does not depend on the choice of such λ and is equal to the symplectic area of a surface filling the curve.

As before, we have

$$y = -\frac{\partial H(x,x')}{\partial x} = \frac{1}{2}\omega(\gamma'(x), \gamma(x') - \gamma(x))$$

and

$$\frac{\partial^2 H(x,x')}{\partial x \partial x'} = -\omega(\gamma'(x), \gamma'(x')).$$

Thus, the twist condition is precisely the condition that γ is symplectically convex.

Now consider q-gons whose vertices lie on γ and that have the winding number p. The symplectic area of a q-gon $(p_1, p_2, \ldots, p_q)$ is

$$\frac{1}{2}\sum_{i=1}^{q} \omega(p_i, p_{i+1}),$$

where the sum is read cyclically, i.e., $p_{q+1} = p_1$. Let $P_{p,q}$ be the greatest symplectic area of such polygons. Then Lemma 2.1 implies a nonplanar generalization of the Dowker theorem: $P_{p,q-1} + P_{p,q+1} < 2P_{p,q}$.

4.5. ***Magnetic billiards.*** Magnetic billiards were introduced in [38]; our main reference is [7].

Let γ be a plane oval. We assume that the magnetic field has constant strength and is perpendicular to the plane. Then the free path of a charge having a fixed energy is an oriented arc of a circle of radius R (the Larmor radius). When the charge hits the boundary γ, it undergoes the billiard reflection, so that the angle of incidence equals the angle of reflection. Unlike the conventional billiards, magnetic billiards is not time-reversible system. It is invariant under simultaneous time and magnetic field reversal, however. Therefore we assume, without loss of generality, that the charge moves in the counterclockwise direction.

Let $k > 0$ be the curvature function of γ, and $k_{\min}$ be its minimal value. The magnetic field is called weak if $1/R < k_{\min}$. We consider only the weak field regime here.

One uses the same coordinates as for conventional billiards: the arc length parameter x on γ and $u = 1 - \cos\alpha$, where α is the angle made by γ and the trajectory at the starting point. That is, the phase space is the space of oriented chords of γ, the vertical foliation consists of the chords with a fixed initial point.

Let φ be the angle made by the arc of the trajectory connecting the points $\gamma(x)$ and $\gamma(x')$ and the chord connecting these two points, see Figure 8. Then, according to [7],

$$\frac{\partial x'}{\partial u} = \frac{|\gamma(x)\gamma(x')|\cos\varphi}{\sin\alpha \sin\alpha'}.$$

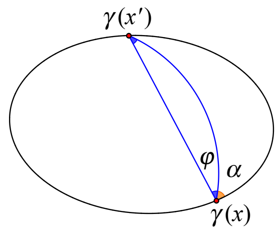

Figure 8. Magnetic billiard. Here α is the (orange) angle made by γ and the arc of the circle connecting $\gamma(x)$ and $\gamma(x')$. The (blue) angle φ is the angle between the arc of the circle and the chord connecting $\gamma(x)$ and $\gamma(x')$.

Therefore magnetic billiards satisfies the twist condition if the angle φ is always acute.

If this assumption holds, then the generating function of the magnetic billiard map is

$$H(x,x') = x' - x - \left(\ell + \frac{1}{R}A\right), \tag{10}$$

where ℓ the length of the arc $\gamma(x)\gamma(x')$ of radius R, and A is the area bounded by this arc and the curve γ and lying on the right (with respect to the orientation) of this arc (see again [7]). When there is no magnetic field, that is, when $R = \infty$, we obtain the generating function of conventional billiard used above. For completeness we also recall, from the appendix in [7], that

$$y = -\frac{\partial H}{\partial x} = 1 - \cos\alpha \in (0,2),$$

which is the same as in the conventional billiard case.

We now show that if the magnetic field is weak then the magnetic billiard map is indeed a twist map. The next statement is contained in [7] as Lemma D1 and the following Corollary. We provide a slightly different proof here for convenience.

Lemma 4.2. *If the magnetic field is weak, magnetic billiard map is a twist map. More precisely, if* $1/R < k_{\min}$, *then* $\varphi < \pi/2$.

Proof. Let $\psi = \pi - \varphi$ be the complementary angle; see Figure 9. We shall prove the equivalent statement: if ψ is acute, then there exists a point on the arc δ of the curve γ from $\gamma(x)$ to $\gamma(x')$ where the curvature $k \le 1/R$.

We use two geometric facts. The first is a lemma due to Schur (see [26], for instance), asserting the following. Let $\gamma_1(s)$ and $\gamma_2(s)$ be two smooth convex arc length parameterized curves of the same length, such that $k_1(s) < k_2(s)$ for all s. Then the chord subtended by γ_1 is greater than that the chord subtended by γ_2 for the same interval of the parameter.

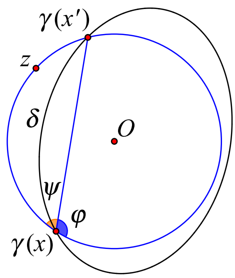

Figure 9. Concerning Lemma 4.2.

The second is the following lemma. If a closed convex curve γ_2 lies inside (the domain bounded by) the closed convex curve γ_1, then the length of γ_1 is greater than or equal to the length of γ_2. This is an easy consequence of the Crofton formula, see, e.g., [23], Lecture 19, and it does not require the curves to be smooth, nor for γ_2 to be contained strictly inside of γ_1.

Now, consider the clockwise oriented arcs of the curve γ and of the circle starting at the point $\gamma(x)$. Consider the first intersection point of these arcs. In Figure 9, this point is $\gamma(x')$ but, in general, it may lie closer to point $\gamma(x)$ on both curves. In order not to complicate the notation, we assume that indeed the first intersection point is $\gamma(x')$.

Then we have two nested convex closed curves: the first is made by the arc of the circle and the chord $\gamma(x')\gamma(x)$, and the second by the arc δ and the same chord. By assumption the second curve lies inside the first. It follows that the arc of the circle has length greater than or equal to that of δ.

Now we argue by contradiction and assume $1/R < k$. Let z be the point on the arc of the circle such that $\gamma(x)z$ has the same length as δ. Since $1/R < k$, Schur's Lemma implies that $|\gamma(x)z| > |\gamma(x')\gamma(x)|$. But the assumption that $\psi < \pi/2$ implies that the center of the circle lies on the right of the chord $\gamma(x')\gamma(x)$; therefore $|\gamma(x')z| < |\gamma(x')\gamma(x)|$ (cf. Figure 9). This is a contradiction. □

Remark 4.3. The same argument proves that if $1/R < k_{\min}$, then a circle of radius R intersects the curve γ in at most two points.

As before, we apply Lemma 2.1 to obtain the following result. Consider curvilinear q-gons with winding number p inscribed in the oval γ whose oriented sides are counterclockwise arcs of radius R with $1/R < k_{\min}$. Let $M_{p,q}$ be the greatest value of

$$P - \frac{1}{R}A,$$

where P is the perimeter of the curvilinear q-gon and A in the area *enclosed* by the curvilinear polygon. Then one has the Dowker-style inequality $M_{p,q-1} + M_{p,q+1} < 2M_{p,q}$.

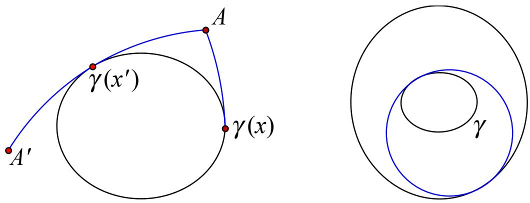

Figure 10. Left: outer magnetic billiard map. Right: its annulus of definition.

Note that the area term in the generating function (10) is the area between the curve γ and the respective arc of the polygon. Since the polygon is closed the area between the polygon and the curve and the area enclosed by the polygon differ by the total area enclosed by γ. This does not change the inequalities.

4.6. ***Outer magnetic billiards.*** The outer magnetic billiard map is defined similarly to the outer billiard map, see Section 3.2, but the tangent lines are replaced by the tangent arcs of circles of a fixed (Larmor) radius, greater than the greatest radius of curvature of the "billiard" curve γ; see Figure 10, left. That is, we are again in the weak magnetic field regime. We assume that γ and the circles are positively oriented and that the orientations agree at the tangency points.

Unlike the outer billiard map, this map is only defined in an annulus whose inner boundary is γ and the outer boundary is the envelope of the Larmor circles tangent to γ which, by Huygens' principle, is equidistant from γ; see Figure 10, right.

This map is area preserving, and the value of the generating function at the point A is the area of the curvilinear triangle bounded by the two Larmor arcs through A and the curve γ. The vertical foliation is given by the forward Larmor half-circles tangent to γ.

Outer magnetic billiard was introduced in [9], where it is shown that this system is isomorphic to magnetic billiard. The correspondence between the latter and the former is given by the map that assigns to the arcs of Larmor circles (inside the billiard table) their centers. The resulting outer magnetic billiard curve is equidistant to the magnetic billiard curve.

A Dowker-style geometric inequality results. Let γ be an oval, and let $N_{p,q}$ denote the minimal area of a circumscribed curvilinear q-gon with winding number p, whose sides are arcs of a fixed radius greater than the greatest curvature radius of γ. Then $N_{p,q-1} + N_{p,q+1} > 2N_{p,q}$. We note that, unlike for inner magnetic billiards, in this case the geometric inequality involves areas only and not a combination of areas and lengths.

4.7. ***Remarks on the theory of interpolating Hamiltonians.*** We conclude with remarks and questions concerning an application of the theory of interpolating

Hamiltonians [30; 34; 37] in convex geometry, namely, to approximation of smooth convex curves by polygons.

It is proved in [30] that the maximal perimeter of the inscribed q-gons $R_{1,q}$ has an asymptotic expansion as $q \to \infty$:

$$R_{1,q} \sim \text{Perimeter}(\gamma) + \sum_{k=1}^{\infty} \frac{c_k}{q^{2k}},$$

with the first nontrivial coefficient

$$c_1 = -\frac{1}{24} \left(\int_\gamma k^{2/3} ds \right)^3,$$

where $k(s)$ is the curvature of γ and s is the arc length parameter. If we read Dowker's inequality as

$$R_{1,q+1} - R_{1,q} \le R_{1,q} - R_{1,q-1},$$

we see that the successive approximation by polygons always improves.

Also the limiting $q \to \infty$ distribution of the vertices of the approximating inscribed q-gons is uniform with respect to the density $k^{2/3} ds$. Likewise for other winding numbers.

Similarly, one has for the minimal area of the circumscribed q-gons

$$Q_{1,q} \sim \text{Area}(\gamma) + \sum_{k=1}^{\infty} \frac{c_k}{q^{2k}},$$

with

$$c_1 = \frac{1}{24} \left(\int_\gamma k^{1/3} ds \right)^3;$$

see [41]. For the maximal area of the inscribed q-gons, one has

$$P_{1,q} \sim \text{Area}(\gamma) + \sum_{k=1}^{\infty} \frac{c_k}{q^{2k}},$$

with

$$c_1 = -\frac{1}{12} \left(\int_\gamma k^{1/3} ds \right)^3;$$

see [1] and [32].

In the last two cases, the limiting $q \to \infty$ distribution of the vertices of the approximating q-gons on the curve γ is uniform with respect to the density $k^{1/3} ds$, that is, uniform with respect to the affine length parameter.

It would be interesting to find explicit expressions for the coefficients c_k in these formulas (see [28; 30] for c_1 and c_2). A closely related problem of describing the coefficients of the expansion of Mather's β-function at zero for Birkhoff billiards is addressed in [40] and, for ellipses, in [8], and for symplectic and outer billiards, very recently, in [6].

There are other ways to measure the quality of approximation of a convex curve by polygons, for example, one can use the area of the symmetric difference of an oval and an approximating polygon, or its analog, replacing area by perimeter. A wealth of results in this directions is available; see [11; 22; 25]. Are these results related to the β-function and the interpolating Hamiltonians of some billiard-like dynamical systems?

Acknowledgements

Albers acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through Germany's Excellence Strategy EXC-2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster), the Transregional Collaborative Research Center CRC/TRR 191 (281071066). Tabachnikov was supported by NSF grant DMS-2005444 and by a Mercator fellowship within the CRC/TRR 191, and he thanks Heidelberg University for its invariable hospitality.

We thank the referee for useful suggestions and comments.

References

[1] P. Albers and S. Tabachnikov, "Introducing symplectic billiards", *Adv. Math.* **333** (2018), 822–867. MR Zbl

[2] P. Albers and S. Tabachnikov, "Symplectically convex and symplectically star-shaped curves: a variational problem", *J. Fixed Point Theory Appl.* **24**:2 (2022), art. id. 27, 24 pp. MR Zbl

[3] P. Albers, L. Costa, and S. Tabachnikov, "Outer length billiards", In preparation. Zbl

[4] S. S. Anisov, "The "area-length" duality and the characteristic 2-chain", *Mat. Zametki* **58**:3 (1995), 445–446. In Russian; translated in *Math. Notes* **58** (1995), 983–984. MR Zbl

[5] L. Baracco and O. Bernardi, "Totally integrable symplectic billiards are ellipses", 2023. Zbl arXiv 2305.19701

[6] L. Baracco, O. Bernardi, and A. Nardi, "Higher order terms of Mather's β-function for symplectic and outer billiards", *J. Math. Anal. Appl.* **537**:2 (2024), art. id. 128353. MR Zbl

[7] N. Berglund and H. Kunz, "Integrability and ergodicity of classical billiards in a magnetic field", *J. Statist. Phys.* **83**:1-2 (1996), 81–126. MR Zbl

[8] M. Bialy, "Mather β-function for ellipses and rigidity", *Entropy* **24**:11 (2022), art. id. 1600, 16 pp. MR Zbl

[9] M. Bialy and A. E. Mironov, "Algebraic non-integrability of magnetic billiards", *J. Phys. A* **49**:45 (2016), art. id. 455101, 18 pp. MR Zbl

[10] M. Bialy, A. E. Mironov, and S. Tabachnikov, "Wire billiards, the first steps", *Adv. Math.* **368** (2020), art. id. 107154, 27 pp. MR Zbl

[11] J. Bochi, "On the approximation of convex bodies by ellipses with respect to the symmetric difference metric", *Discrete Comput. Geom.* **60**:4 (2018), 938–966. MR Zbl

[12] S. V. Bolotin, "Degenerate billiards", pp. 53–71 in *Sovremennye Problemy Mekhaniki*, Tr. Mat. Inst. Steklova **295**, 2016. In Russian; translation in *Proc. Steklov Inst. Math.* **295**:1 (2016), 45–62. MR Zbl

[13] S. V. Bolotin, "Degenerate billiards in celestial mechanics", *Regul. Chaotic Dyn.* **22**:1 (2017), 27–53. MR Zbl

[14] C. Carminati, S. Marmi, D. Sauzin, and A. Sorrentino, "On the regularity of Mather's β-function for standard-like twist maps", *Adv. Math.* **377** (2021), art. id. 107460, 22 pp. MR Zbl

[15] D. DeTemple, "The geometry of circumscribing polygons of minimal perimeter", *J. Geom.* **49**:1-2 (1994), 72–89. MR Zbl

[16] C. H. Dowker, "On minimum circumscribed polygons", *Bull. Amer. Math. Soc.* **50** (1944), 120–122. MR Zbl

[17] C. H. Dowker, "Homology groups of relations", *Ann. of Math.* (2) **56** (1952), 84–95. MR Zbl

[18] H. G. Eggleston, "Approximation to plane convex curves, I: Dowker-type theorems", *Proc. London Math. Soc.* (3) **7** (1957), 351–377. MR Zbl

[19] L. Fejes Tóth, *Lagerungen in der Ebene*, *auf der Kugel und im Raum*, Grundlehren Math. Wiss. **65**, Springer, 1953. MR Zbl

[20] L. Fejes Tóth, "Annäherung von Eibereichen durch Polygone", *Math.-Phys. Semesterber.* **6** (1958/59), 253–261. MR Zbl

[21] J. Féjoz, A. Knauf, and R. Montgomery, "Lagrangian relations and linear point billiards", *Nonlinearity* **30**:4 (2017), 1326–1355. MR Zbl

[22] F. Fodor, "Perimeter approximation of convex discs in the hyperbolic plane and on the sphere", *Discrete Comput. Geom.* **66**:3 (2021), 1190–1201. MR Zbl

[23] D. Fuchs and S. Tabachnikov, *Mathematical omnibus: thirty lectures on classic mathematics*, American Mathematical Society, Providence, RI, 2007. MR Zbl

[24] C. Golé, *Symplectic twist maps: global variational techniques*, Advanced Series in Nonlinear Dynamics **18**, World Scientific, River Edge, NJ, 2001. MR Zbl

[25] P. M. Gruber, "Aspects of approximation of convex bodies", pp. 319–345 in *Handbook of convex geometry*, vol. A, North-Holland, Amsterdam, 1993. MR Zbl

[26] H. W. Guggenheimer, *Differential geometry*, Dover, New York, 1977. MR Zbl

[27] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications **54**, Cambridge University Press, 1995. MR Zbl

[28] M. Ludwig, "Asymptotic approximation of convex curves", *Arch. Math.* (*Basel*) **63**:4 (1994), 377–384. MR Zbl

[29] H. Martini, K. J. Swanepoel, and G. Weiß, "The geometry of Minkowski spaces—a survey, I", *Expo. Math.* **19**:2 (2001), 97–142. MR Zbl

[30] S. Marvizi and R. Melrose, "Spectral invariants of convex planar regions", *J. Differential Geom.* **17**:3 (1982), 475–502. MR Zbl

[31] J. N. Mather and G. Forni, "Action minimizing orbits in Hamiltonian systems", pp. 92–186 in *Transition to chaos in classical and quantum mechanics* (Montecatini Terme, 1991), Lecture Notes in Math. **1589**, Springer, 1994. MR Zbl

[32] D. E. McClure and R. A. Vitale, "Polygonal approximation of plane convex bodies", *J. Math. Anal. Appl.* **51**:2 (1975), 326–358. MR Zbl

[33] J. D. Meiss, "Symplectic maps, variational principles, and transport", *Rev. Modern Phys.* **64**:3 (1992), 795–848. MR Zbl

[34] R. B. Melrose, "Equivalence of glancing hypersurfaces", *Invent. Math.* **37**:3 (1976), 165–191. MR Zbl

[35] J. Molnár, "On inscribed and circumscribed polygons of convex regions", *Mat. Lapok* **6** (1955), 210–218. MR Zbl

[36] J. Pach and P. K. Agarwal, *Combinatorial geometry*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 1995. MR Zbl

[37] V. M. Petkov and L. N. Stoyanov, *Geometry of reflecting rays and inverse spectral problems*, Wiley, Chichester, 1992. MR Zbl

[38] M. Robnik and M. V. Berry, "Classical billiards in magnetic fields", *J. Phys. A* **18**:9 (1985), 1361–1378. MR Zbl

[39] K. F. Siburg, *The principle of least action in geometry and dynamics*, Lecture Notes in Mathematics **1844**, Springer, 2004. MR Zbl

[40] A. Sorrentino, "Computing Mather's β-function for Birkhoff billiards", *Discrete Contin. Dyn. Syst.* **35**:10 (2015), 5055–5082. MR Zbl

[41] S. Tabachnikov, "On the dual billiard problem", *Adv. Math.* **115**:2 (1995), 221–249. MR Zbl

[42] S. Tabachnikov, *Geometry and billiards*, Student Mathematical Library **30**, Amer. Math. Soc., Providence, RI, 2005. MR Zbl

[43] S. Tabachnikov and F. Dogru, "Dual billiards", *Math. Intelligencer* **27**:4 (2005), 18–25. MR Zbl

Received February 25, 2024. Revised June 11, 2024.

PETER ALBERS
INSTITUT FÜR MATHEMATIK
UNIVERSITÄT HEIDELBERG
69120 HEIDELBERG
GERMANY

peter.albers@uni-heidelberg.de

SERGE TABACHNIKOV
DEPARTMENT OF MATHEMATICS
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PA 16802
UNITED STATES

tabachni@math.psu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 330, No. 1, 2024
https://doi.org/10.2140/pjm.2024.330.25

UNKNOTTING VIA NULL-HOMOLOGOUS TWISTS AND MULTITWISTS

SAMANTHA ALLEN, KENAN İNCE, SEUNGWON KIM,
BENJAMIN MATTHIAS RUPPIK AND HANNAH TURNER

This collaborative work was based on the proposed problem and prior work of our senior member, Kenan İnce. After several years of work and the writing and submission of this paper, Kenan passed away unexpectedly. They were an exceptional human, mathematician, and advocate. We dedicate this paper to them.

The *untwisting number* of a knot K is the minimum number of null-homologous twists required to convert K to the unknot. Such a twist can be viewed as a generalization of a crossing change, since a classical crossing change can be effected by a null-homologous twist on 2 strands. While the unknotting number gives an upper bound on the smooth 4-genus, the untwisting number gives an upper bound on the topological 4-genus. The *surgery description number*, which allows multiple null-homologous twists in a single twisting region to count as one operation, lies between the topological 4-genus and the untwisting number. We show that the untwisting and surgery description numbers are different for infinitely many knots, though we also find that the untwisting number is at most twice the surgery description number plus 1.

1. Introduction

Given two knot diagrams D_1, D_2 of knots K_1, K_2 which differ only inside small disks $\Delta \subset D_1$, $\Delta' \subset D_2$ containing at least one crossing, a *local move on* K_1 is the act of replacing Δ with Δ', and hence converting D_1 to D_2. An *unknotting operation* is a local move such that, for any diagram D of a knot K, we may transform D into a diagram of the unknot via a finite sequence of these local moves. A natural question in knot theory is: given an unknotting operation and a knot K, how many such operations are needed to turn K into the unknot? The most common such unknotting operation is a crossing change, which gives rise to the unknotting number $u(K)$. While the unknotting number is quite simple to define, its computation is frequently difficult. For example, Milnor's conjecture about the unknotting number of torus knots was only proven about 25 years later

MSC2020: 57K10, 57K40.
Keywords: 4-manifolds, surgery diagram, unknotting operation, untwisting number.

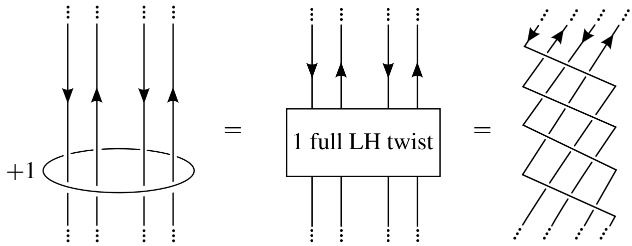

Figure 1. A left-handed null-homologous twist on 4 strands.

by Kronheimer and Mrowka [1993; 1995]. In this paper, we study additional unknotting operations, many of which are generalizations of the crossing change. See Table 1 for an organized list of the related invariants.

One of the primary unknotting operations studied in this paper is a *null-homologous twist*. Mathieu and Domergue [1988] defined this generalization of unknotting number and it was subsequently considered by Livingston [2002]. A null-homologous twist on a knot K is the result of a ± 1-surgery on a null-homologous unknot $U \subset S^3 - K$ bounding a disk D such that $D \cap K = 2k$ points for any $k \in \mathbb{Z}_{\geq 1}$. Diagrammatically, this is the result of adding a full right- or left-handed twist in the twisting region indicated by the unknot U, where $\mathrm{lk}(K, U) = 0$. See Figure 1 for a diagrammatic representation. It is described in [Ince 2016] how a crossing change may be encoded as a null-homologous twist where $D \cap K = 2$. In particular, this implies null-homologous twists are unknotting operations.

The corresponding knot invariant is the *untwisting number* $\mathrm{tu}(K)$, which is defined as the minimum length, taken over all diagrams of K, of a sequence of null-homologous twists beginning at K and resulting in the unknot. This has been the subject of much research in recent years [Baader et al. 2020; Ince 2016; 2017; Livingston 2021; McCoy 2021a; 2021b].

There are many variations of the unknotting number and untwisting number, see Table 1. One variant we will study, due to Nakanishi [2005] (and called the "surgical description number" in that paper), is what we and many other authors call the *surgery description number* $\mathrm{sd}(K)$ of a knot. Again we consider null-homologous twists but now allow any number of full twists to be added in the twisting region; we may call this a *null-homologous m-twist* for $m \in \mathbb{Z}$ to specify the number of twists (with sign) being effected. Then $\mathrm{sd}(K)$ is the minimal number of m-twists necessary to unknot K. (Here, the value of m may change from move to move.)

Another natural variant (due to Murakami [1990]) is the *algebraic unknotting number $u_a(K)$*, the minimum number of crossing changes necessary to turn a given knot into an Alexander polynomial-one knot. Freedman [1982] showed that knots with Alexander polynomial equal to one are topologically slice (in other words, with topological 4-genus $g_4^{\mathrm{top}} = 0$); topologically slice knots are indistinguishable

invariant	definition
$u(K)$	unknotting number of K, i.e., minimal number of crossing changes to unknot
$u_a(K)$	alg. unknotting number, minimal number of crossing changes to Alexander polynomial-one knot
$\mathrm{tu}_a(K)$	alg. untwisting number, minimal number of null-homologous twists to Alexander polynomial-one knot
$\mathrm{tu}(K)$	untwisting number, i.e., the minimal number of null-homologous twists to unknot
$\mathrm{sd}(K)$	surgery description number, i.e., the minimal number of null-homologous multitwists (on the same region of K) to unknot
$\mathrm{sd}_a(K)$	algebraic surgery description number, i.e., the minimal number of null-homologous multitwists (on the same region of K) to Alexander polynomial-one knot
$g_{\mathrm{alg}}(K)$	algebraic genus, i.e., minimal difference in genus between a Seifert surface F for K and a subsurface whose boundary is an Alexander polynomial-one knot

Table 1. Overview of knot invariants appearing in this paper.

from the unknot by *classical invariants*, or knot invariants derived from the Seifert matrix. We consider the similarly defined *algebraic untwisting number* $\mathrm{tu}_a(K)$ and *algebraic surgery description number* $\mathrm{sd}_a(K)$, measuring the number of null-homologous twists or m-twists, respectively, needed to obtain a knot with Alexander polynomial-one, as well.

A tight classical upper bound on the topological 4-genus g_4^{top} of a knot is the *algebraic genus* g_{alg} defined in [Feller and Lewark 2018]. Distinguishing the algebraic genus from other upper bounds on g_4^{top}, such as the algebraic unknotting number, is often achieved by using the bound $g_{\mathrm{alg}} \leq g_3$, where $g_3(K)$ is the 3-genus of K. In Section 3, we provide the first (to our knowledge) known infinite family of knots L_n for which $g_{\mathrm{alg}}(L_n) < u_a(L_n)$ for all $n \in \mathbb{N}$, and since the 3-genus of our examples is large, we do so without using g_3.

The untwisting number connects to recent work of Manolescu and Piccirillo [2023] on candidates for exotic definite 4-manifolds, which uses the concept of *strong H-sliceness* in definite connected sums of $\pm\mathbb{CP}^2$. (See Section 3 for a related definition.) It follows from Proposition 4.1 of [Ince 2017] that, if K can be unknotted using n positive (respectively, negative) nullhomologous twists, then K is strongly topologically H-slice in $X := B^4\#^n \mp \mathbb{CP}^2 \cong \#^n \mp \mathbb{CP}^2$. We use this fact to obstruct knots from having $\mathrm{sd}_a = 1$ in Section 3.

Results. Our main results involve various relationships between the untwisting number and the surgery description number. To start, we give the first known examples (to the authors' knowledge) such that $\mathrm{sd} \neq \mathrm{tu}$. See Section 4 for a description of these knots.

Theorem 1.1. *There are infinitely many knots* $\{K_n\}$ *with* $\mathrm{sd}(K_n) = 1$ *and* $\mathrm{tu}(K_n) = 2$.

This, of course, leads to questions about how far apart the surgery description number and the untwisting number can be.

Question 1.2. Can tu and sd be arbitrarily far apart?

Answering such a question is made more difficult by the close relationships between tu and sd both in definition and in values, demonstrated by the two results:

Theorem 1.3. *Let* $K \subset S^3$ *be a knot. Then* $\mathrm{sd}_a(K) \leq \mathrm{tu}_a(K) \leq 2\,\mathrm{sd}_a(K)$.

Theorem 1.4. *Let* $K \subset S^3$ *be a knot. Then* $\mathrm{sd}(K) \leq \mathrm{tu}(K) \leq 2\,\mathrm{sd}(K) + 1$.

The proof of Theorem 1.3 relies on the work of Duncan McCoy [2021b] relating the untwisting number to the algebraic genus. The proof of Theorem 1.4 is constructive (involving surgery diagrams and Kirby calculus) and allows one to reduce multiple twists in a single region to at most 3 twists in separate regions.

Organization. In Section 2, we give formal definitions of all relevant invariants, as well as some useful prior results. We also prove Theorem 1.3 as a consequence of [McCoy 2021b]. We prove Theorem 1.1 in Section 4 by providing an infinite family of examples where the invariants disagree. Theorem 1.4 is proved in Section 5.

2. Algebraic untwisting invariants

One way to study an unknotting operation is to analyze its impact on the Alexander polynomial of a knot. The effect of an operation on the Alexander polynomial gives rise to *algebraic unknotting operations*:

Definition 2.1. Given an unknotting operation $\mathcal{U}$ and a knot K, the *algebraic* $\mathcal{U}$-*number* $\mathcal{U}_a(K)$ is the minimal number of $\mathcal{U}$-operations that must be performed in order to convert K into a knot with Alexander polynomial-one.

We certainly have that $\mathcal{U}_a(K) \leq \mathcal{U}(K)$ for any unknotting operation $\mathcal{U}$ and knot K. A lower bound on the algebraic unknotting and untwisting numbers is the topological 4-genus. Another (typically tighter) upper bound on the topological 4-genus is the *algebraic genus*, defined by Feller and Lewark [2018].

Definition 2.2. The *algebraic genus* $g_{\mathrm{alg}}(K)$ of a knot K is the minimum difference in genus $g(F) - g(F')$ between a Seifert surface F for K and a subsurface $F' \subset F$ with the property that $\partial F' = K'$ is a knot with $\Delta_{K'}(t) = 1$.

We note that Definition 2.2 implies that a knot K has $g_{\mathrm{alg}}(K) = 0$ if and only if $\Delta_K(t) = 1$. McCoy proves the following useful characterization of the sensitivity of the algebraic genus to null-homologous twisting.

Theorem 2.3 [McCoy 2021b, Theorem 1.1]. *If K and K' are knots and $m, n \in \mathbb{Z}$ are such that a null-homologous m-twist followed by a null-homologous n-twist on K results in K' and $-mn$ is a square, then*

$$|g_{\mathrm{alg}}(K) - g_{\mathrm{alg}}(K')| \leq 1.$$

Proposition 2.4 [McCoy 2021b, Proposition 3.1]. *Given a knot K with $g_{\mathrm{alg}}(K) > 0$, there exists a knot K' with $g_{\mathrm{alg}}(K') = g_{\mathrm{alg}}(K) - 1$ such that K can be obtained from K' by one right-handed and one left-handed null-homologous twist.*

Feller and Lewark [2018] show that for a knot K the algebraic genus and the algebraic unknotting number are related by $g_{\mathrm{alg}}(K) \leq u_a(K) \leq 2g_{\mathrm{alg}}(K)$. We will show that in fact

$$g_{\mathrm{alg}}(K) \leq \mathrm{sd}_a(K) \leq u_a(K) \leq 2g_{\mathrm{alg}}(K) \leq 2\,\mathrm{sd}_a(K) \tag{2.5}$$

and that $\mathrm{sd}_a(K)$ can provide a better lower bound for $u_a(K)$ than $g_{\mathrm{alg}}(K)$. We begin by showing that the algebraic genus is in fact a lower bound on the algebraic surgery description number.

Proposition 2.6. *Let $K \subset S^3$ be a knot. Then $g_{\mathrm{alg}}(K) \leq \mathrm{sd}_a(K)$.*

Proof. Suppose that K is a knot with $\mathrm{sd}_a(K) = k$. Then there exists a sequence of k null-homologous m_i-twists (for $1 \leq i \leq k$) converting K to a knot with Alexander polynomial-one (which by definition has algebraic genus 0). By Theorem 2.3 (with $n = 0$), each of these m_i-twists decreases the algebraic genus by at most 1, whence $g_{\mathrm{alg}}(K) \leq k$. □

Note that in conjunction with the fact that $g_4^{\mathrm{top}} \leq g_{\mathrm{alg}}$ we have that sd_a and sd are upper bounds on the topological 4-genus. Before proving Theorem 1.3, we need to note the following result of İnce.

Theorem 2.7 [Ince 2016, Theorem 1.1]. *Let $K \subset S^3$ be a knot. Then we have $u_a(K) = \mathrm{tu}_a(K)$.*

We are now ready to prove Theorem 1.3: that, $\mathrm{sd}_a(K) \leq \mathrm{tu}_a(K) \leq 2\,\mathrm{sd}_a(K)$ for any knot K.

Proof of Theorem 1.3 and inequality (2.5). Since any single null-homologous twist is an m-twist with $m = \pm 1$, we have $\mathrm{sd}_a(K) \leq \mathrm{tu}_a(K)$ for any knot K. Combining Proposition 2.6 with Feller and Lewark's [2018] result that $u_a(K) \leq 2g_{\mathrm{alg}}(K)$, we have that $u_a(K) \leq 2\,\mathrm{sd}_a(K)$. Theorem 1.3 and inequality (2.5) now follow from Theorem 2.7. □

Note. Borodzik [2019] showed that the minimal number of null-homologous twists *on two strands* needed to convert a knot K into a knot with Alexander polynomial-one is always less than *three* times the algebraic surgery description number. In fact, our Theorem 1.3, together with the fact that a crossing change is a special case of a null-homologous two-strand twist and the fact that $u_a = \mathrm{tu}_a$, refines this upper bound to twice the algebraic surgery description number.

Even though the algebraic unknotting $u_a(K)$ and untwisting numbers $\mathrm{tu}_a(K)$ coincide, the algebraic surgery description number $\mathrm{sd}_a(K)$ can be strictly less than $u_a(K) = \mathrm{tu}_a(K)$; this is the content of Corollary 4.4.

To conclude that the algebraic surgery description number $\mathrm{sd}_a(K)$ can be a better lower bound on the algebraic unknotting number $u_a(K)$ than the algebraic genus $g_{\mathrm{alg}}(K)$, we should show that there is a knot for which $g_{\mathrm{alg}}(K) \neq sd_a(K)$. We provide infinitely many examples with this property in the next section.

3. Infinite families of knots with $g_{\mathrm{alg}} < \mathrm{sd}_a$

A knot $K \subset S^3$ is called *topologically H-slice* in a closed, smooth 4-manifold M if $K \subset \partial(M \setminus B^4)$ bounds a locally flat, properly embedded, null-homologous topological disk in $M \setminus B^4$. In the context of this paper, if a knot K can be converted to a knot which is topologically slice in B^4 via only left-handed or, respectively, only right-handed nullhomologous m-twists, then K is topologically H-slice in $\#_n \pm \mathbb{CP}^2$ for some n. The following proposition is well known and follows from, for instance, [Conway and Nagel 2020, Theorem 3.8]. To interpret their theorem in our setting, consider a knot K (trivially a colored link) which bounds a null-homologous disk D in $\#_m \mathbb{C}P^2$. Here D can be thought of as an annular cobordism from K to the unknot with no double-points.

Proposition 3.1. *If a knot K is topologically H-slice in $\#_m \mathbb{CP}^2$, then for any $\omega \in S^1$ with $\Delta_K(\omega) \neq 0$,*

$$-2m \leq \sigma_K(\omega) \leq 0,$$

where $\sigma_K(\omega)$ is the Levine–Tristram signature function of K.

In particular, the proposition above implies that if the signature function of a knot takes on both positive and negative values, then $\mathrm{sd}_a \neq 1$.

Theorem 3.2. *If K and K' are knots such that*

- $u_a(K) = u_a(K') = 1$,
- *the signature function of K takes a positive value at a nonroot of $\Delta_K(t)$, and*
- *the signature function of K' takes a negative value at a nonroot of $\Delta_{K'}(t)$,*

then $g_{\mathrm{alg}}(K\#K') = 1$. If, in addition, the signature function of $K\#K'$ takes both positive and negative values at a nonroot of $\Delta_{K\#K'}(t)$, then $g_{\mathrm{alg}}(K\#K') < \mathrm{sd}_a(K\#K')$.

Proof. Suppose K and K' are knots which satisfy the assumptions of Theorem 3.2. Now consider the knot $J = K\#K'$. Note that because K and K' have nontrivial signature functions, neither K nor K' has Alexander polynomial-one. So $\Delta_J(t) \neq 1$ and $g_{\text{alg}}(J) \neq 0$. Because $u_a(K) = u_a(K') = 1$, the knots K and K' can be converted into knots with Alexander polynomial-one via a single crossing change. Recall that a crossing change can change the Levine–Tristram signature function by at most ± 2, where the sign depends on the sign of the crossing change (see, for example, [Conway 2021, Proposition 3(1)]). This implies that the crossing changes converting K and K' to knots with trivial Alexander polynomial can be taken to be of opposite signs. So the knot J can be converted into a knot J' with $\Delta_{J'}(t) = 1$ via a sequence of two crossing changes, one positive and one negative. Since a crossing change can be realized by single null-homologous twist, by Theorem 2.3 we have that

$$|g_{\text{alg}}(J) - g_{\text{alg}}(J')| \leq 1.$$

Because $\Delta_{J'}(t) = 1$, we have that $g_{\text{alg}}(J') = 0$. So $g_{\text{alg}}(J) = 1$ as desired.

Now, suppose that K and K' also satisfy that $\sigma_J(\omega) = \sigma_K(\omega) + \sigma_{K'}(\omega)$ takes both positive and negative values. By Proposition 3.1, $J = K\#K'$ is not topologically H-slice in $\#_m \pm \mathbb{CP}^2$ for any $m \in \mathbb{N}$. In particular, J cannot be converted to a topologically slice knot using a single null-homologous m-twist. Thus we have $\text{sd}_a(J) > 1$. □

Theorem 3.3. *There exists an infinite family* $\{K_n\}_{n=2}^{\infty}$ *of prime knots such that* $g_{\text{alg}}(K_n) < \text{sd}_a(K_n)$ *for all* $n \geq 2$.

Note. In fact, since the Levine–Tristram signature and algebraic unknotting number of a knot K are invariants of the S-equivalence class of its Seifert matrix, for any Seifert matrices V, V' satisfying the conditions of Theorem 3.2, there exist infinitely many knots K, K' with Seifert matrices in the S-equivalence classes of V, V', respectively, satisfying the conclusions of the theorem. In particular, our K_n can be chosen to have any adjective (e.g., hyperbolic, quasipositive, ...) for which there are infinitely many representative knots with that property in each S-equivalence class, since our proof relies only on the S-equivalence class of K_n.

In the proof below, we exhibit a concrete family of prime knots via cabling because cabling seems to be of independent interest.

Proof. Suppose that K and K' are knots which satisfy all the assumptions of Theorem 3.2. For example, we can take $K = 10_{32}$ and $K' = -10_{82}$ (see Figure 2). For $n \geq 2$, let $K_n := (K\#K')_{n,1}$, the $(n, 1)$-cable of $K\#K'$. Note that the $(n, 1)$-cable of any knot (where $n \geq 2$) is prime by [Cromwell 2004, Theorem 4.4.1]. Then we have that $g_{\text{alg}}(K_n) \neq 0$ because $\Delta_{K_n}(t) = \Delta_{T(n,1)}(t) \cdot \Delta_{K_1\#K_2}(t^n) \neq 1$ (by [Lickorish 1997, Theorem 6.15] since $g_{\text{alg}}(K_1\#K_2) \neq 0$). On the other hand,

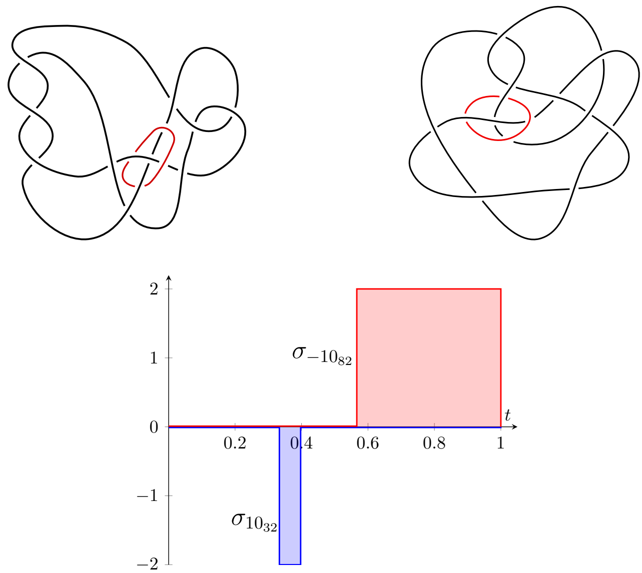

Figure 2. Top: the knots 10_{32} (left) and 10_{82} (right) with red unknots indicating an unknotting crossing change for each. Bottom: the Levine–Tristram signature functions for 10_{32} and -10_{82} (reparametrized so that $\omega = e^{2\pi i t}$).

Feller et al. [2022] tells us how g_{alg} acts under satellite operations. In particular, $g_{\text{alg}}(K_n) \leq g_{\text{alg}}(T(n, 1)) + g_{\text{alg}}(K_1 \# K_2) = 1$, where $T(n, 1)$ denotes the $(n, 1)$-torus knot. So we have that $g_{\text{alg}}(K_n) = 1$.

Also, by [Litherland 1979, Theorem 2], $\sigma_{K_n}(\omega) = \sigma_{T(n,1)}(\omega) + \sigma_{K_1 \# K_2}(\omega^n)$. Since $\sigma_{K_1 \# K_2}(\omega)$ takes both positive and negative values, so does $\sigma_{K_n}(\omega) = \sigma_{K_1 \# K_2}(\omega^n)$. Proposition 3.1 then implies that $\text{sd}_a(K_n) > 1$. □

We remark that, for any knot K with $g_{\text{alg}}(K) = 1$ and $\text{sd}_a(K) \geq 2$, inequality (2.5) implies that $\text{sd}_a(K) = u_a(K) = 2$. In particular, the knots K_n from Theorem 3.3 satisfy $g_{\text{alg}}(K_n) = 1 < 2 = \text{sd}_a(K_n) = u_a(K_n)$ for all $n \geq 2$. A literature search suggests that $\{K_n\}$ is the first known infinite family of knots for which $g_{\text{alg}} < u_a$. Note that, in [Feller and Lewark 2018], the 3-genus is used to distinguish between g_{alg} and u_a for various knots since $g_{\text{alg}}(K) \leq g_3(K)$ while $u_a \leq 2g_3(K)$. In our case, the 3-genus of the K_n grows large, and we use a different strategy for distinguishing between g_{alg} and u_a.

4. Relationships between the surgery description and untwisting numbers

In the last section, we found an infinite family of knots for which $u_a = \text{tu}_a = \text{sd}_a = 2$. Other examples can be found where $u_a = \text{tu}_a = \text{sd}_a$ are abundant. We now endeavor to find examples where the two quantities (and other similar quantities) disagree.

In particular, in this section, we examine the square of inequalities below, and show that each inequality can be strict for infinitely many knots:

$$\begin{array}{ccc} \mathrm{sd}_a & \leq & \mathrm{tu}_a \\ \text{\rotatebox{90}{$\geq$}} & & \text{\rotatebox{90}{$\geq$}} \\ \mathrm{sd} & \leq & \mathrm{tu} \end{array} \tag{4.1}$$

It is easy to find infinitely many knots such that the vertical inequalities in (4.1) are strict; for example, any nontrivial knots with Alexander polynomial-one satisfy $\mathrm{sd}_a < \mathrm{sd}$ and $\mathrm{tu}_a < \mathrm{tu}$. Finding such examples for the horizontal inequalities in (4.1) is more challenging.

It is known that $u(K)$ and $\mathrm{tu}(K)$ can be arbitrarily different [Ince 2016]. In contrast, we can find no examples of knots in the literature with $\mathrm{sd}(K) \neq \mathrm{tu}(K)$. We provide the first known examples below; in fact, we find an infinite family $\{K_n\}$ of knots satisfying the stronger inequality $\mathrm{sd}(K_n) < \mathrm{tu}_a(K_n)$ for all $n \geq 2$. This is the content of Theorem 1.1. The same family provides infinitely many examples where $\mathrm{sd}_a < \mathrm{tu}_a = u_a$.

For the proof of Theorem 1.1, we employ an obstruction to a knot having *algebraic* unknotting number 1 due to Borodzik and Friedl [2015], which in turn generalizes an unknotting number 1 obstruction due to Lickorish [1985]. The obstruction involves the linking pairing on the first homology of the double-branched cover $\Sigma(K)$; see, e.g., [Gordon 1978] for a discussion of the linking pairing.

Theorem 4.2 [Borodzik and Friedl 2015, Theorems 4.5 and 4.6]. *If a knot K can be **algebraically** unknotted by a single crossing change, then there exists a generator h of $H_1(\Sigma(K); \mathbb{Z})$ such that its linking pairing satisfies*

$$l(h, h) = \frac{\pm 2}{\det(K)} \in \mathbb{Q}/\mathbb{Z}.$$

The proof of Theorem 1.1 follows Lickorish's proof that the knot $P(3, 1, 3)$ does not have unknotting number 1 (the main theorem of Lickorish [1985]).

Theorem 1.1. *There are infinitely many knots $\{K_n\}$ with $\mathrm{sd}(K_n) = 1$ and $\mathrm{tu}(K_n) = \mathrm{tu}_a(K_n) = 2$.*

Proof. Suppose the family K_n is the set of pretzel knots of the form $P(10n+3, 1, 3)$; see Figure 3 for two isotopic diagrams and note that the boxed numbers represent *half*-twists. We first note that $\mathrm{sd}(K_n) = 1$ by performing the $+1/2$-surgery (or equivalently a -2-twist) on the curve C indicated in the figure. After the surgery, we obtain the pretzel knots $P(10n + 3, 1, -1)$, all of which are isotopic to the unknot.

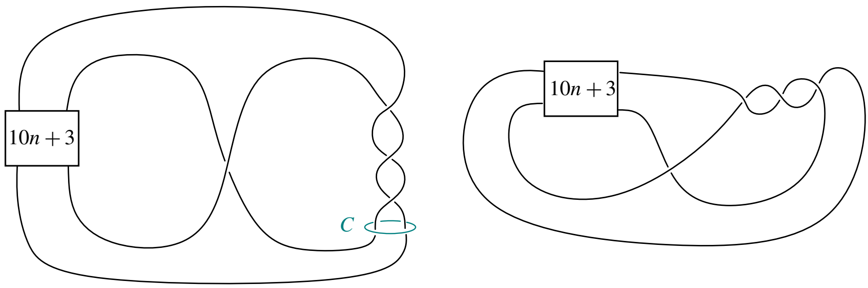

Figure 3. Diagrams of the Pretzel knots $K_n = P(10n+3, 1, 3)$; the boxed numbers represent *half*-twists. Left: a standard diagram for 3-strand pretzel knots, together with an unknotted curve C where a $+1/2$ Dehn surgery can be applied to convert K_n to the unknot. Right: a diagram for the same knot in which it is more clear that the knots are two-bridge. In fact, they have Conway notation $C(10n+3, 1, 3)$.

To conclude that $\mathrm{tu}(K_n) = 2$, it is enough to show that $\mathrm{tu}(K_n) \neq 1$; $\mathrm{tu}(K_n) \leq 2$ since the surgery description move can be effected by two (single) null-homologous twists.

To show that $\mathrm{tu}(K_n) \neq 1$, first recall that $\mathrm{tu}_a(K) \leq \mathrm{tu}(K)$ and that $\mathrm{tu}_a(K) = u_a(K)$ for any knot K. Note that $\Delta_{K_n}(t) \neq 1$ for each $n \geq 2$ (see, e.g., [Lickorish 1997, Example 6.9]), so that $\mathrm{tu}_a(K_n) \neq 0$. We then assume that $\mathrm{tu}_a(K_n) = 1$ for contradiction, and prove that the linking pairing on $H_1(\Sigma(K_n); \mathbb{Z})$ does not satisfy the condition in Theorem 4.2 for any $n \geq 1$.

First, note that the knots K_n are 2-bridge; see Figure 3. Each two-bridge knot has a (nonunique) associated fraction p/q with the property that $\Sigma_2(K) \cong L(p, q)$; see, e.g., [Kawauchi 1996, Chapter 2] for a discussion of two-bridge knots. In fact, $\{K_n\}$ are precisely those 2-bridge knots with continued fraction of the form

$$[10n+3, 1, 3] = 10n+3 + \frac{1}{1+\frac{1}{3}} = \frac{40n+15}{4}.$$

Hence the double-branched covers of these knots $\Sigma(K_n) \cong L(40n+15, 4)$ are lens spaces. So in particular, $\Sigma(K_n)$ can be obtained as surgery on a knot J (in fact the unknot) via $\frac{40n+15}{4}$-surgery. This implies that $H_1(\Sigma(K_n); \mathbb{Z})$ is cyclic of order $40n+15$ generated by μ the image of a meridian of J after surgery, and moreover that $l(\mu, \mu) = \frac{4}{40n+15}$ [Lickorish 1985].

Any generator h of $H_1(\Sigma(K_n))$ is of the form $h = t\mu$ for some integer t. Let h be the generator which must exist according to Theorem 4.2 so that

$$\frac{\pm 2}{40n+15} = l(h, h) = l(t\mu, t\mu) = t^2 \cdot l(\mu, \mu) = \frac{4t^2}{40n+15} \in \mathbb{Q}/\mathbb{Z}. \tag{4.3}$$

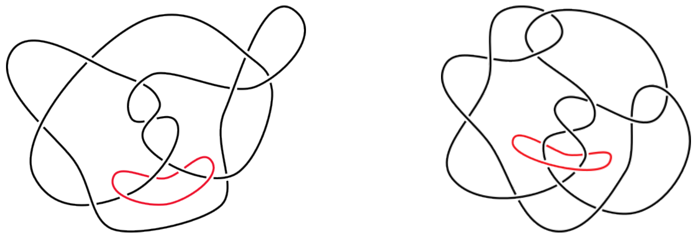

Figure 4. The knots 10_{68} (left) and $11a_{103}$ (right) can be converted to the unknot by inserting two left (resp. right)-handed twists in the regions indicated by the red unknots.

For the two fractions on the far left and far right of (4.3) to be equivalent in $\mathbb{Q}/\mathbb{Z}$, we must have $\pm 2 \equiv 4t^2 \pmod{40n+15}$ so that ± 2 must be a square $\pmod{40n+15}$. We will show that this is not true.

If ± 2 is a square $\pmod{40n+15}$ then it also must be a square $\pmod{a}$ where a is any factor of $40n+15$. In particular, ± 2 must be a square $\pmod 5$. But neither $-2 \equiv 3$ nor 2 are squares $\pmod 5$. This is a contradiction, and hence $\text{tu}_a(K_n) \neq 1$ for each n, which forces $\text{tu}(K_n) \neq 1$. □

Since $\text{sd}_a(K) \leq \text{sd}(K)$ for all knots K, the next corollary immediately follows.

Corollary 4.4. *There are infinitely many knots* $\{K_n\}$ *for which* $\text{sd}_a(K_n) = 1$ *while* $\text{tu}_a(K_n) = u_a(K_n) = 2$.

Note that Corollary 4.4 is the biggest gap we could hope for in the sense that $\text{sd}_a(K) \leq \text{tu}_a(K) = u_a(K) \leq 2\,\text{sd}_a(K)$.

While Theorem 1.1 provides infinitely many examples where $\text{sd} < \text{tu}_a$, one might ask if $\text{sd} \leq \text{tu}_a$ in general. The following theorem provides an answer to this question in the negative.

Theorem 4.5. *The* $(p, 1)$*-cable of the untwisted Whitehead double of any nontrivial knot, which we denote* D_p, *has* $\text{tu}_a(D_p) = u_a(D_p) = 0 < 1 = \text{sd}(D_p)$ *for all* $p \in \mathbb{N}$.

Proof. First, note that the Alexander polynomial of D_p, for any p, is equal to 1 (see the cabling relation in [Lickorish 1997]). Thus $\text{tu}_a(D_p) = 0$. On the other hand, since D_p is not unknotted, we must have that $\text{sd}(D_p) \geq 1$. In fact, one can see that $\text{sd}(D_p) = 1$ by performing a single null-homologous twist about the clasping region in the untwisted Whitehead double. □

Note. To distinguish between sd and tu, obstructions from Heegaard–Floer homology can be used, though this seems feasible only to show that $\text{sd} = 1 < 2 = \text{tu}$ for individual knots. In particular, the sd-moves in Figure 4 show that the knots 10_{68} and $11a_{103}$ have $\text{sd}(K) = 1$, though the facts that $\text{tu}(10_{68}), \text{tu}(11a_{103}) = 2$ are results of [Ince 2017, Theorems 1.3, 1.4].

For all examples we produce with $\mathrm{sd} \neq \mathrm{tu}$ the two invariants in fact only differ by 1. In Section 5 below, we will prove Theorem 1.4 which states that the ratio of tu to sd is at most 3. This leaves open the following question.

Question 4.6. Does there exist a knot K with $\mathrm{sd}(K) = 1$ but $\mathrm{tu}(K) = 3$, or in general so that $\mathrm{tu}(K) = 2\,\mathrm{sd}(K) + 1$?

Note that the techniques used in the proof of Theorem 1.1 cannot be used to obstruct a knot K with $\mathrm{sd}(K) = 1$ from having $\mathrm{tu}(K) \leq 2$ since the algebraic invariants can differ at most by a factor of 2 by Theorem 1.3. It also seems unlikely that the Floer theoretic techniques of Ince [2017] would alone be enough to answer Question 4.6 given the difficulty in obstructing knots from being H-slice in indefinite 4-manifolds [Kjuchukova et al. 2021]. Thus, new techniques are likely needed to answer the question above.

5. An inequality relating surgery description number and untwisting number

In the previous section we asked whether a knot K with $\mathrm{sd}(K) = 1$ and $\mathrm{tu}(K) = 3$ can exist, or more generally, if a knot with $2\,\mathrm{sd}(K) + 1 = \mathrm{tu}(K)$ exists. In this section, we show that the untwisting number is at most twice the surgery description number plus 1.

The following theorem was inspired by the work of Borodzik [2019] on algebraic k-simple knots. In addition, Duncan McCoy suggested the last portion of the proof of Theorem 1.4, improving the upper bound from an earlier version of the paper.

Theorem 1.4. *For any knot K, we have that* $\mathrm{sd}(K) \leq \mathrm{tu}(K) \leq 2\,\mathrm{sd}(K) + 1$.

Note that while the following proof involves a series of Kirby calculus moves, the moves used are slam dunk moves (away from the knot), and handle slides involving only the added components (never the original knot); thus none of the moves alter the isotopy class of the knot. The result is diagrammatic. For a reference on Kirby calculus, see [Gompf and Stipsicz 1999].

Proof. The first inequality is clear from the definitions. To show the second inequality, we will first show that an unknot of framing $\pm 1/(2k+1)$ which is null-homologous in the complement of K can be replaced (via careful Kirby calculus) with two unlinked, null-homologous unknots, one with framing $+1$ and one with framing -1. Thus $2k+1$ full twists in a single twisting region can be realized by a sequence of two full twists (of opposite signs) in some diagram of K. This process (Procedure 1) is described below; an example in the case of five left-handed twists is shown in Figure 5. Throughout, we abuse notation and keep names of unknots unchanged after they have undergone a handle slide.

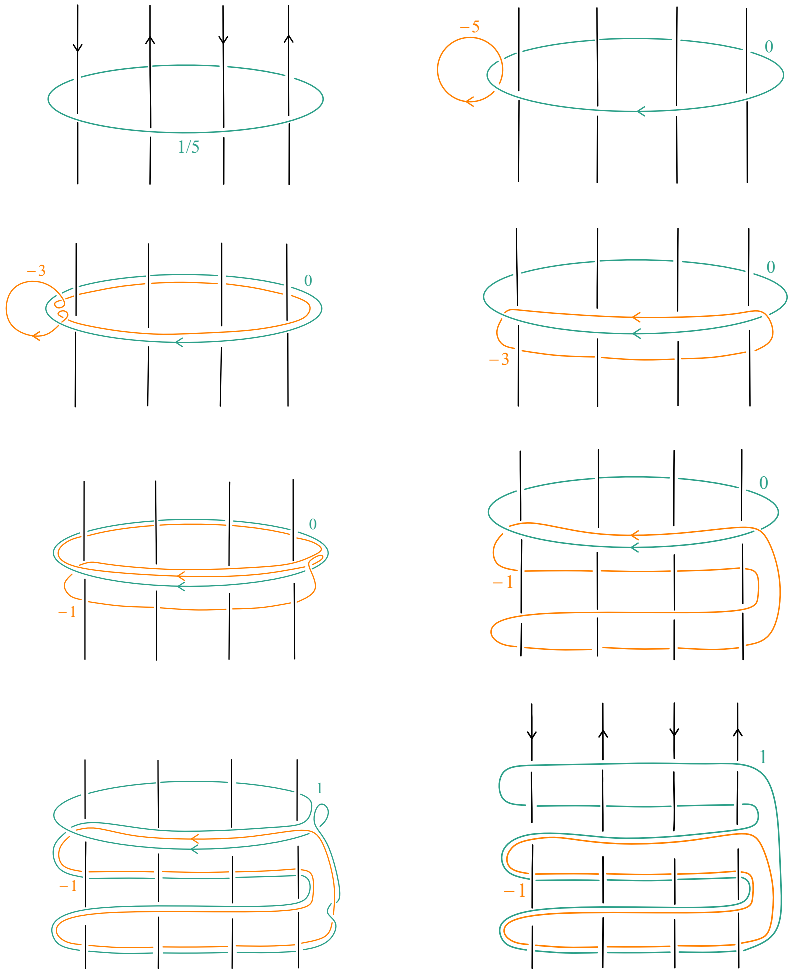

Figure 5. A sequence of Kirby moves which shows that applying 5 parallel null-homologous twists can be obtained by two null-homologous twists. Top row: effecting null-homologous twist(s) (left) and a slam dunk move (right). From second row: handle addition (left) and isotopy (right).

Procedure 1. (1) Use a reverse slam dunk move to view the $\pm 1/(2k+1)$-framed unknot as a 0-framed unknot U_1 geometrically linked once with a $\mp(2k+1)$-framed unknot U_2 as in Figure 5 (top row).

(2) By repeatedly sliding U_2 over U_1, one can change the framing on U_2 to ∓ 1. See Figure 5 (second and third row). Note that, in each handle slide, only the portion of U_2 near U_1 is affected. While this changes how K and U_2 are geometrically linked, the unknots U_1 and U_2 remain linked once.

(3) Finally, slide U_1 over U_2. This has the effect of changing the framing of U_1 by ∓ 1. See Figure 5 (third row, right and bottom row, left). After an isotopy (Figure 5, bottom row, right), it is not hard to see that the resulting U_1 and U_2 are unlinked.

We now show that unknots with framings ± 1 and $\pm 1/(2k)$ which are null-homologous in the complement of K can be replaced (again, via Kirby calculus) with three unlinked, null-homologous unknots, two with framings ± 1 and one with framing ∓ 1. The process is described below; an example is shown in Figure 6.

Procedure 2. (1) Use a reverse slam dunk move to view the $\pm 1/(2k)$-framed unknot as a 0-framed unknot U_1 geometrically linked once with a $\mp(2k)$-framed unknot U_2 as in Figure 6 (top row).

(2) At the beginning of the procedure we assumed we had unknots with framings ± 1 and $\pm 1/(2k)$. Call the unknot with ± 1 framing U_3. Slide U_2 over U_3 with framing ± 1 to change the framing on U_2 by 1. See 6 (middle row, left). At this stage, U_2 is linked with both U_1 and U_3.

(3) Slide U_3 over U_1 in order to unlink U_3 from U_2. The result is that, after an isotopy, U_3 is completely unlinked from U_1 and U_2. In addition, U_1 and U_2 are in position to perform the procedure from the previous paragraph. See Figure 6 (middle row, right and bottom row).

(4) Apply steps (2) and (3) of Procedure 1.

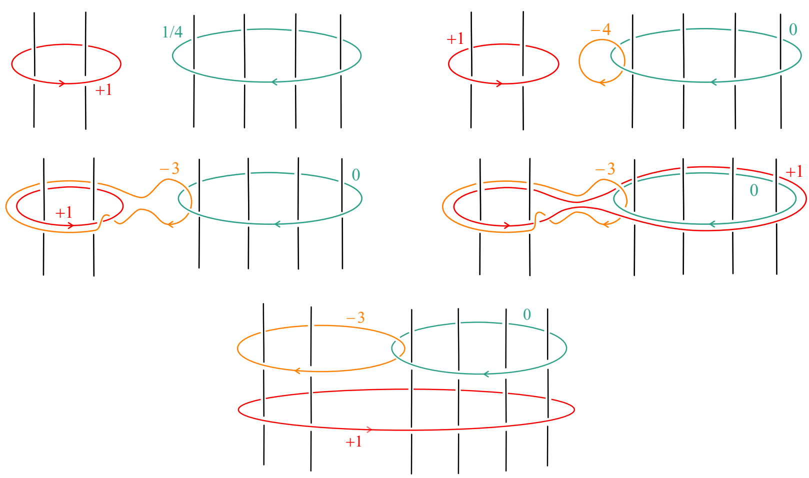

Figure 6. A sequence of Kirby moves to replace +1- and +1/4-framed null-homologous unknots in the knot complement with an unlinked +1-framed component and two components linked once, one with framing 0. Top row: null-homologous unknots in knot complement (left) and a slam dunk move (right). Middle row: handle addition. Bottom: isotopy.

Thus, to see the upper bound, consider the following cases.

- First, if the surgery description number can be realized using only $\pm(2k+1)$-moves (odd numbers of full twists in each twisting region), then we apply Procedure 1 to reduce each $(2k+1)$-move to a $+1$- and -1-move. Thus, in this case, $\mathrm{tu}(K) \leq 2\,\mathrm{sd}(K)$.

- Second, if at least one $\pm(2k)$-move (even number of full twists in a single twisting region) is required to realize the surgery description number, then replace one of the $\pm(2k)$-moves with parallel ±1- and $\pm(2k-1)$-framed unknots. Call the ±1-framed unknot U_3 and now use Procedure 2 with U_3 to reduce each remaining $\pm(2k)$-moves to a $+1$- and -1-move. Thus, $\mathrm{tu}(K) \leq 2\,\mathrm{sd}(K)+1$. □

Note. In the proof of Theorem 1.4, the upper bound of $2\,\mathrm{sd}(K)+1$ can only be sharp when *every* minimal sd-sequence for K involves *only* even numbers of full twists. In all other cases, consider a minimal sd-sequence which involves at least one null-homologous $(2k+1)$-twist for some $k \in \mathbb{Z}$. We may use Procedure 1 on all $\pm1/(2k+1)$-framed unknots to convert each into two ±1-framed unknots, then use Procedure 2 on all $\pm1/(2k)$-framed unknots (if one exists) using one of the ±1-framed unknots obtained via Procedure 1 to build an untwisting sequence of length $2\,\mathrm{sd}(K)$.

Acknowledgements

This work is the product of a research group formed during the American Institute for Mathematics (AIM) virtual Research Community on 4*-dimensional topology*. The paper is based on work completed while Allen, İnce, Kim, and Turner were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the summer of 2022. We want to thank Peter Feller, Stefan Friedl and Duncan McCoy for their valuable comments on an earlier draft. In particular, Duncan McCoy suggested a strengthening of Theorem 1.4. We also thank the referee for their very careful reading and detailed comments. Kim was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2022R1C1C2004559). Ruppik was supported by the Max Planck Institute for Mathematics in Bonn. Turner was supported by NSF grant 1745583.

References

[Baader et al. 2020] S. Baader, I. Banfield, and L. Lewark, "Untwisting 3-strand torus knots", *Bull. Lond. Math. Soc.* **52**:3 (2020), 429–436. MR Zbl

[Borodzik 2019] M. Borodzik, "Untwisting number and Blanchfield pairings", *Osaka J. Math.* **56**:3 (2019), 497–505. MR Zbl

[Borodzik and Friedl 2015] M. Borodzik and S. Friedl, "The unknotting number and classical invariants, I", *Algebr. Geom. Topol.* **15**:1 (2015), 85–135. MR Zbl

[Conway 2021] A. Conway, "The Levine–Tristram signature: a survey", pp. 31–56 in 2019–20 *MATRIX annals*, edited by J. de Gier et al., MATRIX Book Series **4**, Springer, Cham, 2021. Zbl

[Conway and Nagel 2020] A. Conway and M. Nagel, "Stably slice disks of links", *J. Topol.* **13**:3 (2020), 1261–1301. MR Zbl

[Cromwell 2004] P. R. Cromwell, *Knots and links*, Cambridge University Press, Cambridge, 2004. MR Zbl

[Feller and Lewark 2018] P. Feller and L. Lewark, "On classical upper bounds for slice genera", *Selecta Math.* (*N.S.*) **24**:5 (2018), 4885–4916. MR Zbl

[Feller et al. 2022] P. Feller, A. N. Miller, and J. Pinzón-Caicedo, "The topological slice genus of satellite knots", *Algebr. Geom. Topol.* **22**:2 (2022), 709–738. MR Zbl

[Freedman 1982] M. H. Freedman, "The topology of four-dimensional manifolds", *J. Differential Geometry* **17**:3 (1982), 357–453. MR Zbl

[Gompf and Stipsicz 1999] R. E. Gompf and A. I. Stipsicz, *4-manifolds and Kirby calculus*, Graduate Studies in Mathematics **20**, American Mathematical Society, Providence, RI, 1999. MR Zbl

[Gordon 1978] C. M. Gordon, "Some aspects of classical knot theory", pp. 1–60 in *Knot theory* (Plans-sur-Bex, 1977), edited by J.-C. Hausmann, Lecture Notes in Mathematics **685**, Springer, Berlin, 1978. MR Zbl

[Ince 2016] K. Ince, "The untwisting number of a knot", *Pacific J. Math.* **283**:1 (2016), 139–156. MR Zbl

[Ince 2017] K. Ince, "Untwisting information from Heegaard Floer homology", *Algebr. Geom. Topol.* **17**:4 (2017), 2283–2306. MR Zbl

[Kawauchi 1996] A. Kawauchi, *A survey of knot theory*, Birkhäuser Verlag, Basel, 1996. MR Zbl

[Kjuchukova et al. 2021] A. Kjuchukova, A. N. Miller, A. Ray, and S. Sakalli, "Slicing knots in definite 4-manifolds", preprint, 2021. arXiv 2112.14596

[Kronheimer and Mrowka 1993] P. B. Kronheimer and T. S. Mrowka, "Gauge theory for embedded surfaces, I", *Topology* **32**:4 (1993), 773–826. MR Zbl

[Kronheimer and Mrowka 1995] P. B. Kronheimer and T. S. Mrowka, "Gauge theory for embedded surfaces, II", *Topology* **34**:1 (1995), 37–97. MR Zbl

[Lickorish 1985] W. B. R. Lickorish, "The unknotting number of a classical knot", pp. 117–121 in *Combinatorial methods in topology and algebraic geometry* (Rochester, N.Y., 1982), edited by J. R. Harper and R. Mandelbaum, Contemp. Math. **44**, Amer. Math. Soc., Providence, RI, 1985. MR Zbl

[Lickorish 1997] W. B. R. Lickorish, *An introduction to knot theory*, Graduate Texts in Mathematics **175**, Springer, New York, 1997. MR Zbl

[Litherland 1979] R. A. Litherland, "Signatures of iterated torus knots", pp. 71–84 in *Topology of low-dimensional manifolds*, edited by R. Fenn, Lect. Notes Math. **722**, Springer, Berlin, 1979. Zbl

[Livingston 2002] C. Livingston, "The slicing number of a knot", *Algebr. Geom. Topol.* **2** (2002), 1051–1060. MR Zbl

[Livingston 2021] C. Livingston, "Null-homologous unknottings", pp. 59–68 in *Topology and geometry*: *a collection of essays dedicated to Vladimir G. Turaev*, edited by A. Papadopoulos, IRMA Lect. Math. Theor. Phys. **33**, Eur. Math. Soc., Zürich, 2021. MR Zbl

[Manolescu and Piccirillo 2023] C. Manolescu and L. Piccirillo, "From zero surgeries to candidates for exotic definite 4-manifolds", *J. Lond. Math. Soc.* (2) **108**:5 (2023), 2001–2036. MR Zbl

[Mathieu and Domergue 1988] Y. Mathieu and M. Domergue, "Chirurgies de Dehn de pente ± 1 sur certains nœuds dans les 3-variétés", *Math. Ann.* **280**:3 (1988), 501–508. MR Zbl

[McCoy 2021a] D. McCoy, "Gaps between consecutive untwisting numbers", *Glasg. Math. J.* **63**:1 (2021), 59–65. MR Zbl

[McCoy 2021b] D. McCoy, "Null-homologous twisting and the algebraic genus", pp. 147–165 in 2019–20 *MATRIX annals*, edited by J. de Gier et al., MATRIX Book Ser. **4**, Springer, Cham, 2021. MR Zbl

[Murakami 1990] H. Murakami, "Algebraic unknotting operation", *Quest. Answers Gen. Topology* **8**:1 (1990), 283–292. Zbl

[Nakanishi 2005] Y. Nakanishi, "A note on unknotting number, II", *J. Knot Theory Ramifications* **14**:1 (2005), 3–8. MR Zbl

Received July 21, 2023. Revised April 16, 2024.

SAMANTHA ALLEN
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
DUQUESNE UNIVERSITY
PITTSBURGH, PA
UNITED STATES

allens6@duq.edu

KENAN İNCE
DEPARTMENT OF MATHEMATICS
WESTMINSTER COLLEGE
SALT LAKE CITY, UT
UNITED STATES

kince@westminstercollege.edu

SEUNGWON KIM
DEPARTMENT OF MATHEMATICS
SUNGKYUNKWAN UNIVERSITY
SUWON
SOUTH KOREA

seungwon.kim@skku.edu

BENJAMIN MATTHIAS RUPPIK
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF
DÜSSELDORF
GERMANY

benjamin.ruppik@hhu.de

HANNAH TURNER
SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GA
UNITED STATES

hannah.turner@math.gatech.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 330, No. 1, 2024
https://doi.org/10.2140/pjm.2024.330.43

$\mathbb{R}$-MOTIVIC v_1-PERIODIC HOMOTOPY

EVA BELMONT, DANIEL C. ISAKSEN AND HANA JIA KONG

We compute the v_1-periodic $\mathbb{R}$-motivic stable homotopy groups. The main tool is the effective slice spectral sequence. Along the way, we also analyze $\mathbb{C}$-motivic and η-periodic v_1-periodic homotopy from the same perspective.

1. Introduction

The computation of the stable homotopy groups of spheres is a difficult but central problem of stable homotopy theory. There is much that we do not know about stable homotopy. However, the v_1-periodic stable homotopy groups (also known as the homotopy groups of the spectrum J) are completely understood, and they have interesting number-theoretic properties.

The goal of this article is to explore v_1-periodic stable homotopy in the $\mathbb{R}$-motivic context. This choice of ground field represents a middle ground between the well-understood $\mathbb{C}$-motivic situation and the much more difficult situation of an arbitrary field, in which arithmetic necessarily enters into the picture.

From our perspective, the field $\mathbb{R}$ introduces just one piece of arithmetic: the failure of -1 to have a square root. This leads to complications in $\mathbb{R}$-motivic homotopical computations, but they can be managed with care and attention to detail.

Classically, v_1-periodic homotopy is detected by the connective spectrum j^{top}, which is defined to be the fiber of a map

$$\mathrm{ko}^{\mathrm{top}} \xrightarrow{\psi^3-1} \Sigma^4\mathrm{ksp}^{\mathrm{top}},$$

where $\mathrm{ko}^{\mathrm{top}}$ is the connective real K-theory spectrum, $\mathrm{ksp}^{\mathrm{top}}$ is the connective symplectic K-theory spectrum, and ψ^3 is an Adams operation. (A superscript $^{\mathrm{top}}$ indicates that we are discussing the classical context here, not the motivic context.)

Daniel C. Isaksen was supported by NSF grant DMS-1904241. Hana Jia Kong was supported by the National Science Foundation under Grant DMS-1926686. This article answers a question posed by Mark Behrens to Isaksen in 2010 at the Conference on Homotopy Theory and Derived Algebraic Geometry, Fields Institute, Toronto, Canada. The authors thank William Balderrama, Robert Bruner, and John Rognes for helpful discussions.

MSC2020: primary 55Q10; secondary 14F42, 55Q50, 55T99.

Keywords: $\mathbb{R}$-motivic image of J, $\mathbb{R}$-motivic $K(1)$-local sphere, effective slice spectral sequence, motivic stable homotopy group, v_1-periodicity.

In fact, $\mathrm{ko}^{\mathrm{top}}$ itself is the more natural target for the map $\psi^3 - 1$. However, the fiber of $\mathrm{ko}^{\mathrm{top}} \xrightarrow{\psi^3-1} \mathrm{ko}^{\mathrm{top}}$ has a minor defect. It has some additional homotopy classes in stems -1, 0, and 1 that do not correspond to homotopy classes for the sphere spectrum. In other words, the map from S^0 to this fiber is not surjective in homotopy. If we change the target of $\psi^3 - 1$ from $\mathrm{ko}^{\mathrm{top}}$ to its 3-connective cover $\Sigma^4 \mathrm{ksp}^{\mathrm{top}}$, this problem disappears, and the map from S^0 to the fiber is onto in homotopy.

It is possible to mimic these constructions in motivic stable homotopy theory [5]. At the prime 2, one can define the motivic connective spectrum j to be the fiber of a map $\mathrm{ko} \xrightarrow{\psi^3-1} \Sigma^{4,2}\mathrm{ksp}$, where ko is the very effective connective Hermitian K-theory spectrum, ksp is defined in terms of very effective covers of ko, and ψ^3 is a motivic lift of an Adams operation.

However, from a computational perspective, this definition of j introduces apparently unnecessary complications. It is possible to compute the homotopy of $\mathbb{R}$-motivic j using the techniques that appear later in this article. However, the computation is slightly messy, involving some exceptional differentials and exceptional hidden extensions in low dimensions. In any case, the homotopy of the $\mathbb{R}$-motivic sphere does not surject onto the homotopy of $\mathbb{R}$-motivic j. In other words, the main rationale for using ksp in the first place does not apply in the motivic situation.

On the other hand, the computation of the homotopy of the $\mathbb{R}$-motivic fiber of $\mathrm{ko} \xrightarrow{\psi^3-1} \mathrm{ko}$ is much cleaner. Moreover, it tells us just as much about v_1-periodic $\mathbb{R}$-motivic homotopy as j. In other words, it has all of the computational advantages of j, while avoiding some unfortunate complications.

Consequently, here we will be solely concerned with the fiber of $\mathrm{ko} \xrightarrow{\psi^3-1} \mathrm{ko}$. We use the notation L for this fiber in order to avoid confusion with the traditional meaning of j. The symbol L is meant to draw a connection to the classical $K(1)$-local sphere $L_{K(1)}S^0$, which is the fiber of $\mathrm{KO}^{\mathrm{top}} \xrightarrow{\psi^3-1} \mathrm{KO}^{\mathrm{top}}$. Our main result is a computation of the homotopy of L.

Theorem 1.1. *The homotopy of the $\mathbb{R}$-motivic spectrum L is depicted in Figures 13–19 via the E_∞-page of the effective spectral sequence, including all hidden extensions by ρ, h, and η.*

The proof of Theorem 1.1 appears in Section 5. See especially Theorem 5.12 and Proposition 5.13.

Beware that the homotopy of the $\mathbb{R}$-motivic spheres does not surject onto the homotopy of $\mathbb{R}$-motivic L. It is possible that we may have not yet constructed the "correct" motivic version of the classical connective spectrum j^{top}. These considerations raise questions about vector bundles and the motivic Adams conjecture. We make no attempt to study these more geometric issues.[1]

[1] After the first version of this article was released, some of these issues have been addressed in [2].

We claim to compute the v_1-periodic $\mathbb{R}$-motivic stable homotopy groups, but this claim deserves some clarification. We do not use an intrinsic definition of v_1-periodic $\mathbb{R}$-motivic homotopy, although such a definition could probably be formulated in terms of the motivic $K(1)$-local sphere. See [7] for some progress on motivic $K(1)$-localization.

Rather, we merely compute the homotopy of L, and we observe that it detects large-scale structure in the stable homotopy of the $\mathbb{R}$-motivic sphere, which was described in a range in [9]. In other words, we have a practical description of $\mathbb{R}$-motivic v_1-periodic homotopy, not a theoretical one.

The careful reader may object that our approach with effective spectral sequences is long-winded and unnecessarily complicated. In fact, the homotopy of L could be determined by direct analysis of the long exact sequence associated to the defining fiber sequence for L. However, there is a disadvantage in this direct approach. We find that the effective filtration is useful additional information about the homotopy of L that helps us understand the computation. The effective filtration is part of the "higher structure" of the homotopy of L. For example, some subtle phenomena, such as hidden multiplicative extensions, can only shift into higher effective filtration, so detailed knowledge of effective filtrations of homotopy classes can rule out possibilities that may otherwise be difficult to analyze. Another example occurs with Toda brackets, which may be computable using effective differentials. While we have no immediate uses for this higher structure, we know from experience that it inevitably becomes important in deeper homotopical analyses.

1A. ***Charts.*** We provide on pages 73–82 charts that display the effective spectral sequences for ko and L, as well as their $\mathbb{C}$-motivic counterparts. We consider these charts to be the central achievement of this article. We encourage the reader to rely heavily on them. In a sense, they provide an illustrated guide to our computations.

Caution must be exercised in the comparison to [9] since the Adams filtrations and effective filtrations are different. As in [9], our charts consider each coweight separately; we have found that this is a practical way of studying $\mathbb{R}$-motivic homotopy groups. Periodicity by τ^4 (which is not a permanent cycle, but should be thought of as a periodicity operator in coweight 4) allows us to give a fairly compact depiction of the homotopy of L in coweights congruent to 0, 1, and 2 modulo 4; see Figures 13, 14, and 15.

The homotopy of L in coweights congruent to 3 modulo 4 is much more interesting but harder to describe. See Figures 17 and 18.

1B. ***Completions.*** We are computing exclusively in the 2-complete context. This simplifies all questions surrounding convergence of spectral sequences. Also, the final computational 2-complete answers are easier to state than their 2-localized or integral counterparts.

We generally omit completions from our notation for brevity. For example, we write $\mathbb{Z}$ for the 2-adic integers, and we write KO for the 2-completed $\mathbb{R}$-motivic Hermitian K-theory spectrum.

Section 2C discusses these topics in slightly more detail.

1C. ***Regarding the element 2.*** When passing from the effective E_∞-page to stable homotopy groups, one must choose homotopy elements that are represented by each element of the E_∞-page. For the element 2 in the E_∞-page, there is more than one choice in $\pi_{0,0}$ because of the presence of elements in the E_∞-page in higher effective filtration.

From the perspective of abelian groups, the element $2 = 1 + 1$ is the obvious choice of homotopy element. However, there is another element h, also detected by 2 in the effective spectral sequence, that turns out to be a much more convenient choice. The difference between h and 2 in homotopy is detected by the element ρh_1 in higher filtration (to be discussed later). Experience has shown that the motivic stable homotopy groups are easier to describe in terms of h than in terms of 2. For example, we have the relations $\mathsf{h}\rho = 0$ and $\mathsf{h}\eta = 0$, where ρ and η are homotopy elements detected by ρ and h_1 respectively. However, neither 2ρ nor 2η are zero. Because of the presence of elements in higher filtration, the homotopy elements ρ and η are not uniquely defined by the effective E_∞-page elements that detect them. However, the mentioned relations hold for all choices. In this discussion, the exact definitions of ρ and η are less important than the observation that they satisfy nicer relations with respect to h than with respect to 2.

There are two additional reasons why the element h plays a central role. First, it corresponds to the hyperbolic plane under the isomorphism between motivic $\pi_{0,0}$ and the Grothendieck–Witt group of symmetric bilinear forms [25]. Second, it plays the role of the zeroth Hopf map, in the sense that the Steenrod operations on its cofiber are simpler than the Steenrod operations for the cofiber of 2.

Consequently, instead of describing motivic stable homotopy groups as a module over the 2-adic integers $\mathbb{Z}$ (i.e., in terms of the action of 2), it is easier to describe the homotopy groups in terms of the action of h.

1D. ***Future directions.*** Our work points toward several open problems.

Problem 1.2. Compute motivic v_1-periodic homotopy over an arbitrary base field. Using [5], one can define L as the fiber of the map $\psi^3 - 1$, and it is conceivable that one could carry out the effective spectral sequence for L in this level of generality, similar to the kind of computations that appear in [27] and [28]. See Section 1E for further discussion. For prime fields of characteristic not two, some explicit computations were carried out in [22].

Problem 1.3. Recompute the homotopy of L using the $\mathbb{R}$-motivic Adams spectral sequence. This would be a useful comparison object for further computations with the Adams spectral sequence for the $\mathbb{R}$-motivic sphere. The classical Adams spectral sequence for j^{top} was studied by Davis [14], but it was only recently computed completely by Bruner and Rognes [12]. We are proposing a motivic analogue of their results.

Problem 1.4. Carry out the effective spectral sequence for the $\mathbb{R}$-motivic sphere in a range. These computations would serve as a useful companion to $\mathbb{R}$-motivic Adams spectral sequence computations [9]. The idea is to build on the techniques that are developed here.

Problem 1.5. Compute the v_1-periodic C_2-equivariant stable homotopy groups. More precisely, carry out the C_2-effective spectral sequence for a C_2-equivariant version of L. The details will be similar to but more complicated than the computations in this article. See [21] for the effective approach to the C_2-equivariant version of ko. Alternatively, one might compute the v_1-periodic C_2-equivariant stable homotopy groups by periodicizing the v_1-periodic $\mathbb{R}$-motivic groups with respect to τ, as considered by Behrens and Shah [8].

Recall that the $\mathbb{R}$-motivic and C_2-equivariant stable homotopy groups are isomorphic in a range [10]. Consequently, we anticipate that some version of the structure described here appears in the C_2-equivariant context as well.

In the equivariant context, we mention Balderrama's [6] computation of the homotopy groups of the Borel C_2-equivariant $K(1)$-local sphere, using techniques that are entirely different from ours. Roughly speaking, Balderrama computes the $\tau^4 v_1^4$-periodicization of our result. The effective E_∞ charts in Figures 13–19 possess an obvious regularity every 8 stems, and Balderrama's computation sees that regular pattern.

Problem 1.6.[2] Study $K(1)$-localization in the motivic context, which ought to be something like localization with respect to KGL/2. Compute $K(1)$-local motivic homotopy. This would provide an intrinsic definition of v_1-periodic homotopy that would improve upon the practical computational perspective of this article.

A guide to the motivic situation could lie in the work of Balderrama [6] and Carrick [13] on equivariant localizations.

1E. *Towards v_1-periodic homotopy over general base fields.* Our explicit computations point the way towards a complete computation of the v_1-periodic motivic stable homotopy groups over arbitrary fields. The situation here is analogous to the η-periodic $\mathbb{R}$-motivic computations of [16], which foreshadowed the more general η-periodic computations of [32], [26], and [5].

[2]After the first version of this article was released, some progress has occurred in [7].

Problem 1.7. Let k be an arbitrary field of characteristic different from 2. Let $GW(k)$ be the Grothendieck–Witt ring of symmetric bilinear forms over k. Describe the 2-primary homotopy groups of the k-motivic spectrum L in terms of the cokernels and kernels of multiplication by various powers of 2 and of h on $GW(k)$.

Problem 1.7 is stated only in terms of 2-primary computations because that is the most interesting part. We expect that the generalization to odd primes is straightforward.

The exact powers of 2 and h that are required in Problem 1.7 depend not only on the coweight but also on the stem. Figures 17 and 18 show that $2^{v(j)+3}$ is the relevant power of 2 in most stems in coweight $4j-1$. Here $v(j)$ is the 2-adic valuation of j, i.e., largest number v such that 2^v divides j. In coweight $4j-1$ and stem $4i-1$, we see larger powers of 2, as well as powers of h.

Similar observations apply to the kernels that contribute to coweight $4i$.

1F. *Outline.* Section 2 contains some background information that we will need to get started on our computations. We briefly discuss convergence of the effective spectral sequences that we will use. We recall some results of Bachmann–Hopkins [5] about motivic Adams operations and of Ananyevskiy–Röndigs–Østvær [1] about the slices of ko.

In Section 2, we have taken some care to eliminate details that we do not use. In other words, Section 2 describes the minimal hypotheses necessary in order to carry out our computations.

Section 3 considers $\mathbb{C}$-motivic computations, which play two roles in our work. First, they serve as a warmup to the more intricate $\mathbb{R}$-motivic computations. Second, the comparison between $\mathbb{R}$-motivic and $\mathbb{C}$-motivic homotopy is a necessary ingredient for our computations. In this section, we describe the effective spectral sequence for $\mathrm{ko}^{\mathbb{C}}$. This material is well-known, since it is the same (up to regrading) as the $\mathbb{C}$-motivic Adams–Novikov spectral sequence for $\mathrm{ko}^{\mathbb{C}}$, which is nearly the same as the classical Adams–Novikov spectral sequence for $\mathrm{ko}^{\mathrm{top}}$. We then use the fiber sequence

$$L^{\mathbb{C}} \longrightarrow \mathrm{ko}^{\mathbb{C}} \xrightarrow{\psi^3-1} \mathrm{ko}^{\mathbb{C}}$$

in order to determine the E_1-page of the effective spectral sequence for $L^{\mathbb{C}}$.

We next completely analyze the effective spectral sequence for the η-periodicization $L^{\mathbb{C}}[\eta^{-1}]$. The η-periodic spectral sequence is significantly simpler than the unperiodicized spectral sequence. We note the close similarity between the homotopy of $L^{\mathbb{C}}[\eta^{-1}]$ and the computations of Andrew–Miller [3].

The η-periodic effective differentials completely determine the unperiodicized effective differentials for $L^{\mathbb{C}}$. Finally, we determine hidden extensions in the effective E_∞-page for $L^{\mathbb{C}}$.

Section 3 completely computes the homotopy of $L^{\mathbb{C}}$, but the effective spectral sequence is not necessarily the simplest way of obtaining the computation. Nevertheless, we have chosen this approach because of its relationship to our later $\mathbb{R}$-motivic computations.

Section 4 analyzes the effective spectral sequence for $\mathbb{R}$-motivic ko, including all differentials and hidden extensions. The E_1-page is readily determined from the work of Ananyevskiy–Röndigs–Østvær [1] on the slices of ko. We draw particular attention to the formula

$$(\tau h_1)^2 = \tau^2 \cdot h_1^2 + \rho^2 \cdot v_1^2. \tag{1-1}$$

This formula has a major impact on the shape of the answers that we obtain. In a sense, our work merely draws algebraic conclusions from (1-1) and η-periodic information. The hidden extensions in the effective E_∞-page for ko are easily determined by comparison to the $\mathbb{C}$-motivic case, using the relationship between $\mathbb{C}$-motivic and $\mathbb{R}$-motivic homotopy that is described in [8, Corollary 1.9].

Our computation of the homotopy of $\mathbb{R}$-motivic ko is not original. See [21] for a C_2-equivariant analogue of the effective spectral sequence for ko. The $\mathbb{R}$-motivic computation can be extracted from the C_2-equivariant computation by dropping the "negative cone" elements. Also, Hill [17] computed the Adams spectral sequence for ko, although the $\mathbb{R}$-motivic spectrum ko had not yet been constructed at the time.

The next step, undertaken in Section 4B, is to analyze the effect of ψ^3 on the effective spectral sequence of ko. This follows from a straightforward comparison to the classical case, together with careful bookkeeping. In turn, this leads to a complete understanding of the effective E_1-page of L, which is described in Section 5A. Again, this is mostly a matter of careful bookkeeping.

Section 5B completely analyzes the effective spectral sequence for η-periodic $L[\eta^{-1}]$. This information is essentially already well-known, either from [16] or from Ormsby–Röndigs [26], although those references do not specifically mention L.

As in the $\mathbb{C}$-motivic situation of Section 3, η-periodic information yields all that we need to know about the unperiodic situation, including all multiplicative relations in the effective E_1-page for L (see Section 5C) and all differentials (see Sections 5D and 5E). We again emphasize the significance of (1-1) in carrying out the details. Finally, Section 5F studies hidden extensions in the effective E_∞-page for L. As for ko, these hidden extensions follow by comparison to the $\mathbb{C}$-motivic case.

1G. *Notation.* We use the following conventions.

- $v(n)$ is the 2-adic valuation of n, i.e., the largest integer v such that 2^v divides n.
- Except in Section 2, everything is implicitly 2-completed. For example, S is actually the 2-complete $\mathbb{R}$-motivic sphere spectrum, and $\mathbb{Z}$ is the 2-adic integers.

- $s_*(X)$ are the slices of a motivic spectrum X.
- $E_r(X)$ is the E_r-page of the effective spectral sequence for a motivic spectrum X.
- We find the effective slice filtration to be slightly inconvenient for our purposes. We prefer to use the "Adams–Novikov filtration", which equals twice the effective filtration minus the stem.
- Coweight equals the stem minus the motivic weight.
- Elements in $E_r(X)$ are tri-graded. We write $E_r^{s,f,w}(X)$ to denote the part with topological dimension s, Adams–Novikov filtration f, and motivic weight w.
- We use unadorned symbols for $\mathbb{R}$-motivic spectra. For example, ko is the very effective cover of the $\mathbb{R}$-motivic Hermitian K-theory spectrum.
- $X^{\mathbb{C}}$ is the $\mathbb{C}$-motivic extension-of-scalars spectrum of an $\mathbb{R}$-motivic spectrum X.
- X^{top} is the Betti realization of an $\mathbb{R}$-motivic spectrum X.
- S is the $\mathbb{R}$-motivic sphere spectrum.
- KO is the $\mathbb{R}$-motivic spectrum that represents Hermitian K-theory (also known as KQ).
- ko is the very effective connective cover of KO.
- HA is the $\mathbb{R}$-motivic Eilenberg–Mac Lane spectrum on the group A.
- ψ^3 is an Adams operation. We use the same symbol in the $\mathbb{R}$-motivic, $\mathbb{C}$-motivic, and classical situations.
- L is the fiber of $\text{ko} \xrightarrow{\psi^3-1} \text{ko}$.
- $\Sigma^{s,w}X$ is a (bigraded) suspension of a motivic spectrum X.
- $\pi_{*,*}(X)$ are the bigraded stable homotopy groups of an $\mathbb{R}$-motivic or $\mathbb{C}$-motivic spectrum.
- Recall that ϵ is the motivic homotopy class that is represented by the twist map $S \wedge S \to S \wedge S$, where S is the motivic sphere spectrum. Let h be the element $1-\epsilon$, which corresponds to the hyperbolic plane under the isomorphism between $\pi_{0,0}(S)$ and the Grothendieck-Witt ring $GW(\mathbb{R})$ [25].
- The element ρ belongs to the $\mathbb{R}$-motivic homology of a point. It is the class represented by -1 in the Milnor K-theory of $\mathbb{R}$. Since ρ survives all of the spectral sequences under consideration, we use the same symbol for the corresponding homotopy class. However, there is a choice of homotopy class represented by ρ because of the presence of elements in higher filtration. There is an inconsistency in the literature about this choice. Following [4], we define ρ such that $\epsilon = \rho\eta - 1$, or equivalently $2 = \rho\eta + \mathsf{h}$.

We frequently use names for indecomposables that consist of more than one symbol. For example, Theorem 2.1 discusses the indecomposable element v_1^2 of the effective E_1-page for $\mathrm{ko}^{\mathbb{C}}$. These longer names are slightly more cumbersome. This is especially the case when we consider products. We will use expressions of the form $x \cdot y$ for clarity.

On the other hand, our names are particularly convenient because they reflect the origins of the elements in terms of the spectral sequences that we use. For example, consider the indecomposable element $2v_1^2$ of the effective E_∞-page for $\mathrm{ko}^{\mathbb{C}}$, as discussed in Theorem 3.3 (see also Figure 2). This name reflects the element's origin in the effective E_1-page. It also illuminates relations such as

$$2v_1^2 \cdot 2v_1^2 = 4 \cdot v_1^4$$

However, one must be careful about possible error terms in such formulas; see especially (1-1).

2. Background

In this section only, we write ko for the integral version of the very effective cover of the Hermitian K-theory spectrum, and we use the usual decorations to indicate localizations and completions of ko. After that, ko is assumed to be 2-completed.

2A. ***The effective slices of* ko.** We recall the structure of the effective slices of ko.

Theorem 2.1 [1, Theorem 17]. *The slices of* ko *are*

$$s_*(\mathrm{ko}) = H\mathbb{Z}[h_1, v_1^2]/(2h_1),$$

where v_1^2 *and* h_1 *have degrees* $(4, 0, 2)$ *and* $(1, 1, 1)$ *respectively.*

We explain the expression in Theorem 2.1. Each monomial of degree (s, f, w) contributes a summand of $\Sigma^{s,w} HA$ in the $\left(\frac{s+f}{2}\right)$-th slice. Here HA is the motivic Eilenberg–Mac Lane spectrum associated to A. The abelian group A is $\mathbb{F}_2$ when the monomial is 2-torsion, and is $\mathbb{Z}$ when the monomial is torsion free. We list the first three slices as examples:

$$\begin{aligned} s_0(\mathrm{ko}) &= H\mathbb{Z}\{1\}, \\ s_1(\mathrm{ko}) &= \Sigma^{1,1} H\mathbb{F}_2\{h_1\}, \\ s_2(\mathrm{ko}) &= \Sigma^{2,2} H\mathbb{F}_2\{h_1^2\} \vee \Sigma^{4,2} H\mathbb{Z}\{v_1^2\}. \end{aligned}$$

Beware that the multiplicative structure of $s_*(\mathrm{ko})$ is not completely captured by the notation in Theorem 2.1. The essential multiplicative relation is (1-1), which follows immediately from the general formulas in [1].

Remark 2.2. The calculation of the slices of the motivic sphere spectrum, due to Röndigs, Spitzweck, and Østvær [27], is commonly expressed at the prime 2 as

$$s_*(S) = H\mathbb{Z} \otimes \mathrm{Ext}^{*,*}_{BP_*BP}(BP_*, BP_*).$$

Analogously, Theorem 2.1 says that

$$s_*(\mathrm{ko}) = H\mathbb{Z} \otimes \mathrm{Ext}^{*,*}_{BP_*BP}(BP_*, BP_*(\mathrm{ko}^{\mathrm{top}})).$$

However, we do not know of a general theorem relating the slices of a motivic spectrum with the Adams–Novikov E_2-page for its topological counterpart.

2B. ***The Adams operation ψ^3 and the spectrum L.*** Bachmann and Hopkins [5] constructed a motivic analogue of the classical Adams operation ψ^3. We summarize the results that we need.

Theorem 2.3 [5]. *There is a unital ring map $\psi^3 : \mathrm{ko}\left[\frac{1}{3}\right] \to \mathrm{ko}\left[\frac{1}{3}\right]$ whose Betti realization is the classical Adams operation ψ^3.*

Proof. There is a unital ring map $\psi^3 : \mathrm{KO}\left[\frac{1}{3}\right] \to \mathrm{KO}\left[\frac{1}{3}\right]$ [5, Theorem 3.1], which is an E_∞-map. Its Betti realization is also an E_∞-map whose action on the classical Bott element is multiplication by 81. These properties uniquely characterize the classical Adams operation.

Now apply very effective covers, and the result about ko follows formally. □

The original result is more general in more than one sense. First, it works over general base schemes in which 2 is invertible, while we only use the construction over $\mathbb{R}$. Second, its values are computed more precisely than just compatibility with the classical values.

Corollary 2.4.

(1) *$\psi^3 : \pi_{*,*}(\mathrm{ko}_2^\wedge) \to \pi_{*,*}(\mathrm{ko}_2^\wedge)$ is a ring map.*

(2) *If x is in the image of the unit map $\pi_{*,*}(S_2^\wedge) \to \pi_{*,*}(\mathrm{ko}_2^\wedge)$, then $\psi^3(x) = x$.*

(3) *There is a commutative diagram*

$$\begin{array}{ccc} \pi_{*,*}(\mathrm{ko}_2^\wedge) & \xrightarrow{\psi^3} & \pi_{*,*}(\mathrm{ko}_2^\wedge) \\ \downarrow & & \downarrow \\ \pi_*((\mathrm{ko}^{\mathrm{top}})_2^\wedge) & \xrightarrow[\psi^3]{} & \pi_*((\mathrm{ko}^{\mathrm{top}})_2^\wedge), \end{array}$$

where the vertical maps are Betti realization homomorphisms.

Proof. These are computational consequences of Theorem 2.3. Part (1) follows from the fact that ψ^3 is a ring map. Part (2) follows from the fact that ψ^3 is unital. Part

(3) follows from the fact that the Betti realization of the motivic Adams operation is the classical Adams operation. □

Remark 2.5. Corollary 2.4 can also be stated in a localized sense rather than completed sense, but we will not need that.

Definition 2.6. Let L be the fiber of the map $\mathrm{ko}\left[\frac{1}{3}\right] \xrightarrow{\psi^3-1} \mathrm{ko}\left[\frac{1}{3}\right]$.

Note that our definition of L is already localized; we do not consider an integral version. Except for this section, L is assumed to be 2-completed.

The most important point for us is that there is a fiber sequence

$$L^{\wedge}_2 \longrightarrow \mathrm{ko}^{\wedge}_2 \xrightarrow{\psi^3-1} \mathrm{ko}^{\wedge}_2$$

of completed spectra since completion preserves fiber sequences.

2C. *Convergence of the effective spectral sequence.* The *effective spectral sequence* for a motivic spectrum X denotes the spectral sequence associated to the effective slice filtration of X. We refer to [23; 27] for details on the construction and properties of this spectral sequence.

The effective slice filtration [31] has truncations $f^q(X)$ and quotients (i.e., slices) $s_q(X)$. The E_1-page of the effective spectral sequence is $\pi_{*,*}(s_*(X))$. In good cases, it converges to the homotopy groups of a completion of X. We also use the very effective slice filtration [30], but only to define ko.

The slice functors do not necessarily commute with completions, i.e., $s_*(X)^{\wedge}_2$ and $s_*(X^{\wedge}_2)$ are not always equivalent. Consequently, we must carefully define the spectral sequences that we use to study completed spectra. On the other hand, the effective slices do interact nicely with localizations [29, Corollary 4.6].

Theorem 2.7. *There are strongly convergent spectral sequences*

$$E_1^{s,f,w}(\mathrm{ko}) = \pi_{s,w}\big(s_{\frac{s+f}{2}}(\mathrm{ko})^{\wedge}_2\big) \Longrightarrow \pi_{s,w}(\mathrm{ko}^{\wedge}_2)$$

and

$$E_1^{s,f,w}(L) = \pi_{s,w}\big(s_{\frac{s+f}{2}}(L)^{\wedge}_2\big) \Longrightarrow \pi_{s,w}(L^{\wedge}_2),$$

with differentials $d_r : E_r^{s,f,w} \to E_r^{s-1,f+2r+1,w}$.

We remind the reader that our grading of the effective spectral sequence is different than the standard grading in the literature. Briefly, s represents the topological stem, f represents the Adams–Novikov filtration (not the effective filtration), and w represents the motivic weight. See Section 1G for more discussion.

Proof. We discuss the spectral sequence for ko in detail; most of the argument for L is the same.

Consider the effective slice tower

$$f^0(\mathrm{ko}) \leftarrow f^1(\mathrm{ko}) \leftarrow f^2(\mathrm{ko}) \leftarrow \cdots.$$

Now take the 2-completion of this tower to obtain

$$f^0(\mathrm{ko})^\wedge_2 \leftarrow f^1(\mathrm{ko})^\wedge_2 \leftarrow f^2(\mathrm{ko})^\wedge_2 \leftarrow \cdots.$$

The resulting layers are the same as $s_*(\mathrm{ko})^\wedge_2$ since completion respects cofiber sequences. Beware that this is not necessarily the same as the slice tower of the completion $\mathrm{ko}^\wedge_2$, since slices do not interact nicely with completions. The associated spectral sequence of this tower is the one described in the statement of the theorem.

It remains to determine the target of the completed spectral sequence. The limit of the uncompleted slice tower of ko is equivalent to its η-completion [27], [1], i.e.,

$$\operatorname{holim} f^n(\mathrm{ko}) \simeq \mathrm{ko}^\wedge_\eta.$$

Completion respects limits, so the limit $\operatorname{holim}(f^n(\mathrm{ko})^\wedge_2)$ of the completed slice tower is equivalent to $(\mathrm{ko}^\wedge_\eta)^\wedge_2$, which is equivalent to $\mathrm{ko}^\wedge_2$ by [18, Theorem 1]. Consequently, the completed effective spectral sequence of ko converges to the homotopy of $\mathrm{ko}^\wedge_2$, as desired.

Strong convergence follows from [11, Theorem 7.1], which has a technical hypothesis involving derived E_∞-pages. For ko, this technical hypothesis follows directly from the computations of Section 4. For L, the technical hypothesis follows directly from the computations in Sections 5D and 5E. □

Remark 2.8. By construction, we have a fiber sequence

$$s_*(L)^\wedge_2 \longrightarrow s_*(\mathrm{ko})^\wedge_2 \xrightarrow{\psi^3-1} s_*(\mathrm{ko})^\wedge_2,$$

which yields a long exact sequence

$$\cdots \longrightarrow E_1^{s,f,w}(L) \longrightarrow E_1^{s,f,w}(\mathrm{ko}) \xrightarrow{\psi^3-1} E_1^{s,f,w}(\mathrm{ko}) \longrightarrow \cdots.$$

This long exact sequence will be our main tool for computing $E_1(L)$ in Section 5A.

3. $\mathbb{C}$-motivic computations

In this section, we carry out a preliminary computation of the effective spectral sequences for $\mathrm{ko}^{\mathbb{C}}$ and $L^{\mathbb{C}}$. We also consider the η-periodic spectral sequences. We are primarily interested in $\mathbb{R}$-motivic computations, but we will need to compare our $\mathbb{R}$-motivic computations to their $\mathbb{C}$-motivic counterparts.

3A. ***The effective spectral sequence for*** $\mathbf{ko}^{\mathbb{C}}$**.** We review the effective spectral sequence for $\mathrm{ko}^{\mathbb{C}}$.

coweight	(s, f, w)	x	$d_1(x)$	$\psi^3(x)$
0	$(1, 1, 1)$	h_1		h_1
1	$(0, 0, -1)$	τ		τ
2	$(4, 0, 2)$	v_1^2	τh_1^3	$9v_1^2$

Table 1. Multiplicative generators for $E_1(\mathrm{ko}^{\mathbb{C}})$.

Proposition 3.1. *The effective spectral sequence for* $\mathrm{ko}^{\mathbb{C}}$ *takes the form*

$$E_1(\mathrm{ko}^{\mathbb{C}}) = \mathbb{Z}[\tau, h_1, v_1^2]/2h_1.$$

Proof. This follows from Theorem 2.1 by taking stable homotopy groups. There are no possible error terms to complicate the multiplicative structure. □

Table 1 lists the generators of $E_1(\mathrm{ko}^{\mathbb{C}})$. Figure 1 depicts $E_1(\mathrm{ko}^{\mathbb{C}})$ graphically.

Proposition 3.2. *Table 1 gives the values of the effective d_1 differential on the multiplicative generators of $E_1(\mathrm{ko}^{\mathbb{C}})$.*

Proof. The $\mathbb{C}$-motivic effective spectral sequence is identical to the $\mathbb{C}$-motivic Adams–Novikov spectral sequence up to reindexing. This claim does not appear to be cleanly stated in the literature, but it is a computational consequence of the weight 0 result of [24, Theorem 1]. Alternatively, there is only one pattern of effective differentials that computes the motivic stable homotopy groups of $\mathrm{ko}^{\mathbb{C}}$, which were previously described using the $\mathbb{C}$-motivic Adams spectral sequence [20]. □

Theorem 3.3. *The E_∞-page of the effective spectral sequence for* $\mathrm{ko}^{\mathbb{C}}$ *takes the form*

$$E_\infty(\mathrm{ko}^{\mathbb{C}}) = \frac{\mathbb{Z}[\tau, h_1, 2v_1^2, v_1^4]}{2h_1, \tau h_1^3, (2v_1^2)^2 = 4 \cdot v_1^4}.$$

Proof. For degree reasons, there can be no higher differentials in the effective spectral sequence for $\mathrm{ko}^{\mathbb{C}}$. □

Table 2 lists the multiplicative generators of $E_\infty(\mathrm{ko}^{\mathbb{C}})$. Figure 2 depicts $E_\infty(\mathrm{ko}^{\mathbb{C}})$ in graphical form.

Remark 3.4. There are no possible hidden extensions in $E_\infty(\mathrm{ko}^{\mathbb{C}})$ for degree reasons. Therefore, Theorem 3.3 describes $\pi_{*,*}(\mathrm{ko}^{\mathbb{C}})$ as a ring.

3B. *The effective E_1-page for $L^{\mathbb{C}}$.* Our next goal is to describe the effective E_1-page $E_1(L^{\mathbb{C}})$. First we must study the values of ψ^3 on $\mathrm{ko}^{\mathbb{C}}$.

Lemma 3.5. *The map $E_\infty(\mathrm{ko}^{\mathbb{C}}) \to E_\infty(\mathrm{ko}^{\mathbb{C}})$ induced by ψ^3 on effective E_∞-pages takes the values shown in Table 2.*

coweight	(s, f, w)	x	$\psi^3(x)$
0	$(1, 1, 1)$	h_1	h_1
1	$(0, 0, -1)$	τ	τ
2	$(4, 0, 2)$	$2v_1^2$	$9\cdot 2v_1^2$
4	$(8, 0, 4)$	v_1^4	$81v_1^4$

Table 2. Multiplicative generators for $E_\infty(\mathrm{ko}^{\mathbb{C}})$.

Proof. All values follow immediately by comparison along Betti realization to the values of classical ψ^3. □

Lemma 3.6. *The map $E_1(\mathrm{ko}^{\mathbb{C}}) \to E_1(\mathrm{ko}^{\mathbb{C}})$ induced by ψ^3 on effective E_1-pages takes the values shown in Table 1.*

Proof. The values of ψ^3 on $E_1(\mathrm{ko}^{\mathbb{C}})$ are compatible with the values of ψ^3 on $E_\infty(\mathrm{ko}^{\mathbb{C}})$, as shown in Table 2 (see also Lemma 3.5). This immediately yields all values. □

In order to describe $E_1(L^{\mathbb{C}})$, we need some elementary number theory.

Definition 3.7. Let $v(n)$ be the 2-adic valuation of n, i.e., the exponent of the largest power of 2 that divides n.

Lemma 3.8.
$$v(3^n - 1) = \begin{cases} 1 & \text{if} \quad v(n) = 0, \\ 2 + v(n) & \text{if} \quad v(n) > 0, \end{cases}$$

Proof. Let $n = 2^a \cdot b$, where b is an odd number, so $v(n) = a$. Then

$$3^n - 1 = \big(1 + 3^{2^a} + (3^{2^a})^2 + \cdots + (3^{2^a})^{b-1}\big)(3-1)\prod_{i=0}^{a-1}(1 + 3^{2^i}).$$

The first factor is odd, so it does not contribute to the 2-adic valuation. The factor $(1 + 3^{2^i})$ has valuation 1 if $i > 0$, and it has valuation 2 if $i = 0$. □

Proposition 3.9. *The chart in Figure 3 depicts the effective E_1-page of $L^{\mathbb{C}}$.*

Proof. The long exact sequence

$$\cdots \longrightarrow E_1(L^{\mathbb{C}}) \longrightarrow E_1(\mathrm{ko}^{\mathbb{C}}) \xrightarrow{\psi^3 - 1} E_1(\mathrm{ko}^{\mathbb{C}}) \longrightarrow \cdots$$

induces a short exact sequence

$$0 \longrightarrow \Sigma^{-1}C \longrightarrow E_1(L^{\mathbb{C}}) \longrightarrow K \longrightarrow 0,$$

where C and K are the cokernel and kernel of $E_1(\mathrm{ko}^{\mathbb{C}}) \xrightarrow{\psi^3 - 1} E_1(\mathrm{ko}^{\mathbb{C}})$ respectively. The cokernel and kernel can be computed directly from the information given in Table 1 (see also Lemma 3.6).

The kernel is additively generated by all multiples of h_1 in $E_1(\mathrm{ko}^{\mathbb{C}})$, together

coweight	(s, f, w)	generator
1	$(0, 0, -1)$	τ
$2k$	$(4k+1, 1, 2k+1)$	$h_1 v_1^{2k}$
$2k-1$	$(4k-1, 1, 2k)$	ιv_1^{2k}

Table 3. Multiplicative generators for $E_1(L^{\mathbb{C}})$: $k \geq 0$.

with the elements τ^k for $k \geq 0$.

The cokernel C is nearly the same as $E_1(\mathrm{ko}^{\mathbb{C}})$ itself. We must impose the relations $(3^{2k} - 1)v_1^{2k} = 0$ for all $k > 0$. Lemma 3.8 says that $3^{2k} - 1$ equals $2^{\nu(2k)+2} \cdot u$, where u is an odd number, i.e., a unit in our 2-adic context. Therefore, the relation $(3^{2k} - 1)v_1^{2k} = 0$ is equivalent to the relation $2^{\nu(2k)+2} v_1^{2k} = 0$. □

Table 3 lists some elements of the effective E_1-page of $L^{\mathbb{C}}$. In fact, these elements are multiplicative generators for $E_1(L^{\mathbb{C}})$. By inspection, all elements of $E_1(L^{\mathbb{C}})$ are of the form $\tau^a h_1^b x$, for some x in the table.

We use the same notation for elements of $E_1(L^{\mathbb{C}})$ and their images in $E_1(\mathrm{ko}^{\mathbb{C}})$. On the other hand, we define the elements ιx of $E_1(L^{\mathbb{C}})$ by the property that they are the image of x under the map $\iota : \Sigma^{-1} E_1(\mathrm{ko}) \to E_1(L)$. For example, the element 1 of $E_1(\mathrm{ko})$ maps to ι.

Remark 3.10. Our choice of notation for elements of $E_1(L^{\mathbb{C}})$ is helpful for the particular analysis at hand. The generators of $E_1(L^{\mathbb{C}})$ also have traditional names from the perspective of the Adams–Novikov spectral sequence. Namely, $h_1 v_1^{2k}$ and ιv_1^{2k} correspond to α_{2k+1} and $\alpha_{2k/\nu(8k)}$ respectively. However, the α-family perspective is not so helpful for us.

3C. ***The effective spectral sequence of $L^{\mathbb{C}}[\eta^{-1}]$.*** Next, we describe the effective spectral sequence of $L^{\mathbb{C}}[\eta^{-1}]$.

In the η-periodic context, the element h_1 is a unit, so its powers are inconsequential for computational purposes, and have been removed from all η-periodic formulas. The appropriate powers of h_1 can be easily reconstructed from the degrees of elements (although this reconstruction is typically not necessary).

Proposition 3.11. *The effective E_1-page for $L^{\mathbb{C}}[\eta^{-1}]$ is given by*

$$E_1(L^{\mathbb{C}}[\eta^{-1}]) = \mathbb{F}_2[h_1^{\pm 1}, \tau, v_1^2, \iota]/\iota^2.$$

Proof. The functors s_* commute with homotopy colimits [29, Corollary 4.6]. Therefore, we can just invert h_1 in $E_1(\mathrm{ko}^{\mathbb{C}})$ to obtain

$$E_1(\mathrm{ko}^{\mathbb{C}}[\eta^{-1}]) = \mathbb{F}_2[h_1^{\pm 1}, \tau, v_1^2].$$

See Proposition 3.1 (and Figure 1) for the description of $E_1(\mathrm{ko}^{\mathbb{C}})$.

The map $E_1(\mathrm{ko}^{\mathbb{C}}[\eta^{-1}]) \xrightarrow{\psi^3-1} E_1(\mathrm{ko}^{\mathbb{C}}[\eta^{-1}])$ is trivial because $(\psi^3-1)(v_1^{2k})$ is a multiple of 2, as shown in Table 1 (see also Lemma 3.6). Therefore, the long exact sequence

$$\cdots \longrightarrow E_1(L^{\mathbb{C}}[\eta^{-1}]) \longrightarrow E_1(\mathrm{ko}^{\mathbb{C}}[\eta^{-1}]) \xrightarrow{\psi^3-1} E_1(\mathrm{ko}^{\mathbb{C}}[\eta^{-1}]) \longrightarrow \cdots$$

implies that $E_1(L^{\mathbb{C}}[\eta^{-1}])$ splits as

$$E_1(\mathrm{ko}^{\mathbb{C}}[\eta^{-1}]) \oplus \Sigma^{-1} E_1(\mathrm{ko}^{\mathbb{C}}[\eta^{-1}]).$$

This establishes the additive structure of $E_1(L[\eta^{-1}])$, as well as most of the multiplicative structure.

The relation $\iota^2 = 0$ is immediate because no nonzero values for ι^2 are possible. □

Proposition 3.12. *In the effective spectral sequence for $L^{\mathbb{C}}[\eta^{-1}]$, we have $d_1(v_1^2) = \tau$. The effective differentials are zero on all other multiplicative generators on all pages.*

Proof. The value of $d_1(v_1^2)$ in $E_1(L^{\mathbb{C}}[\eta^{-1}])$ follows by comparison of effective spectral sequences along the maps $L^{\mathbb{C}} \to L^{\mathbb{C}}[\eta^{-1}]$ and $L^{\mathbb{C}} \to \mathrm{ko}^{\mathbb{C}}$. Table 1 (see also Proposition 3.2) gives the value of $d_1(v_1^2)$ in $E_1(\mathrm{ko}^{\mathbb{C}})$. □

Remark 3.13. The effective spectral sequence for $L^{\mathbb{C}}[\eta^{-1}]$ is very close to the effective spectral sequence for the η-periodic sphere $S^{\mathbb{C}}[\eta^{-1}]$. The effective spectral sequence for $S^{\mathbb{C}}[\eta^{-1}]$ is the same (up to reindexing) as the motivic Adams–Novikov spectral sequence for $S^{\mathbb{C}}[\eta^{-1}]$. This motivic Adams–Novikov spectral sequence is analyzed in [3]. The element ι is not present in $E_1(S^{\mathbb{C}}[\eta^{-1}])$, but its multiples $\iota(v_1^2)^k$ are present.

3D. *Effective differentials for $L^{\mathbb{C}}$.*

Proposition 3.14. *Table 4 gives the values of the effective d_1 differentials on the multiplicative generators of $E_1(L^{\mathbb{C}})$. There are no higher differentials in the effective spectral sequence for $L^{\mathbb{C}}$.*

coweight	(s, f, w)	x	$d_1(x)$
1	$(0, 0, -1)$	τ	
$4k$	$(8k+1, 1, 4k+1)$	$h_1 v_1^{4k}$	
$4k+2$	$(8k+5, 1, 4k+3)$	$h_1 v_1^{4k+2}$	$\tau h_1^3 \cdot h_1 v_1^{4k}$
$4k-1$	$(8k-1, 1, 4k)$	ιv_1^{4k}	
$4k+1$	$(8k+3, 1, 4k+2)$	ιv_1^{4k+2}	$\tau h_1^3 \cdot \iota v_1^{4k}$

Table 4. Effective d_1 differentials for $L^{\mathbb{C}}$: $k \geq 0$.

Proof. All of these differentials follow immediately from the effective d_1 differentials for $L^{\mathbb{C}}[\eta^{-1}]$, which are determined by Proposition 3.12.

For degree reasons, there are no possible higher differentials. □

Theorem 3.15. *The E_∞-page of the effective spectral sequence for $L^{\mathbb{C}}$ is depicted in Figure 4.*

Proof. Because there are no higher effective differentials for $L^{\mathbb{C}}$, we obtain the effective E_∞-page immediately from the effective d_1 differentials in Table 4 (see also Proposition 3.14). □

3E. *Hidden extensions in $E_\infty(L^{\mathbb{C}})$.*

Proposition 3.16. *In the effective spectral sequence for $L^{\mathbb{C}}$, the elements $h_1 v_1^{4k}$ do not support hidden* h *extensions for all $k \geq 0$.*

Proof. The elements $h_1 v_1^{4k}$ detect elements in $\pi_{*,*}L^{\mathbb{C}}$ that are in the image of the homotopy $\pi_{*,*}S^{\mathbb{C}}$ of the $\mathbb{C}$-motivic sphere. In the $\mathbb{C}$-motivic sphere, these v_1-periodic elements are annihilated by h. □

Remark 3.17. The proof of Proposition 3.16 appeals to knowledge of the homotopy of the $\mathbb{C}$-motivic sphere. In fact, one can avoid this by use of Toda brackets in the homotopy of $L^{\mathbb{C}}$. Namely, in the homotopy of $L^{\mathbb{C}}$, the E_∞-page element $h_1 v_1^{4k+4}$ detects an element in the bracket $\langle \mathsf{h}^3\sigma, \mathsf{h}, \alpha \rangle$, where α is detected by $h_1 v_1^{4k}$ and σ is detected by ιv_1^4. By induction,

$$\langle \mathsf{h}^3\sigma, \mathsf{h}, \alpha \rangle \mathsf{h} = \mathsf{h}^3 \cdot \sigma \langle \mathsf{h}, \alpha, \mathsf{h} \rangle = \mathsf{h}^3 \cdot \sigma \cdot \tau\eta \cdot \alpha = 0.$$

Proposition 3.18. *In the effective spectral sequence for $L^{\mathbb{C}}$, there are hidden* h *extensions from $\iota 4 v_1^{4k+2}$ to $\tau h_1^2 \cdot h_1 v_1^{4k}$ for all $k \geq 0$.*

Proof. Recall that $\tau\eta^2 = \langle \mathsf{h}, \eta, \mathsf{h} \rangle$ in the homotopy of the $\mathbb{C}$-motivic sphere [19, Table 7.23]. If α is a homotopy element of $L^{\mathbb{C}}$ such that $\mathsf{h}\alpha$ is zero, then

$$\alpha \cdot \tau\eta^2 = \alpha \langle \mathsf{h}, \eta, \mathsf{h} \rangle = \langle \alpha, \mathsf{h}, \eta \rangle \mathsf{h}.$$

In particular, let α be detected by $h_1 v_1^{4k}$. Note that $\mathsf{h}\alpha = 0$ by Proposition 3.16. Then $\tau h_1^2 \cdot h_1 v_1^{4k}$ detects a homotopy element that is divisible by h, so $\tau h_1^2 \cdot h_1 v_1^{4k}$ must be the target of a hidden h extension. There is only one possible source for this extension. □

4. The effective spectral sequence for ko

We now study the effective spectral sequence for $\mathbb{R}$-motivic ko.

Proposition 4.1. *The effective spectral sequence for* ko *takes the form*

$$E_1(\mathrm{ko}) = \frac{\mathbb{Z}[\rho, \tau^2, h_1, \tau h_1, v_1^2]}{2\rho, 2h_1, 2\cdot\tau h_1, (\tau h_1)^2 = \tau^2 \cdot h_1^2 + \rho^2 \cdot v_1^2}.$$

coweight	(s, f, w)	x	$d_1(x)$	$\psi^3(x)$	image in $E_1(\mathrm{ko}[\eta^{-1}])$
0	$(-1, 1, -1)$	ρ		ρ	ρ
0	$(1, 1, 1)$	h_1		h_1	1
1	$(1, 1, 0)$	τh_1		τh_1	$\tau \cdot h_1$
2	$(0, 0, -2)$	τ^2	$\rho^2 \cdot \tau h_1$	τ^2	$\tau^2 + \rho^2 \cdot v_1^2 \cdot h_1^{-2}$
2	$(4, 0, 2)$	v_1^2	$\tau h_1 \cdot h_1^2$	$9 v_1^2$	v_1^2

Table 5. Multiplicative generators for $E_1(\mathrm{ko})$.

Proof. The additive structure follows from Theorem 2.1 by taking stable homotopy groups. We need that the homotopy groups of $\mathbb{R}$-motivic $H\mathbb{Z}$ are

$$H\mathbb{Z}_{*,*} = \mathbb{Z}[\tau^2, \rho]/2\rho,$$

and the homotopy groups of $\mathbb{R}$-motivic $H\mathbb{F}_2$ are

$$(H\mathbb{F}_2)_{*,*} = \mathbb{F}_2[\tau, \rho].$$

The multiplicative structure is mostly also immediate from Theorem 2.1. As explained in [21], our formula for $(\tau h_1)^2$ is equivalent to the formula $\eta^2 \xrightarrow{\delta} \sqrt{\alpha}$ given in [1, p. 1029]. □

Table 5 lists the generators of $E_1(\mathrm{ko})$. Figure 5 depicts $E_1(\mathrm{ko})$ graphically.

Proposition 4.2. *Table 5 gives the values of the effective d_1 differential on the multiplicative generators of $E_1(\mathrm{ko})$.*

Proof. The value of $d_1(\tau^2)$ follows from [1, Theorem 20] and $\mathbb{R}$-motivic Steenrod algebra actions. Then the value of $d_1(v_1^2)$ follows from (1-1).

Alternatively, there is only one pattern of effective differentials that computes the motivic stable homotopy groups of ko, which were previously computed with the $\mathbb{R}$-motivic Adams spectral sequence [17]. □

The entire d_1 differential in the effective spectral sequence for ko can easily be deduced from Proposition 4.2 and the Leibniz rule.

Theorem 4.3. *The E_∞-page of the effective spectral sequence for* ko *is depicted in Figures 6, 7, and 8.*

Proof. The Leibniz rule, together with the values in Table 5 (see also Proposition 4.2), completely determines the effective d_1 differential on $E_1(\mathrm{ko})$. The E_2-page can then be determined directly. However, the computation is not entirely straightforward. Of particular note is the differential

$$d_1(\tau^2 \cdot \tau h_1 \cdot v_1^2) = \tau^4 \cdot h_1^4 + \rho^4 \cdot v_1^4,$$

which yields the relation

$$\tau^4 \cdot h_1^4 = \rho^4 \cdot v_1^4 \tag{4-1}$$

in $E_2(\text{ko})$.

For degree reasons, there can be no higher differentials in the effective spectral sequence for ko. □

For legibility, Figures 6, 7, and 8 display $E_\infty(\text{ko})$ in three different charts separated by coweight modulo 4. There is no chart for coweights 3 mod 4 because $E_\infty(\text{ko})$ is zero in those coweights.

Figure 9 illustrates part of the analysis of the d_1 differentials and the determination of $E_2(\text{ko})$; it is meant to be representative, not thorough. The chart shows some of the elements in coweights 1 and 2 mod 4, together with the d_1 differentials that relate these elements. In this chart, one can see that $\tau^2 \cdot h_1^2 + \rho^2 \cdot v_1^2$ survives to $E_2(\text{ko})$. This element survives to $E_\infty(\text{ko})$. It is labeled $(\tau h_1)^2$ in Figure 8, in accordance with (1-1).

Remark 4.4. There is an alternative, slightly more structured, method for obtaining $E_\infty(\text{ko})$. One can filter $E_1(\text{ko})$ by powers of τh_1 and obtain a spectral sequence that converges to $E_2(\text{ko})$. In this spectral sequence, we have the relation $\tau^2 \cdot h_1^2 = \rho^2 \cdot v_1^2$. There are differentials $d_1(\tau^2) = \rho^2 \cdot \tau h_1$ and $d_1(v_1^2) = h_1^2 \cdot \tau h_1$. Then there is a higher differential $d_3(\tau^2 \cdot v_1^2) = (\tau h_1)^3$. None of this is essential to our study, but the interested reader may wish to carry out the details.

Table 6 lists the multiplicative generators of $E_\infty(\text{ko})$. It is possible to give a complete list of relations. However, the long list is not so helpful for understanding the structure of $E_\infty(\text{ko})$. The charts in Figures 6, 7, and 8 are more useful for this purpose.

coweight	(s, f, w)	x	$\psi^3(x)$
0	$(-1, 1, -1)$	ρ	ρ
0	$(1, 1, 1)$	h_1	h_1
1	$(1, 1, 0)$	τh_1	τh_1
2	$(0, 0, -2)$	$2\tau^2$	$2\tau^2$
2	$(4, 0, 2)$	$2v_1^2$	$9 \cdot 2v_1^2$
4	$(0, 0, -4)$	τ^4	τ^4
4	$(4, 0, 0)$	$2\tau^2 v_1^2$	$9 \cdot 2\tau^2 v_1^2$
4	$(8, 0, 4)$	v_1^4	$81 v_1^4$

Table 6. Multiplicative generators for $E_\infty(\text{ko})$.

coweight	source	type	target	(s, f, w)
2	$2v_1^2$	ρ	$(\tau h_1)^2 h_1$	$(3, 3, 1)$
4	$2\tau^2 v_1^2$	ρ	$\tau^4 \cdot h_1^3$	$(3, 3, -1)$
4	$2\tau^2 v_1^2$	η	$\rho^3 \cdot v_1^4$	$(5, 3, 1)$
2	$2\tau^2$	η	$\rho(\tau h_1)^2$	$(1, 3, -1)$
1	τh_1	h	$\rho \cdot \tau h_1 \cdot h_1$	$(1, 3, 0)$
2	$(\tau h_1)^2$	h	$\rho(\tau h_1)^2 h_1$	$(2, 4, 0)$

Table 7. Hidden extensions in $E_\infty(\text{ko})$.

Proposition 4.5. *Table 7 lists some hidden extensions by ρ, h, and η in the effective spectral sequence for* ko. *All other hidden extensions by ρ, h, and η are v_1^4-multiples and τ^4-multiples of these.*

Proof. Recall from [8, Corollary 1.9] that the homotopy of ko/ρ is isomorphic to the homotopy of $\text{ko}^{\mathbb{C}}$. Therefore, we completely understand the homotopy of ko/ρ from Theorem 3.3 and Figure 2.

The hidden ρ extensions follow from inspection of the long exact sequence associated to the cofiber sequence

$$\Sigma^{-1,-1}\text{ko} \xrightarrow{\rho} \text{ko} \longrightarrow \text{ko}/\rho.$$

The map $\text{ko} \to \text{ko}/\rho$ takes the elements $\tau^4 \cdot h_1^3$ and $(\tau h_1)^2 h_1$ to zero because there are no possible targets in the homotopy of ko/ρ. Therefore, those two elements must receive hidden ρ extensions, and there is only one possibility in both cases.

The relation $\tau^4 \cdot h_1^4 = \rho^4 \cdot v_1^4$ (see (4-1)) then implies that $2\tau^2 v_1^2$ also supports an h_1 extension.

The map $\text{ko}/\rho \to \Sigma^{0,-1}\text{ko}$ takes τ^3 and $\tau^3 h_1$ to $2\tau^2$ and $\rho(\tau h_1)^2$ respectively. There is an h_1 extension connecting τ^3 and $\tau^3 h_1$ in ko/ρ, so there must be a hidden η extension from $2\tau^2$ to $\rho(\tau h_1)^2$.

The hidden h extension on τh_1 follows from the analogous hidden extension in the homotopy groups of the $\mathbb{R}$-motivic sphere [15] [9], using the unit map $S \to \text{ko}$. Alternatively, this hidden extension is computed in [17, Proposition 4.3] in the context of the $\mathbb{R}$-motivic Adams spectral sequence for ko.

Finally, multiply by τh_1 to obtain the hidden h extension on $(\tau h_1)^2$.

For degree reasons, there are no other possible hidden extensions to consider. □

Remark 4.6. We have completely analyzed the E_∞-page of the effective spectral sequence for ko, but this is not quite the same as completely describing the homotopy of ko. In particular, one must choose an element of $\pi_{*,*}\text{ko}$ that is represented by each multiplicative generator of $E_\infty(\text{ko})$ (see Table 6). In some cases, there is more than one choice because of the presence of elements in higher filtration in the E_∞-page. The choices of ρ, h_1, τh_1, and τ^4 can be made arbitrarily; the ring structure is unaffected by these choices. The elements $2\tau^2$ and $2v_1^2$ are already well-defined because there are no elements in higher filtration. Finally, the choices of $2\tau^2 v_1^2$ and v_1^4 can then be uniquely specified by the relations $\rho \cdot 2\tau^2 v_1^2 = \tau^4 \cdot h_1^3$ and $\rho^4 \cdot v_1^4 = \tau^4 \cdot h_1^4$.

4A. *η-periodic* ko. Later we will need some information about the η-periodic spectrum $\mathrm{ko}[\eta^{-1}]$. As in Section 3C, powers of h_1 are inconsequential for computational purposes in the η-periodic context. Consequently, we have removed these powers from all η-periodic formulas.

Proposition 4.7. *The effective E_1-page for* ko *is given by*

$$E_1(\mathrm{ko}[\eta^{-1}]) = \mathbb{F}_2[h_1^{\pm 1}, \tau, \rho, v_1^2].$$

Moreover, the periodicization map $\mathrm{ko} \to \mathrm{ko}[\eta^{-1}]$ *induces the map on effective E_1-pages whose values are given in Table 5.*

The first part of Proposition 4.7 was first proved in [1, Theorem 19], although the notation is different.

Proof. The functors s_* commute with homotopy colimits [29, Corollary 4.6]. Therefore, we can just invert h_1 in the description of $E_1(\mathrm{ko})$ given in Proposition 4.1 (see also Figure 5).

After inverting h_1, the relation $2h_1$ in $E_1(\mathrm{ko})$ implies that $2 = 0$ in $E_1(\mathrm{ko}[\eta^{-1}])$. This gives that

$$E_1(\mathrm{ko}[\eta^{-1}]) = \frac{\mathbb{F}_2[h_1^{\pm 1}, \rho, \tau^2, \tau h_1, v_1^2]}{\tau^2 = h_1^{-2}(\tau h_1)^2 + h_1^{-2} \cdot \rho^2 \cdot v_1^2}.$$

Because of the relation, the generator τ^2 is redundant.

The values of the periodicization map given in Table 5 are immediate from the algebraic analysis of the previous paragraph. □

Remark 4.8. Table 5 gives an unexpected value for τ^2. Recall that τ^2 is indecomposable in $E_1(\mathrm{ko})$, so there is no inconsistency. The unexpected value arises from (1-1).

4B. *The Adams operation ψ^3 in effective spectral sequences.* Our goal in this section is to study ψ^3 as a map of effective spectral sequences. This will allow us to compute the E_1-page of the effective spectral sequence for L.

Lemma 4.9. *The map $E_\infty(\mathrm{ko}) \to E_\infty(\mathrm{ko})$ induced by ψ^3 on effective E_∞-pages takes the values shown in Table 6.*

Proof. Corollary 2.4(2) gives the values of ψ^3 on ρ, h_1, and τh_1.

The value of ψ^3 on τ^4 is determined immediately by comparison along Betti realization to the classical value $\psi^3(1) = 1$. The computation is greatly simplified by ignoring terms in higher effective filtration. Similarly, the value of ψ^3 on $2\tau^2$ is determined by the classical value $\psi^3(2) = 2$.

The remaining values in Table 6 are also determined by comparison along Betti realization to the classical values $\psi^3(2v_1^2) = 9 \cdot 2v_1^2$ and $\psi^3(v_1^4) = 81v_1^4$. □

Lemma 4.10. *The map $E_1(\mathrm{ko}) \to E_1(\mathrm{ko})$ induced by ψ^3 on effective E_1-pages takes the values shown in Table 5.*

Proof. The values of ψ^3 on $E_1(\mathrm{ko})$ are compatible with the values of ψ^3 on $E_\infty(\mathrm{ko})$, as shown in Table 6. This immediately yields the value of ψ^3 on ρ, h_1, and τh_1.

The value of $\psi^3((\tau^2)^2)$ must be $(\tau^2)^2$ by compatibility with the value of $\psi^3(\tau^4)$ in $E_\infty(\mathrm{ko})$. Then the relation $\psi^3((\tau^2)^2) = (\psi^3(\tau^2))^2$ implies that $\psi^3(\tau^2) = \tau^2$.

Similarly, the value of $\psi^3((v_1^2)^2)$ must be $81(v_1^2)^2$ by compatibility with the value of $\psi^3(v_1^4)$ in $E_\infty(\mathrm{ko})$. Then the relation $\psi^3((v_1^2)^2) = (\psi^3(v_1^2))^2$ implies that $\psi^3(v_1^2) = 9v_1^2$. □

Remark 4.11. Since ψ^3 is a ring homomorphism, all values of ψ^3 on $E_1(\mathrm{ko})$ are readily determined by the values on multiplicative generators given in Table 5. In particular, for all $k \geq 0$,

$$\psi^3(v_1^{2k}) = 9^k v_1^{2k}.$$

Remark 4.12. Table 5 implies that $\psi^3(v_1^4) = 81v_1^4$. The careful reader will notice that this expression appears to be simpler than the analogous formula in [5, Theorem 3.1(2)]. The difference is explained by the fact that we are working only up to higher effective filtration. In particular, our formulas do not reflect the difference between the homotopy elements 2 and h, since their difference is detected by ρh_1 in higher effective filtration. This also means that our formulas are less precise, but that has no consequence for our computational results.

5. The effective spectral sequence for L

5A. ***The effective E_1-page of L.*** In this section we compute the E_1-page of the effective spectral sequence for L.

The fiber sequence $L \to \mathrm{ko} \xrightarrow{\psi^3-1} \mathrm{ko}$ induces a fiber sequence

$$s_*L \longrightarrow s_*\mathrm{ko} \xrightarrow{\psi^3-1} s_*\mathrm{ko}$$

on slices. Upon taking homotopy groups, we obtain a long exact sequence

$$\cdots \longrightarrow E_1(L) \longrightarrow E_1(\mathrm{ko}) \xrightarrow{\psi^3-1} E_1(\mathrm{ko}) \longrightarrow \cdots.$$

Table 5 (see also Lemma 4.10) gives us complete computational knowledge of the map $E_1(\mathrm{ko}) \to E_1(\mathrm{ko})$. This allows us to compute $E_1(L)$.

Proposition 5.1. *The chart in Figure 10 depicts the effective E_1-page of L.*

Proof. The long exact sequence

$$\cdots \longrightarrow E_1(L) \longrightarrow E_1(\mathrm{ko}) \xrightarrow{\psi^3-1} E_1(\mathrm{ko}) \longrightarrow \cdots$$

induces a short exact sequence

$$0 \longrightarrow \Sigma^{-1}C \longrightarrow E_1(L) \longrightarrow K \longrightarrow 0,$$

where C and K are the cokernel and kernel of $E_1(\mathrm{ko}) \xrightarrow{\psi^3-1} E_1(\mathrm{ko})$. The cokernel and kernel can be computed directly from the information given in Lemma 4.10. See also Remark 4.11.

The kernel consists of all elements in $E_1(\mathrm{ko})$ with the exception of the integer multiples of $\tau^{2j} \cdot v_1^{2k}$ for $j \geq 0$ and $k > 0$.

The cokernel C is nearly the same as $E_1(\mathrm{ko})$ itself. We must impose the relations $(3^{2k}-1)v_1^{2k} = 0$ for all $k > 0$. Lemma 3.8 says that $3^{2k}-1$ equals $2^{\nu(2k)+2} \cdot u$, where u is an odd number, i.e., a unit in our 2-adic context. Therefore, the relation $(3^{2k}-1)v_1^{2k} = 0$ is equivalent to the relation $2^{\nu(2k)+2}v_1^{2k} = 0$. □

Table 8 lists some elements of the effective E_1-page of L. In fact, by inspection these elements are multiplicative generators for $E_1(L)$.

We use the same notation for elements of $E_1(L)$ and their images in $E_1(\mathrm{ko})$. On the other hand, we define the element ιx of $E_1(L)$ to be the image of x under the map $\iota : \Sigma^{-1}E_1(\mathrm{ko}) \to E_1(L)$. For example, the element 1 of $E_1(\mathrm{ko})$ maps to ι in $E_1(L)$.

5B. ***The effective spectral sequence for $L[\eta^{-1}]$.*** In Section 5A, we determined the effective E_1-page of L. The next steps in the analysis of the effective spectral sequence for L are to determine the multiplicative structure of $E_1(L)$ (see Section 5C) and to determine the effective differentials (see Sections 5D and 5E).

coweight	(s, f, w)	generator	image in $E_1(L[\eta^{-1}])$
2	$(0, 0, -2)$	τ^2	$\tau^2+\rho^2\cdot v_1^2$
$2k+1$	$(4k+1, 1, 2k)$	$\tau h_1 v_1^{2k}$	$\tau(v_1^2)^k$
$2k$	$(4k-1, 1, 2k-1)$	ρv_1^{2k}	$\rho(v_1^2)^k$
$2k$	$(4k+1, 1, 2k+1)$	$h_1 v_1^{2k}$	$(v_1^2)^k$
$2k-1$	$(4k-1, 1, 2k)$	ιv_1^{2k}	$\iota(v_1^2)^k$

Table 8. Multiplicative generators for $E_1(L)$: $k \geq 0$.

Before doing so, we collect some information on the η-periodicization $L[\eta^{-1}]$. We will study $L[\eta^{-1}]$ by comparing to the more easily understood $\mathrm{ko}[\eta^{-1}]$.

As in Sections 3C and 4A, powers of h_1 are inconsequential for computational purposes in the η-periodic context. Consequently, we have removed these powers from all η-periodic formulas.

Proposition 5.2. *The effective E_1-page for $L[\eta^{-1}]$ is given by*

$$E_1(L[\eta^{-1}]) = \mathbb{F}_2[h_1^{\pm 1}, \tau, \rho, v_1^2, \iota]/\iota^2.$$

the periodicization map $L \to L[\eta^{-1}]$ induces the map $E_1(L) \to E_1(L[\eta^{-1}])$ whose values are given in Table 8.

Proof. As in Proposition 4.7, we can just invert h_1 in the additive description of $E_1(L)$ given in Proposition 5.1.

The map $E_1(\mathrm{ko}[\eta^{-1}]) \xrightarrow{\psi^3-1} E_1(\mathrm{ko}[\eta^{-1}])$ is trivial because $(\psi^3 - 1)(h_1) = 0$, as shown in Table 5 (see also Lemma 4.10). Therefore, the long exact sequence

$$\cdots \longrightarrow E_1(L[\eta^{-1}]) \longrightarrow E_1(\mathrm{ko}[\eta^{-1}]) \xrightarrow{\psi^3-1} E_1(\mathrm{ko}[\eta^{-1}]) \longrightarrow \cdots$$

splits as

$$E_1(L[\eta^{-1}]) \cong E_1(\mathrm{ko}[\eta^{-1}]) \oplus \Sigma^{-1} E_1(\mathrm{ko}[\eta^{-1}]).$$

With Proposition 4.7, this establishes the additive structure of $E_1(L[\eta^{-1}])$, as well as most of the multiplicative structure.

The relation $\iota^2 = 0$ is immediate because no nonzero values for ι^2 are possible. □

Remark 5.3. As in Remark 4.8, Table 8 gives an unexpected value for τ^2, which arises from (1-1). Also, the last column of Table 8 leaves out of h_1 for readability.

Remark 5.4. Note that $E_1(L[\eta^{-1}])$ is very close to the effective E_1-page for the η-periodic sphere $S[\eta^{-1}]$ [27, Theorem 2.32] [26, Theorem 2.3]. The element ι is not present in $E_1(S[\eta^{-1}])$, but the elements ιv_1^{2k} are present.

Proposition 5.5. *Some values of the differentials in the effective spectral sequence of $L[\eta^{-1}]$ are:*

(1) $d_1(v_1^2) = \tau$.

(2) $d_{n+1}(v_1^{2^n}) = \rho^{n+1} \cdot \iota v_1^{2^n}$ *for* $n \geq 2$.

The effective differentials are zero on all other multiplicative generators on all pages.

Following our convention throughout this section, we have omitted the powers of h_1 from the formulas in Proposition 5.5.

Proof. The d_1 differential follows from [27, Lemma 4.2] or [26, Theorem 2.6].

To study the higher differentials, consider the map $S[\eta^{-1}] \to L[\eta^{-1}]$. This map induces an isomorphism on stable homotopy groups, except in coweight -1. This follows from a minor adjustment to [5, Theorem 1.1]. The adjustment arises from the fact that our $L[\eta^{-1}]$ is the fiber of $\mathrm{ko}[\eta^{-1}] \xrightarrow{\psi^3-1} \mathrm{ko}[\eta^{-1}]$, while [5, Theorem 1.1] refers to the fiber of $\mathrm{ko}[\eta^{-1}] \xrightarrow{\psi^3-1} \Sigma^{8,4}\mathrm{ko}[\eta^{-1}]$.

The homotopy of $S[\eta^{-1}]$ is completely computed in [16], so the homotopy of $L[\eta^{-1}]$ is known (except in coweight -1). There is only one pattern of differentials that is compatible with the known values for $L[\eta^{-1}]$. □

Remark 5.6. In the language of [26, Section 4], Proposition 5.5 establishes the profile of the η-periodic effective spectral sequence over ℝ.

5C. *Multiplicative relations for $E_1(L)$.* In this section, we will completely describe the product structure on $E_1(L)$. We do not need all of this structure for our later computations, but we include it for completeness.

Proposition 5.7. *Table 9 lists some products in $E_1(L)$.*

Proof. All of these products are detected by $E_1(L[\eta^{-1}])$, which is described in Proposition 5.2. We need the values of the periodicization map $E_1(L) \to E_1(L[\eta^{-1}])$ given in Table 8. □

5D. *The effective d_1 differential for L.* Our next task is to compute the differentials in the effective spectral sequence for L.

Proposition 5.8. *Table 10 gives the values of the effective d_1 differential on the multiplicative generators of $E_1(L)$.*

Proof. All of these differentials follow immediately from the effective d_1 differentials for $L[\eta^{-1}]$, which are all determined by Proposition 5.5(1) Beware that the exact values of the map $E_1(L) \to E_1(L[\eta^{-1}])$, as shown in Table 8, are important.

For example, consider the differential on the element $\tau h_1 v_1^{4k+2}$. It maps to $\tau(v_1^2)^{2k+1}$ in $E_1(L[\eta^{-1}])$ (up to h_1 multiples, which as usual we ignore in the η-periodic situation). The η-periodic differential on this latter element is $\tau^2(v_1^2)^{2k}$.

	ρv_1^{2j}	$h_1 v_1^{2j}$	$\tau h_1 v_1^{2j}$	ιv_1^{2j}
ρv_1^{2k}	$\rho \cdot \rho v_1^{2j+2k}$			
$h_1 v_1^{2k}$	$\rho \cdot h_1 v_1^{2j+2k}$	$h_1 \cdot h_1 v_1^{2j+2k}$		
$\tau h_1 v_1^{2k}$	$\rho \cdot \tau h_1 v_1^{2j+2k}$	$h_1 \cdot \tau h_1 v_1^{2j+2k}$	$\tau^2 \cdot h_1 \cdot h_1 v_1^{2j+2k} + \rho \cdot \rho v_1^{2j+2k+2}$	
ιv_1^{2k}	$\rho \cdot \iota v_1^{2j+2k}$	$h_1 \cdot \iota v_1^{2j+2k}$	$\tau h_1 \cdot \iota v_1^{2j+2k}$	0

Table 9. Products in $E_1(L)$: $j \geq 0$ and $k \geq 0$.

coweight	(s, f, w)	x	$d_1(x)$
2	$(0, 0, -2)$	τ^2	$\rho^2 \cdot \tau h_1$
$4k$	$(8k-1, 1, 4k-1)$	ρv_1^{4k}	
$4k+2$	$(8k+3, 1, 4k+1)$	ρv_1^{4k+2}	$\rho h_1^2 \cdot \tau h_1 v_1^{4k}$
$4k$	$(8k+1, 1, 4k+1)$	$h_1 v_1^{4k}$	
$4k+2$	$(8k+5, 1, 4k+3)$	$h_1 v_1^{4k+2}$	$h_1^3 \cdot \tau h_1 v_1^{4k}$
$4k+3$	$(8k+5, 1, 4k+2)$	$\tau h_1 v_1^{4k+2}$	$\tau^2 \cdot h_1^3 \cdot h_1 v_1^{4k} + \rho^2 h_1 \cdot h_1 v_1^{4k+2}$
$4k+1$	$(8k+1, 1, 4k)$	$\tau h_1 v_1^{4k}$	
$4k+1$	$(8k+3, 1, 4k+2)$	ιv_1^{4k+2}	$\tau h_1 \cdot h_1^2 \cdot \iota v_1^{4k}$
$4k-1$	$(8k-1, 1, 4k)$	ιv_1^{4k}	

Table 10. Effective d_1 differentials for L: $k \geq 0$.

Finally, we need to find an element of $E_1(L)$ in the correct degree whose η-periodicization is $\tau^2(v_1^2)^{2k}$, The only possibility is $\tau^2 \cdot h_1^3 \cdot h_1 v_1^{4k} + \rho^2 h_1 \cdot h_1 v_1^{4k+2}$. □

Remark 5.9. All d_1 differentials in $E_1(L)$ can be deduced from the information in Table 10 and the Leibniz rule, but the computations can be complicated by the multiplicative relations of Table 9. For example,

$$d_1(\tau^2 \cdot \tau h_1 v_1^2) = \rho^2 \cdot \tau h_1 \cdot \tau h_1 v_1^2 + \tau^2(\tau^2 \cdot h_1^4 + \rho^2 h_1 \cdot h_1 v_1^2) = \tau^4 \cdot h_1^4 + \rho^4 \cdot v_1^4.$$

Having completely analyzed the slice d_1 differentials for $E_1(L)$, it is now possible to compute the E_2-page of the slice spectral sequence for L.

Proposition 5.10. *The E_2-page of the effective spectral sequence for L is depicted in Figures 11, 12, 14, and 15.*

For legibility, Figures 11, 12, 14, and 15 display $E_2(L)$ in four different charts separated by coweight modulo 4. Note that Figures 14 and 15 also serve as E_∞-page charts in coweights 1 and 2 modulo 4 because there are no higher differentials that affect these coweights.

Proof. The Leibniz rule, together with the values in Table 10, completely determines the effective d_1 differential on $E_1(L)$. The E_2-page can then be determined directly. However, as in the proof of Theorem 4.3, the computation is not entirely straightforward.

It turns out that the d_1 differential preserves the image of the map $\Sigma^{-1}E_1(\text{ko}) \to E_1(L)$. Moreover, it turns out that all d_1 differentials with values in the image of $\Sigma^{-1}E_1(\text{ko}) \to E_1(L)$ also have source in this image. (This is not for formal reasons; in fact, the higher effective differentials do not have this property.) Consequently, the determination of the E_2-page splits into two separate computations: one for the image of $\Sigma^{-1}E_1(\text{ko}) \to E_1(L)$, and one for the cokernel of the same map.

In more concrete terms, we can determine $E_2(L)$ by first considering only elements of the form ιx, and then separately considering only elements that are not of this form.

The d_1 differential on the image of $\Sigma^{-1}E_1(\mathrm{ko}) \to E_1(L)$ is identical to the d_1 differential for ko discussed in Section 4. The d_1 differential on the cokernel of $\Sigma^{-1}E_1(\mathrm{ko}) \to E_1(L)$ is similar to the d_1 differential on $E_1(\mathrm{ko})$, but slightly different. The difference is created by the absence of the elements v_1^{2k} in $E_1(L)$. □

5E. *Higher differentials.* We now consider the higher differentials in the effective spectral sequence for L.

By inspection of the charts for $E_2(L)$, the only possible higher differentials have source in coweight congruent to 0 modulo 4 and value in coweight congruent to 3 modulo 4. In other words, in coweights congruent to 1 and 2 modulo 4, we have that $E_2(L)$ equals $E_\infty(L)$.

It turns out that there are many higher differentials. In fact, nearly all of the elements in $E_2(L)$ in coweight congruent to 0 modulo 4 support differentials. While it is possible to write down explicit formulas for all of these differentials, the formulas would be cumbersome and not so helpful. Rather, we give a more qualitative description of the differentials because it is more useful for computation.

Proposition 5.11. *Consider the elements of $E_2(L)$ in coweights congruent to* 0 *modulo* 4 *that belong to the cokernel of the map $\Sigma^{-1}E_2(\mathrm{ko}) \to E_2(L)$.*

(1) *The only permanent cycles are the multiples of* 1, *the multiples of $2\tau^{4k}$ for $k \geq 0$, and $\rho^a h_1^b$ for all $a \geq 0$ and $b \geq 0$.*
(2) *Excluding the elements listed in* (1), *if an element has coweight congruent to 2^{r-1} modulo 2^r, then it supports a d_r differential.*

Proposition 5.11 may seem imprecise because it does not give the values of the differentials. However, there is only one nonzero possible value in every case, so there is no ambiguity.

Proof. These differentials follow immediately from the η-periodic differentials of Proposition 5.5, together with multiplicative relations in $E_2(L)$.

For example, consider the element $\tau^8 \cdot \rho v_1^{12}$ in coweight 20, which is congruent to 2^2 modulo 2^3. Using Table 8, we find that this element maps to $\rho^9(v_1^2)^{10}$ in $E_2(L[\eta^{-1}])$. Here we are using that τ^2 is zero in $E_2(L[\eta^{-1}])$ since it is hit by an η-periodic d_1 differential. Proposition 5.5 says that this element supports an η-periodic d_3 differential. It follows that $\tau^8 \cdot \rho v_1^{12}$ also supports a d_3 differential. □

Theorem 5.12. *The E_∞-page of the effective spectral sequence for L is depicted in Figures 13, 14, 15, 16, 17, 18, and 19.*

Proof. The E_∞-page can be deduced directly from the higher differentials described in Proposition 5.11. □

The E_∞-page in coweights congruent to 3 modulo 4 is by far the most complicated case. Figures 17, 18, and 19 display $E_\infty(L)$ in coweights congruent to 3 modulo 8, 7 modulo 16, and 15 modulo 32 respectively.

In each case (and more generally in coweights congruent to $2^{n-1}-1$ modulo 2^n, we see similar patterns with minor variations. The lower boundary of each chart takes the same shape. The upper boundary of the τ-periodic portion of each chart also takes the same shape. However, the filtration jump between the lower and upper boundaries increases linearly with n.

In addition to the τ-periodic portion of each chart, there are also τ-torsion, η-periodic regions. These consist of bands of infinite h_1-towers of width n that repeat every 2^{n+1} stems. The first such band starts at $\iota v_1^{2^{n-1}}$.

5F. ***Hidden extensions.*** Our last goal is to compute hidden extensions by ρ, h, and η. See [19, Section 4.1] for a precise definition of a hidden extension. Fortunately, none of the complications associated with crossing extensions occur in our situation.

Proposition 5.13. *Table 11 lists some hidden extensions by ρ, h, and η in the effective spectral sequence for L.*

Proof. The last column of Table 11 indicates the reason for each hidden extension. Some of the hidden extensions follow from the analogous extensions for ko given in Table 7, using the maps $\Sigma^{-1}\mathrm{ko} \to L$ and $L \to \mathrm{ko}$.

coweight	source	type	target	(s, f, w)	proof
0	$\iota \cdot \tau h_1$	h	$\iota \cdot \rho h_1 \cdot \tau h_1$	$(0, 2, 0)$	$\Sigma^{-1}\mathrm{ko} \to L$
1	τh_1	h	$\rho h_1 \cdot \tau h_1$	$(1, 1, 0)$	$L \to \mathrm{ko}$
1	$\iota(\tau h_1)^2$	h	$\iota \cdot \rho h_1 (\tau h_1)^2$	$(1, 3, 0)$	$\Sigma^{-1}\mathrm{ko} \to L$
1	$\iota \cdot 2\tau^2$	η	$\iota \cdot \rho (\tau h_1)^2$	$(-1, 1, -2)$	$\Sigma^{-1}\mathrm{ko} \to L$
1	$\iota 2 v_1^2$	ρ	$\iota \cdot h_1 (\tau h_1)^2$	$(3, 1, 2)$	$\Sigma^{-1}\mathrm{ko} \to L$
1	$\iota 4 v_1^2$	h	$h_1^2 \cdot \tau h_1$	$(3, 1, 2)$	L/ρ
2	$(\tau h_1)^2$	h	$\rho h_1 (\tau h_1)^2$	$(2, 2, 0)$	$L \to \mathrm{ko}$
3	$\iota 4 \tau^2 v_1^2$	h	$(\tau h_1)^3$	$(3, 1, 0)$	L/ρ
3	$\iota 2 \tau^2 v_1^2$	ρ	$\iota \tau^4 \cdot h_1^3$	$(3, 1, 0)$	$\Sigma^{-1}\mathrm{ko} \to L$
3	$\iota 2 \tau^2 v_1^2$	η	$\rho^3 \cdot \iota v_1^4$	$(3, 1, 0)$	$\Sigma^{-1}\mathrm{ko} \to L$
2	$2\tau^2$	η	$\rho (\tau h_1)^2$	$(0, 0, -2)$	$L \to \mathrm{ko}$
3	$(\tau h_1)^3$	h	$\iota \tau^4 \cdot \rho^2 h_1^6$	$(3, 3, 0)$	L/ρ
5	$\iota v_1^4 \cdot 8\tau^2$	h	$\rho^2 \cdot \tau h_1 v_1^4$	$(7, 1, 2)$	L/ρ

Table 11. Hidden extensions in $E_\infty(L)$.

Other extensions follow from the long exact sequence associated to the cofiber sequence

$$\Sigma^{-1,-1}L \xrightarrow{\rho} L \longrightarrow L/\rho.$$

Here we need that the homotopy of L/ρ is isomorphic to the homotopy of $L^{\mathbb{C}}$, as shown in [8, Corollary 1.9]. For example, the hidden h extensions of Proposition 3.18 give hidden h extensions in L/ρ, which then imply the hidden extension from $\iota 4v_1^2$ to $h_1^2 \cdot \tau h_1$. □

Remark 5.14. The hidden extensions in Table 11 are τ^4-periodic in the following sense. If we take the source and target of each extension in $E_1(L)$ and multiply by τ^4, then we obtain permanent cycles that are related by a hidden extension. For example, the hidden h extension from τh_1 to $\rho h_1 \cdot \tau h_1$ generalizes to a family of hidden extensions from $\tau^{4k+1}h_1$ to $\rho h_1 \cdot \tau^{4k+1}h_1$ for all $k \geq 0$.

Remark 5.15. Similarly to the τ^4-periodicity discussed in Remark 5.14, most of the hidden extensions in Table 11 are v_1^4-periodic as well. For example, the hidden h extension from τh_1 to $\rho h_1 \cdot \tau h_1$ generalizes to a family of hidden extensions from $\tau h_1 v_1^{4k}$ to $\rho h_1 \cdot \tau h_1 v_1^{4k}$ for all $k \geq 0$. There are three exceptions, which appear below the horizontal divider at the bottom of the table. These exceptions are discussed in more detail in Remarks 5.16, 5.17, and 5.18.

Remark 5.16. The hidden η extension from $2\tau^2$ to $\rho(\tau h_1)^2$ is τ^4-periodic as in Remark 5.14, but it is not v_1^4-periodic. The elements $2\tau^2 v_1^{4k}$ are not permanent cycles for $k \geq 1$.

Remark 5.17. The hidden h extension from $\iota v_1^4 \cdot 8\tau^2$ to $\rho^2 \cdot \tau h_1 v_1^4$ is v_1^4-periodic, but the situation is slightly more complicated than in Remark 5.15. For all k, $\rho^2 \cdot \tau h_1 v_1^{4k}$ receives a hidden h extension from an appropriate multiple of $\iota v_1^{4k} \cdot 2\tau^2$. For example, as shown in Figure 14, there is a hidden h extension from $\iota v_1^{4k} \cdot 16\tau^2$ to $\rho^2 \cdot \tau h_1 v_1^8$.

Remark 5.18. The hidden h extension from $(\tau h_1)^3$ to $\iota\tau^4 \cdot \rho^2 h_1^6$ is v_1^4-periodic, but the situation is more complicated than in Remarks 5.15 and 5.17. For all $k \geq 0$, the element $(\tau h_1)^2 \tau h_1 v_1^{4k}$ supports a hidden h extension to the element of $E_\infty(L)$ of highest filtration in the appropriate degree. For example, as shown in Figure 18, there is a hidden h extension from $(\tau h_1)^2 \cdot \tau^5 h_1$ to $\iota\tau^8 \cdot \rho^3 h_1^7$. Figures 17, 18, and 19 show several extensions of this type.

6. Charts

We explain the notation used in the charts.

- The horizontal coordinate is the stem s. The vertical coordinate is the Adams-Novikov filtration f (see Section 1G for further discussion).

- Black or green circles represent copies of $\mathbb{F}_2$, periodicized by some power of τ. The relevant power of τ varies from chart to chart.
- Black or green unfilled boxes represent copies of $\mathbb{Z}$ (the 2-adic integers), periodicized by some power of τ. The relevant power of τ varies from chart to chart.
- Black or green boxes containing a number n represent copies of $\mathbb{Z}/2^n$, periodicized by some power of τ. The relevant power of τ varies from chart to chart.
- Red unfilled boxes represent copies of $\mathbb{Z}$ (the 2-adic integers) that are not τ^k-periodic for any k.
- Green objects represent elements in the image of the map $E_1(\Sigma^{-1}\mathrm{ko}) \to E_1(L)$ (or $E_1(\Sigma^{-1}\mathrm{ko}^{\mathbb{C}}) \to E_1(L^{\mathbb{C}})$). Beware that the color refers to the E_1-page origin of the element, not the properties of the homotopical element that it detects. For example, in Figure 4, the element τh_1^3 detects an element in $\pi_{3,2}L^{\mathbb{C}}$ that maps to zero in $\pi_{3,2}\mathrm{ko}^{\mathbb{C}}$, so it is in the image of $\pi_{4,2}\mathrm{ko}^{\mathbb{C}}$. Nevertheless, the element is colored black because it is not in the image on E_1-pages.
- Black objects represent elements in the cokernel of the map $E_1(\Sigma^{-1}\mathrm{ko}) \to E_1(L)$ (or $E_1(\Sigma^{-1}\mathrm{ko}^{\mathbb{C}}) \to E_1(L^{\mathbb{C}})$). In other words, they are detected by the map $L \to \mathrm{ko}$ (or $L^{\mathbb{C}} \to \mathrm{ko}^{\mathbb{C}}$). As in the previous paragraph, beware of the distinction between E_1-page origins and homotopical properties.
- Lines of slope 1 represent h_1-multiplications.
- Black or green arrows of slope 1 represent infinite sequences of elements that are τ^k-periodic for some $k > 0$ and are connected by h_1-multiplications.
- Red arrows of slope 1 represent infinite sequences of elements that are connected by h_1-multiplications and are not τ^k-periodic for any k.
- Lines of slope -1 represent ρ-multiplications.
- Dashed lines of slope -1 represent ρ-multiplications whose values are multiples of τ^k for some $k > 0$. For example, in Figure 6, we have $\rho \cdot \rho^3 v_1^4$ equals $\tau^4 \cdot h_1^4$.
- Black or green arrows of slope -1 represent infinite sequences of elements that are τ^k-periodic for some $k > 0$ and are connected by ρ-multiplications.
- Light blue lines of slope -3 represent effective d_1 differentials.
- Dashed light blue lines of slope -3 represent effective d_1 differentials that hit multiples of τ^k, for some $k > 0$. For example, the dashed line in Figure 1 indicates that $d_1(v_1^2)$ equals τh_1^3.
- Dark blue lines indicate hidden extensions by h, ρ, or h_1.

- Dashed dark blue lines indicate hidden extensions whose value is a multiple of τ^k for some $k > 0$. For example, in Figure 4, there is a hidden h extension from $\iota 4 v_1^2$ to τh_1^3.

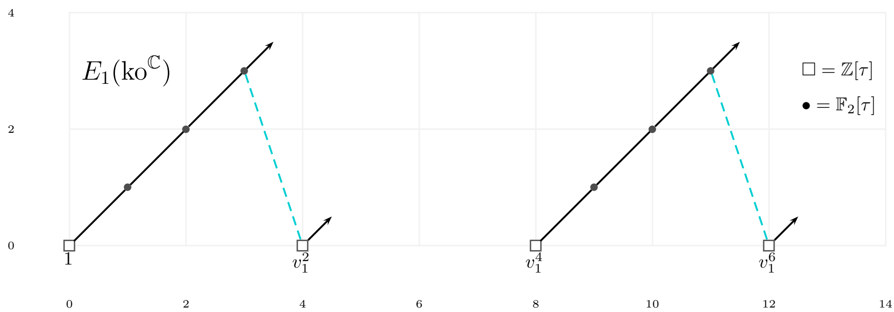

Figure 1. The E_1-page of the effective spectral sequence for $\mathrm{ko}^{\mathbb{C}}$.

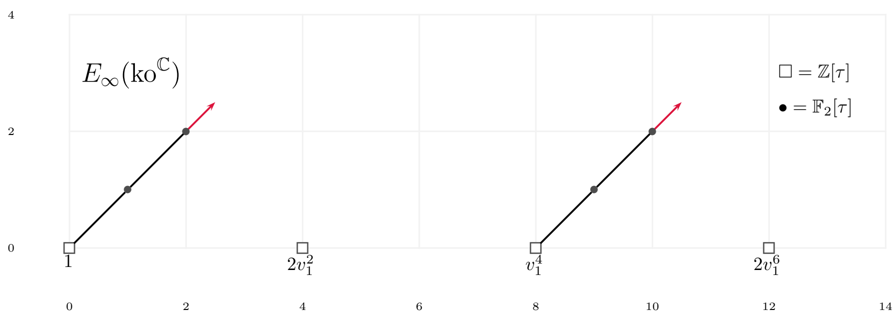

Figure 2. The E_∞-page of the effective spectral sequence for $\mathrm{ko}^{\mathbb{C}}$.

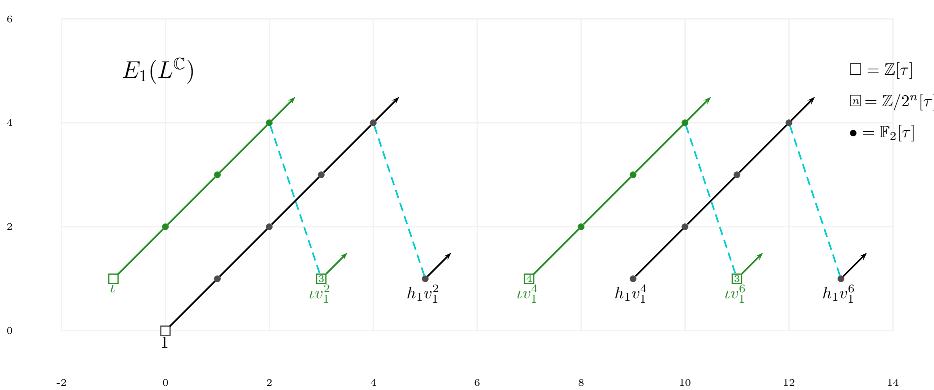

Figure 3. The E_1-page of the effective spectral sequence for $L^{\mathbb{C}}$.

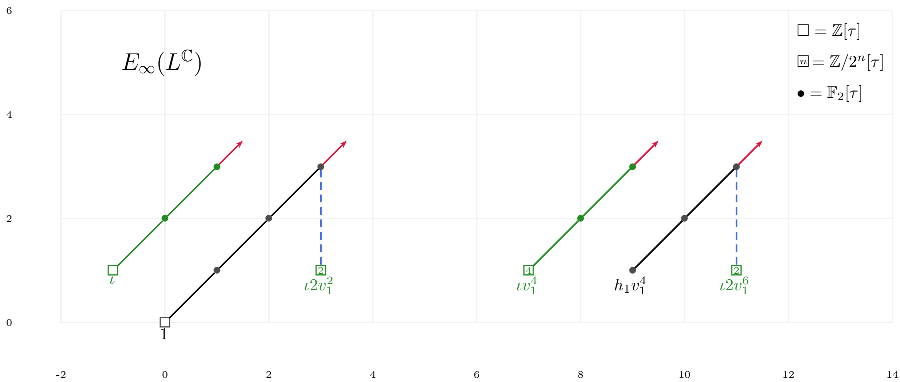

Figure 4. The E_∞-page of the effective spectral sequence for $L^{\mathbb{C}}$.

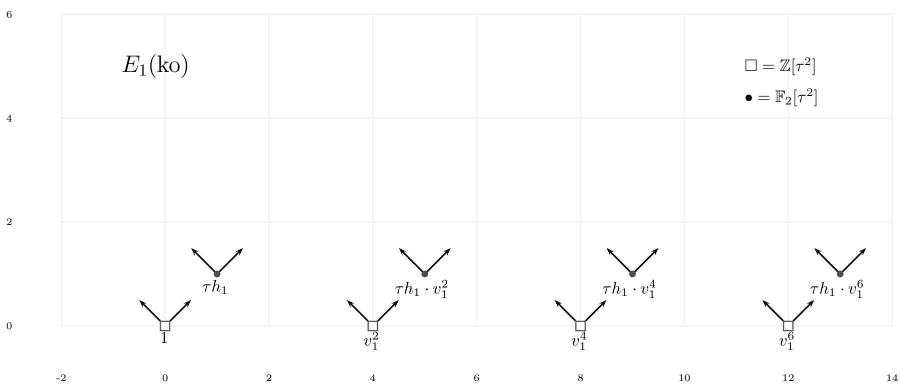

Figure 5. The E_1-page of the effective spectral sequence for ko.

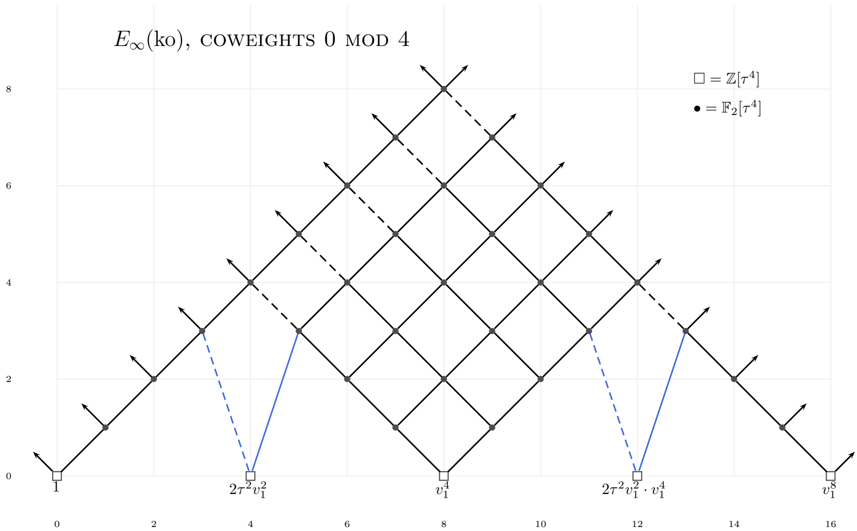

Figure 6. The E_∞-page of the effective spectral sequence for ko in coweights 0 mod 4.

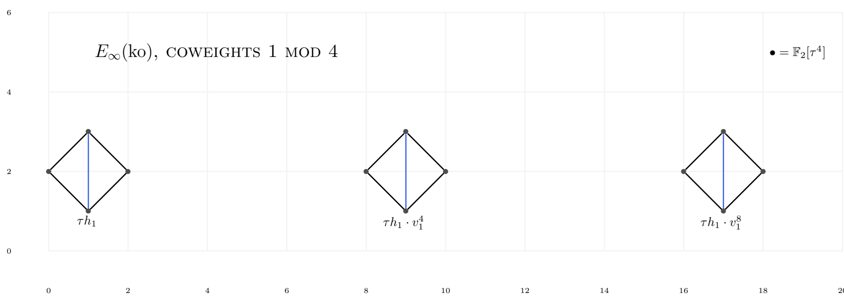

Figure 7. The E_∞-page of the effective spectral sequence for ko in coweights 1 mod 4.

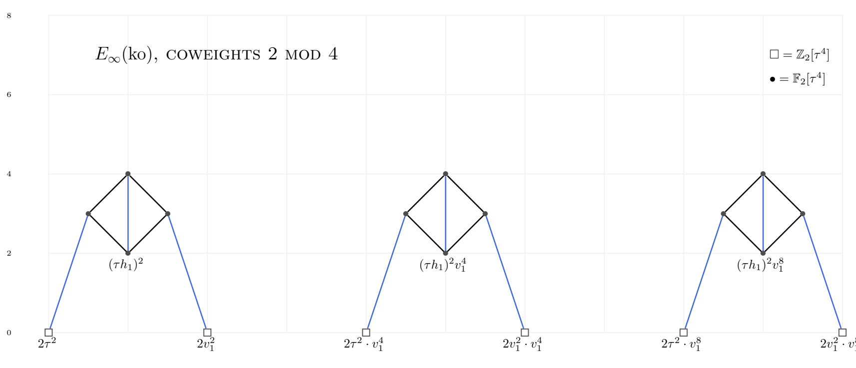

Figure 8. The E_∞-page of the effective spectral sequence for ko in coweights 2 mod 4.

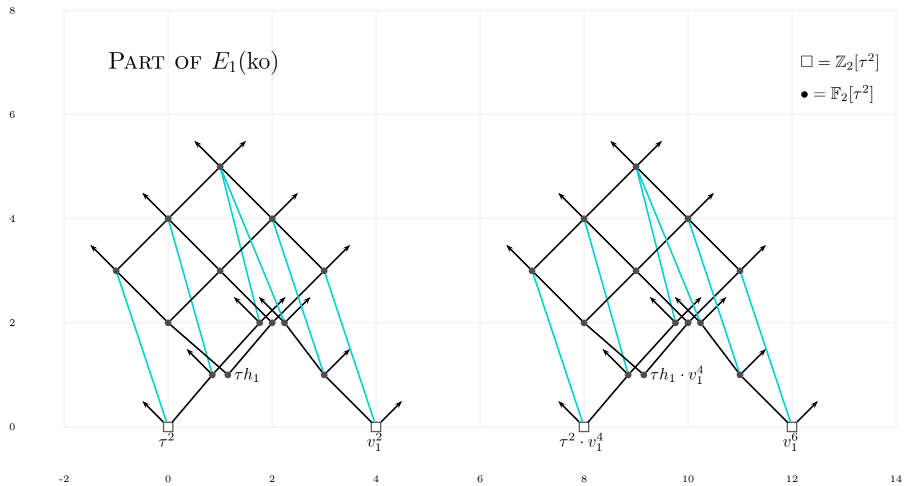

Figure 9. Some differentials in the effective spectral sequence for ko.

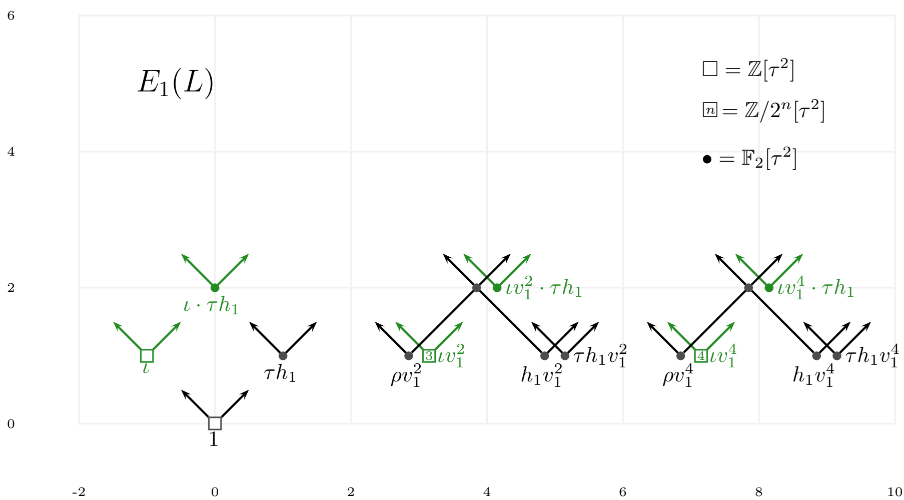

Figure 10. The E_1-page of the effective spectral sequence for L.

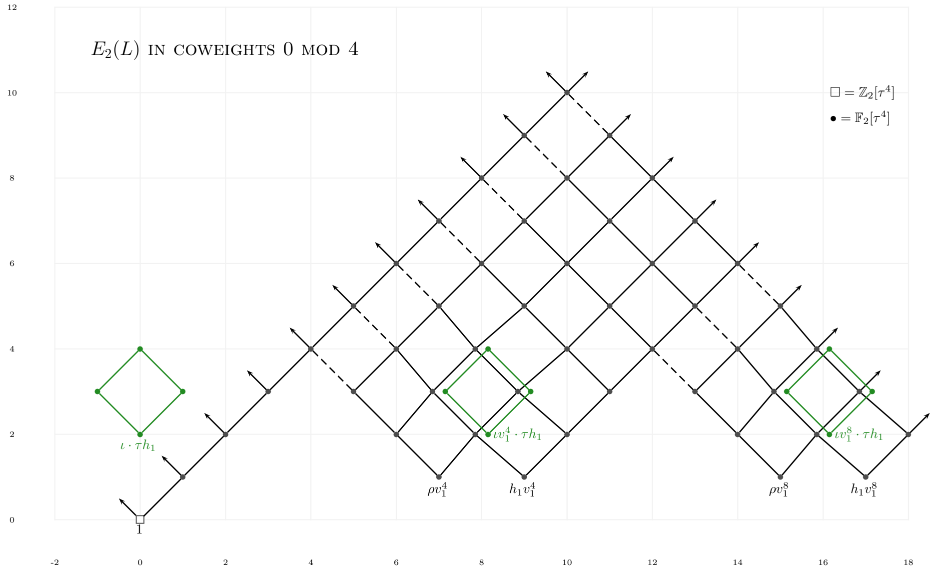

Figure 11. The E_2-page of the effective spectral sequence for L in coweights 0 mod 4.

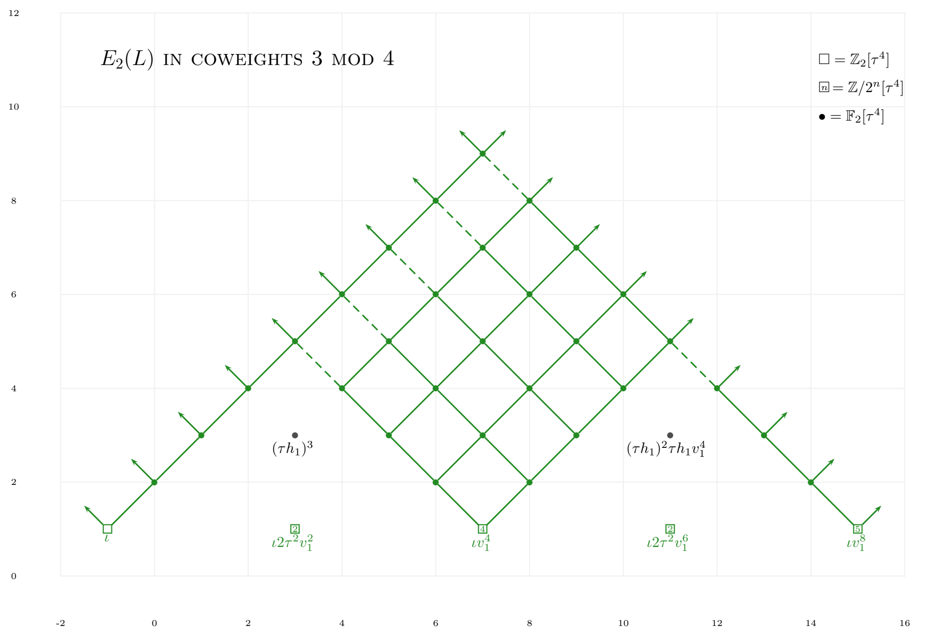

Figure 12. The E_2-page of the effective spectral sequence for L in coweights 3 mod 4.

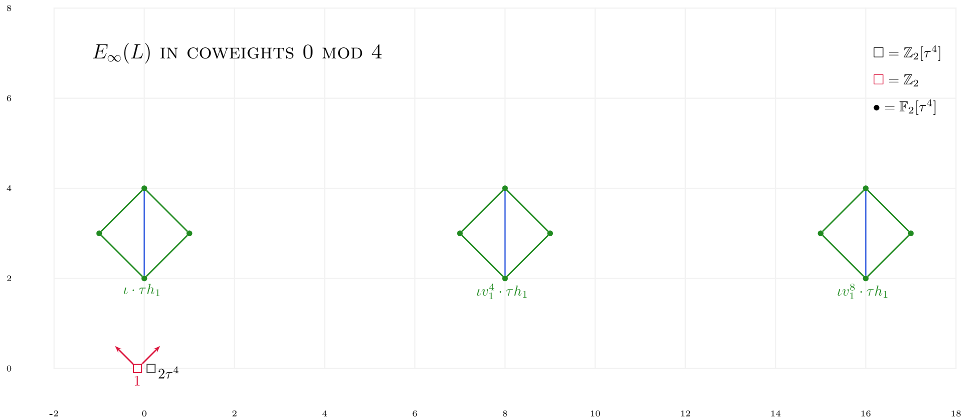

Figure 13. The E_∞-page of the effective spectral sequence for L in coweights 0 mod 4.

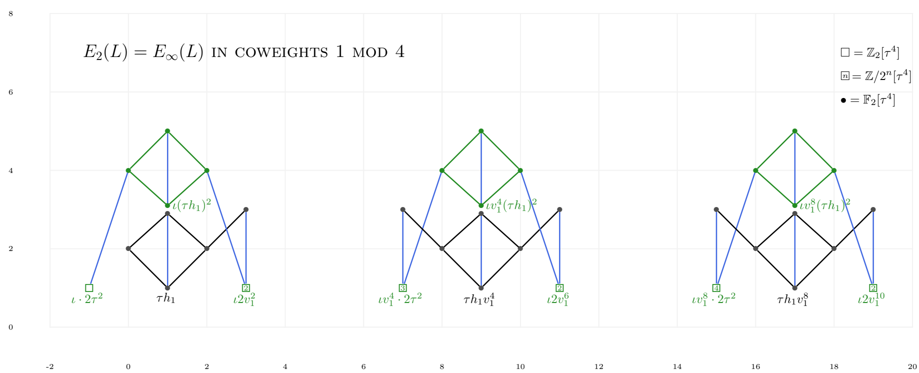

Figure 14. The E_∞-page of the effective spectral sequence for L in coweights 1 mod 4.

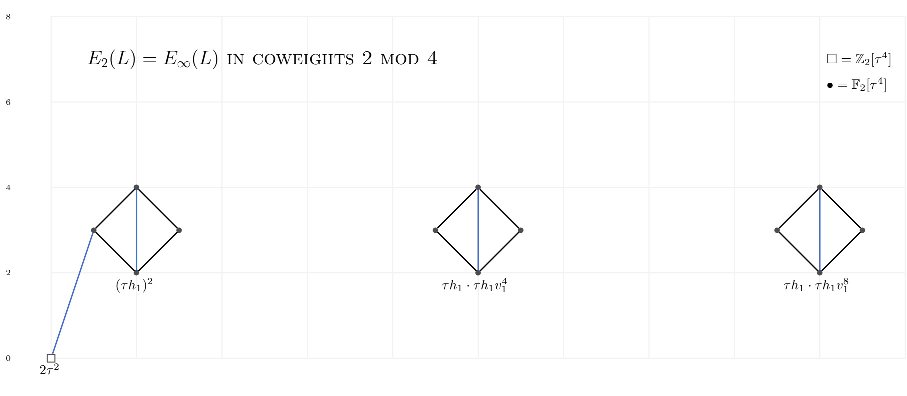

Figure 15. The E_∞-page of the effective spectral sequence for L in coweights 2 mod 4.

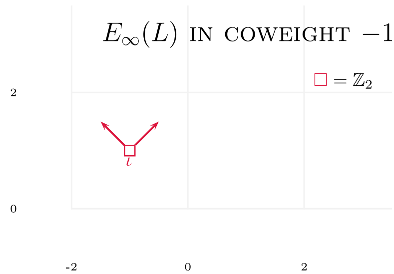

Figure 16. The E_∞-page of the effective spectral sequence for L in coweight -1.

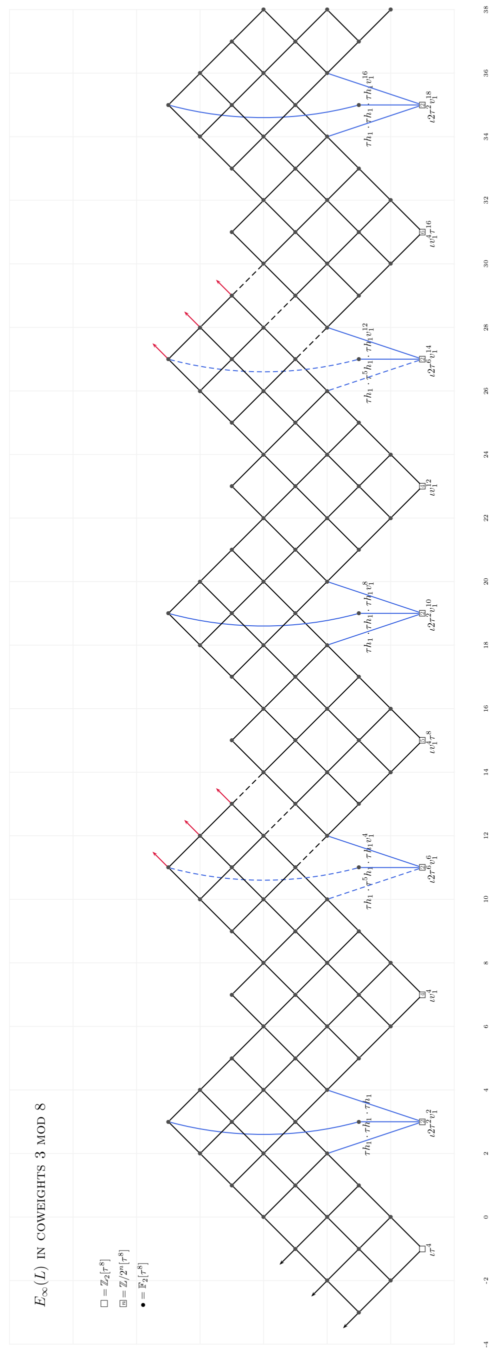

Figure 17. The E_∞-page of the effective spectral sequence for L in coweights 3 mod 8.

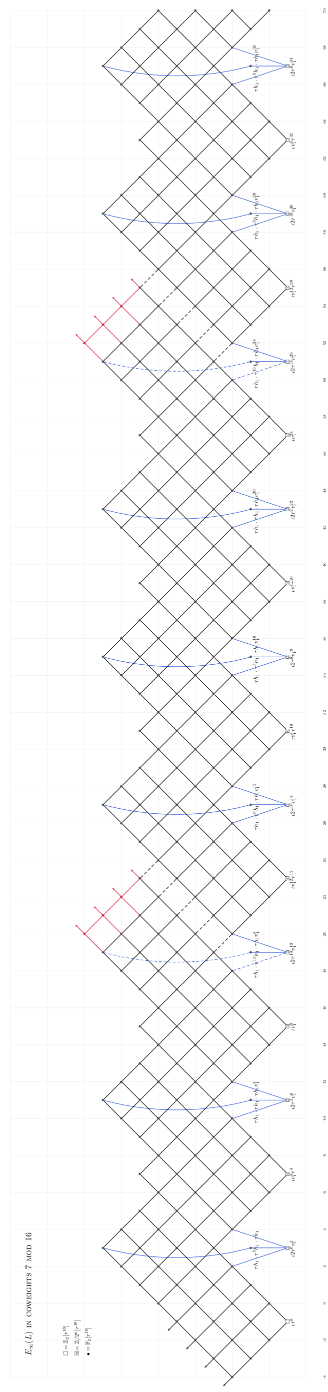

Figure 18. The E_∞-page of the effective spectral sequence for L in coweights 7 mod 16.

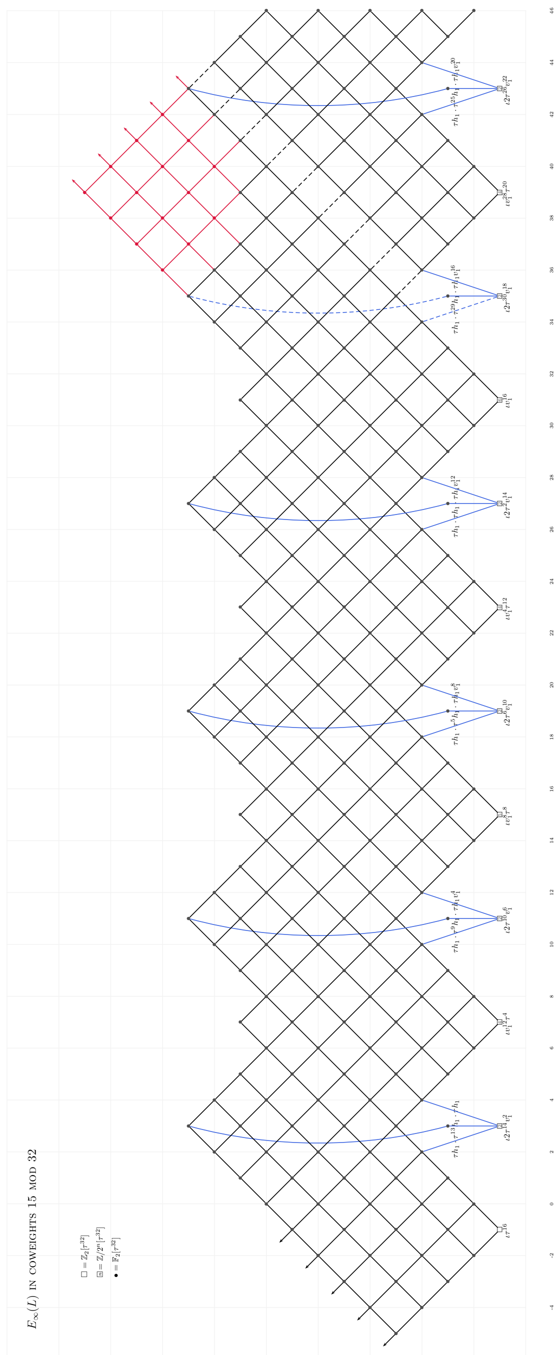

Figure 19. The E_∞-page of the effective spectral sequence for L in coweights 15 mod 32.

References

[1] A. Ananyevskiy, O. Röndigs, and P. A. Østvær, "On very effective hermitian K-theory", *Math. Z.* **294**:3-4 (2020), 1021–1034. MR Zbl

[2] A. Ananyevskiy, E. Elmanto, O. Röndigs, and M. Yakerson, "The motivic Adams conjecture", 2023. Zbl arXiv 2310.00974

[3] M. Andrews and H. Miller, "Inverting the Hopf map", *J. Topol.* **10**:4 (2017), 1145–1168. MR Zbl

[4] T. Bachmann, "Motivic and real étale stable homotopy theory", *Compos. Math.* **154**:5 (2018), 883–917. MR Zbl

[5] T. Bachmann and M. J. Hopkins, "η-periodic motivic stable homotopy theory over fields", 2020. Zbl arXiv 2005.06778

[6] W. Balderrama, "The C_2-equivariant $K(1)$-local sphere", 2021. Zbl arXiv 2103.13895

[7] W. Balderrama, K. Ormsby, and J. D. Quigley, "A motivic analogue of the K(1)-local sphere spectrum", 2023. arXiv 2307.13512

[8] M. Behrens and J. Shah, "C_2-equivariant stable homotopy from real motivic stable homotopy", *Ann. K-Theory* **5**:3 (2020), 411–464. MR Zbl

[9] E. Belmont and D. C. Isaksen, "$\mathbb{R}$-motivic stable stems", *J. Topol.* **15**:4 (2022), 1755–1793. MR Zbl

[10] E. Belmont, B. J. Guillou, and D. C. Isaksen, "C_2-equivariant and $\mathbb{R}$-motivic stable stems II", *Proc. Amer. Math. Soc.* **149**:1 (2021), 53–61. MR

[11] J. M. Boardman, "Conditionally convergent spectral sequences", pp. 49–84 in *Homotopy invariant algebraic structures* (Baltimore, 1998), Contemp. Math. **239**, Amer. Math. Soc., Providence, RI, 1999. MR Zbl

[12] R. R. Bruner and J. Rognes, "The Adams spectral sequence for the image-of-J spectrum", *Trans. Amer. Math. Soc.* **375**:8 (2022), 5803–5827. MR Zbl

[13] C. Carrick, "Smashing localizations in equivariant stable homotopy", *J. Homotopy Relat. Struct.* **17**:3 (2022), 355–392. MR Zbl

[14] D. M. Davis, "The cohomology of the spectrum bJ", *Bol. Soc. Mat. Mexicana* (2) **20**:1 (1975), 6–11. MR Zbl

[15] D. Dugger and D. C. Isaksen, "Low-dimensional Milnor–Witt stems over $\mathbb{R}$", *Ann. K-Theory* **2**:2 (2017), 175–210. MR

[16] B. J. Guillou and D. C. Isaksen, "The η-inverted $\mathbb{R}$-motivic sphere", *Algebr. Geom. Topol.* **16**:5 (2016), 3005–3027. MR Zbl

[17] M. A. Hill, "Ext and the motivic Steenrod algebra over $\mathbb{R}$", *J. Pure Appl. Algebra* **215**:5 (2011), 715–727. MR Zbl

[18] P. Hu, I. Kriz, and K. Ormsby, "Convergence of the motivic Adams spectral sequence", *J. K-Theory* **7**:3 (2011), 573–596. MR Zbl

[19] D. C. Isaksen, "Stable stems", *Mem. Amer. Math. Soc.* **262**:1269 (2019), viii+159. MR

[20] D. C. Isaksen and A. Shkembi, "Motivic connective K-theories and the cohomology of A(1)", *J. K-Theory* **7**:3 (2011), 619–661. MR Zbl

[21] H. J. Kong, "The C_2-effective spectral sequence for C_2-equivariant connective real K-theory", *Tunis. J. Math.* **5**:4 (2023), 627–662. MR Zbl

[22] H. J. Kong and J. D. Quigley, "The slice spectral sequence for a motivic analogue of the connective $K(1)$-local sphere", preprint, 2023. arXiv 2209.08603

[23] M. Levine, "Convergence of Voevodsky's slice tower", *Doc. Math.* **18** (2013), 907–941. MR Zbl

[24] M. Levine, "The Adams–Novikov spectral sequence and Voevodsky's slice tower", *Geom. Topol.* **19**:5 (2015), 2691–2740. MR Zbl

[25] F. Morel, "On the motivic π_0 of the sphere spectrum", pp. 219–260 in *Axiomatic, enriched and motivic homotopy theory*, NATO Sci. Ser. II Math. Phys. Chem. **131**, Kluwer, Dordrecht, 2004. MR Zbl

[26] K. Ormsby and O. Röndigs, "The homotopy groups of the η-periodic motivic sphere spectrum", *Pacific J. Math.* **306**:2 (2020), 679–697. MR Zbl

[27] O. Röndigs, M. Spitzweck, and P. A. Østvær, "The first stable homotopy groups of motivic spheres", *Ann. of Math.* (2) **189**:1 (2019), 1–74. MR Zbl

[28] O. Röndigs, M. Spitzweck, and P. A. Østvær, "The second stable homotopy groups of motivic spheres", 2021. Zbl arXiv 2103.17116

[29] M. Spitzweck, "Relations between slices and quotients of the algebraic cobordism spectrum", *Homology Homotopy Appl.* **12**:2 (2010), 335–351. MR Zbl

[30] M. Spitzweck and P. A. Østvær, "Motivic twisted K-theory", *Algebr. Geom. Topol.* **12**:1 (2012), 565–599. MR Zbl

[31] V. Voevodsky, "Open problems in the motivic stable homotopy theory, I", pp. 3–34 in *Motives, polylogarithms and Hodge theory, I: Motives and polylogarithms* (Irvine, CA, 1998), Int. Press Lect. Ser. **3**, International Press, Somerville, MA, 2002. MR Zbl

[32] G. M. Wilson, "The eta-inverted sphere over the rationals", *Algebr. Geom. Topol.* **18**:3 (2018), 1857–1881. MR Zbl

Received November 10, 2023. Revised May 21, 2024.

EVA BELMONT
DEPARTMENT OF MATHEMATICS, APPLIED MATHEMATICS, AND STATISTICS
CASE WESTERN RESERVE UNIVERSITY
CLEVELAND, OH 44106
UNITED STATES
eva.belmont@case.edu

DANIEL C. ISAKSEN
DEPARTMENT OF MATHEMATICS
WAYNE STATE UNIVERSITY
DETROIT, MI 48202
UNITED STATES
isaksen@wayne.edu

HANA JIA KONG
ZHEJIANG UNIVERSITY
HANGZHOU, ZHEJIANG 310058
CHINA
hana.jia.kong@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 330, No. 1, 2024
https://doi.org/10.2140/pjm.2024.330.85

HIGHER-GENUS QUANTUM K-THEORY

YOU-CHENG CHOU, LEO HERR AND YUAN-PIN LEE

We prove genus g invariants in quantum K-theory are determined by genus zero invariants of a smooth stack in the spirit of K. Costello's result in Gromov–Witten theory.

0. Introduction

Let X be a smooth quasiprojective variety over $\mathbb{C}$. Let $\overline{M}_{g,R}(X,\beta)$ be the space of genus g, R-pointed stable maps to X with degree β. The perfect obstruction theory on $\overline{M}_{g,R}(X,\beta)$ [Behrend and Fantechi 1997] endows the moduli stack with a "virtual structure sheaf" $\mathcal{O}^{\mathrm{vir}}_{\overline{M}_{g,R}(X,\beta)}$ [Lee 2004].

Let $\alpha_i \in K^\circ(X)$ and L_i be the universal cotangent line bundles. When the insertion

$$\Omega := \sum_I a_I \prod_{i=1}^R L_i^{k_i} \otimes ev_i^* \alpha_i$$

has an action by S_R, the permutation-equivariant pushforward

$$\sum_j (-1)^j H^j(\overline{M}_{g,R}(X,\beta), \mathcal{O}^{\mathrm{vir}} \otimes \Omega) \tag{1}$$

is an element in the Grothendieck group of S_R-representations with $\oplus$, i.e., virtual representations. We can also take a subgroup of S_R instead. These are by definition the *permutation-equivariant quantum K-invariants*.

The main theorem of this paper is the following.

Theorem 0.1 (see Theorem 3.1). *Genus g quantum K-invariants on X can be computed from permutation-equivariant genus zero quantum K-invariants on*

$$[\mathrm{Sym}^{g+1} X] = [X^{g+1}/S_{g+1}].$$

A similar statement holds for X a smooth DM stack with projective coarse moduli space.

MSC2020: 14N35, 19E08.

Keywords: quantum K-theory, Costello's pushforward formula.

We believe the higher-genus *permutation-equivariant* quantum K-theory of X can also be computed from genus zero theory of $[\mathrm{Sym}^{g+1} X]$ by extending our methods.

When the target, X itself or $[\mathrm{Sym}^{g+1} X]$, is a Deligne–Mumford stack, the definition of $\overline{M}_{g,R}(X, \beta)$ involves twisted/orbifold curves and twisted stable maps. The domain curves are families of pointed nodal curves with cyclic gerbe structures at the marked points and nodes, such that the gerbe structures at the nodes are *balanced*. This means they are locally stack quotients of the node $R[x, y]/(xy)$ by the antidiagonal action

$$\zeta.(x, y) := (\zeta x, \zeta^{-1} y), \quad \zeta \in \mu_r.$$

Note that we do *not* require the μ_r-gerbe structures at the marked points to be trivial in families as in [Costello 2006]. *Twisted stable maps* are *representable* morphisms from twisted curves to the target with finite automorphisms. $\overline{M}_{g,R}(X, \beta)$ are the moduli stacks of (twisted) stable maps with the discrete data g, R, β.

Hence, the marked points are no longer literal "points", but gerbes. Due to the nontrivial gerbes at the marked points, the evaluation maps have the natural codomain a partially rigidified inertia stack $\bar{\mathrm{I}}(X)$, instead of the inertia stack IX. This has been done in [Chen and Ruan 2002; 2004] and [Abramovich et al. 2008] in the context of cohomology and Chow groups.

We only need one class pulled back from $\bar{\mathrm{I}}(-)$ as opposed to ordinary K-theory, which comes from $[\mathrm{Sym}^k X]$ for some $k \leq g+1$ in Section 3B. We do not need the full K-theory of $\bar{\mathrm{I}}([\mathrm{Sym}^{g+1} X])$.

The appearance of permutation-equivariant K-theory is quite natural, not simply a "technical clutch". In cohomological Gromov–Witten theory, we often rely on the fact that the substacks ("strata") appearing in "common operations" (e.g., fixed-point loci in torus localizations or the components of inertia stacks of the moduli) are *variants* of known quantities in the sense of induction. These variants can be identified with the actual known quantities in Gromov–Witten theory by simple modifications. For example, for the purpose of computing Gromov–Witten invariants, we have

$$\int_{[\overline{M}/S_n]} \cdots = \frac{1}{n!} \int_{[\overline{M}]} \pi^*(\dots).$$

These equalities are no longer true in quantum K-theory. In fact, we have

$$\chi([\overline{M}/S_n], \dots) = \chi_{S_n}(\overline{M}, \pi^*(\dots))^{S_n}.$$

We note that K-theory on $[\overline{M}/S_n]$ can be identified with the S_n-equivariant K-theory on $\overline{M}$, and $\chi_{S_n}(-)^{S_n}$ is the pushforward in the S_n-equivariant theory, i.e., the S_n-invariant part $(-)^{S_n}$ of alternating sum of sheaf cohomologies viewed as S_n representations $\chi_{S_n}(-)$. This necessitates permutation-equivariant quantum K-theory.

Quantum K-theory has already been defined for stacks; see, e.g., [Tonita and Tseng 2013] and [Zhang, Section 2.4]. A comparison of the quantum K-theories with trivial and nontrivial gerbes at marked points can be found in [Zhang, Remark 2.8]. See also [Abramovich et al. 2002, Sections 4.4, 4.5] in the cohomological context. We allow nontrivial gerbes and recall the basic definitions in Section 1E.

Quantum K-invariants are roughly Gromov–Witten invariants computed in K-theory instead of cohomology or Chow groups. The idea of computing genus g Gromov–Witten invariants of any smooth projective variety X in terms of genus zero quantum K-invariants of quotient stack $[\mathrm{Sym}^{g+1} X]$ goes back to M. Kontsevich and was independently obtained in K. Costello's thesis [Costello 2006]. This paper can be considered as a K-theoretic version of this circle of ideas.

The calculation of genus zero quantum K-theoretic invariants is simpler and self-contained, while the higher-genus invariants necessarily involve invariants of lower genus. Genus-0 quantum K-theory is much better understood, with additional *finite difference* structure in addition to the usual D-module structure.

Quantum K-theory has connections with modern enumerative geometry, integrable systems, representation theory, geometric combinatorics and theoretical physics. Its influence on theoretical physics is largely its relation to 3-dimensional topological field theory. See the pioneering works of N. Nekrasov, H. Jockers, P. Mayr etc. [Jockers and Mayr 2019; 2020]. For its connection to representation theory, see [Okounkov 2017]. At the very onset of the quantum K-theory, it was intimately connected to integrable systems. See, for example, [Givental and Lee 2003]. It has also inspired much progress in geometric combinatorics through works like [Buch and Mihalcea 2011; Buch et al. 2013; 2020]. Most of these works are in genus zero. We hope that our algorithm will prove useful in the further development of higher-genus quantum K-theory.

We work exclusively with schemes, stacks, etc. locally of finite-type over the complex numbers $\mathbb{C}$. In particular, they are locally noetherian.

1. Higher-genus quantum K-invariants

Let C' be a general genus g smooth curve with a general divisor B of degree $d = g + 1$. There is exactly one ramified cover $f : C' \to \mathbb{P}^1$ of degree d with ramification divisor $B = f^*\infty$ over infinity by Riemann–Roch (see [Costello 2006, Lemma 6.0.1] and [Herr and Wise 2023, Theorem 3.12]. The following facts come from Costello [2006].

- This entails a birational map between moduli spaces (Lemma 1.16).
- By adding stack structure $\widetilde{\cdot}$, we can make $f : \widetilde{C}' \to \widetilde{\mathbb{P}^1}$ a finite étale cover. This is pulled back along a map $\widetilde{\mathbb{P}^1} \to BS_d$ from a stacky genus zero curve to the moduli space BS_d of finite étale degree d covers.

We can similarly interpolate between genus g and genus zero maps to a fixed smooth, quasiprojective target X.

Remark 1.1. Write $\langle d\rangle = \{1, 2, \dots, d\}$ for the ordered set of d elements. An S_d-torsor $P \to X$ is equivalent to the data of a finite étale degree d cover

$$T' := (P \times \langle d\rangle)/S_d \to X.$$

The universal d-sheeted cover

$$(\mathrm{pt} \times \langle d\rangle)/S_d \to BS_d$$

can be noncanonically identified with the map $BS_{d-1} \to BS_d$ induced by any of the d inclusions $S_{d-1} \subseteq S_d$.

Consider a twisted, representable stable map $\widetilde{C}' \to X$ together with a finite étale degree d cover $\widetilde{C}' \to \widetilde{C}$ of a stacky curve $\widetilde{C}$ of genus zero. To promote $\widetilde{C}'$ to a marked curve, we need only order the fibers of the marked points of $\widetilde{C}$.

Our data is pulled back from the universal finite étale degree d cover mapping to X:

$$\begin{array}{ccc} \widetilde{C}' & \xrightarrow{\quad\quad} & [(X^d \times \langle d\rangle)/S_d] \\ \big\downarrow & \ulcorner & \big\downarrow \\ \widetilde{C} & \xrightarrow{\quad\quad} & [\mathrm{Sym}^d X] = [X^d/S_d] \end{array}$$

and the whole diagram has finitely many automorphisms over the right arrow if and only if the map $\widetilde{C} \to [\mathrm{Sym}^d X]$ is stable.

Definition 1.2. The stack $\widetilde{\mathcal{K}}_{0,n}([\mathrm{Sym}^d X])$ parameterizes families $\widetilde{C} \to S$ of twisted curves of genus zero with n marked points and a representable map $\widetilde{C} \to [\mathrm{Sym}^d X]$ together with an ordering of the fibers over the marked points. The marked points of $\widetilde{C}$ may be nontrivial gerbes over S.

The stack $\widetilde{\mathcal{K}}_{0,n}([\mathrm{Sym}^d X])$ equivalently parametrizes families of ramified d-sheeted covers $C' \to C$ together with maps $C' \to X$ that have finitely many automorphisms. All ramification points are marked and the fibers above the marked points of C are all the marked points of C'. By "ordering of the fibers", we mean that the fibers of $C' \to C$ over each marked point of C must be ordered, a torsor for a product of symmetric groups. We later consider variants where less of the marked points of $C' \to C$ are ordered; see Figure 2.

Our twisted/stacky stable maps and curves are different from [Costello 2006]. For families of curves $\widetilde{C} \to S$ over a base scheme S, the i-th marked point of $\widetilde{C}$ may be a *nontrivial* μ_{r_i} gerbe for $r_i \in \mathbb{Z}_{\geq 1}$. We fix the orders r_i later.

We want to apply the K-theoretic version of Costello's pushforward formula [Chou et al. 2023, Theorem 2.7] to a square from [Herr and Wise 2023, Section 3.2] introduced in Section 1A:

(2)
$$\begin{array}{ccc} \widetilde{\mathcal{K}}_\Xi(S^d X) & \xrightarrow{\;q\;} & \overline{M}_{g,R}(X) \\ {\scriptstyle \pi'}\big\downarrow & \ulcorner & \big\downarrow{\scriptstyle \pi} \\ \widetilde{\mathfrak{K}}_\Xi(BS_d) & \xrightarrow[\;p\;]{} & \mathfrak{M}_{g,R} \end{array}$$

The stacks $\overline{M}_{g,R}(X)$ of stable maps and $\mathfrak{M}_{g,R}$ of prestable curves are standard. We do not fix a curve class β, so this space is a disjoint union over choices of β.

We denote $\widetilde{\mathcal{K}}_\Xi(S^d X) \subseteq \widetilde{\mathcal{K}}_{0,n}([\mathrm{Sym}^d X])$ a substack with appropriate discrete invariants fixed in Section 1B. The stack $\widetilde{\mathfrak{K}}_\Xi(BS_d)$, denoted $\widetilde{\mathfrak{M}}_{0,n}(BS_d)$ in [Herr and Wise 2023, before Lemma 3.6], is approximately the stack of prestable maps $\widetilde{C} \to BS_d$ from the genus zero twisted base curves parameterized in $\widetilde{\mathcal{K}}_\Xi(S^d X)$.

The obstruction theory for π' is pulled back from π. The problem is that p is of degree
$$e = k!(g!)^{\#J}(g!)^k,$$
while the K-theoretic virtual pushforward formula so far only applies to birational maps. We decompose p as a finite étale torsor of degree e composed with a birational map to which the pushforward formula applies.

1A. ***Costello's square* (2).** We describe (2). Write $\langle d\rangle = \{1, 2, \dots, d\}$. A subset $A \subseteq \langle \ell\rangle$ will be fixed later; the symbols $\overline{M}_{g,R}(X)$, $\mathfrak{M}_{g,R}$ refer to moduli stacks of ordinary stable maps and prestable curves with $R = \ell - \#A$ marked points. We do not fix the curve class β for simplicity. We assume $R \geq 1$.

The substack $\widetilde{\mathcal{K}}_\Xi(S^d X) \subseteq \widetilde{\mathcal{K}}_{0,n}([\mathrm{Sym}^d X])$ parametrizes stable maps of genus zero curves to $[\mathrm{Sym}^d X]$, identified with triples $C \leftarrow C' \to X$ above. The Ξ refers to fixed discrete invariants (see Section 1B): ramification profiles of $C' \to C$, the numbers n and ℓ of marked points for C and C', the genus of C', and the degree d of $C' \to C$. The number $\ell \leq dn$ is the sum of the degrees of the fibers in the ramification profiles. These invariants satisfy Riemann–Hurwitz to ensure that the space is nonempty:

(3)
$$2g - 2 = -2d + \sum_{P \in C'} (e_P - 1).$$

The functor q forgets the marked points $A \subseteq \langle \ell\rangle$ of C' and then takes the *stabilization* $\overline{C}' \to X$ of the resulting map $C' \to X$.

The map π' forgets the stable map to X. To make the diagram commute, π' must remember the stabilization $\overline{C}'$ of $C' \to X$. Define the stack $\widetilde{\mathfrak{K}}_\Xi(BS_d)$ of triples $C \leftarrow C' \to D$, where $C' \to C$ is a ramified cover of type Ξ and $C' \to D$ a partial stabilization after forgetting $A \subseteq \langle \ell\rangle$. The map p sends this triple to D. The square (2) is cartesian and p is proper by Lemma 3.9 and Corollary 3.7 of [Herr and Wise 2023], respectively.

Remark 1.3. The degree $e = k!(g!)^{\#J}(g!)^k$ differs from both [Herr and Wise 2023, Theorem 3.12] and [Costello 2006, Lemma 6.0.1]. Our use of nontrivial gerbes instead of trivialized gerbes accounts for the difference from [Herr and Wise 2023]. Costello's version is reconciled in Remark 3.15 of loc. cit. Our degree can be computed using the proof of Theorem 3.12 of loc. cit. or by taking into account the degrees of the universal gerbes.

The degree e is the order of a group $\Gamma = (S_g)^J \times S_g \wr S_k$ that reorders marked points of $C' \to C$ discussed in Section 1D.

The stabilization $\overline{C}' \to X$ was omitted in [Costello 2006], leading to nonproper moduli stacks or noncommutative diagrams. This could be rectified using Costello's technology of weighted graphs instead of our partial stabilizations.

1B. ***Specifying*** Ξ**.** We unpack our discrete data:

$$\Xi = \left\{ \begin{array}{c} g(C') = g,\ g(C) = 0,\ d = g+1, \\ \langle \ell \rangle \to \langle n \rangle \text{ is } \langle k \rangle \times \langle d \rangle \overset{pr_1}{\longmapsto} \langle k \rangle,\ J \times \langle d \rangle \overset{pr_1}{\longmapsto} J,\ I \mapsto \infty, \\ \forall j \in J,\ r_j = 1,\ \forall j \in \langle k \rangle,\ r_j = 2,\ \gamma : I \to \mathbb{Z}_{\geq 1},\ r_\infty = \operatorname{lcm}(\gamma(i)), \\ \bigsqcup_J B\mu_1 = * \to BS_d,\ \bigsqcup_{i \in \langle k \rangle} B\mu_2 \overset{\phi}{\longrightarrow} BS_d. \end{array} \right\}$$

See Figure 1 for an example which is Figure 2 in [Herr and Wise 2023].

Ramification profiles are specified by an action of μ_r on an unordered set of size d. Take a small loop around $p \in C$, and its lifts to C' identify which of the d sheets come together over p. Encode this action in a map $B\mu_r \to BS_d$ up to isomorphism.

Remark 1.4. The category of maps $BG \to BH$ has:

- Objects: homomorphisms $f : G \to H$.
- Morphisms $f_1 \to f_2$: elements $h \in H$ which conjugate one morphism to another $f_1 = h f_2 h^{-1}$. They are all isomorphisms.

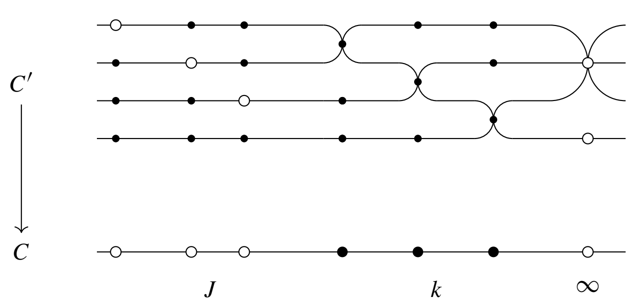

Figure 1. A cover in Ξ, $g = 3$, $d = 4$. Marked points are black if forgotten $A \subseteq \langle \ell \rangle$ and white if remembered under the map to $\mathfrak{M}_{g,R}$. The space $\widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d)$ forgets the ordering on the black marked points.

Objects of the category $\mathrm{Hom}(B\mu_r, BS_d)$ can be identified with actions $\mu_r \circlearrowright \langle d\rangle$. Isomorphisms between two such actions are relabelings of the set $\langle d\rangle$ of d elements. So an isomorphism class of functors $B\mu_r \to BS_d$ is an action of μ_r on an unlabeled set with d elements.

The action $\mu_r \circlearrowright \langle d\rangle$ contains the information of a ramification point $C' \to C$ of a map of curves. The stacky quotient $[\langle d\rangle/\mu_r]$ is the fiber of $\widetilde{C}' \to \widetilde{C}$ over the point $B\mu_r \in C$. To extract the set-theoretic fiber, we take the coarse moduli space $\langle d\rangle/\mu_r$. This gives an unlabeled set with some number of elements between 1 and d. In families, $B\mu_r$ is allowed to be a nontrivial gerbe.

A point $p \in C$ is *simply ramified* if its fiber consists of $d-1$ points, where exactly two of the d sheets come together and the other points in the fiber are unramified. This corresponds to the action $\mu_r \circlearrowright \langle d\rangle$ with one 2-cycle and the rest of the points fixed, up to reordering $\langle d\rangle$.

Let $k \geq 0$ be an integer, $g = g(C')$ be the genus of C', and fix the degree $d = g+1$. Divide the n marked points of C into three sets:

- $\{\infty\}$: write $I \subseteq C'$ for the fiber over this point $\infty \in C$. This point has ramification described by a map $B\mu_{r_\infty} \to BS_d$ or function $\gamma : I \to \mathbb{Z}_{\geq 1}$. That is, $\gamma(i)$ is the size of the stabilizer of i in the corresponding action $\mu_{r_\infty} \circlearrowright \langle d\rangle$.
- J: these points $J \subseteq C$ have no ramification.
- $\langle k\rangle$: these points have simple ramification.

This gives partitions

$$\langle n\rangle = J \sqcup \langle k\rangle \sqcup \{\infty\}, \quad \langle \ell\rangle = J \times \langle g+1\rangle \sqcup \langle kg\rangle \sqcup I.$$

The map $\langle \ell\rangle \to \langle n\rangle$ on marked points is compatible with these partitions.

The j-th marked point of C is a μ_{r_j}-gerbe, where $r_j = 1$ at unramified points $j \in J$, $r_j = 2$ for the k simple ramification points and ∞ has $r_\infty = \mathrm{lcm}(\gamma(i))$ the least common multiple of the ramification function γ on I. These data are subject to a constraint easier seen with trivialized gerbes: the sum $B\mathbb{Z} \to BS_d$ of the classifying space maps from all the composites $\mathbb{Z} \to \mu_r \to S_d$ be zero, lest the space be empty. This corresponds to the presentation of $\pi_1(\mathbb{P}^1 \setminus \langle n\rangle)$ via generators whose product is trivial.

Let $A \subseteq \langle \ell\rangle$ consist of all of $\langle kg\rangle$, none of I, and a subset of $J \times \langle g+1\rangle$ such that $J \times \langle g+1\rangle \setminus A \to J$ is a bijection. Note that the set $\langle \ell\rangle \setminus A = I \sqcup J$ has R elements.

Take $k = \#I + 3g - 1$ so that all the dimensions agree [Herr and Wise 2023, Theorem 3.12]:

$$\dim \widetilde{\mathfrak{K}}_{\Xi}(BS_d) = \dim \mathfrak{M}_{g,R}.$$

We can now prove our main equality between virtual fundamental classes. We first recall their definition in K-theory.

1C. *K-theory.* Let Y be a finite-type noetherian algebraic stack. The K-theory of Y is the K-theory of a category of lis-ét sheaves on Y, for which there are two main options:

- $K_\circ(Y)$: coherent sheaves, otherwise known as G-theory $G(Y)$.
- $K^\circ(Y)$: locally free sheaves of finite rank.

We work with $\mathbb{Q}$-coefficients, tensoring these groups up to $\mathbb{Q}$-vector spaces

$$K_\circ(Y) = K_\circ(Y) \otimes \mathbb{Q}, \quad K^\circ(Y) = K^\circ(Y) \otimes \mathbb{Q}.$$

These groups are generated by classes $[F]$ of coherent/locally free lis-ét sheaves F on Y, modulo relations $[F'] + [F''] = [F]$ for each exact sequence

$$0 \to F' \to F \to F'' \to 0.$$

See [Chou et al. 2023, Section 1] for discussion.

The groups K°, $K_\circ$ coincide on Y whenever every coherent sheaf F admits a finite resolution by locally free sheaves. Under certain hypotheses on Y, this is equivalent to Y being a quotient stack [Edidin et al. 2001, Remark 2.15].

Let $f : X \to Y$ be a map between finite-type noetherian algebraic stacks. Pullback and pushforward of sheaves sometimes induce maps on K° and $K_\circ$.

- K°: pullback f^* always exists and pushforward f_* makes sense when $X \to Y$ is finite étale.
- $K_\circ$: pullback f^* exists when f is flat. Armed with a perfect obstruction theory, we can also define a pullback $f^!$ even if f is not flat.

If f is proper and of DM-type, define the pushforward f_* on $K_\circ$-theory as the alternating sum

$$f_* F := \sum_i (-1)^i R^i f_* F. \tag{4}$$

We must check that this sum is finite.

When the map $f : X \to Y$ is clear from context, we write $\beta|_X = f^*\beta$ for classes $\beta \in K^\circ(Y)$ or $\beta \in K_\circ(Y)$ without risk of confusion.

Lemma 1.5. *Let $p : X \to Y$ be a proper, DM-type morphism between finite-type noetherian algebraic stacks. The pushforward*

$$p_* : K_\circ(X) \to K_\circ(Y).$$

of (4) is well defined on $K_\circ$-theory.

Proof. We argue that the sum (4) is finite. Write $\overline{X}$ for the relative coarse moduli space of the map p and N for a number larger than the dimensions of the fibers of $\overline{X} \to Y$. This is possible using quasicompactness of Y.

We claim $R^i p_* F$ vanishes for $i > N$ for any coherent sheaf F. The claim is étale local in Y and so is the formation of the relative coarse moduli space $\overline{X}$, so we can assume Y is an affine scheme. The map $t : X \to \overline{X}$ is finite flat, so pushforward is exact $Rt_* = t_*$. This reduces to the representable case $\overline{X} \to Y$. Because Y is affine, it results from dimensional vanishing [Stacks 2005–, 0A4R]. □

Pushforward from a proper DM stack to a point is denoted χ.

Example 1.6. If the morphism $p : X \to Y$ is not of DM-type, the pushforward need not be well defined. Take $B\mathbb{G}_m \to \text{pt}$. The cohomology of $B\mathbb{G}_m$ is freely generated as a ring by the first Chern class of the universal line bundle and does not vanish in any degree.

Example 1.7. Let G be a finite group. Sheaf pushforward along $p : BG \to \text{pt}$ sends a complex G-representation V to the invariant subspace V^G. The pushforward is then the alternating sum of group cohomology

$$\chi(V) = \sum (-1)^i [H^i(G, V)].$$

Because we work over $\mathbb{C}$, the structure sheaf of pt is $\mathcal{O}_{pt} = \mathbb{C}$. Likewise G-representations V on $BG = [\operatorname{Spec} \mathbb{C}/G]$ are complex representations, and the order of the group $\#G$ is invertible in V. The group cohomology therefore vanishes:

$$H^i(G, V) = 0, \quad i \neq 0.$$

The alternating sum is just the first term $\chi(V) = [V^G]$.

The projection formula holds in both K° and $K_\circ$, where defined

$$f_*(\alpha \otimes f^*\beta) = f_*\alpha \otimes \beta. \tag{5}$$

This results from the formula on the level of sheaves [Stacks 2005–, 08EU].

The main K-theory classes we are interested in are the fundamental class $[\mathcal{O}_X] \in K_\circ(X)$ and the *virtual fundamental class* (also known as *virtual structure sheaf*) [Lee 2004, Section 2.3; Qu 2018, Definition 2.2; Chou et al. 2023, Definition 1.2]. Consider a map $f : X \to M$ from a DM stack X to a smooth stack M endowed with a perfect obstruction theory $C_{X/M} \subseteq E$. The virtual fundamental class $[\mathcal{O}_X^{\text{vir}}]$ is the image of the structure sheaf of the normal cone $[\mathcal{O}_{C_{X/M}}]$ under the isomorphism [Chou et al. 2023, Remark 1.6]

$$[\mathcal{O}_X^{\text{vir}}] = \sigma^*[\mathcal{O}_{C_{X/M}}], \quad \sigma^* : K_\circ(E) \simeq K_\circ(X).$$

Example 1.8. Let $\pi : Y = BG \times X \to X$ be a trivial gerbe for a finite group G. Suppose X has a perfect obstruction theory over some M and Y is given the induced perfect obstruction theory. Then the virtual fundamental class pulls back $\pi^*[\mathcal{O}_X^{\text{vir}}] = [\mathcal{O}_Y^{\text{vir}}]$.

Example 1.7 describes π_* as taking G-invariants of a representation. Then we have $\pi_*\pi^*[\mathcal{O}_X] = [\mathcal{O}_X]$. Using the projection formula, this implies that the virtual fundamental class also pushes forward:

$$\pi_*[\mathcal{O}_Y^{\text{vir}}] = \pi_*(\pi^*[\mathcal{O}_X^{\text{vir}}] \otimes [\mathcal{O}_Y]) = [\mathcal{O}_X^{\text{vir}}] \otimes \pi_*\pi^*[\mathcal{O}_X] = [\mathcal{O}_X^{\text{vir}}].$$

Proposition 1.9. *Let $\pi : \mathcal{G} \to X$ be a gerbe banded by a finite group G. The base X is a scheme or algebraic stack which we emphasize lies over $\mathbb{C}$. Then the structure sheaf pushes forward to the structure sheaf, both as sheaves and in K-theory*:

$$R\pi_*\mathcal{O}_{\mathcal{G}} = \pi_*\mathcal{O}_{\mathcal{G}} = \mathcal{O}_X, \quad \pi_*[\mathcal{O}_{\mathcal{G}}] = [\mathcal{O}_X] \in K_\circ(X).$$

The same holds for virtual fundamental classes if $\mathcal{G}$ is given the induced perfect obstruction theory from X:

$$\pi_*[\mathcal{G}]^{\text{vir}} = [X]^{\text{vir}} \quad \textit{in } K_\circ(X).$$

Proof. The statement on sheaves implies that on K-theoretic classes and is local in X. We can then assume that $\mathcal{G}$ is trivial, fitting in a pullback square:

$$\begin{array}{ccc} \mathcal{G} & \longrightarrow & BG \\ \downarrow & \ulcorner & \downarrow \\ X & \longrightarrow & \text{pt} \end{array}$$

Example 1.7 covers the case of $BG \to \text{pt}$, and the general case results from cohomology and base change applied to this square.

The statement on virtual fundamental classes results from Example 1.8. □

Remark 1.10. The proof of Proposition 1.9 does not work for schemes over $\mathbb{Z}$. The groups $H^i(G, V)$ for $i \neq 0$ are torsion, and so are the sheaves $R^i\pi_*V$ for any coherent sheaf on $\mathcal{G}$. But this does not mean they vanish in $K_\circ \otimes \mathbb{Q}$. Tensoring $- \otimes \mathbb{Q}$ kills K-theoretic classes that are torsion in the group law on K-theory, not the classes of sheaves that themselves are torsion.

The trivial gerbe $[\operatorname{Spec} \mathbb{Z}/G] \to \operatorname{Spec} \mathbb{Z}$ satisfies Example 1.7 and Proposition 1.9, because the classes of torsion groups vanish in the K-theory of the integers. But this statement does not localize.

We need two related theorems on the behavior of (virtual) fundamental classes under pushforward. These extend Hironaka's theorem and Costello's theorem, respectively.

Theorem 1.11 (Hironaka's pushforward theorem [Chou et al. 2023, Proposition 2.3]). *Let $p : X \to Y$ be a proper birational map of smooth DM stacks. The pushforward of the fundamental class of X is that of Y in K-theory*:

$$p_*[\mathcal{O}_X] = [\mathcal{O}_Y] \quad \textit{in } K_\circ(Y).$$

Our Costello-type pushforward theorem was originally in the more general context of log geometry. We remove log structures in our citation for simplicity.

Theorem 1.12 (Costello's pushforward theorem [Chou et al. 2023, Theorem 2.7]). *Consider a pullback square of algebraic stacks*

$$\begin{array}{ccc} X & \xrightarrow{p} & Y \\ \downarrow & \ulcorner & \downarrow \\ M & \xrightarrow{q} & N \end{array}$$

with X, Y DM stacks and M, N smooth. Suppose $Y \to N$ is equipped with a perfect obstruction theory and $X \to M$ is given the pullback perfect obstruction theory. If the map q is proper birational, the pushforward of the virtual class of X is that of Y

$$p_*[\mathcal{O}_X^{\mathrm{vir}}] = [\mathcal{O}_Y^{\mathrm{vir}}] \quad \textit{in } K_\circ(Y).$$

To use these theorems, it is important that the relevant maps are proper and birational. Birational maps $f : X \to Y$ of stacks must have an open dense subset of each X and Y that are isomorphic.

Remark 1.13. For stacks, *pure degree one* [Herr and Wise 2023, Definition 2.3] and birational are not the same. The map $p : B\mathbb{Z}/2 \sqcup B\mathbb{Z}/2 \to \mathrm{pt}$ is pure degree one but not birational. The pushforward of the fundamental class is not the fundamental class:

$$p_*[\mathcal{O}_{B\mathbb{Z}/2 \sqcup B\mathbb{Z}/2}] = 2 \cdot [\mathcal{O}_{\mathrm{pt}}] \quad \text{in } K_\circ(\mathrm{pt}).$$

If a morphism of schemes is of pure degree one, it is birational. More generally, if $X \to Y$ is a morphism of stacks of pure degree one inducing a *representable* morphism $U \to V$ on open dense substacks $U \subseteq X$, $V \subseteq Y$, it is birational.

1D. ***An intermediary stack.*** The moduli stack $\widetilde{\mathfrak{K}}_\Xi(BS_d)$ parametrizes a triple $C \leftarrow C' \to D$ of curves over any base S. We introduce a variant $\widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d)$ to describe how virtual classes push forward in Proposition 1.17.

The functor $p : \widetilde{\mathfrak{K}}_\Xi(BS_d) \to \mathfrak{M}_{g,n}$ sends such a triple to D. The map p is proper of degree $e = k!(g!)^{\#J}(g!)^k$ [Herr and Wise 2023, Section 3]. This map forgets everything about $C' \to C$, including the ordering of the forgotten marked points under $C' \to D$.

Let $A \subseteq \langle \ell \rangle$ be the points forgotten under p as in Figure 1. Let $\widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d)$ be the space similar to $\widetilde{\mathfrak{K}}_\Xi(BS_d)$, but where the marked points $A \subseteq \langle \ell \rangle$ are *unordered*. The forgetful map $\widetilde{\mathfrak{K}}_\Xi(BS_d) \to \widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d)$ is a torsor under

$$\Gamma := (S_g)^J \times S_g \wr S_k.$$

The group $S_g \wr S_k := S_k \ltimes (S_g)^k$ is the wreath product. There is a short exact sequence

$$1 \to S_g^k \to S_g \wr S_k \to S_k \to 1$$

and a section $S_k \dashrightarrow S_g \wr S_k$ of the quotient. We may view $S_g \wr S_k$ as a subgroup of S_{gk} by choosing an identification of $\langle gk \rangle$ with $\langle g \rangle \times \langle k \rangle$.

The k copies of S_g reorder the unramified points in the fibers with simple ramification points, while S_k reorders the fibers themselves and their images $\langle k \rangle \subseteq C$.

Example 1.14. Isomorphisms of curves in $\widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d)$ need not stabilize the unordered marked points. For example, $\mathbb{P}^1$ with three unordered points has automorphism group S_3 by interchanging the points $0, 1, \infty$. The moduli space of genus zero curves with three unordered points is then BS_3.

These choices of ordering certain marked points can also be made on the moduli of stable maps to the stack $[\mathrm{Sym}^d X]$.

Definition 1.15. Let $\widetilde{\mathcal{K}}_\Xi(S^d X) \subseteq \widetilde{\mathcal{K}}_{0,n}([\mathrm{Sym}^d X])$ be the moduli space of representable stable maps to $[\mathrm{Sym}^d X]$ with discrete invariants Ξ. This parameterizes étale, d-sheeted covers $\widetilde{C}' \to \widetilde{C}$ with minimal stack structure together with stable maps $\widetilde{C}' \to X$. The curves may have nontrivial gerbes at marked points. All the marked points of C' and C are ordered.

Define $\mathcal{K}^*_\Xi(S^d X)$ analogously to $\widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d)$ by forgetting the ordering on the marked points of C' corresponding to $A \subseteq \langle \ell \rangle$ (see Figure 2).

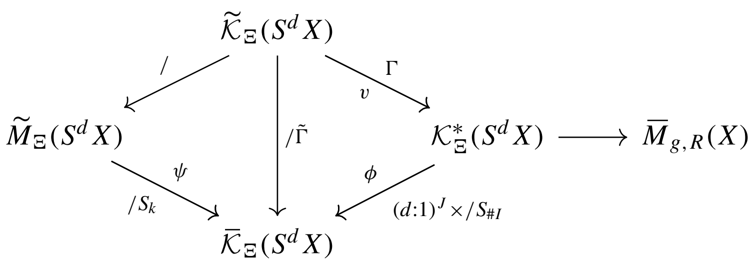

Figure 2. The stacks of stable maps to a fixed target X. $\overline{M}_{g,R}(X)$ is the ordinary space of stable maps to X. The rest are spaces of stacky genus zero maps $\widetilde{C} \to [\mathrm{Sym}^d X]$. These can be interpreted as ramified finite covers $C' \to C$ of nonstacky curves together with a map $C' \to X$ satisfying a stability condition. The difference between $\widetilde{M}_\Xi(S^d X)$, $\widetilde{\mathcal{K}}_\Xi(S^d X)$, $\mathcal{K}^*_\Xi(S^d X)$, $\overline{\mathcal{K}}_\Xi(S^d X)$ lies in which points of C', C are ordered. The maps "$/G$" between them are quotients by various groups G reordering the marked points. The one exception is $\mathcal{K}^*_\Xi(S^d X) \to \overline{\mathcal{K}}_\Xi(S^d X)$, which is a quotient followed by a $d^{\#J}$-sheeted cover (denoted $(d:1)^J$).

Extend (2) to the cartesian diagram

$$
\begin{array}{ccccc}
 & & \overset{q}{\frown} & & \\
\widetilde{\mathcal{K}}_{\Xi}(S^d X) & \xrightarrow{\ v\ } & \mathcal{K}^*_{\Xi}(S^d X) & \xrightarrow{\ w\ } & \overline{M}_{g,R}(X) \\
\big\downarrow{\scriptstyle \pi'} & \ulcorner & \big\downarrow & \ulcorner & \big\downarrow{\scriptstyle \pi} \\
\widetilde{\mathfrak{K}}_{\Xi}(BS_d) & \xrightarrow{\ v\ } & \widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d) & \xrightarrow{\ \omega\ } & \mathfrak{M}_{g,R} \\
 & & \underset{p}{\smile} & &
\end{array} \tag{6}
$$

Lemma 1.16. *The map $\omega : \widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d) \to \mathfrak{M}_{g,R}$ is proper and birational.*

Proof. Fix a generic smooth D and prescribed ramification divisor $B = \sum_{i \in I} d(i)[i]$ over ∞ specified by Ξ. The proof of [Herr and Wise 2023, Theorem 3.12] shows that there is exactly one cover $C' \to C$ with ramification in B and $C' \to D$ a partial stabilization. The map is thus proper and of pure degree one, but this is not yet sufficient by Remark 1.13.

We argue that ω is generically representable, hence birational. The proof of [Herr and Wise 2023, Theorem 3.12] shows that if $D \in \mathfrak{M}_{g,R}(X)$ is general, the preimage under ω is exactly one cover $C' \to C$ with $C' = D$. Consider automorphisms

$$
\begin{array}{ccc}
C' & \overset{\sim}{\text{——}} & C' \\
\big\downarrow & & \big\downarrow \\
C & \overset{\sim}{\text{——}} & C
\end{array}
$$

of the map $C' \to C$. These form a subgroup of automorphisms of C' because $C' \to C$ is an epimorphism. Since the map $\underline{\mathrm{Aut}}(C' \to C) \to \underline{\mathrm{Aut}}(D)$ is injective, the map is generically representable and hence birational. □

Hironaka's pushforward theorem (see Theorem 1.11) equates their fundamental classes:

$$\omega_*[\mathcal{O}_{\widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d)}] = [\mathcal{O}_{\mathfrak{M}_{g,n}}] \quad \text{in } K_\circ(\mathfrak{M}_{g,n}).$$

Costello's pushforward theorem (see Theorem 1.12) likewise equates the virtual fundamental classes:

Proposition 1.17. *The fundamental class pushes forward along the map w in (6):*

$$\omega_*[\mathcal{O}_{\widetilde{\mathfrak{K}}^*_{\Xi^*}(BS_d)}] = [\mathcal{O}_{\mathfrak{M}_{g,n}}] \quad \textit{in } K_\circ(\mathfrak{M}_{g,n}).$$

As a result, the virtual fundamental class pushes forward the same way:

$$w_*[\mathcal{O}^{\mathrm{vir}}_{\mathcal{K}^*_{\Xi}(S^d X)}] = [\mathcal{O}^{\mathrm{vir}}_{\overline{M}_{g,R}(X)}] \quad \textit{in } K_\circ(\overline{M}_{g,R}(X)).$$

The wreath product $S_g \wr S_k$ arises naturally as the automorphism group of the projection $\langle g\rangle \times \langle k\rangle \to \langle k\rangle$ of marked points of $C' \to C$:

Remark 1.18 (thanks to J. Rufus Lawrence). The iterated stack-theoretic symmetric product $[\mathrm{Sym}^k[\mathrm{Sym}^g X]]$ is isomorphic to the global quotient

$$[\mathrm{Sym}^k[\mathrm{Sym}^g X]] \simeq [X^{gk}/(S_g \wr S_k)].$$

To make sense of $[\mathrm{Sym}^k -]$ applied to a stack $\mathscr{S}$, use its functor of points

$$[\mathrm{Sym}^k \mathscr{S}](T) = \left\{ \begin{array}{ccc} T' & \longrightarrow & \mathscr{S} \\ {\scriptstyle k:1}\downarrow & & \\ T & & \end{array} \right\},$$

where $T' \to T$ is a k-sheeted cover that is part of the moduli.

Consider the evaluation map corresponding to a point $C' \to C$ of $\mathcal{K}^*_{\Xi}(S^d X)$ over a base T. Upon ordering the k-marked points i of C, we get a map to $[\mathrm{Sym}^g X]$ corresponding to each i. Ordering the k-marked points of C entails a k-sheeted cover of T with a map to $[\mathrm{Sym}^g X]$. These evaluation maps are precisely

$$\mathcal{K}^*_{\Xi}(S^d X) \to [\mathrm{Sym}^k[\mathrm{Sym}^g X]] = [X^{gk}/(S_g \wr S_k)].$$

Two more stacks $\widetilde{M}_{\Xi}(S^d X)$, $\overline{\mathcal{K}}_{\Xi}(S^d X)$. The stack $\widetilde{\mathcal{K}}_{\Xi}(S^d X)$ is the simplest because all the marked points of C' and C are ordered, but we will not actually use it for our theorem. The variant $\mathcal{K}^*_{\Xi}(S^d X)$ above is virtually birational to $\overline{M}_{g,R}(X)$. We need two more variants, completing Figure 2.

Definition 1.19. Let $\widetilde{M}_{\Xi}(S^d X)$ be the moduli space of representable twisted stable maps $C \to [\mathrm{Sym}^d X]$. It is the same as $\widetilde{\mathcal{K}}_{\Xi}(S^d X)$, except the marked points of C' are not ordered. The only difference from twisted stable maps $C \to [\mathrm{Sym}^d X]$ in the literature is the nontrivial gerbes.

The quotient maps $\widetilde{\mathcal{K}}_{\Xi}(S^d X) \to \widetilde{M}_{\Xi}(S^d X)$, $\widetilde{\mathcal{K}}_{\Xi}(S^d X) \to \mathcal{K}^*_{\Xi}(S^d X)$ forget different marked points, so there is not a map between them. Define $\overline{\mathcal{K}}_{\Xi}(S^d X)$ to forget all the marked points of both, so only the points of C that are not in $\langle k \rangle \subseteq C$ are ordered.

As shown in Figure 2, there are quotient maps from $\widetilde{\mathcal{K}}_{\Xi}(S^d X)$ to all the others $\widetilde{M}_{\Xi}(S^d X)$, $\overline{\mathcal{K}}_{\Xi}(S^d X)$, $\mathcal{K}^*_{\Xi}(S^d X)$ by various groups reordering marked points. The map $\widetilde{M}_{\Xi}(S^d X) \to \overline{\mathcal{K}}_{\Xi}(S^d X)$ quotients by S_k, while $\mathcal{K}^*_{\Xi}(S^d X) \to \overline{\mathcal{K}}_{\Xi}(S^d X)$ is a composite of many d-sheeted covers indexed by J and a quotient by $S_{\#I}$.

Remark 1.20. Remark that $R = 1$ when $\#I = 1$ and $J = \varnothing$. In that case, $k = 3g$. The map $\mathcal{K}^*_{\Xi}(S^d X) \to \overline{\mathcal{K}}$ is an isomorphism precisely when one of the conditions holds:

- $R = 1$.
- $\#I = 1$ and $g = 0$.

See Figure 2.

Remark 1.21. The finite étale maps

$$\begin{array}{ccc} \widetilde{\mathcal{K}}_{\Xi}(S^d X) & \xrightarrow{\ v\ } & \mathcal{K}^*_{\Xi}(S^d X) \\ \downarrow & & \downarrow \phi \\ \widetilde{M}_{\Xi}(S^d X) & \xrightarrow{\ \psi\ } & \overline{\mathcal{K}}_{\Xi}(S^d X) \end{array}$$

in Figure 2 all equate virtual fundamental classes under pullback:

$$v^*[\mathcal{O}^{\mathrm{vir}}_{\mathcal{K}^*_{\Xi}(S^d X)}] = [\mathcal{O}^{\mathrm{vir}}_{\widetilde{\mathcal{K}}_{\Xi}(S^d X)}], \text{ etc.}$$

1E. *Gromov–Witten invariants in the $K_\circ$-theory of stacks.* Quantum $K_\circ$-theoretic invariants have been defined variously in the literature [Tonita and Tseng 2013] [Zhang, Section 2.4]. Our definition parallels [Abramovich et al. 2008], adding in ψ classes. Our invariants differ by a scaling factor due to conventions over whether gerbes at marked points are trivialized; see [Tonita and Tseng 2013; Zhang, Remark 2.8] or the original [Abramovich et al. 2002, Sections 4.4, 4.5] for comparison. We allow nontrivial gerbes.

For any moduli space $\mathcal{K}$ of stable maps $C \to Y$ from n-pointed curves, there is an evaluation map

$$\mathrm{ev} : \mathcal{K} \to Y^n.$$

If Y is a stack and we take representable twisted stable maps $\widetilde{C} \to Y$ in say $\mathcal{K} = \overline{M}_{g,n}(Y)$, the "points" of $\widetilde{C}$ are not quite points, but μ_r-banded gerbes. The evaluation map doesn't produce points of Y, but cyclic gerbes mapping representably to Y. Cyclic gerbes representably embedded in Y form the *rigidified cyclotomic inertia* stack $\bar{\mathrm{I}}(Y)$ [Abramovich et al. 2008, Section 3], so the evaluation map is

$$\mathrm{ev} : \mathcal{K} \to (\bar{\mathrm{I}}(Y))^n.$$

The stack $\bar{\mathrm{I}}(Y)$ is closely related to the inertia stack IY. The universal gerbe over $\bar{\mathrm{I}}(Y)$ can be identified with representable maps from the trivial gerbe

$$\mathrm{Hom}^{\mathrm{rep}}(B\mu_r, Y).$$

Fixing an isomorphism $\mu_r \simeq \mathbb{Z}/r$ over $\mathbb{C}$, we get a map to the inertia stack

$$\begin{array}{ccc} \mathrm{Hom}^{\mathrm{rep}}(B\mu_r, Y) & \longrightarrow & IY \\ \downarrow & & \\ \bar{\mathrm{I}}(Y) & & \end{array}$$

by composing $B\mathbb{Z} \to B\mu_r \to Y$. See Section 3C for a worked example.

Instead of pulling back $K_\circ$-theoretic classes from Y, we pull back from $K_\circ(\bar{\mathrm{I}}(Y))$. In our case, $Y = [\mathrm{Sym}^d X]$ and our evaluation map is

$$ev_{C,\infty} : \widetilde{\mathcal{K}}_{\Xi}(S^d X) \to \bar{\mathrm{I}}([\mathrm{Sym}^d X]).$$

Now we add in ψ classes. Write $\mathcal{U} \to \mathfrak{M}_{g,n}$ for the universal curve and L_i for the conormal bundle at the i-th marked section σ_i:

$$L_i := N_{\sigma_i} \simeq \sigma_i^* T_{\mathcal{U}/\mathfrak{M}_{g,n}}.$$

Use the same notation for their pullback to any moduli space with prestable curves, for example, $\mathcal{K}_{g,n}(Y)$, $\overline{M}_{g,n}$. The classes of $L_i \in K^\circ(\mathfrak{M}_{g,n})$ and their pullbacks to various moduli stacks are referred to as ψ classes.

Beware that maps between moduli spaces involving stabilization do not have the same ψ classes. We compare the ψ class of a map $C \to X$ with the stabilization C^{st} in Section 2.

The stack $\mathcal{K}_{g,n}(Y)$ supports an obstruction theory relative to $\mathfrak{M}_{g,n}$ that lets us define virtual fundamental classes $\mathcal{O}^{\mathrm{vir}}_{\mathcal{K}_{g,n}(Y)}$ in $K_\circ(\mathcal{K}_{g,n}(Y))$. Classes

$$\alpha_1, \dots, \alpha_n \quad \text{in } K_\circ(\bar{\mathrm{I}}(Y))$$

and exponents $e_1, \dots, e_n$ give rise to a (descendent) *Gromov–Witten invariant*

$$\langle \alpha_1 L_1^{e_1}, \dots, \alpha_n L_n^{e_n} \rangle = \chi\Big(\mathcal{O}^{\mathrm{vir}}_{\mathcal{K}_{g,n}(Y)} \otimes \prod ev_i^* \alpha_i \otimes L_i^{e_i}\Big) \quad \text{in } \mathbb{Q}.$$

We are equally interested in power series of these invariants.

When $Y = [\mathrm{Sym}^d X]$, there are two twisted curves $\widetilde{C}' \to \widetilde{C}$ and hence two evaluation maps and two sets of ψ classes. The ψ classes of $\widetilde{C}'$ are the same as the marked points of $\widetilde{C}$ below. The main technical problem in Theorem 3.1 will be converting between classes pulled back along the evaluation map of $\widetilde{C}'$ and that of $\widetilde{C}$. Genuine Gromov–Witten invariants have classes $ev^*\alpha$ pulled back from the evaluation map of $\widetilde{C}$, not that of $\widetilde{C}'$. This convention parrots [Abramovich et al. 2008].

1F. ***Permutation-equivariant K-theory.*** The K-theory of $\overline{\mathcal{K}}_\Xi(S^d X)$ is equivalent to permutation-equivariant K-theory of $\widetilde{M}_\Xi(S^d X)$, as in [Givental 2017]. Ordinary quantum K-theory entails "correlators" defined as the integrals:

$$\langle \alpha_1 L_1^{m_1}, \dots, \alpha_n L_n^{m_n} \rangle := \chi\Big(\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{K}}_\Xi(S^d X)} \otimes \prod ev_i^*(\alpha_i) L_i^{m_i}\Big) \quad \text{in } K_\circ(\mathrm{pt}) = \mathbb{Q}.$$

How can we compute Euler characteristics of $\overline{\mathcal{K}}_\Xi(S^d X)$ by working on $\widetilde{M}_\Xi(S^d X)$? Take the pullback square

$$\begin{array}{ccccc} \widetilde{M}_\Xi(S^d X) & \longrightarrow & \mathrm{pt} & & \\ {\scriptstyle \psi}\downarrow & \ulcorner & \downarrow & & \\ \overline{\mathcal{K}}_\Xi(S^d X) & \longrightarrow & BS_k & \xrightarrow{-_{S_k}} & \mathrm{pt} \end{array}$$

Denote pushforward along $\overline{\mathcal{K}}_\Xi(S^d X) \to BS_k$ by $\chi_{S_k}(-)$. This remembers the S_k-representation on the virtual vector space $\chi(\psi^*(-))$. To get the ordinary Euler characteristic $\chi(-)$ on $\overline{\mathcal{K}}_{\Xi}(S^d X)$, we have to take the virtual S_k-invariants of the virtual representation $\chi_{S_k}(-)$:

$$\chi(-) = (\chi_{S_k}(-))^{S_k}.$$

These should be derived invariants, but we work over $\mathbb{Q}$. The higher-group cohomologies of S_k valued in a representation all vanish, so the distinction is moot.

The class $\psi^*(-)$ on $\widetilde{M}_{\Xi}(S^d X)$ will be a Gromov–Witten invariant where the insertions at the marked points forgotten under ψ are identical. In our case, we only care about insertions away from those forgotten marked points. We will only integrate classes at ordered marked points of both $\overline{\mathcal{K}}_{\Xi}(S^d X)$ and $\widetilde{M}_{\Xi}(S^d X)$.

Even if we insert away from the permuted points $A \subseteq \langle \ell \rangle$, the action of Γ still nontrivially permutes the sections:

Example 1.22. Consider S_3 acting on $\overline{M}_{0,4} = \mathbb{P}^1$ by permuting the last three marked points. The generating function of the S_3-invariant quantum K-invariants on $\overline{M}_{0,4}$ with only L_1 can be calculated

$$\chi\left([\overline{M}_{0,4}/S_3], \frac{1}{1-q_1 L_1}\right) = \frac{1}{(1-q_1^2)(1-q_1^3)}.$$

Indeed, $L_1 = \mathscr{O}(1)$, $\overline{M}_{0,4} = \mathbb{P}^1$, and $H^{\geq 1}(\overline{M}_{0,4}, L_1^d) = 0$ for all $d \geq 0$. The case $d = 1$ has the sections the linear functions on $\mathbb{P}^1$, which are never S_3-invariant (up to Möbius transformations). It is easy to see that $d = 2$ and $d = 3$ have invariant sections. In fact, $[\overline{M}_{0,4}/S_3] = \mathbb{P}(2, 3)$ and the formula follows [Lee and Qu 2014].

1G. *Grothendieck–Riemann–Roch (GRR).* This expository section explains why Grothendieck–Riemann–Roch (GRR) does not reduce equivariant Euler characteristics to ordinary ones on stacks, the way it would for schemes.

Using Grothendieck–Riemann–Roch for schemes, one would expect an equality

$$[\mathcal{O}^{\text{vir}}_{\widetilde{\mathcal{K}}_{\Xi}(S^d X)}] \stackrel{?}{=} [\mathcal{O}^{\text{vir}}_{\mathcal{K}^*_{\Xi}(S^d X)}]^{\oplus \#\Gamma} \quad \text{in } K_\circ(\mathcal{K}^*_{\Xi}(S^d X))_{\mathbb{Q}}.$$

Coupled with the projection formula, this would reduce permutation equivariant integrals to ordinary ones.

For schemes, this holds. Let $\pi : P \to X$ be a G-torsor with P, X schemes for some finite group G. GRR gives a commutative square

$$\begin{array}{ccc} K_\circ(P) & \longrightarrow & A_*(P)_{\mathbb{Q}} \\ {\scriptstyle \pi_*}\downarrow & & \downarrow{\scriptstyle \pi_*} \\ K_\circ(X) & \longrightarrow & A_*(X)_{\mathbb{Q}}. \end{array}$$

The Todd classes cancel out since $P \to X$ is étale and $T_X|_P = T_P$, so one can take the horizontal arrows as the Chern character isomorphisms. Since the pushforward in Chow groups gives $\pi_*[P] = \#G \cdot [X]$ and the horizontal isomorphisms send 1 to 1, we have

$$\pi_*[\mathcal{O}_P] = [\mathcal{O}_X^{\oplus \#G}]. \tag{7}$$

These formulas do not hold for stacks!

Take $P = \mathrm{pt}$, $X = BG$. The analogous GRR square

$$\begin{array}{ccc} K_\circ(\mathrm{pt}) & \longrightarrow & A_*(\mathrm{pt})_\mathbb{C} \\ \downarrow & & \downarrow \\ K_\circ(BG) & \longrightarrow & A_*(BG)_\mathbb{C} \end{array} \tag{8}$$

still commutes [Edidin 2013, Section 5], and the Todd class terms vanish. But the lower horizontal arrow is not multiplicative and does not send 1 to 1! One cannot identify $\pi_*\mathcal{O}_{\mathrm{pt}}$ and $\mathcal{O}_{BG}^{\oplus \#G}$.

Example 1.23. Let $G = \mathbb{Z}/2$ and consider the quotient map $\mathrm{pt} \to BG$. The inertia stack is

$$IBG = BG \sqcup \mathrm{pt},$$

so its rational Chow groups are $A_*(IBG) = \mathbb{C}^{\oplus 2}$. The GRR square for the quotient $\pi : \mathrm{pt} \to BG$ is then

$$\begin{array}{ccc} K_\circ(\mathrm{pt}) & \xrightarrow{ch(-)} & A_*(\mathrm{pt}) \\ \downarrow \pi_* & & \downarrow \pi_* \\ K_\circ(BG) & \xrightarrow{ch(-)} & A_*(IBG) \simeq \mathbb{C}^{\oplus 2}. \end{array}$$

The Todd classes are trivial here.

The Chern character of the trivial representation $\mathcal{O}_{BG}$ is $(1, 1)$. By GRR [Edidin 2013, Theorem 5.4], the Chern character of $\pi_*\mathcal{O}_{\mathrm{pt}}$ is

$$ch(\pi_*\mathcal{O}_{\mathrm{pt}}) = (2, 0) \quad \text{in } A_*(IBG) \simeq \mathbb{C}^{\oplus 2}.$$

We can see that $\pi_*\mathcal{O}_{\mathrm{pt}} \neq \mathcal{O}_{BG}^2$.

Remark 1.24. One can define a Borel equivariant $K_\circ$ theory for stacks in which formula (7) holds using [Noohi 2012]. One can equip them with virtual fundamental classes and study Borel equivariant quantum K-theory.

The problem with (8) is that $\mathrm{pt} \to BG$ introduces stack structure. It is representable, but points in BG have more automorphisms than pt does. We want the opposite of representable, that the automorphism groups of points surject. We can prove a version of (7) in this setting [Herr and Lee].

2. Stabilization and ψ classes

2A. *Costello's lemma for stabilizing ψ classes.* We recall the following three categories defined in [Costello 2006, Section 3]: Γ^u, which contains the label of nodal curves; Γ^t, which contains the label of twisted nodal curves; and Γ^c, which contains labels of twisted marked curves $\mathcal{C}$ and $\mathcal{C}'$ with an étale morphism $\mathcal{C}' \to \mathcal{C}$. They are related by the diagram

$$\Gamma^c \overset{s,t}{\rightrightarrows} \Gamma^t \xrightarrow{r} \Gamma^u,$$

where r, t stand for source and target of the étale map $\mathcal{C}' \to \mathcal{C}$ and r maps $\mathcal{C}$ to its coarse moduli space. These categories will depend on a semigroup A. Furthermore, one can relate the functors between graphs into morphisms of stacks of moduli of curves by applying the functor $\mathfrak{M}$.

Given $\eta \in \Gamma^c$ consider the diagram

$$\mathfrak{M}_\eta \xrightarrow{s} \mathfrak{M}_{s(\eta)} \xrightarrow{r} \mathfrak{M}_{r(s(\eta))} \xrightarrow{\pi} \mathfrak{M}_{\nu(I)},$$

where $I \subsetneq T(s(\eta))$ is a chosen finite subset such that after removing tails in I, $s(\eta)$ remains stable. Let $\nu(I)$ be obtained from $r(s(\eta))$ by removing the tails in I and denote the resulting contraction map by π.

To compare the pullback of ψ classes via s, r and π, we define the notation $S(e, t, I)$ as follows: Consider the diagram

$$\mathfrak{M}_\gamma \to \mathfrak{M}_{r(s(\eta))} \xrightarrow{\pi} \mathfrak{M}_{\nu(I)},$$

where $\gamma \to r(s(\eta))$ is a contraction of Γ^u. Now let $t \in T(\gamma) \setminus I$ and $e \in E(\gamma)$. Let

$$S(e, t, I) := \begin{cases} 1 & \text{if } t \text{ is in a vertex of } \gamma_e \text{ contracted after forgetting the tails } I, \\ 0, & \text{otherwise.} \end{cases}$$

Here γ_e is obtained by contracting all edges of γ except e.

Costello described the pullback of ψ classes in Chow groups.

Lemma 2.1 [Costello 2006, Section 4.1]. *For each $t \in T(\nu(I))$, we have*

$$s^* r^* \pi^*(\psi_t) = m(t)\psi_t - \sum_{f:\gamma\to\eta} S(f, t, I)[\mathfrak{M}_f],$$

where the sum is over $f : \gamma \to \eta$ in Γ^c with $\#E(t(\gamma)) = \#E(f) = 1$, $\mathfrak{M}_f \hookrightarrow \mathfrak{M}_\eta$ is the closed substack supported on the image of f, and

$$S(f, t, I) := \sum_{e \in E(s(\gamma))} m(e)\, S(e, t, I)$$

is the corresponding multiplicity.

A similar result holds in K-theory:

Lemma 2.2. *For each $t \in T(\nu(I))$, we have*

$$s^* r^* \pi^*(L_t) = L_t^{\otimes m(t)} \otimes \mathcal{O}\Big(- \sum_{f:\gamma \to \eta} S(f, t, I)\, \mathfrak{M}_f \Big),$$

with $f : \gamma \to \eta$ and $S(f, t, I)$ defined in previous lemma.

We would like to write $\mathcal{O}(k\mathfrak{M}_f)$ in terms of a torsion sheaf.

2B. *Explicit formulas for stabilizing ψ classes in K-theory.* Before addressing stable maps to a target V, we work on the moduli of curves. Let $\pi : \overline{M}_{g,m+n} \to \overline{M}_{g,m}$. We introduce some notations below.

Decoration. Decorations index trees of rational curves to be contracted under forgetting points and stabilizing (see Figure 3).

For the special case $m = 1$, we denote a decoration of degree r as

$$\underline{a} = (a_{1,1}, \ldots, a_{1,n_1}) \ldots (a_{r,1}, \ldots, a_{r,n_r}).$$

We further assume that

$$\{a_{1,1}, \ldots, a_{1,n_1}, \ldots, a_{r,1}, \ldots, a_{r,n_r}\} \subset \{2, 3, \ldots, n+1\} = [2, n+1]$$

and

$$a_{i,1} < a_{i,2} < \cdots < a_{i,n_i} \quad \text{for all } i.$$

For the general case, a corresponding decoration is denoted by

$$\underline{a} = (\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_m),$$

where each $\underline{a}_i$ is a decoration in the special case $m = 1$ and their disjoint set union forms a subset of $[m+1, m+n]$. We also denote it by $\underline{a}$ if no confusion may occur.

Degeneration strata. Given $m = 1$ and a decoration $\underline{a}$, we define the corresponding degeneration strata of codimension r on $\overline{M}_{g,1+n}$ as

$$\overline{M}_{g,1+n} \supset D_{1,\underline{a}} := \Big(\bigcap_{1 \le i \le r} D_{1a_{1,1}\ldots a_{1,n_1} a_{2,1} \ldots a_{i,n_i}} \Big),$$

where $D_{abcde\ldots}$ is the divisor with markings $abcde\ldots$ lying on the rational tail. This is the closure of the locus where the curves have rational tails indexed by $\underline{a}$.

For the general case, given a decoration $\underline{a}$, we define the corresponding stratum

$$\overline{M}_{g,m+n} \supset D_{\langle m \rangle, \underline{a}} := \bigcap_{1 \le i \le m} D_{i,\underline{a}_i}.$$

Normal bundle. Given $m = 1$ and a strata $D_{1,\underline{a}}$ as above, we have

$$\mathcal{O}^2_{D_{1,\underline{a}}} = \lambda_{-1}(\oplus_{i=1}^r L_{1i})\, \mathcal{O}_{D_{1,\underline{a}}}.$$

Here L_{1i} is the normal bundle of the i-th node. More precisely, it can be described as

$$\mathcal{O}^2_{D_{1,a_{1,1}\dots a_{i,n_i}}} = \lambda_{-1}(\tilde{L_i})\mathcal{O}_{D_{1,a_{1,1}\dots a_{i,n_i}}}.$$

Then we define

$$L_{1i} := \tilde{L_i}|_{D_{1,\underline{a}}}.$$

For the general case, we denote by L_{ji} the normal bundle of the i-th node of the tail containing the marking j for $1 \le j \le m$.

Type. Given $m = 1$ and a decoration $\underline{a}$ or its corresponding strata $D_{A_1\underline{a}}$, we define its *type* to be 0 if

$$a_{1,1} < a_{j,k} \quad \text{for any } (j,k) \neq (1,1),$$

we define its type to be $\underline{l} = (l_1 \dots l_s)$ if

$$\begin{aligned} &a_{1,1} > a_{2,1} > \cdots > a_{l_1-1,1},\\ &a_{l_1,1} > a_{l_1+1,1} > \cdots > a_{l_2-1,1},\\ &\quad\vdots\\ &a_{l_s,1} > a_{l_s+1,1} > \cdots > a_{r,1} \end{aligned}$$

and

$$a_{1,1} > a_{l_1,1} > \cdots > a_{l_s,1}.$$

Given a general decoration $\underline{a} = (\underline{a}_1, \dots, \underline{a}_m)$, we define its type on each $\underline{a}_i$ as in the special case and also denote it by $\underline{l}$ if no confusion may occur. We also say it is of type 0 if $\underline{a}_i$ is of type 0 for all i.

Polynomial corresponding to decoration. Given $m = 1$ and a decoration $\underline{a}$ of type (0), we define its corresponding polynomial with r variables to be

$$F_{1,\underline{a}}(x_1, \dots, x_r) = \left(1 - \prod_{i=1}^r x_i\right).$$

For a partition $\underline{a}$ of type $(l_1, \dots, l_s)$, we define its corresponding polynomial to be

$$F_{1,\underline{a}}(x_1, \dots, x_r) = \sum_{j=0}^s \left(1 - \prod_{i=l_j}^{l_{j+1}-1} x_i\right).$$

Here we set $l_0 = 1$ and $l_{s+1} = r + 1$.

For the general case, given a decoration $\underline{a} = (\underline{a}_1, \dots, \underline{a}_m)$, we define $F_{i,\underline{a}_i}$ as in the special case for all $i = 1, \dots, m$.

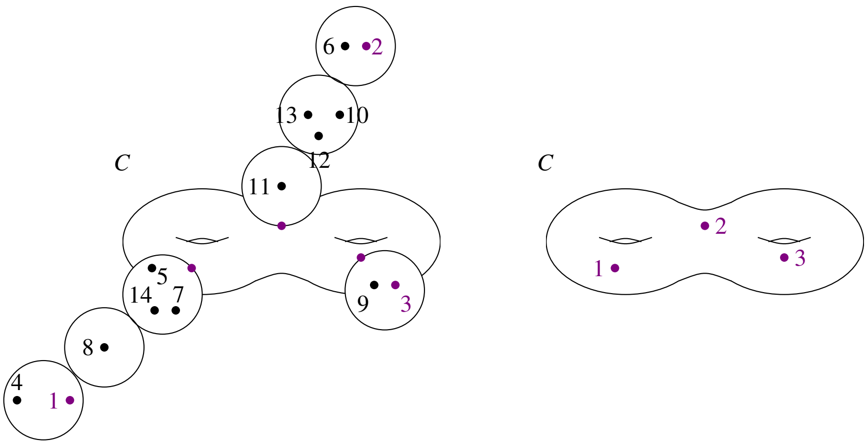

Figure 3. Rational tails corresponding to the decoration $\underline{a} = (\underline{a}_1, \underline{a}_2, \underline{a}_3)$ of type 0, with $\underline{a}_1 = (a_{111})(a_{121})(a_{131}, a_{132}, a_{133}) = (4)(8)(5, 7, 14)$, $\underline{a}_2 = (a_{211})(a_{221}, a_{222}, a_{223})(a_{231}) = (6)(10, 12, 13)(11)$, $\underline{a}_3 = (a_{311}) = (9)$. The smallest numbered marked point on each rational tail (except 1, 2, and 3) must be on the $\mathbb{P}^1$ farthest from the main component to be type 0.

Difference operator. We define the difference operator δ on a (multi)variable polynomial $F(\underline{x}) = F(x_1, \dots, x_n)$ as

$$\delta(F(\underline{x})) := \frac{F(\underline{x}) - \sum_i F(\underline{x})|_{x_i=1} + \cdots + (-1)^n F(\underline{x})|_{x_1=\cdots=x_n=1}}{(1-x_1)\cdots(1-x_n)}.$$

Lemma 2.3. *Let $\pi : \overline{M}_{g,m+n} \to \overline{M}_{g,m}$. Then we have*

$$\pi^*(L_1) = L_1 + \sum_{\underline{a}:\text{type } 0} (-1)^{\operatorname{codim} D_{\underline{a}}} \mathcal{O}_{D_{\underline{a}}} \in K^0(\overline{M}_{g,m+n}).$$

Proof. We decompose π as

$$\overline{M}_{g,m+n} \xrightarrow{\pi_{m+n}} \overline{M}_{g,m+n-1} \to \cdots \to \overline{M}_{g,m+1} \xrightarrow{\pi_{m+1}} \overline{M}_{g,m}$$

and compute $\pi^*(L_1) = \pi^*_{m+n} \dots \pi^*_{m+1}(L_1)$ step by step.

Given a stratum $D \subset \overline{M}_{g,m+n}$ of type 0, we consider the inclusion-exclusion formula

$$\pi^*_{m+n} \mathcal{O}_{\pi_{m+n}(D)} = \sum_{i=1}^{r} \mathcal{O}_{D_i} - \sum_{i<j} \mathcal{O}_{D_i \cap D_j} + \dots,$$

where $\pi^*_{m+n}(\pi_{m+n}(D)) = \bigcup_{i=1}^{r} D_i$ with D_i irreducible strata.

Note that $\mathcal{O}_D$ will show up in either the first or the second term of the right-hand side depending on whether $\mathrm{ft}_{n+m}(D)$ is stable or not. Since $\pi_{n+m}(D)$ is still of type 0. The lemma follows by induction. $\square$

Remark 2.4. This expression is not symmetric with respect to indices $m+1, \ldots,$ $m+n$ since we chose a special order of pull-backs. Different choices of pull-back order will result in different expressions. Nevertheless, any expression will give the same element in $K^0(\overline{M}_{g,n+m})$.

Lemma 2.5. *Let* $\pi : \overline{M}_{g,1+n} \to \overline{M}_{g,1}$, $\underline{a}$ *be a decoration, and* G *be any power series. Then*

$$\operatorname{Coeff}(\mathcal{O}_{D_{1,\underline{a}}}, \pi^* G(L_1)) = \delta(G(L_1 - F_{1,\underline{a}})),$$

where $F_{1,\underline{a}} = F_{1,\underline{a}}(L_{11}, \ldots, L_{1,\deg \underline{a}})$ *with* L_{1j} *defined above. When applying* δ*-operator on the right-hand side, we view* $G(L_1 - F_{1,\underline{a}})$ *as power series with variables* $\{L_{ij}\}$ *and view* L_1 *as constant.*

Proof. We start with a special case $m = 1$ and $n = 2$. By previous lemma, we have

$$\pi^* L_1 = L_1 - \mathcal{O}_{D_{12}} - \mathcal{O}_{D_{13}} - \mathcal{O}_{D_{123}} + \mathcal{O}_{D_{(12)(3)}}.$$

Let $G(x)$ be any power series. To compute $\operatorname{Coeff}(\mathcal{O}_{D_{(12)(3)}}, G(\pi^* L_1))$, we introduce the following process:

Write $x = \mathcal{O}_{D_{12}}$ and $y = \mathcal{O}_{D_{123}}$ and hence $\mathcal{O}_{D_{(12)(3)}} = xy$. Now we have

$$\begin{aligned}
&\operatorname{Coeff}(\mathcal{O}_{D_{(12)(3)}}, G(\pi^* L_1)) \\
&\quad = \frac{G(L_1 - x - y + xy) + G(L_1 - y) + G(L_1 - x) + G(L_1)}{xy}\Big|_{x=1-L_{11}, y=1-L_{12}}. \\
&\quad = \delta\big(G(L_1 - (1 - L_{11} L_{12}))\big).
\end{aligned}$$

Some remarks are in order:

- L_{11} and L_{12} are characterized by $x^2 = (1 - L_{11})\, x$ and $y^2 = (1 - L_{12})\, y$.
- The second equality above follows from the definition of δ. Here we view $G(L_1 - (1 - L_{11} L_{12}))$ as power series in L_{11} and L_{12}.
- $1 - L_{11} L_{12}$ is exactly the polynomial $F_{1,(12)(3)}(L_{11}, L_{12})$, i.e., the polynomial corresponding to the strata $D_{(12)(3)}$.

For general case, we can compute the coefficient of $D_{1,\underline{a}}$ using the above computation process. It suffices to find the polynomial corresponding to $D_{1,\underline{a}}$. To find the polynomial, we compute

$$-\sum_{\substack{\underline{a}' : \text{type } 0 \\ D_{1,\underline{a}'} \supset D_{1,\underline{a}}}} (-1)^{\operatorname{codim} D_{1,\underline{a}'}} \mathcal{O}_{D_{1,\underline{a}'}} \cdot \mathcal{O}_{D_{1,\underline{a}'}}\big|_{D_{1,\underline{a}}} = F_{1,\underline{a}}(L_{11}, \ldots, L_{1 \deg \underline{a}})\, \mathcal{O}_{D_{1,\underline{a}}}.$$

A direct computation shows that it is exactly the polynomial we defined above. □

Theorem 2.6. *Let $\pi : \overline{M}_{g,m+n} \to \overline{M}_{g,m}$ and $G = G(x_1, \dots, x_m)$ be any power series of m variables. Then*

$$\pi^* G(L_1, \dots, L_m) = G(L_1, \dots, L_m) + \sum_{\underline{a}:\text{all decorations}} \mathcal{O}_{D_{\langle m\rangle,\underline{a}}} \delta(G(L_1 - F_{1,\underline{a}_1}, \dots, L_m - F_{m,\underline{a}_m})),$$

where $F_{i,\underline{a}_i} = F_{i,\underline{a}_i}(L_{i1}, \dots, L_{i \deg \underline{a}_i})$ with L_{ij} defined above. When applying the δ-operator on the right-hand side, we view

$$G(L_1 - F_{1,\underline{a}_1}, \dots, L_m - F_{m,\underline{a}_m})$$

as power series with variables $\{L_{ij}\}$ and view L_i as constant.

Proof. It follows from the definition that

$$D_{\langle m\rangle,\underline{a}} := \bigcap_{1 \le i \le m} D_{i,\underline{a}_i},$$

and the following observation that if $F(\underline{x}_1, \dots, \underline{x}_r) = \prod_{i=1}^r F_i(\underline{x}_i)$, then

$$\delta(F(\underline{x}_1, \dots, \underline{x}_r)) = \prod_{i=1}^r \delta(F_i(\underline{x}_i)).$$

Here $\underline{x}_i$ could be multiindices. □

Corollary 2.7. *Let $\pi : \overline{M}_{g,m+n} \to \overline{M}_{g,m}$. Then we have*

$$\pi^* e^{\sum_{i=1}^m r_i L_i} = e^{\sum_{i=1}^m r_i L_i} + \sum_{\underline{a}:\text{all decorations}} \mathcal{O}_{D_{\langle m\rangle,\underline{a}}} \prod_{i=1}^m \delta(e^{r_i(L_i - F_{i,\underline{a}_i})}).$$

Use the same assumptions as in the previous theorem when applying δ-operator on the right-hand side.

Proof. Take G to be $e^{\sum_{i=1}^m r_i L_i}$ and apply the previous theorem. Notice that if $F(\underline{x}_1, \dots, \underline{x}_r) = \prod_{i=1}^r F_i(\underline{x}_i)$, then

$$\delta(F(\underline{x}_1, \dots, \underline{x}_r)) = \prod_{i=1}^r \delta(F_i(\underline{x}_i)).$$

Here $\underline{x}_i$ could be multiindices. □

2C. ***Application to the map $w : \mathcal{K}^*_{\Xi}(S^d X) \to \overline{M}_{g,R}(X)$.*** We continue to write $R = \ell - \#A = I \sqcup J$ for the number of marked points of D. Consider the map $p : \widetilde{\mathfrak{K}}_{\Xi}(BS_d) \to \mathfrak{M}_{g,R}$ sending a triple $C \leftarrow C' \to D$ to D as before.

Write M_i, L_i', and L_i for the cotangent line bundles on D, C' and C respectively. Let $\iota : \langle R\rangle \subseteq \langle \ell\rangle$ be the inclusion of marked points such that $\iota(i) \in C'$ maps to $i \in D$. Write $\bar{\iota}(i)$ for the corresponding point of C.

Given $f:\gamma\to\eta$ lying over D, we define $F_{f,\bar\iota(i)}$ as

$$F_{f,\bar\iota(i)} = F_{i,\underline{a}_i}(L_{i1}^{m(1)},\dots,L_{i\deg\underline{a}_i}^{m(\deg\underline{a}_i)}).$$

The power m is given by the ramification between nodes, $F_{i,\underline{a}_i}$ is defined in the point target case, and $\underline{a}_i$ is the combinatorial type of f. The definition of type is exactly the same as the point case.

Proposition 2.8. *Let $G = G(x_1,\dots,x_R)$ be any power series of R variables. Then*

$$\begin{aligned} &p^*(G(M_1,\dots,M_R)) \\ &= G(L_{\bar\iota(1)}^{m(t_1)},\dots,L_{\bar\iota(R)}^{m(t_R)}) + \sum_{f:\gamma\to\eta} \mathcal{O}_{\mathfrak{M}_f}\,\delta(G(L_{\bar\iota(1)}^{m(t_R)} - F_{f,\bar\iota(1)},\dots,L_{\bar\iota(R)}^{m(t_R)} - F_{f,\bar\iota(R)})),\end{aligned}$$

where δ only applies on variables $\{L_{ij}\}$.

Proof. The proof is similar to the point case. We only need to take care of the difference coming from ramification points.

For the power on marked points, note that $p^*M_i = (L_i')^{m(t_i)} +$ torsion part and $L_i' = L_{\bar\iota(i)}$ since $\widetilde{C}'\to\widetilde{C}$ is étale.

For the definition of $F_{f,\bar\iota(i)}$, if D is a divisor, note that

$$\mathcal{O}_{mD} := \mathcal{O} - \mathcal{O}(-mD) = \mathcal{O} - (\mathcal{O}-\mathcal{O}_D)^m = \delta(L_e^m)\,\mathcal{O}_D,$$

where L_e is characterized by $\mathcal{O}_D^2 = (1-L_e)\mathcal{O}_D$. This explains the power in the definition of $F_{f,\bar\iota(i)}$. □

Remark 2.9. The cover $\widetilde{C}'\to\widetilde{C}$ is étale, so any marked point $i\in\widetilde{C}'$ and its image $s\in\widetilde{C}$ will have the same ψ classes $L_i' = L_s$. Proposition 2.8 writes the pullback $p^*G(M_1,\dots,M_R)$ as a power series in L_i'; plugging in L_s for each L_i' in the fiber of $s\in\widetilde{C}$ rewrites this pullback as a power series $H(L_1,\dots,L_n)$. This power series is invariant under $\underline{\mathrm{Aut}}(\widetilde{C}'|\widetilde{C})$, giving an analogous formula on $\mathcal{K}_{\Xi}^*(S^dX)$:

$$\omega^*G(M_1,\dots,M_R) = H^*(L_1,\dots,L_n).$$

Remark 2.10. The map $\phi:\mathcal{K}_{\Xi}^*(S^dX)\to\overline{\mathcal{K}}_{\Xi}(S^dX)$ is a finite étale map. The ψ classes of $\mathcal{K}_{\Xi}^*(S^dX)$ are pulled back from those of $\overline{\mathcal{K}}_{\Xi}(S^dX)$. If $H(\vec{L})$ is a power series with coefficients $\alpha_I\in K^\circ(\mathcal{K}_{\Xi}^*(S^dX))$ in K-theory, write $H^\phi(\vec{L})$ for the power series on $\overline{\mathcal{K}}_{\Xi}(S^dX)$ with coefficients the pushforwards $\phi_*\alpha_I$ of the coefficients of $H(\vec{L})$. If $\beta\in K_\circ(\overline{\mathcal{K}}_{\Xi}(S^dX))$, $\alpha\in K^\circ(\mathcal{K}_{\Xi}^*(S^dX))$, the projection formula equates

$$\phi_*(\phi^*\beta\otimes H(\vec{L})\otimes\alpha) = \beta\otimes H^\phi(\vec{L})\otimes\phi_*\alpha.$$

Remark 2.11. We sketch how to compute the coefficients $\mathcal{O}_{\mathfrak{M}_f}$ in terms of divisors on the moduli space of curves, and then how to deal with those divisors in quantum K-theory. This makes the power series $H^\phi(\vec{L})$ computable.

A map of covers $f:\gamma\to\eta$ induces a map $f_*:\mathfrak{M}_\gamma\to\mathfrak{M}_\eta$ with image a closed substack $\mathfrak{M}_f$. The map $f_*:\mathfrak{M}_\gamma\to\mathfrak{M}_f$ is a combination of a finite étale torsor for the automorphisms $\underline{\mathrm{Aut}}(f|\eta)$ of the source over the target and a gerbe part for the stacky points. By Proposition 1.9, the gerbe part does not affect the pushforward. The pushforward is then related to the regular representation of the sheaf $\underline{\mathrm{Aut}}(f|\eta)$.

After the above reductions, it remains to explain that quantum K-invariants involving the torsion structure sheaf $\mathcal{O}_S$ supported on boundary strata S can be written in terms of the "usual" quantum K-invariants. In the cohomological Gromov–Witten theory, this is achieved by the splitting axiom, as the boundary stratum consist of the substacks indexed by the "dual graphs" exhibiting the imposed nodes of the general curves. In quantum K-theory a parallel splitting axiom is also available, albeit in a more sophisticated form. See [Givental 2000] and the genus reduction and splitting axioms in [Lee 2004, Section 4.3] for details. One will then have to push these coefficients forward as outlined in Section 3A.

3. Main theorem

Let $\alpha_1,\dots,\alpha_R\in K^\circ(X)$ be classes and form

$$\alpha=\alpha_1\boxtimes\cdots\boxtimes\alpha_R=\alpha_1|_{X^R}\otimes\cdots\otimes\alpha_R|_{X^R}\in K^\circ(X^R).$$

Let $G(\vec{M})$ a power series in the ψ classes of $\overline{M}_{g,R}(X)$ with coefficients arbitrary classes in $K^\circ(\overline{M}_{g,R}(X))$.

Form the Gromov–Witten invariant

$$\chi\Big(\mathcal{O}^{\mathrm{vir}}_{\overline{M}_{g,R}(X)}\otimes ev^*\prod\alpha_i G(\vec{M})\Big)\quad\text{in } K_\circ(\mathrm{pt})=\mathbb{Q}.$$

Write $H(\vec{L})$ for the power series in the ψ-classes L_i which is equal to $\omega^*G(\vec{M})$ by Remark 2.9. Likewise write $H^\phi(\vec{L})$ with the power series with coefficients given by the pushforwards of those of $H(\vec{L})$ as in Remark 2.10.

Theorem 3.1. *Gromov–Witten invariants on $\overline{M}_{g,R}(X)$ are equal to S_k-invariant Euler characteristics on the space $\widetilde{M}_\Xi(S^dX)$ of stable genus zero maps to* $[\mathrm{Sym}^dX]$:

$$\begin{aligned}\chi(\mathcal{O}^{\mathrm{vir}}_{\overline{M}_{g,R}(X)}\otimes ev^*\alpha G(\vec{M}))&=\chi(\mathcal{O}^{\mathrm{vir}}_{\mathcal{K}^*_\Xi(S^dX)}\otimes ev^*\alpha H(\vec{L}))\\&=\chi(\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{K}}_\Xi(S^dX)}\otimes\phi_*(ev^*\alpha)H^\phi(\vec{L}))\\&=\chi_{S_k}\big(\mathcal{O}^{\mathrm{vir}}_{\widetilde{M}_\Xi(S^dX)}\otimes\psi^*\big(\phi_*(ev^*\alpha)H^\phi(\vec{L})\big)\big)^{S_k}.\end{aligned}$$

Proof. For the first equality, apply the projection formula for $w:\mathcal{K}^*_\Xi(S^dX)\to\overline{M}_{g,R}(X)$ and the equality of virtual fundamental classes from Proposition 1.17. The evaluation maps are compatible and the power series $H(\vec{L})$ is designed to be the pullback $w^*G(\vec{M})$.

The second equality results from the projection formula and the pullback

$$\phi^*[\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{K}}_\Xi(S^d X)}] = [\mathcal{O}^{\mathrm{vir}}_{\mathcal{K}^*_\Xi(S^d X)}].$$

The power series $H^\phi(\vec{L})$ applies ϕ_* to the coefficients, so we are using the projection formula for each monomial of $H(\vec{L})$ one at a time.

The third equality results from the S_k-quotient $\psi : \widetilde{M}_\Xi(S^d X) \to \overline{\mathcal{K}}_\Xi(S^d X)$ as in Section 1F. The pullback

$$\begin{array}{ccc} \widetilde{M}_\Xi(S^d X) & \longrightarrow & \mathrm{pt} \\ \psi\downarrow & \ulcorner & \downarrow \\ \overline{\mathcal{K}}_\Xi(S^d X) & \longrightarrow & BS_k \end{array}$$

equates the underlying vector space of $\chi_{S_k}(V)$ with the pullback $\chi(\psi^* V)$. The pushforward map $BS_k \to \mathrm{pt}$ then takes the quotient $(-)^{S_k}$ by S_k. □

We have reduced Gromov–Witten invariants on $\overline{M}_{g,R}(X)$ to some equivariant Euler characteristics on $\widetilde{M}_\Xi(S^d X)$. But are these Euler characteristics actually Gromov–Witten invariants?

Lemma 3.2. *The equivariant Euler characteristic*

$$\chi\Big(\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{K}}_\Xi(S^d X)} \otimes \phi_*\Big(ev^* \prod \alpha_i\Big)\, H^\phi(\vec{L})\Big)$$

is a "Gromov–Witten invariant" in genus zero. In other words, the class $\phi_* ev^* \prod \alpha_i$ can be described as the pullback of a class via the evaluation map of $\overline{\mathcal{K}}_\Xi(S^d X)$.

We spend the rest of the section proving Lemma 3.2. This lemma was essentially left to the reader in [Costello 2006], although it is simpler in Chow groups than in K-theory. Reducing the part $H^\phi(\vec{L})$ is left to the reader, following Remark 2.11 and the process we outline for the evaluation classes.

3A. ***Turning equivariant Euler characteristics on $\widetilde{M}_\Xi(S^d X)$ into proper Gromov–Witten invariants.*** We need to show that $\phi_* ev^* \alpha$ is pulled back from the evaluation map on $\overline{\mathcal{K}}_\Xi(S^d X)$. We first show it is pulled back from a natural map

$$\overline{\mathcal{K}}_\Xi(S^d X) \to ([\mathrm{Sym}^d X])^J \times [\mathrm{Sym}^{\#I} X].$$

In the covers $C' \to C$ parameterized by $\overline{\mathcal{K}}_\Xi(S^d X)$, none of the marked points of C' are ordered. Write $\mathcal{Q}_j \to \overline{\mathcal{K}}_\Xi(S^d X)$ for $j \in J$ for the d-sheeted cover of preimages of $j \in C$ in C'. Likewise, let $\mathcal{P} \to \overline{\mathcal{K}}_\Xi(S^d X)$ be the $S_{\#I}$-torsor ordering the preimages in C' of $\infty \in C$. The product over $\overline{\mathcal{K}}_\Xi(S^d X)$ of all these d-sheeted covers and the $S_{\#I}$-torsor is $\mathcal{K}^*$:

$$\mathcal{K}^*_\Xi(S^d X) = \prod_{\overline{\mathcal{K}}_\Xi(S^d X)} \mathcal{Q}_j \times_{\overline{\mathcal{K}}_\Xi(S^d X)} \mathcal{P} \to \overline{\mathcal{K}}_\Xi(S^d X).$$

Considering the map $\mathcal{Q}_j \to X$ as a d-sheeted cover of $\overline{\mathcal{K}}$, it is parameterized by a map to $[\mathrm{Sym}^d X]$:

$$\begin{array}{ccc} \mathcal{Q}_j & \longrightarrow & X \times [\mathrm{Sym}^{d-1} X] \\ \big\downarrow & \ulcorner & \big\downarrow w_j \\ \overline{\mathcal{K}}_\Xi(S^d X) & \longrightarrow & [\mathrm{Sym}^d X] \end{array}$$

The equivariant map $\mathcal{P} \to X^I$ is also parameterized by a map to a symmetric product stack, but with different total space:

$$\begin{array}{ccc} \mathcal{P} & \longrightarrow & X^{\# I} \\ \big\downarrow & \ulcorner & \big\downarrow w_I \\ \overline{\mathcal{K}}_\Xi(S^d X) & \longrightarrow & [\mathrm{Sym}^{\# I} X] \end{array}$$

There is then a pullback square

$$\begin{array}{ccccccc} \mathcal{K}^*_\Xi(S^d X) & \longrightarrow & \prod \mathcal{Q}_j \times \mathcal{P} & \longrightarrow & X^J \times ([\mathrm{Sym}^{d-1} X])^J \times X^{\# I} & \longrightarrow & X^R \\ \big\downarrow \phi & \ulcorner & \big\downarrow & \ulcorner & \big\downarrow w & & \\ \overline{\mathcal{K}}_\Xi(S^d X) & \xrightarrow{\Delta} & \overline{\mathcal{K}}_\Xi(S^d X)^R & \longrightarrow & ([\mathrm{Sym}^d X])^J \times [\mathrm{Sym}^{\# I} X] & & \end{array} \tag{9}$$

Take classes $\alpha_1, \cdots, \alpha_R \in K^\circ(X)$. Write

$$\alpha := \alpha_1 \boxtimes \cdots \boxtimes \alpha_R$$

for the tensor product of the pullback of these classes to $X^J \times ([\mathrm{Sym}^{d-1} X])^J \times X^I$. Pullback and pushforward in K° theory along cartesian squares commute (5), so the resulting classes on $\overline{\mathcal{K}}_\Xi(S^d X)$ are the same:

$$\phi_*(\alpha|_{\mathcal{K}^*_\Xi(S^d X)}) = (w_* \alpha)|_{\overline{\mathcal{K}}_\Xi(S^d X)} \quad \text{in } K^\circ(\overline{\mathcal{K}}_\Xi(S^d X)). \tag{10}$$

3B. *Evaluation maps.* We have shown that $\phi_* ev^* \alpha$ is pulled back from the map $\overline{\mathcal{K}}_\Xi(S^d X) \to ([\mathrm{Sym}^d X])^J \times [\mathrm{Sym}^{\# I} X]$. We need to compare this with the natural evaluation map on $\overline{\mathcal{K}}_\Xi(S^d X)$.

Gromov–Witten invariants are certain integrals of K theoretic classes pulled back from the evaluation maps as defined in Section Section 1E. To make sense of this, we need to be pedantic about the correct evaluation map for each target stack.

For spaces parameterizing multiple curves C_1, C_2 such as Hurwitz stacks, there is more than one evaluation map. We default to the evaluation maps of the base curve of the cover to be correct.

Example 3.3. The space $\widetilde{M}_\Xi(S^d X)$ parameterizes ramified covers $C' \to C$ together with a map $C' \to X$. There are n ordered marked points of C and ℓ unordered points of C'. The resulting evaluation map is

$$\mathrm{ev} : \widetilde{M}_\Xi(S^d X) \to \bar{\mathrm{I}}([\mathrm{Sym}^d X])^n.$$

Example 3.4. The space $\bar{\mathcal{K}}_\Xi(S^d X)$ is similar to $\widetilde{M}_\Xi(S^d X)$ but harder, because the k gerbey points on C are not even ordered. That part of the evaluation map lands in a symmetric stack of a symmetric stack

$$\mathrm{ev} : \bar{\mathcal{K}}_\Xi(S^d X) \to \bar{\mathrm{I}}\big(([\mathrm{Sym}^d X])^J \times [\mathrm{Sym}^k[\mathrm{Sym}^d X]] \times [\mathrm{Sym}^d X]\big).$$

The evaluation maps fit in a commutative square:

$$\begin{array}{ccc} \widetilde{M}_\Xi(S^d X) & \xrightarrow{\mathrm{ev}} & \bar{\mathrm{I}}\big(([\mathrm{Sym}^d X])^J \times ([\mathrm{Sym}^d X])^k \times [\mathrm{Sym}^d X]\big) \\ \downarrow & & \downarrow \\ \bar{\mathcal{K}}_\Xi(S^d X) & \xrightarrow{\mathrm{ev}} & \bar{\mathrm{I}}\big(([\mathrm{Sym}^d X])^J \times [\mathrm{Sym}^k[\mathrm{Sym}^d X]] \times [\mathrm{Sym}^d X]\big) \end{array}$$

We won't need classes on the middle factor $\bar{\mathrm{I}}([\mathrm{Sym}^k[\mathrm{Sym}^d X]])$. Write ev' for the projection away from this factor on the evaluation map of $\bar{\mathcal{K}}_\Xi(S^d X)$:

$$\mathrm{ev}' : \bar{\mathcal{K}}_\Xi(S^d X) \to \bar{\mathrm{I}}\big(([\mathrm{Sym}^d X])^J \times [\mathrm{Sym}^d X]\big).$$

Remark 3.5. The 2-functor $\bar{\mathrm{I}}(-)$ does not distribute over products because of the representability requirement. For example, the identity map on $B\mathbb{Z}/2 \times B\mathbb{Z}/2$ is representable, but it doesn't factor through a representable map to either factor. It is more accurate to say the evaluation map lands in the product of $\bar{\mathrm{I}}(-)$ applied to each factor, so it is a product of the evaluation maps for each marked point.

Our ramification points are μ_r-banded gerbes mapping to BS_d. Given a map $B\mu_r \to BS_d$, we can extract the set theoretic fiber of the stacky point of $\widetilde{C}' \to \widetilde{C}$ by taking the set-theoretic quotient $\langle d\rangle/\mu_r$ of the corresponding action as in Remark 1.4.

More generally, we have a μ_r-gerbe $\mathcal{G} \to T$ with a map $\mathcal{G} \to [\mathrm{Sym}^d X]$. This means a finite étale cover $\tilde{\mathcal{G}} \to \mathcal{G}$ of degree d. The coarse moduli space of $\mathcal{G}$ is T, and that of $\tilde{\mathcal{G}}$ is a finite étale cover $T' \to T$. The degree k of this cover is some number less than d. The map $\tilde{\mathcal{G}} \to X$ factors through T' because X is a scheme. This procedure gives a map

$$c : \bar{\mathrm{I}}([\mathrm{Sym}^d X]) \to \coprod_{k \le d} [\mathrm{Sym}^k X], \tag{11}$$

sending $\mathcal{G} \to [\mathrm{Sym}^d X]$ to $T' \to X$. We describe this map in detail in the next Section Section 3C.

Because we fixed the discrete data Ξ for $\bar{\mathcal{K}}_\Xi(S^d X)$, we know which component of each factor of $\bigsqcup_{k\le d}[\mathrm{Sym}^k X]$ it maps to

$$\begin{array}{ccc} \bar{\mathcal{K}}_\Xi(S^d X) & \longrightarrow & ([\mathrm{Sym}^d X])^J \times [\mathrm{Sym}^{\#I} X] \\ {\scriptstyle \mathrm{ev}'}\downarrow & & \downarrow{\scriptstyle \mathrm{inc}} \\ \bar{\mathrm{I}}\big(([\mathrm{Sym}^d X])^J \times [\mathrm{Sym}^d X]\big) & \xrightarrow{\ c\ } & \big(\bigsqcup_k [\mathrm{Sym}^k X]\big)^{J\cup\infty} \end{array}$$

Let β be a class in $K^\circ\big(([\mathrm{Sym}^d X])^J \times [\mathrm{Sym}^{\#I} X]\big)$. Because the map inc is an inclusion of components, we have

$$\mathrm{inc}^* \, \mathrm{inc}_* \, \beta = \beta.$$

Then the pullback $\beta|_{\bar{\mathcal{K}}_\Xi(S^d X)}$ is the same as the class

$$\mathrm{ev}'^* \, c^* \, \mathrm{inc}_* \, \beta \quad \text{in } K^\circ(\bar{\mathcal{K}}_\Xi(S^d X)).$$

Proof of Lemma 3.2. Take $\beta = w_* \alpha$ above. Then

$$\phi_*(\alpha|_{\mathcal{K}^*_\Xi(S^d X)}) = (w_* \alpha)|_{\bar{\mathcal{K}}_\Xi(S^d X)} = \mathrm{ev}'^* \, c^* \, \mathrm{inc}_* \, w_* \, \alpha,$$

using equation (10) and the discussion immediately above. This expresses the factor $\phi_*(\alpha|_{\mathcal{K}^*_\Xi(S^d X)})$ in the Gromov–Witten invariant as a class pulled back via the evaluation map on $\bar{\mathcal{K}}_\Xi(S^d X)$. □

3C. *Describing the map c.* We describe the map c in detail using the inertia stack $\mathrm{I}([\mathrm{Sym}^d X])$. This section is purely expository, and an example is given at the end.

The inertia stack of a quotient stack $[Y/G]$ is the disjoint union of the quotients

$$I([Y/G]) = \bigsqcup_{g\in G \text{ conj classes}} [Y^g/C_g]$$

of the fixed locus Y^g by the centralizer $C_g \subseteq G$.

For $[\mathrm{Sym}^d X]$, we have $Y = X^d$ and $G = S_d$. Conjugacy classes of S_d are indexed by cycle types, the multiset of lengths of cycles. For example,

$$g_1 := (12)(34)(5) \in S_5 \mapsto \{2, 2, 1\}, \quad g_2 := (234)(761)(5)(89) \in S_9 \mapsto \{3, 3, 1, 2\}.$$

Let N_s be the number of cycles of length s, $N := \sum N_s$ the total number of cycles, and t the cardinality of the set of distinct lengths in the cycle type. We include all cycles of length one N_1, so $\sum i N_i = d$. For g_1, g_2 above,

$$g_1 \mapsto N_1 = 1, N_2 = 2, t = 2, \quad g_2 \mapsto N_1 = 1, N_2 = 1, N_3 = 2, t = 3.$$

The centralizer of a cycle type g is

$$C_g = S_{N_1} \times S_{N_2} \rtimes (\mathbb{Z}/2)^{N_2} \times S_{N_3} \rtimes (\mathbb{Z}/3)^{N_3} \cdots \times S_{N_t} \rtimes (\mathbb{Z}/t)^{N_t}.$$

The fixed locus $(X^d)^g \subseteq X^d$ is the multidiagonal

$$(X^d)^g = \{(x_1, \cdots, x_d) \mid x_i = x_j \text{ if } i, j \text{ in the same cycle}\},$$

so $(X^d)^g \cong X^N$. For g_1, g_2, we have

$$(X^d)^{g_1} = \{(x, x, y, y, z)\} \in X^5, \quad (X^d)^{g_2} = \{(x, x, x, y, y, y, z, w, w)\} \subseteq X^9.$$

Write $H_i = S_{N_i} \rtimes (\mathbb{Z}/i)^{N_i} = \mathbb{Z}/i \wr S_{N_i}$. There is an exact sequence

$$0 \to (\mathbb{Z}/i)^{N_i} \to H_i \to S_{N_i} \to 1$$

and a splitting $S_{N_i} \subseteq H_i$. The subgroups $\mathbb{Z}/i$ act trivially on the diagonal fixed locus $X \subseteq X^{N_i}$, so the stack quotient is a trivial gerbe

$$[X/(\mathbb{Z}/i)] = X \times B\mathbb{Z}/i.$$

The quotient $X^N/\prod H_i$ is the product of symmetric products

$$X^N/\prod H_i = \prod_i [\mathrm{Sym}^{N_i}[X/(\mathbb{Z}/i)]] = \prod_i [\mathrm{Sym}^{N_i}(X \times B\mathbb{Z}/i)].$$

On each component, there is a map to a single symmetric product

$$\prod [\mathrm{Sym}^{N_i}(X \times B\mathbb{Z}/i)] \to \prod [\mathrm{Sym}^{N_i} X] \xrightarrow{c'} [\mathrm{Sym}^N X].$$

The second map c' takes t covers $T'_i \to T$ of degrees N_i and assembles them into one cover $T' = \bigsqcup T'_i \to T$ of degree N. This gives a map

$$\bar{c} : I([\mathrm{Sym}^d X]) \to \bigsqcup_{N \le d} [\mathrm{Sym}^N X]. \tag{12}$$

The reader can check this coincides with the map c defined in (11).

Lemma 3.6. *There is a commutative diagram*

$$\begin{array}{ccc} I([\mathrm{Sym}^d X]) & \overset{\bar{c}}{\searrow} & \\ \downarrow & & \\ \bar{\mathrm{I}}([\mathrm{Sym}^d X]), & \xrightarrow[c]{} & \bigsqcup_{k \le d} [\mathrm{Sym}^k X] \end{array}$$

where c is the map (11) and $\bar{c}$ is (12).

We explain the above from the point of view of covers. A T-point of $I([\mathrm{Sym}^d X])$ is a map $T \times B\mu_r \to [\mathrm{Sym}^d X]$, which is a d-sheeted cover $P \to T \times B\mu_r$. Write P_0

for the pullback d-sheeted cover of T:

$$\begin{array}{ccc} P_0 & \longrightarrow & P \\ \downarrow & \ulcorner & \downarrow \\ T & \longrightarrow & T \times B\mu_r \end{array}$$

so $P = [P_0/\mu_r]$.

Fix a generator $\mathbb{Z}/r \simeq \mu_r$. If T is a geometric point, $P_0 \simeq \langle d \rangle$ and $\mu_r \circlearrowright P_0$ is an element $\sigma \in S_d$. Its order is the lcm of the cycle lengths, which divides r; r is equal to the lcm when the map $T \times B\mu_r \to [\mathrm{Sym}^d X]$ is representable. Reordering $P_0 \simeq \langle d \rangle$ conjugates σ, so σ is well defined as a conjugacy class.

Even if T is not a geometric point, this defines a locally constant function

$$T \mapsto S_d/_{ad} S_d, \quad t \mapsto [\sigma]$$

from T to the conjugacy classes of S_d. This decomposes $I([\mathrm{Sym}^d X])$ into components corresponding to cycle type, or partitions of d.

Given a partition $d = \sum i N_i$ corresponding to the cycle type of $\sigma \in S_d$, a point $T \to I([\mathrm{Sym}^d X])$ factors through the corresponding component if its fibers at geometric points are isomorphic to $\langle d \rangle / \sigma$. The N_i different orbits of i points may be interchanged in families over T, and each such orbit may vary in a $B\mathbb{Z}/i$ family. The collection of orbits of i points is parameterized by the stack

$$[\mathrm{Sym}^{N_i}(B\mathbb{Z}/i)].$$

We have $\prod [\mathrm{Sym}^{N_i}(X \times B\mathbb{Z}/i)]$ instead because we also need a map to X.

The case $d = 5$. If $d = 5$, the seven cycle types/partitions are

$$5, \quad 4+1, \quad 3+2, \quad 3+1+1, \quad 2+2+1, \quad 2+1+1+1, \quad 1+1+1+1+1.$$

The corresponding components of $I([\mathrm{Sym}^5 X])$ are

$$X \times B\mathbb{Z}/5, \quad X^2 \times B\mathbb{Z}/4, \quad X^2 \times B\mathbb{Z}/3 \times B\mathbb{Z}/2, \quad X/(\mathbb{Z}/3) \times X^2/S_2,$$
$$X \times X^2/(S_2 \rtimes (\mathbb{Z}/2)^2), \quad X/\mathbb{Z}/2 \times X^3/S_3, \quad X^5/S_5.$$

Project away from the cyclic gerbes $B\mathbb{Z}/i$:

$$X, \quad X^2, \quad X^2, \quad X \times [\mathrm{Sym}^2 X], \quad [\mathrm{Sym}^2 X] \times X, \quad X \times [\mathrm{Sym}^3 X], \quad [\mathrm{Sym}^5 X].$$

For example, consider a trivial gerbe $b : B\mu_4 \to [\mathrm{Sym}^5 X]$ mapping to the symmetric product. This corresponds to a 5-sheeted cover $P \to B\mu_4$, counting stacky multiplicity.

The composite $\mathrm{pt} \to B\mu_4 \to [\mathrm{Sym}^5 X] \to BS_5$ parameterizes a 5-sheeted cover of the point, which we trivialize and view as $\langle 5 \rangle$. The action of μ_4 on $\langle 5 \rangle$ can be viewed as a map $\mu_4 \to S_5$. Choose a generator to identify $\mu_4 \simeq \mathbb{Z}/4$ and let $\sigma \in S_5$ be the image of the generator.

The element σ has order dividing 4. Take, for example,

$$\sigma = g_1 = (12)(34)(5).$$

Its order 2 is not 4, so the classifying map b is not representable.

The stack quotient of $\langle 5 \rangle$ by $\mathbb{Z}/4 \cdot \sigma$ is a disjoint union

$$B\mu_2 \sqcup B\mu_2 \sqcup B\mu_4.$$

This is the 5-sheeted cover $P \to B\mu_4$. The corresponding point $\mathrm{pt} \to I([\mathrm{Sym}^5 X])$ factors through the component of the partition $2+2+1$, i.e., $X \times X^2/(S_2 \rtimes (\mathbb{Z}/2)^2)$ above.

The cover comes with a map $P \to X$. The map $\bar{c}$ takes coarse moduli spaces of $P \to B\mu_4$, obtaining $\langle 3 \rangle \to \mathrm{pt}$. The map $P \to X$ factors through $\langle 3 \rangle$ because X is a scheme with no stack structure. This sends the component $X \times X^2/(S_2 \rtimes (\mathbb{Z}/2)^2)$ to $[\mathrm{Sym}^3 X]$.

4. Elliptic curves example

We apply our theorem in the case of elliptic curves, with $g = R = 1$, $d = 2$, and $X = \mathrm{pt}$. The symmetric product is the classifying stack $[\mathrm{Sym}^d X] = BS_d$. The results are reassuring but not surprising. See Figure 4.

We know the map w is proper and birational. For elliptic curves, more is true.

Lemma 4.1. *The map* $w : \widetilde{\mathcal{K}}^*_{\Xi^*}(BS_d) \to \overline{M}_{1,1}$ *is an isomorphism.*

Proof. The map on coarse moduli spaces is an isomorphism $\mathbb{P}^1 \;=\; \mathbb{P}^1$. It remains to show the stack structure is the same; i.e., the automorphism groups of the admissible covers $C' \to C$ are the same as that of the stabilization $D = C'^{\mathrm{st}}$. We already checked this for a generic elliptic curve in the proof of Lemma 1.16.

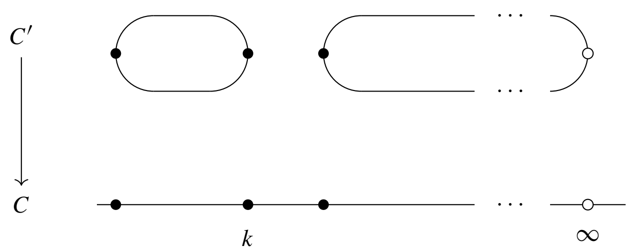

Figure 4. An elliptic curve C' and its double cover of $C = \mathbb{P}^1$ simply ramified at four points: Ξ with $g = 1$, $d = 2$, $k = 0$. Marked points are white if remembered and black if forgotten under the map $p : \widetilde{\mathfrak{K}}_\Xi(BS_d) \to \mathfrak{M}_{1,1}$.

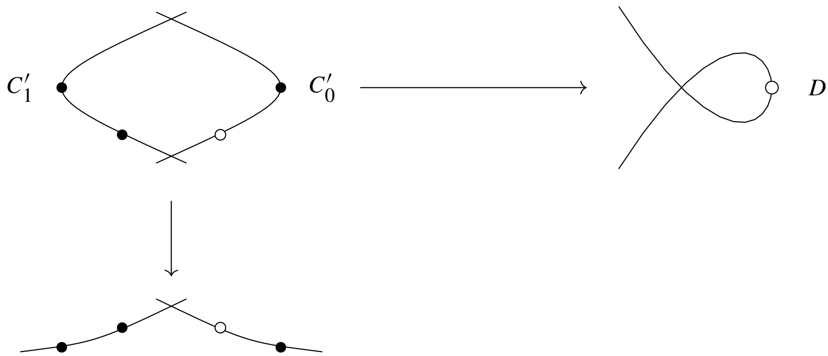

Figure 5. The admissible cover $C' \to C$ which stabilizes to the nodal cubic $C'^{\mathrm{st}} = D$. The components of C' are labeled C_0', C_1' and they each doubly cover a component of C. The pair of nodes P is the intersection $C_0' \cap C_1'$ of the components.

We want to show $\underline{\mathrm{Aut}}(C' \to C) = \underline{\mathrm{Aut}}(D)$. Again, it helps that automorphisms of the covering map are a subgroup of automorphisms of the source $\underline{\mathrm{Aut}}(C' \to C) \subseteq \underline{\mathrm{Aut}}(C')$. These automorphisms must send ramification points to ramification points, and it will be clear that they also send the preimage $I \in C'$ of infinity to itself.

For all the smooth elliptic curves D, the source is already stable $C' = D$ and the unique map $C' \to C = \mathbb{P}^1$ is the quotient by the elliptic involution. This identifies $\underline{\mathrm{Aut}}(C' \to C) \simeq \underline{\mathrm{Aut}}(D)$ for all smooth elliptic curves with automorphism group $\mathbb{Z}/2$. The smooth curves $j = 0$, 1728 remain, as does the singular cubic.

The curve $j = 1728$ has the equation

$$y^2 = x^3 - x.$$

The automorphism group $\mathbb{Z}/4$ is generated by $(x, y) \mapsto (-x, iy)$. This commutes with the automorphism $x \mapsto -x$ of $\mathbb{P}^1$. Likewise $j = 0$ has the equation

$$y^2 = x^3 - 1.$$

Letting ζ be a sixth root of unity, the automorphism group is generated by the map $(x, y) \mapsto (\zeta^2 x, \zeta^3 y)$. This also commutes with an automorphism of $\mathbb{P}^1$.

For the singular elliptic curve D, the preimage is an admissible cover $C' \to C$ with both source and target reducible. See Figure 5. Each consists of two components, labeled C_i', C_i for $i = 0, 1$. Assume that the point in $I \subseteq C'$ lies on C_0'.

The map restricts to two double covers $C_i' \to C_i$ of $\mathbb{P}^1$'s. The preimage $P \subseteq C'$ of the singular point of C is a pair of nodes joining C_0', C_1'.

Any automorphism of C' must restrict to an automorphism of each component because of the marked point. Any automorphism φ of C' restricts to an automorphism of the pair of nodes P which determines φ. This is because φ must preserve at least three points on each component, the nodes and the ramification points. All such automorphisms lie over C, so $\underline{\mathrm{Aut}}(C' \to C) = \underline{\mathrm{Aut}}(P) = \mathbb{Z}/2$. This is the same automorphism group of the singular stabilization $D = C'^{\mathrm{st}}$, the nodal cubic. $\square$

The map $\phi : \mathcal{K}^*_\Xi(BS_d) \to \bar{\mathcal{K}}_\Xi(BS_d)$ is also an isomorphism by Remark 1.20, so Theorem 3.1 merely says that quantum K invariants on $\overline{M}_{1,1} = \mathbb{P}(4, 6)$ are equivariant Euler characteristics on its natural S_3-cover by $\widetilde{M}_\Xi(BS_d)$.

This cover is pulled back from the quotient $\mathbb{P}^1 \to \mathbb{P}(2, 3)$ by S_3 acting on $\lambda \in \mathbb{P}^1$. This map classically parameterizes the Legendre form

$$y^2 = x(x-1)(x-\lambda)$$

of an elliptic curve. The map $\widetilde{M}_\Xi(BS_d) \to \bar{\mathcal{K}}_\Xi(BS_d) = \overline{M}_{1,1}$ then fits in a pullback square

$$\begin{array}{ccc} \widetilde{M}_\Xi(BS_d) & \longrightarrow & \overline{M}_{1,1} \\ \downarrow & \ulcorner & \downarrow \\ \mathbb{P}^1 & \xrightarrow{/S_3} & \mathbb{P}(2,3) \end{array}$$

with vertical arrows μ_2-gerbes.

Acknowledgements

The authors wish to thank Sarah Arpin, Peter Bakic, Sebastian Bozlee, Renzo Cavalieri, Kevin Costello, Dan Edidin, Adeel Khan, Rufus Lawrence, Sam Molcho, Hsian-Hua Tseng, Jonathan Wise, and the Mathoverflow community for their help and support. This project originated in the work of Herr with Wise [Herr and Wise 2023].

Chou and Lee wish to acknowledge the partial supports from Academia Sinica, National Science and Technology Council, Simons Foundation and the University of Utah. Herr thanks Leiden University, University of Utah and the National Science Foundation for its support through RTG grant #1840190.

References

[Abramovich et al. 2002] D. Abramovich, T. Graber, and A. Vistoli, "Algebraic orbifold quantum products", pp. 1–24 in *Orbifolds in mathematics and physics* (Madison, WI, 2001), Contemp. Math. **310**, Amer. Math. Soc., Providence, RI, 2002. MR Zbl

[Abramovich et al. 2008] D. Abramovich, T. Graber, and A. Vistoli, "Gromov–Witten theory of Deligne–Mumford stacks", *Amer. J. Math.* **130**:5 (2008), 1337–1398. MR Zbl

[Behrend and Fantechi 1997] K. Behrend and B. Fantechi, "The intrinsic normal cone", *Invent. Math.* **128**:1 (1997), 45–88. MR Zbl

[Buch and Mihalcea 2011] A. S. Buch and L. C. Mihalcea, "Quantum K-theory of Grassmannians", *Duke Math. J.* **156**:3 (2011), 501–538. MR Zbl

[Buch et al. 2013] A. S. Buch, P.-E. Chaput, L. C. Mihalcea, and N. Perrin, "Finiteness of cominuscule quantum K-theory", *Ann. Sci. Éc. Norm. Supér.* (4) **46**:3 (2013), 477–494. MR Zbl

[Buch et al. 2020] A. S. Buch, S. Chung, C. Li, and L. C. Mihalcea, "Euler characteristics in the quantum K-theory of flag varieties", *Selecta Math.* (*N.S.*) **26**:2 (2020), art. id. 29. MR Zbl

[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), Contemp. Math. **310**, Amer. Math. Soc., Providence, RI, 2002. MR Zbl

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

[Chou et al. 2023] Y.-C. Chou, L. Herr, and Y.-P. Lee, "The log product formula in quantum K-theory", *Math. Proc. Cambridge Philos. Soc.* **175**:2 (2023), 225–252. MR Zbl

[Costello 2006] K. Costello, "Higher genus Gromov–Witten invariants as genus zero invariants of symmetric products", *Ann. of Math.* (2) **164**:2 (2006), 561–601. MR Zbl

[Edidin 2013] D. Edidin, "Riemann–Roch for Deligne–Mumford stacks", pp. 241–266 in *A celebration of algebraic geometry*, Clay Math. Proc. **18**, Amer. Math. Soc., Providence, RI, 2013. MR Zbl

[Edidin et al. 2001] D. Edidin, B. Hassett, A. Kresch, and A. Vistoli, "Brauer groups and quotient stacks", *Amer. J. Math.* **123**:4 (2001), 761–777. MR Zbl

[Givental 2000] A. Givental, "On the WDVV equation in quantum K-theory", *Michigan Math. J.* **48** (2000), 295–304. MR Zbl

[Givental 2017] A. Givental, "Permutation-equivariant quantum K-theory, I: Definitions. Elementary K-theory of $\overline{\mathcal{M}}_{0,n}/S_n$", *Mosc. Math. J.* **17**:4 (2017), 691–698. MR Zbl

[Givental and Lee 2003] A. Givental and Y.-P. Lee, "Quantum K-theory on flag manifolds, finite-difference Toda lattices and quantum groups", *Invent. Math.* **151**:1 (2003), 193–219. MR Zbl

[Herr and Lee] L. Herr and Y. P. Lee, "Quantum K-theory and log geometry", preprint. In preparation.

[Herr and Wise 2023] L. Herr and J. Wise, "Costello's pushforward formula: errata and generalization", *Manuscripta Math.* **171**:3-4 (2023), 621–642. MR Zbl

[Jockers and Mayr 2019] H. Jockers and P. Mayr, "Quantum K-theory of Calabi–Yau manifolds", *J. High Energy Phys.* **2019** (2019), art. id. 11. MR Zbl

[Jockers and Mayr 2020] H. Jockers and P. Mayr, "A 3d gauge theory/quantum K-theory correspondence", *Adv. Theor. Math. Phys.* **24**:2 (2020), 327–457. MR Zbl

[Lee 2004] Y.-P. Lee, "Quantum K-theory, I: Foundations", *Duke Math. J.* **121**:3 (2004), 389–424. MR Zbl

[Lee and Qu 2014] Y.-P. Lee and F. Qu, "Euler characteristics of universal cotangent line bundles on $\overline{\mathcal{M}}_{1,n}$", *Proc. Amer. Math. Soc.* **142**:2 (2014), 429–440. MR Zbl

[Noohi 2012] B. Noohi, "Homotopy types of topological stacks", *Adv. Math.* **230**:4-6 (2012), 2014–2047. MR Zbl

[Okounkov 2017] A. Okounkov, "Lectures on K-theoretic computations in enumerative geometry", pp. 251–380 in *Geometry of moduli spaces and representation theory*, IAS/Park City Math. Ser. **24**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

[Qu 2018] F. Qu, "Virtual pullbacks in K-theory", *Ann. Inst. Fourier* (*Grenoble*) **68**:4 (2018), 1609–1641. MR Zbl

[Stacks 2005–] "The Stacks project", 2005–, available at https://stacks.math.columbia.edu.

[Tonita and Tseng 2013] V. Tonita and H.-H. Tseng, "Quantum orbifold Hirzebruch–Riemann–Roch theorem in genus zero", preprint, 2013. arXiv 1307.0262

[Zhang] M. Zhang, "Quantum K-theory of toric stacks", preprint, available at https://tinyurl.com/ZhangWb.

Received June 2, 2023. Revised April 14, 2024.

YOU-CHENG CHOU
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI
TAIWAN
bensonchou@gate.sinica.edu.tw

LEO HERR
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF LEIDEN
LEIDEN
NETHERLANDS
herr@math.utah.edu

YUAN-PIN LEE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI
TAIWAN
ypleemath@gate.sinica.edu.tw

PACIFIC JOURNAL OF MATHEMATICS
Vol. 330, No. 1, 2024
https://doi.org/10.2140/pjm.2024.330.123

UNKNOTTED CURVES ON GENUS-ONE SEIFERT SURFACES OF WHITEHEAD DOUBLES

SUBHANKAR DEY, VERONICA KING, COLBY T. SHAW,
BÜLENT TOSUN AND BRUCE TRACE

We consider homologically essential simple closed curves on Seifert surfaces of genus-one knots in S^3, and in particular those that are unknotted or slice in S^3. We completely characterize all such curves for most twist knots: they are either positive or negative braid closures; moreover, we determine exactly which of those are unknotted. A surprising consequence of our work is that the figure-eight knot admits infinitely many unknotted essential curves up to isotopy on its genus-one Seifert surface, and those curves are enumerated by Fibonacci numbers. On the other hand, we prove that many twist knots admit homologically essential curves that cannot be positive or negative braid closures. Indeed, among those curves, we exhibit an example of a slice but not unknotted homologically essential simple closed curve. We continue our investigation of unknotted essential curves for arbitrary Whitehead doubles of nontrivial knots, and obtain that there is precisely one unknotted essential simple closed curve in the interior of a double's standard genus-one Seifert surface. As a consequence we obtain many new examples of 3-manifolds that bound contractible 4-manifolds.

1. Introduction

Suppose $K \subseteq S^3$ is a genus g knot with Seifert surface Σ_K. Let b be a curve in Σ_K which is *homologically essential* — that is, it is not separating Σ_K — and *a simple closed curve* — that is, it has one component and does not intersect itself. Furthermore, we will focus on those that are *unknotted* or *slice* in S^3 — that is, each bounds a disk in S^3 or B^4. In this paper we seek to make progress on the following problem:

Problem. *Characterize and, if possible, list all such curves b for the pair (K, Σ_K), where K is a genus-one knot and Σ_K its Seifert surface.*

Our original motivation for studying this problem comes from the intimate connection between unknotted or slice homologically essential curves on a Seifert

MSC2010: 57K33,57K43,32E20.
MSC2020: primary 57K30; secondary 57K10.
Keywords: unknotted curves on Seifert surfaces, contractible 4-manifolds.

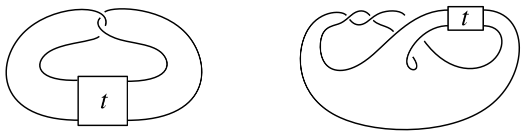

Figure 1. On the left is the twist knot K_t, where the box contains t full right-handed twists if $t \in \mathbb{Z}_{>0}$, and $|t|$ full left-handed twists if $t \in \mathbb{Z}_{<0}$. On the right is the standard Seifert surface for K_t.

surface of a genus-one knot and 3-manifolds that bound contractible 4-manifolds. We defer the detailed discussion of this connection to Section 1B, where we also provide some historical perspective. For now, however, we will focus on getting a hold on the stated problem above for a class of genus-one knots, and as we will make clear in the next few results, this problem is already remarkably interesting and fertile on its own.

1A. ***Main results.*** A well-studied class of genus-one knots is that of the so-called twist knot $K = K_t$, which is described by the diagram on the left of Figure 1 (cf. [Casson and Gordon 1978, page 182]). We note that with this convention K_{-1} is the right-handed trefoil $T_{2,3}$ and K_1 is the figure-eight knot 4_1. We will consider the genus-one Seifert surface Σ_K for $K = K_t$, as depicted on the right of Figure 1.

Theorem 1.1. *Let $t \leq 2$. Then the genus-one Seifert surface Σ_K of $K = K_t$ admits infinitely many homologically essential, unknotted curves if and only if $t = 1$, that is, if and only if K is the figure-eight knot 4_1.*

Indeed, we can be more precise and characterize all homologically essential, simple closed curves on Σ_K, from which Theorem 1.1 follows easily. To state this we recall an essential simple closed curve c on Σ_K can be represented (almost uniquely) by a pair of nonnegative integers (m, n), where m is the number of times $c = (m, n)$ runs around the left band and n is the number of times it runs around the right band in Σ_K. Moreover, since c is connected, we can assume $\gcd(m, n) = 1$. Finally, to uniquely describe c, we call c an ∞ *curve* if its orientation switches from one band to the other or a *loop curve* if it has the same orientation on both bands (see Figure 9).

Theorem 1.2. *Let $K = K_t$ be a twist knot and Σ_K its Seifert surface, as in Figure 1. Then:*

(1) *For $K = K_t$ with $t \leq -1$, we can characterize all homologically essential simple closed curves on Σ_K as the closures of negative braids in Figure 10. In the case of the right-handed trefoil $K_{-1} = T_{2,3}$, exactly 6 of these (see Figure 2) are unknotted in S^3. For $t < -1$, exactly 5 of these (see Figure 4) are unknotted in S^3.*

(2) *For $K = K_1 = 4_1$, we can characterize all homologically essential simple closed curves on Σ_K as the closures of braids in Figure 15. A curve on this surface is unknotted in S^3 if and only if it is*

(a) *a trivial curve* $(1, 0)$ *or* $(0, 1)$,
(b) *an* ∞ *curve in the form of* (F_{i+1}, F_i), *or*
(c) *a loop curve in the form of* (F_i, F_{i+1}), *where* F_i *represents the* i*-th Fibonacci number (see Figure 3).*

For twist knots $K = K_t$ with $t > 1$ the situation is more complicated. Under further hypothesis on the parameters m, n we can obtain results similar to those in Theorem 1.2, and these will be enough to extend the theorem entirely to the case of $K = K_2$, so-called Stevedore's knot 6_1 (here we use the KnotInfo database [Livingston and Moore 2024] for identifying small knots and their various properties). More precisely we have:

Theorem 1.3. *Let* $K = K_t$ *be a twist knot and* Σ_K *its Seifert surface, as in Figure 1. Then*:

(1) *When* $t > 1$ *and* $m < n$, *we can characterize all homologically essential simple closed curves on* Σ_K *as the closures of positive braids in Figure 24*(a),(b). *Exactly* 5 *of these* (*see Figure 4*) *are unknotted in* S^3.
(2) *When* $t > 1$ *and* $m > n$:
   (a) *If* $m - tn > 0$, *then we can characterize all homologically essential simple closed curves on* Σ_K *as the closures of negative braids in Figures 28 and 31. Exactly* 5 *of these* (*see Figure 4*) *are unknotted in* S^3.
   (b) *If* $m - n < n$ *and the curve is an* ∞ *curve*, *then we can characterize all homologically essential simple closed curves on* Σ_K *as the closures of positive braids in Figure 29. Exactly* 5 *of these* (*see Figure 4*) *are unknotted in* S^3.
(3) *For* $K = K_2 = 6_1$, *we can characterize all homologically essential simple closed curves on* Σ_K *as the closures of positive or negative braids. Exactly* 5 *of these* (*see Figure 4*) *are unknotted in* S^3.

What Theorem 1.3 cannot cover is the case $t > 2$, $m > n$ and $m - tn < 0$ or when $m - n < n$ and the curve is a loop curve. Indeed in this range *not* every homologically essential curve is a positive or negative braid closure. For example, when $(m, n) = (5, 2)$ and $t = 3$ one obtains that the corresponding essential ∞ curve, as a smooth knot in S^3, is the knot $m(5_2)$ (see Figure 34 in Section 5 for a verification of this), and for $(m, n) = (7, 3)$ and $t = 3$, the corresponding knot is 10_{132}; both of these are known (e.g., via the KnotInfo database [Livingston and Moore 2024]) not to be positive braid closures — coincidentally, these knots are not unknotted or slice. Moreover we can explicitly demonstrate (see below) that if one removes the assumption of "∞" from part (2)(b) in Theorem 1.3, then the conclusion claimed there fails for certain loop curves when $t > 2$. A natural question is then whether for knots $K = K_t$ with $t > 2$, $m > n$ and $m - tn < 0$ or an $m - n < n$ loop curve,

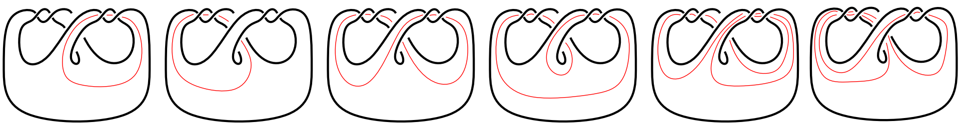

Figure 2. It can easily be shown that these six curves, from left to right, (0, 1), (1, 0), (1, 1) ∞, (1, 1) loop, (1, 2) ∞ and (2, 1) ∞, on Σ_K are unknotted in S^3. One can easily check that the other (1, 2) and (2, 1) curves (that is, the (1, 2) loop and (2, 1) loop curves) both yield the left-handed trefoil $T_{2,-3}$, and hence they are not unknotted in S^3.

there exists unknotted or slice curves on Σ_K other than those listed in Figure 4. A follow-up question will be whether there exists slice but not unknotted curves on Σ_K for some $K = K_t$. We can answer the latter question in the affirmative as follows:

Theorem 1.4. *Let $K = K_t$ be a twist knot with $t > 2$ and Σ_K its Seifert surface, as in Figure 1, and consider the loop curve (m, n) with $m = 3$, $n = 2$ on Σ_K. Then this curve, as a smooth knot in S^3, is the pretzel knot $P(2t-5, -3, 2)$. This knot is never unknotted but it is slice (exactly) when $t = 4$, in which case this pretzel knot is also known as the curious knot 8_{20}.*

Remark 1.5. We note that the choices of m, n values made in Theorem 1.4 are somewhat special in that they yielded an infinite family of pretzel knots, and that it includes a slice but not unknotted curve. Indeed, by using the work of Rudolph [1993], we can show (see Proposition 3.8) that the loop curve (m, n) with $m-n=1$, $n > 2$ and $t > 4$ on Σ_K, as a smooth knot in S^3, is never slice. The calculation gets

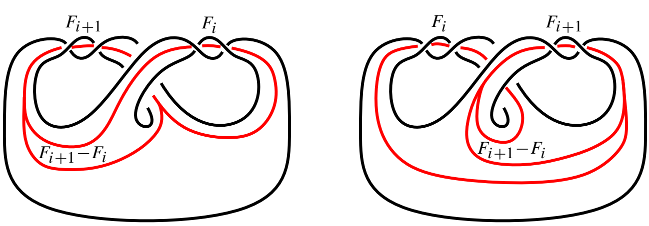

Figure 3. The two infinite families of unknotted curves for the figure-eight knot in S^3. The letters on parts of our curve or in certain locations stands for the number of strands at that particular curve or location. For example, for the (m, n) ∞ curve on the left we will show in Section 3B via explicit isotopies how, starting with the known unknotted (1, 1) ∞ curve, we can recursively obtain the following sequence of unknotted curves: $(1, 1) \sim (3, 2) \sim (8, 5) \sim (21, 13) \sim (55, 34) \sim \cdots$.

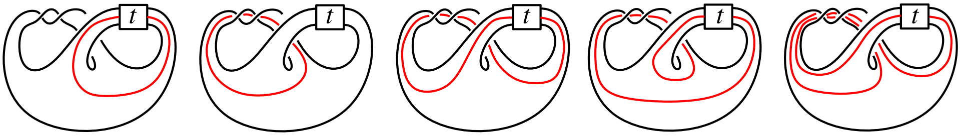

Figure 4. These five curves, from left to right, (0, 1), (1, 0), (1, 1) ∞, (1, 1) loop and (2, 1) ∞, on Σ_K, where $K = K_t$, $t \neq 1, -1$, are unknotted curves in S^3.

quickly complicated once $m - n > 1$, and it stays an open problem if in this range one can find other slice but not unknotted curves.

We can further generalize our study of unknotted essential curves on minimal-genus Seifert surfaces of genus-one knots for the Whitehead doubles of nontrivial knots. We first introduce some notation. Letting P be the twist knot K_t embedded (where $t = 0$ is allowed) in a solid torus $V \subset S^3$ and K denote an arbitrary knot in S^3, we identify a tubular neighborhood of K with V in such a way that the longitude of V is identified with the longitude of K coming from a Seifert surface. The image of P under this identification is a knot, $D^{\pm}(K, t)$, called the positive/negative t-twisted Whitehead double of K. In this situation the knot P is called the pattern for $D^{\pm}(K, t)$, and K is referred to as the companion. Figure 5 depicts the positive -3-twisted Whitehead double of the left-handed trefoil, $D^{+}(T_{2,-3}, -3)$. If one takes K to be the unknot, then $D^{+}(K, t)$ is nothing but the twist knot K_t.

Theorem 1.6. *Let K denote a nontrivial knot in S^3. Suppose that Σ_K is a standard genus-one Seifert surface for the Whitehead double of K. Then there is precisely one unknotted homologically essential, simple closed curve in the interior of Σ_K.*

1B. ***From unknotted curves to contractible 4-manifolds.*** The problem of finding unknotted homologically essential curves on a Seifert surface of a genus-one knot is

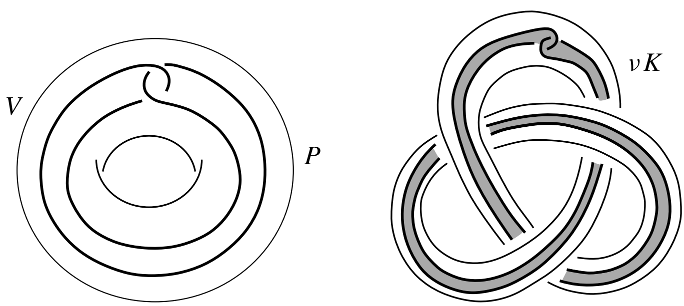

Figure 5. On the left is the solid torus $V \subset S^3$ and the pattern twist knot P (in this case $t = 0$). On the right is the positive -3-twisted Whitehead double of the left-handed trefoil, and its standard genus-one Seifert surface.

interesting on its own, but it is also useful for studying some essential problems in low-dimensional topology. We expand on one of these problems a little more. An important and still open question in low-dimensional topology asks: *which homology* 3*-sphere*[1] *bounds a homology* 4*-ball or contractible* 4*-manifold* (see [Kirby 1978, Problem 4.2])? This problem can be traced back to the famous Whitney embedding theorem and other important subsequent results due to Hirsch [1961], Wall [1965], and Rohlin [1965]. Since then, the research towards understanding this problem has stayed active. It has been shown that many infinite families of homology spheres do bound contractible 4-manifolds [Casson and Harer 1981; Fickle 1984; Stern 1978; Zeeman 1965] and at the same time many powerful techniques and homology cobordism invariants, mainly coming from Floer and gauge theories [Manolescu 2016; Fintushel and Stern 1985; Rohlin 1952], have been used to obtain constraints. See [Şavk 2024] for a detailed recent survey on various constructions and obstructions mentioned above.

In our case, using our main results, we will be able to list some more homology spheres that bound contractible 4-manifolds. This is because of the following theorem of Fickle [1984, Theorem 3.1], which was one of the main motivations for the research in this paper.

Theorem 1.7 (Fickle). *Let K be a knot in S^3 that has a genus-one Seifert surface F with a primitive element $[b] \in H_1(F)$ such that the curve b is unknotted in S^3. If b has self-linking s, then the homology* 3*-sphere obtained by* $1/(s \pm 1)$ *Dehn surgery on K bounds a contractible*[2] 4*-manifold.*

Theorem 1.7 was generalized (along with a somewhat more accessible proof of Fickle's theorem) by Etnyre and Tosun [2023, Theorem 1] to genus-one knots in the boundary of a homology 4-ball W, where the assumption on the curve b is relaxed so that b is slice in W. This will be useful (see Corollary 1.9 below) for applying to the slice but not unknotted curve/knot found in Theorem 1.4.

We also want to take the opportunity to highlight an interesting and still open conjecture [Fickle 1984, page 481, Conjecture] attributed to Fintushel and Stern.

Conjecture 1.8 (Fintushel and Stern). *Let K be a knot in the boundary of a homology* 4*-ball W which has genus-one Seifert surface with a primitive element $[b] \in H_1(F)$ such that b is slice in W. If b has self-linking s, then the homology* 3*-sphere obtained by* $1/k(s \pm 1)$, $k \geq 0$, *Dehn surgery on K bounds a homology* 4*-ball.*

[1] A homology 3-sphere/4-ball is a closed, oriented, smooth 3-/4-manifold having the integral homology groups of S^3/B^4.

[2] Indeed, this contractible manifold is a *Mazur-type* manifold, namely, it is a contractible 4-manifold that has a single handle for each of the indices 0, 1 and 2, where the 2-handle is attached along a knot that links the 1-handle algebraically once. This condition yields a trivial fundamental group.

Corollary 1.9. *Let K_t be a nontrivial twist knot. Then the homology spheres obtained by*

(1) $\pm\frac{1}{2}$ *Dehn surgery on* $K_1 = 4_1$,

(2) $-\frac{1}{2}$ *and* $-\frac{1}{4}$ *Dehn surgeries on* $K_{-1} = T_{2,3}$,

(3) $-\frac{1}{2}$ *and* $1/(t \pm 1)$ *and* $1/((t-2) \pm 1)$ *Dehn surgeries on* K_t *for* $t \neq \pm 1$,

(4) $\frac{1}{2}$ *Dehn surgery on* K_4

bound contractible 4*-manifolds.*

Corollary 1.10. *The homology spheres obtained by* $-\frac{1}{2}$ *Dehn surgery on* $D^+(K, t)$ *each bound a contractible* 4*-manifold.*

Remark 1.11. The 3-manifolds in Corollary 1.9(2) are the Brieskorn spheres $\Sigma(2, 3, 13)$ and $\Sigma(2, 3, 25)$; they were identified by Casson–Harer and Fickle to bound contractible 4-manifolds. Also, it was known already that the result of $\frac{1}{2}$ Dehn surgery on the figure-eight knot bounds a contractible 4-manifold (see [Tosun 2022, Theorem 18 and Figure 6]), and from this we obtain the result in Corollary 1.9(1), as the figure-eight knot is an amphichiral knot. The result in Corollary 1.10 also follows from [Fickle 1984, Theorem 3.6].

Remark 1.12. It is known that the result of $\frac{1}{n}$ Dehn surgery on a slice knot $K \subset S^3$ bounds a contractible 4-manifold. To see this, note that at the 4-manifold level with this surgery operation what we are doing is removing a neighborhood of the slice disk from B^4 (the boundary at this stage is zero surgery on K) and then attaching a 2-handle to a meridian of K with framing $-n$. Now, simple algebraic topology arguments show that this resulting 4-manifold is contractible.

It is a well-known result [Casson and Gordon 1978] that a nontrivial twist knot $K = K_t$ is slice if and only if $K = K_2$ (Stevedore's knot 6_1). So, by arguments above, we already know that result of $\frac{1}{n}$ surgery on K_2 bounds a contractible 4-manifold for any integer n. But, interestingly, we do not recover this by using Theorem 1.3.

Organization. The paper is organized as follows. In Section 2 we set some basic notation and conventions that will be used throughout the paper. Section 3 contains the proofs of Theorems 1.2, 1.3 and 1.4. Our main goal will be to organize, case by case, essential simple closed curves on genus-one Seifert surfaces Σ_K, through sometimes lengthy isotopies, into explicit positive or negative braid closures. Once this is achieved we use a result due to Cromwell that says the Seifert algorithm applied to the closure of a positive/negative braid closure gives a minimal-genus surface. This together with some straightforward calculations will help us to determine the unknotted curves exactly. But sometimes it will not be obvious how or even possible to reduce an essential simple closed curve to a positive or negative braid closure (see Sections 3B, 3C and 3D and Figure 34 in Section 5).

Further analyzing these cases will yield interesting phenomenon listed in Theorems 1.3 and 1.4. Section 4 contains the proof of Theorem 1.6. Finally, Section 5 contains the proofs of Corollaries 1.9 and 1.10 and some final remarks.

2. Preliminaries

In this section, we set some notation and make preparations for the proofs in the next three sections. In Figure 6 we record some basic isotopies/conventions that will be repeatedly used during proofs. Most of these are evident, but for the reader's convenience we explain how the moves in parts (a) and (f) work in Figures 7 and 8. We remind the reader that letters on parts of our curve, as in part (e) of Figure 6, or in a certain location, are used to denote the number of strands that particular curve has.

Recall also an essential, simple closed curve on Σ_K can be represented by a pair of nonnegative integers (m, n), where m is the number of times it runs around the left band and n is the number of times it runs around the right band in Σ_K, and since we are dealing with connected curves we must have that m, n are relatively prime.

We have two cases: $m > n$ or $n > m$. For an (m, n) curve with $m > n$, after the m strands pass under the n strands on the Seifert surface, the curve can be split into two sets of strands. For this case, assume that the top set is made of n strands. They must connect to the n strands going over the right band, leaving the other set to be made of $m - n$ strands. Now, we can split the other side of the set of m

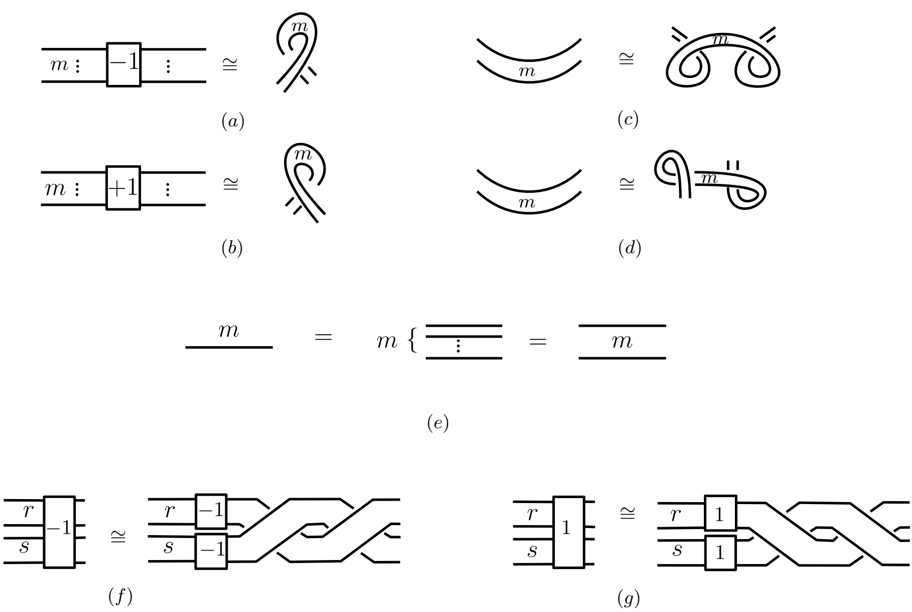

Figure 6. Various isotopies.

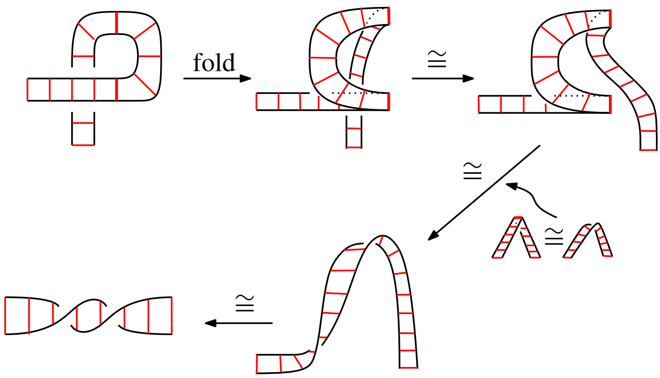

Figure 7. Diagrammatic proof of the move in Figure 6(a). As indicated, the passage from the top right figure to the bottom right is via "smoothing" a creased edge.

strands into two sections. The $m-n$ strands on the right can only go to the bottom of these two sections, because otherwise the curve would have to intersect itself on the surface. This curve is referred to as an (m,n) ∞ curve. See Figure 9(a). The other possibility for an (m,n) curve with $m>n$ has n strands in the bottom set instead, and they loop around to connect with the n strands going over the right band. This leaves the other to have $m-n$ strands. We can split the other side of the set of m strands into two sections. The $m-n$ strands on the right can only go to the top of these two sections, because again otherwise the curve would have to intersect itself on the surface. The remaining subsection must be made of n strands and connect to the n strands going over the right band. This curve is referred to as an (m,n) loop curve. See Figure 9(b). The case of an (m,n) curve with $n>m$ is similar. See Figure 9(c),(d).

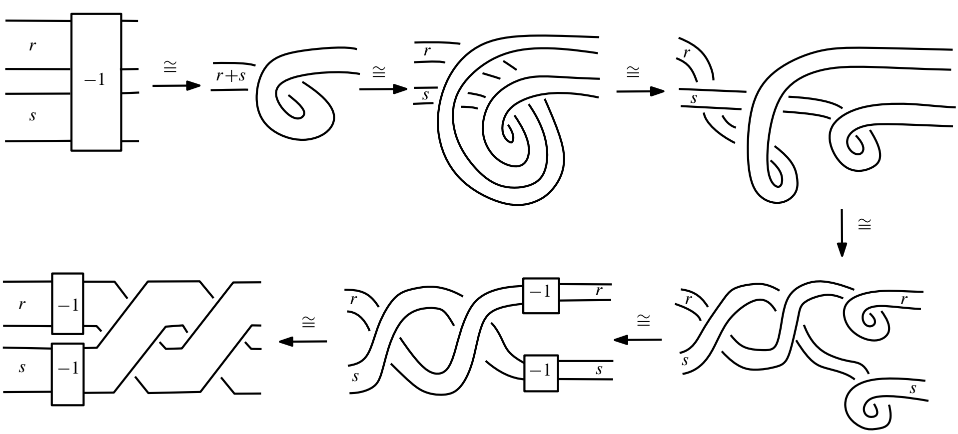

Figure 8. Diagrammatic proof of the move in Figure 6(f).

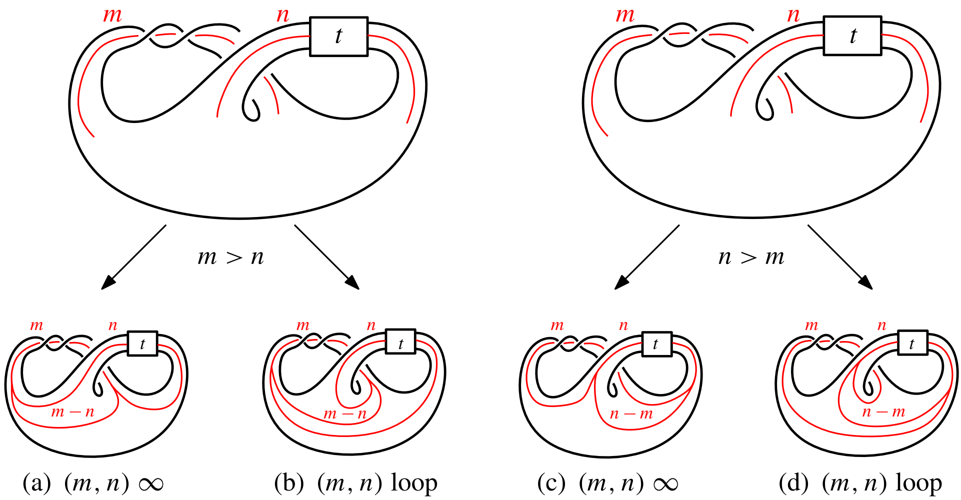

Figure 9. Possibilities for an essential, simple closed curve (m, n) on Σ_K.

3. Twist knots

In this section we provide the proofs of Theorems 1.2, 1.3 and 1.4. We do this in four parts. Sections 3A and 3B contain all technical details of Theorem 1.2, Section 3C contains details of Theorem 1.3, and Section 3D contains Theorem 1.4 .

3A. ***Twist knot with*** $t < 0$**.** In this section we consider twist knots $K = K_t$, $t \leq -1$. This in particular includes the right-handed trefoil K_{-1}.

Proposition 3.1. *All essential, simple closed curves on* Σ_K *can be characterized as the closure of one of the negative braids in Figure 10.*

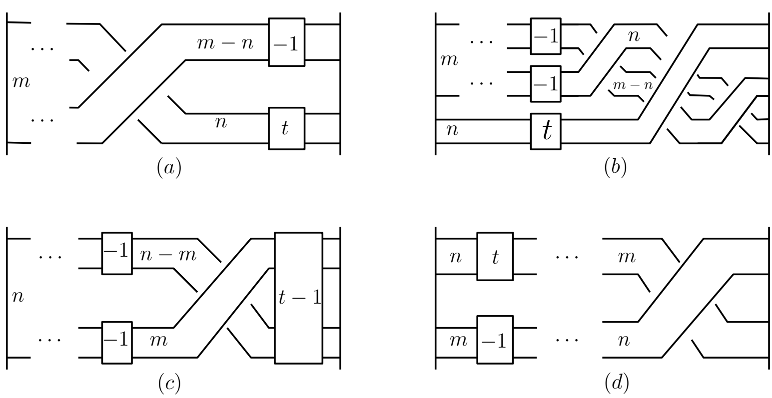

Figure 10

Proof. It suffices to show all possible curves for an arbitrary m and n such that $\gcd(m,n) = 1$ are the closures of either braid in Figure 10. As mentioned earlier we will deal with cases where both $m, n \geq 1$ since cases involving 0 are trivial. There are four cases to consider. The arguments for each of these will be quite similar, and so we will explain the first case in detail and refer to the rather self-explanatory drawings/figures for the remaining cases.

Case 1: (m,n) ∞ curve with $m > n > 0$. This case is explained in Figure 11. The picture on top left is the (m,n) curve of interest. The next picture to its right is the (m,n) curve where we ignore the surface it sits on and use the convention from Figure 6(e). The next picture is an isotopy where we push the split between n strands and $m-n$ strands along the dotted blue arc. The next picture is obtained by simple isotopy. The passage from the top right picture to the bottom right is via Figure 6(c). The passage from the bottom right to the figure on its left is obtained by pushing $m-n$ strands around along the green arc. The goal here is to put the curve in a braid closure position. Finally, by applying simple isotopies and Figure 6(a) repeatedly we replace all the loops with full negative twists. Note that we moved the full negative twist on $m-n$ strands clockwise fashion around to bring it in the bottom of the figure. This gives the picture on the bottom left, which is the closure of the negative braid depicted in Figure 10(a).

Case 2: (m,n) loop curve with $m > n > 0$. By a series of isotopies, as indicated in Figure 12, the (m,n) curve in this case can be simplified to the knot depicted on the right of Figure 12, which is the closure of negative braid in Figure 10(b).

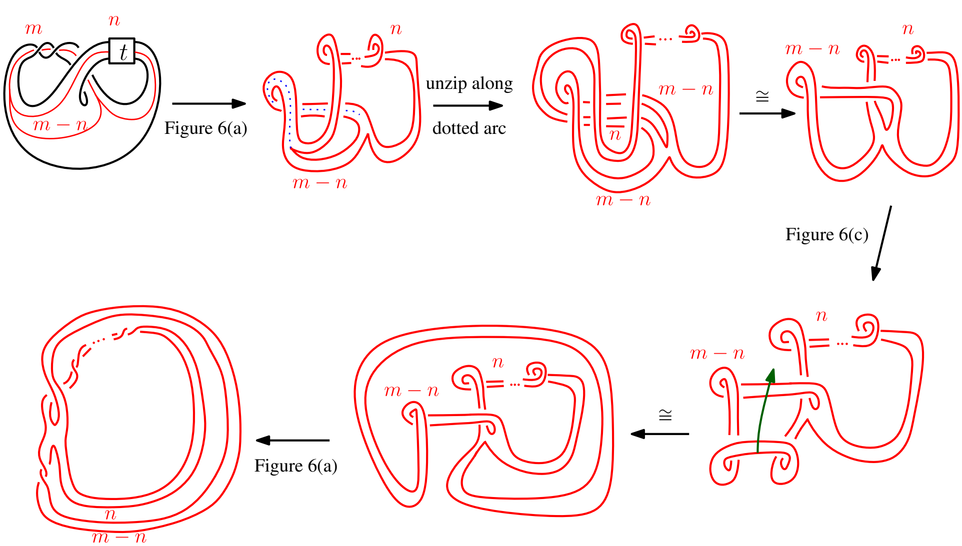

Figure 11

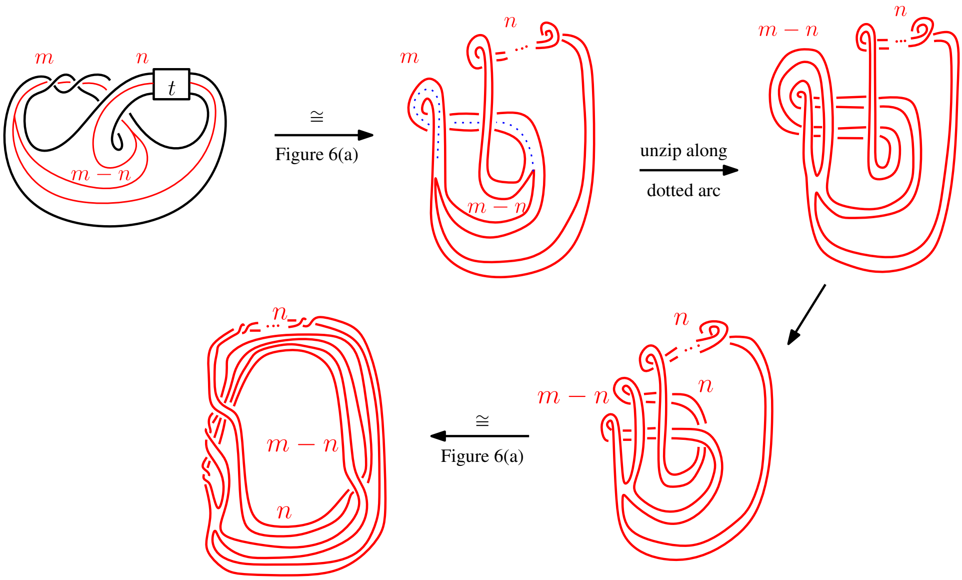

Figure 12

Case 3: (m, n) ∞ curve with $n > m > 0$. By a series of isotopies, as indicated in Figure 13, the (m, n) curve in this case can be simplified to the knot depicted on the bottom left of Figure 13, which is the closure of negative braid in Figure 10(c).

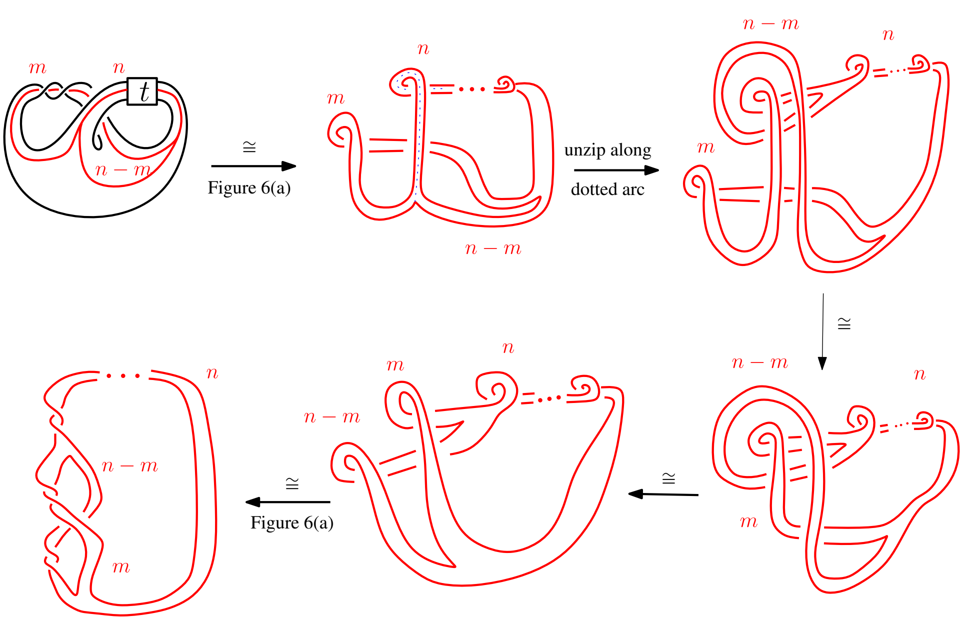

Figure 13

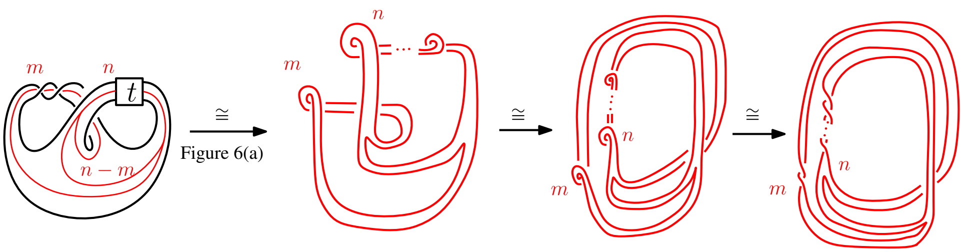

Figure 14

Case 4: (m,n) loop curve with $n>m>0$. By a series of isotopies, as indicated in Figure 14, the (m,n) curve in this case can be simplified to the knot depicted on the right of Figure 14, which is the closure of negative braid in Figure 10(d). □

Next, we determine which of the curves in Proposition 3.1 are unknotted. It is a classic result due to Cromwell [1989] (see also [Stoimenow 2003, Corollary 4.2]) that the Seifert algorithm applied to the closure of a positive braid gives a minimal-genus surface.

Proposition 3.2. *Let β be a braid as in Figure 10 and $K=\hat{\beta}$ be its closure. Let $s(K)$ be the number of Seifert circles and $l(K)$ be the number of crossings in each braid diagram. Then $(s(K),l(K))$ is equal to*

$$\big(m,\ |t|n(n-1)+(m-n)(m-n-1)+n(m-n)\big) \qquad \textit{for } \beta \textit{ as in Figure 10(a)},$$

$$\big(m+n,\ (|t|+1)n(n-1)+(m-n)(m-n-1)+nm+2n(m-n)\big) \qquad \textit{for } \beta \textit{ as in Figure 10(b)},$$

$$\big(n,\ (|t-1|)n(n-1)+(n-m)(n-m-1)+m(m-1)+m(n-m)\big) \qquad \textit{for } \beta \textit{ as in Figure 10(c)},$$

$$\big(m+n,\ |t|n(n-1)+m(m-1)+nm\big) \qquad \textit{for } \beta \textit{ as in Figure 10(d)}.$$

Proof. Consider the braid β as in Figure 10(a). Clearly, it has m Seifert circles as β has m strands. Next, we will analyze the three locations in which crossings occur. First consider the t negative full twists on n strands. Since each strand crosses over the other $n-1$ strands, we obtain $|t|n(n-1)$ crossings. Second, the negative full twist on $m-n$ strands produces additional $(m-n)(m-n-1)$ crossings. Lastly, notice the part of β where $m-n$ strands overpass the other n strands, and so for each strand in $m-n$ strands we obtain an additional n crossings. Hence, for $K=\hat{\beta}$, we calculate

$$l(\hat{\beta})=|t|n(n-1)+(m-n)(m-n-1)+n(m-n).$$

The calculations for the other cases are similar. □

We can now prove the first part of Theorem 1.2.

Proof of Theorem 1.2(a). Proposition 3.1 proves the first half of our theorem. To determine there are exactly six unknotted curves when $t = -1$ and five when $t < -1$, let B be the set containing the six and five unknotted curves as in Figures 2 and 4, respectively. It suffices to show an essential, simple closed curve c on Σ_K, where $c \notin B$, cannot be unknotted in S^3. We know by Proposition 3.1 that c is the closure of one of the braids in Figure 10 in S^3, where $m, n \geq 1$ and $\gcd(m, n) = 1$. We show, case by case, that the Seifert surface obtained via the Seifert algorithm for curves $c \notin B$ in each case has positive genus, and hence it cannot be unknotted.

- Let $c = (m, n)$ be the closure of the negative braid as in Figure 10(a) and Σ_c its Seifert surface obtained by the Seifert algorithm. There are m Seifert circles, and by Proposition 3.2,

$$l(c) = |t|n(n-1) + (m-n)(m-n-1) + n(m-n).$$

Hence

$$g(\Sigma_c) = \tfrac{1}{2}(1 + l - s) = \tfrac{1}{2}\big(m(m-n-2) + n(|t|(n-1)+1) + 1\big).$$

If $m = n+1$, then we get $g(\Sigma_c) = \frac{1}{2}|t|n(n-1)$, which is positive as long as $n > 1$; note that when $c = (2, 1)$ we indeed get an unknotted curve. If $m > n+1$, then $g(\Sigma_c) \geq \frac{1}{2}\big(n(|t|(n-1)+1)+1\big) > 0$ as long as $n > 0$. So, $c \notin B$ is not an unknotted curve as long as $m > n \geq 1$.

- Let $c = (m, n)$ be the closure of the negative braid as in Figure 10(b) and Σ_c its Seifert surface obtained by the Seifert algorithm. There are $n + m$ Seifert circles, and by Proposition 3.2,

$$l(c) = (|t|+1)n(n-1) + (m-n)(m-n-1) + nm + 2n(m-n).$$

Hence

$$g(\Sigma_c) = \tfrac{1}{2}\big(m(m+n-2) + n(|t|(n-1)-1) + 1\big).$$

One can easily see that this quantity is always positive as long as $n \geq 1$. So, $c \notin B$ is not an unknotted curve when $m > n \geq 1$.

- Let $c = (m, n)$ be the closure of the negative braid as in Figure 10(c) and Σ_c its Seifert surface obtained by the Seifert algorithm. There are n Seifert circles, and by Proposition 3.2,

$$l(c) = (|t|-1)n(n-1) + (n-m)(n-m-1) + m(m-1) + m(n-m).$$

Hence

$$g(\Sigma_c) = \tfrac{1}{2}\big(n(|t|(n-1) - m - 1) + m^2 + 1\big).$$

This is always positive as long as $m \geq 1$ and $|t| \neq 1$; note that when $c = (1, 2)$ and $|t| = 1$ we indeed get an unknotted curve. So, $c \notin B$ is not an unknotted curve when $n > m \geq 1$.

• Let $c = (m, n)$ be the closure of the negative braid as in Figure 10(d) and Σ_c its Seifert surface obtained by the Seifert algorithm. There are $n + m$ Seifert circles, and by Proposition 3.2,

$$l(c) = |t|n(n-1) + m(m-1) + nm.$$

Hence

$$g(\Sigma_c) = \tfrac{1}{2}\big(|t|n(n-1) + m(m-2) + n(m-1) + 1\big).$$

One can easily see that this quantity is always positive as long as $m \geq 0$. So, $c \notin B$ is not an unknotted curve when $n > m \geq 1$.

This completes the first part of Theorem 1.2. □

3B. *Figure-eight knot.* The case of figure-eight knot is certainly the most interesting one. It is rather surprising, even to the authors, that there exists a genus-one knot with infinitely many unknotted curves on its genus-one Seifert surface. As we will see, understanding homologically essential curves for the figure-eight knot will be similar to what we did in the previous section. The key difference develops in Cases 2 and 4 below where we show how, under certain conditions, a homologically essential (m, n) ∞ (resp. (m, n) loop) curve can be reduced to the homologically essential $(m-n, 2n-m)$ ∞ (resp. $(2m-n, n-m)$ loop) curve, and how this recursively produces infinitely many distinct homology classes that are represented by the unknot, and we will show that certain Fibonacci numbers can be used to describe these unknotted curves. Finally we will show for the figure-eight knot this is the only way that an unknotted curve can arise. Adapting the notation developed thus far we start characterizing homologically essential simple closed curves on the genus-one Seifert surface Σ_K of the figure-eight knot K.

Proposition 3.3. *All essential, simple closed curves on Σ_K can be characterized as the closure of one of the braids in Figure 15 (note the first and third braids from the left are negative and positive braids, respectively).*

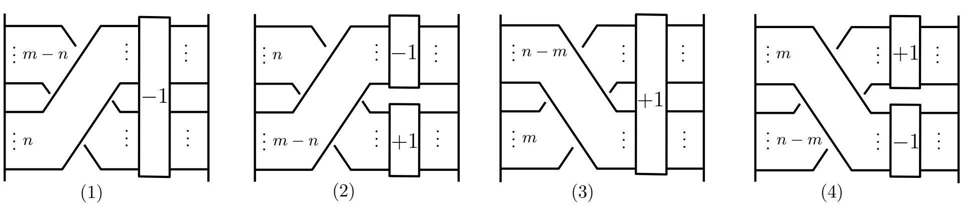

Figure 15. Braid representations of curves on Σ_K, where K is the figure-eight knot. From left to right: (m, n) loop curve with $m > n$; (m, n) ∞ curve with $m > n$; (m, n) ∞ curve with $n > m$; (m, n) loop curve with $n > m$.

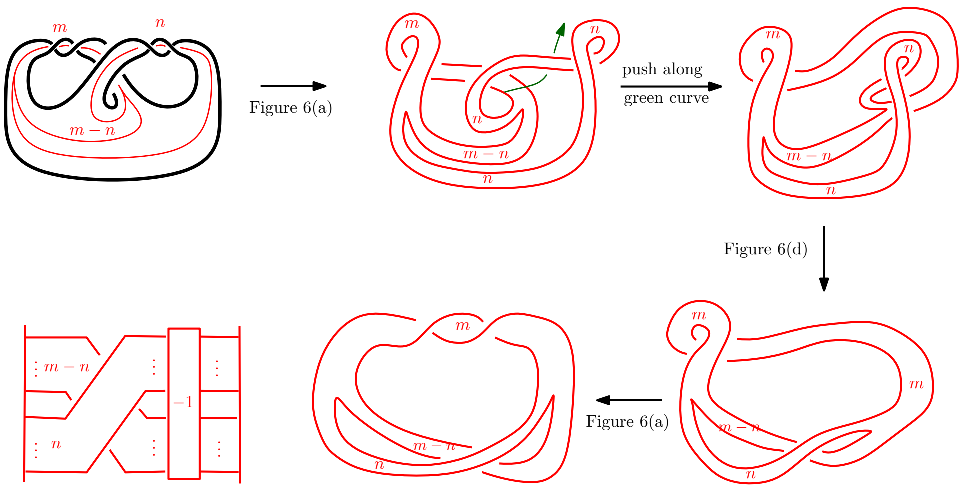

Figure 16

Proof. The curves $(1, 0)$ and $(0, 1)$ are clearly unknots. Moreover, because $\gcd(m, n) = 1$, the only curve with $n = m$ is the $(1, 1)$ curve, which is also the unknot in S^3. For the rest of the arguments below, we will assume $n > m$ or $m > n$. There are four cases to consider:

Case 1: (m, n) loop curve with $m > n > 0$. This curve can be turned into a negative braid following the process in Figure 16. The reader will observe that the process here is very similar to those in the previous section. We mention that the passage from the middle figure on the top to the one on its right is obtained by pushing the m strands along the green curve till it is clear from a positive loop of n strands.

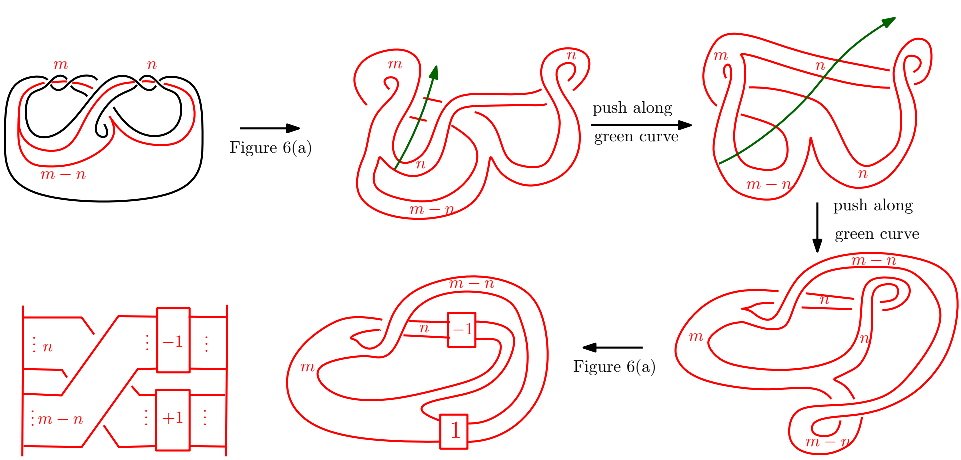

Figure 17

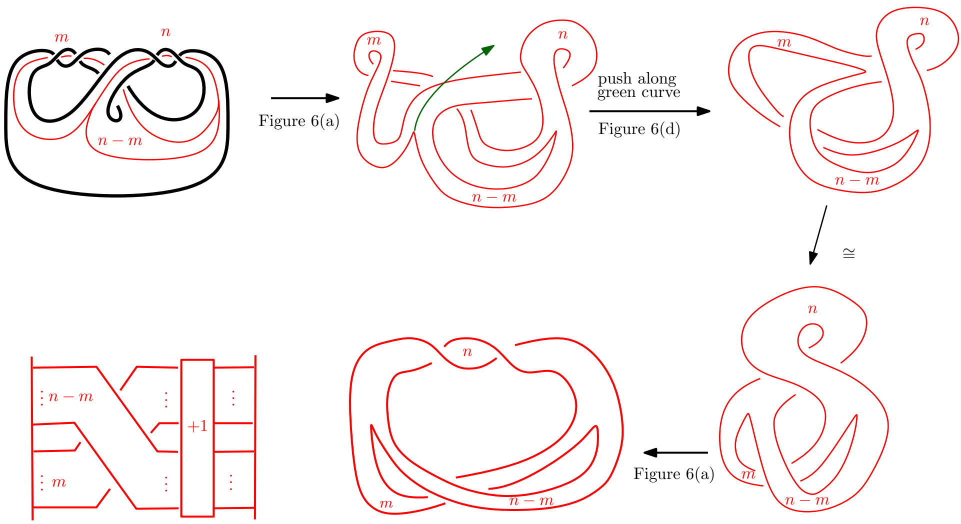

Figure 18

Finally, the middle curve on the bottom is our final curve, which is the closure of the negative braid to its left.

Case 2: (m, n) ∞ curve with $m > n > 0$. As mentioned at the beginning, this case (and Case 4) are much more involved and interesting (in particular the subcases of Cases 2c and 4c; see the proof Proposition 3.4). Following the process as in Figure 17, the curve can be isotoped as in the bottom right of that figure, which is the closure of the braid on its left — that is, the second braid from the left in Figure 15.

Case 3: (m, n) ∞ curve with $n > m > 0$. This curve can be turned into a positive braid following the process in Figure 18.

Case 4: (m, n) loop curve with $n > m > 0$. This curve can be turned into the closure of a braid following the process in Figure 19. □

We next determine which of these curves are unknotted:

Proposition 3.4. *A homologically essential curve c characterized as in Proposition 3.3 is unknotted if and only if it is* (a) *a trivial curve* $(1, 0)$ *or* $(0, 1)$, (b) *an* ∞ *curve in the form of* (F_{i+1}, F_i), *or* (c) *a loop curve in the form of* (F_i, F_{i+1}).

Proof. Let c denote one of these homologically essential curve listed in Proposition 3.3. We will analyze the unknottedness of c in four separate cases.

Case 1: Suppose $c = (m, n)$ is the closure of the negative braid in the bottom left of Figure 16. Note the minimal Seifert surface of c, Σ_c, has $(n)(m-n) + (m)(m-1)$ crossings and m Seifert circles. Hence

$$g(\Sigma_c) = \tfrac{1}{2}(n(m-n) + (m-1)^2).$$

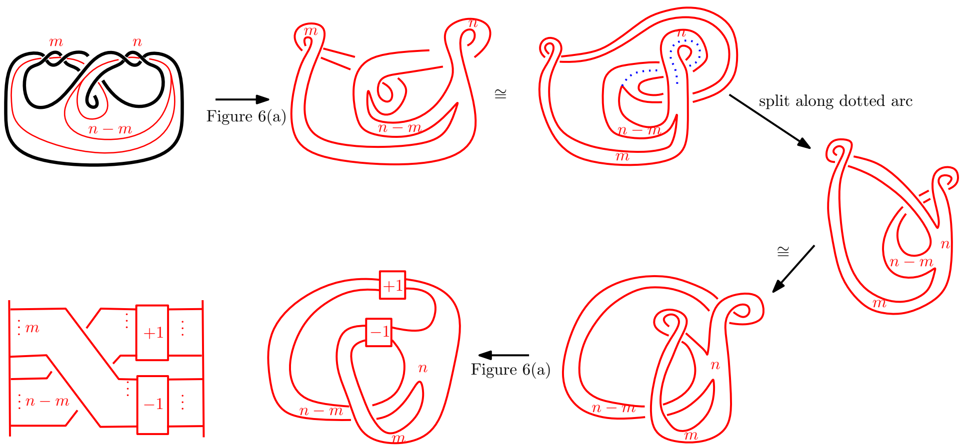

Figure 19

This is a positive integer for all m, n with $m > n$. So c is never unknotted in S^3 as long as long $m > n > 0$.

Case 2: Suppose c is of the form in the bottom right of Figure 17. Since this curve is not a positive or negative braid closure, we cannot directly use Cromwell's result as in Case 1 or the previous section. There are three subcases to consider.

Case 2a: $m - n = n$. Because m and n are relatively prime integers, we must have that $m = 2$ and $n = 1$, and we can easily see that this (2, 1) curve is unknotted.

Case 2b: $m - n > n$. This curve can be turned into a negative braid following the process in Figure 20. More precisely, we start, on the top left of that figure, with the curve appearing on the bottom right of Figure 17. We extend the split along the dotted blue arc and isotope m strands to reach the next figure. We note that this splitting can be done since, by assumption, $m - 2n > 0$. Then using Figure 6(a) and further isotopy we reach the final curve on the bottom right of Figure 20, which is obviously the closure of the negative braid depicted on the bottom left of that picture.

The minimal Seifert surface coming from this negative braid closure contains $m - n$ circles and $(m-2n)n + (m-n)(m-n-1)$ twists. Hence

$$g(\Sigma_c) = \tfrac{1}{2}((m-2n)n + (m-n)(m-n-2) + 1).$$

This a positive integer for all integers m, n with $m - n > n$. So, c is not unknotted in S^3.

Case 2c: $m - n < n$. We organize this curve some more. We start, on the top left of Figure 21, with the curve that is on the bottom left of Figure 17. We extend the split along the dotted blue arc and isotope $m - n$ strands to reach the next figure. After some isotopies we reach the curve on the bottom left of Figure 21. In other

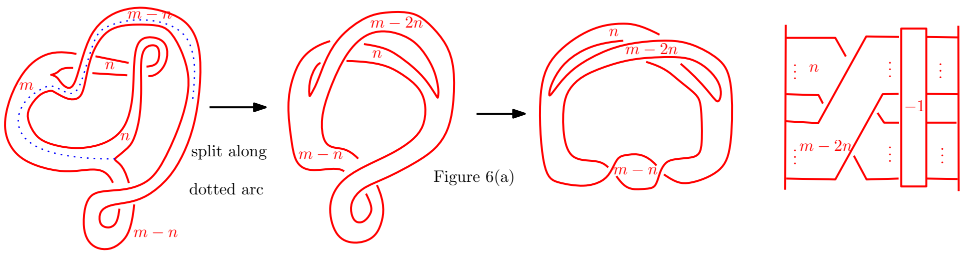

Figure 20

words, this subcase of Case 2c leads to a reduced version of the original picture (top left curve in Figure 17), in the sense that the number of strands over either handle is less than the number of strands in the original picture.

This case can be further subdivided depending on the relationship between $2n-m$ and $m-n$, but this braid (or rather its closure) will turn into a $(m-n, 2n-m)$ ∞ curve when $m-n > 2n-m$:

Case 2c(i): $2n-m=m-n$. This simplifies to $3n=2m$. Because $\gcd(m, n)=1$, this will only occur for $m=3$ and $n=2$, and the resulting curve is a $(1, 1)$ ∞ curve. In other words here we observed that the $(3, 2)$ curve has been reduced to a $(1, 1)$ curve.

Case 2c(ii): $2n-m > m-n$. This means that we are dealing with a curve under Case 3, and we will see that all curves considered there are positive braid closures.

Case 2c(iii): $2n-m < m-n$. Here, we remain under Case 2. So for $m > n > m-n$, the (m, n) ∞ curve is isotopic to the $(m-n, 2n-m)$ ∞ curve. This isotopy series will be denoted by $(m, n) \sim (m-n, 2n-m)$. Equivalently, there is a series of isotopies such that $(m-n, 2n-m) \sim (m, n)$. If (k, l) denotes a curve at one stage of this isotopy, then $(k, l) \sim ((k+l)+k, k+l)$. So, starting with $k=l=1$, we

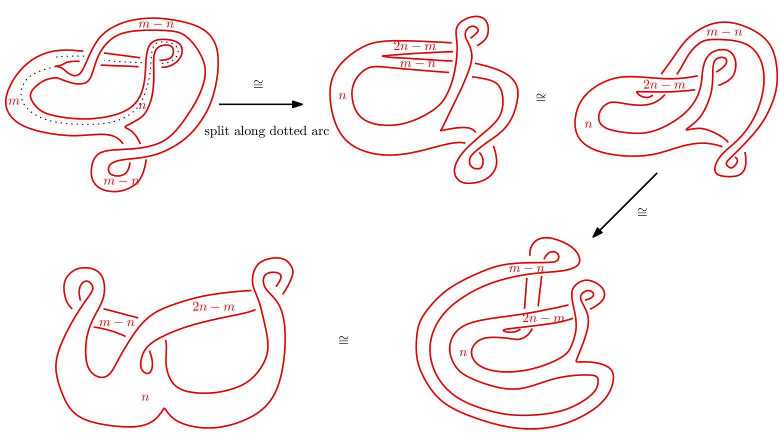

Figure 21

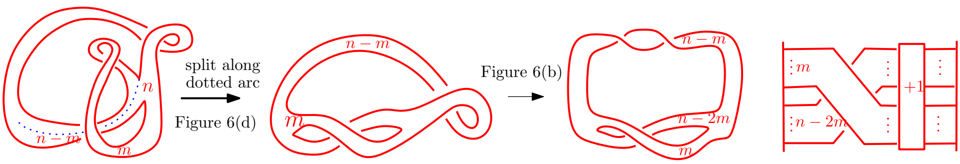

Figure 22

recursively obtain

$$(1, 1) \sim (3, 2) \sim (8, 5) \sim (21, 13) \sim (55, 34) \sim \cdots .$$

In a similar fashion, if we start with $k = 2,\ l = 1$, we obtain

$$(2, 1) \sim (5, 3) \sim (13, 8) \sim (34, 21) \sim (89, 55) \sim \cdots .$$

Notice every curve c above is of the form $c = (F_{i+1}, F_i)$, $i \in \mathbb{Z}_{>0}$, where F_i denotes the i-th *Fibonacci number*. We will call these *Fibonacci curves*. We choose (1, 1) and (2, 1) because they are known unknots. As a result, this relation generates an infinite family of homologically distinct simple closed curves on Σ_K that are unknotted in S^3.

Case 3: Suppose a curve, c, is of the form in Figure 15(3), which is the closure of the positive braid depicted in the bottom left of Figure 18. An argument similar to that applied to Case 1 can be used to show c is never unknotted in S^3.

Case 4: Suppose c is of the form as in the bottom right of Figure 19. Similar to Case 2, there are three subcases to consider.

Case 4a: $m = n - m$. Then $2m = n$. Because $\gcd(m, n) = 1$, $m = 1$ and $n = 2$, resulting in the unknot.

Case 4b: $n - m > m$. Then $n - 2m > 0$, and following the isotopies in Figure 22, the curve can be changed into the closure of the positive braid depicted on the bottom right of that figure.

Identical to Case 2b, the curve c in this case is never unknotted in S^3.

Case 4c: $m > n - m$. Then $2m - n > 0$, and we can split the m strands into two: a set of $n - m$ strands and a set of $2m - n$ strands.

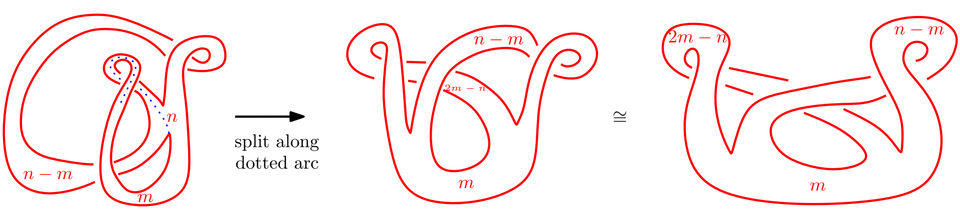

Figure 23

This case can be further subdivided depending on the relationship between $n-m$ and $2m-n$, but this braid will turn into a $(2m-n, n-m)$ loop curve when $n-m > 2m-n$:

Case 4c(i): $2m-n = n-m$. This simplifies to $3m = 2n$. Because $\gcd(m, n) = 1$, this will only occur for $m = 2$ and $n = 3$, and the resulting curve is a (1, 1) loop curve.

Case 4c(ii): $n-m < 2m-n$. This means that we are dealing with a curve under Case 1, and we saw that all curves considered there are negative braid closures.

Case 4c(iii): $n-m > 2m-n$. Here, we remain under Case 4. So for $n > m > n-m$, an (m, n) loop curve has the following isotopy series: $(m, n) \sim (2m-n, n-m)$. If (k, l) denotes a curve at one stage of this isotopy, then the reverse also holds: $(k, l) \sim (k+l, (k+l)+l)$. As a result, much like Case 2c, we can generate two infinite families of unknotted curves in S^3:

$$(1, 1) \sim (2, 3) \sim (5, 8) \sim (13, 21) \sim (34, 55) \sim \cdots,$$
$$(1, 2) \sim (3, 5) \sim (8, 13) \sim (21, 34) \sim (55, 89) \sim \cdots.$$

Notice every curve c is of the form $c = (F_i, F_{i+1})$, $i \in \mathbb{Z}_{>0}$. Finally, we show that this is the only way one can get unknotted curves. That is, we claim:

Lemma 3.5. *If a homologically essential curve c on Σ_K for $K = 4_1$ is unknotted, then it must be a Fibonacci curve.*

Proof. From above, it is clear that if our curve c is Fibonacci, then it is unknotted. So it suffices to show if a curve is not Fibonacci then it is not unknotted. We will demonstrate this for loop curves under Case 4. Let c be a loop curve that is not Fibonacci but is unknotted. Since it is unknotted, it fits into either Case 4a or 4c. But the only unknotted curve from Case 4a is a (1, 1) curve which is a Fibonacci curve, so c must be under Case 4c. By our isotopy relation, $(m, n) \sim (2m-n, n-m)$. So, the curve can be reduced to a minimal form, say (a, b), where $(a, b) \neq (1, 1)$ and $(a, b) \neq (2, 1)$. We will now analyze this reduced curve (a, b):

- If $a = b$, then $(a, b) = (1, 1)$, a contradiction.
- If $a > b$, then (a, b) is under Case 1; none of those are unknotted.
- If $b-a < a < b$, then (a, b) is still under Case 4c, and not in reduced form, a contradiction.
- If $a < b-a < b$, then (a, b) is under Case 4b; none of those are unknotted.
- If $b-a = a < b$, then $(a, b) = (2, 1)$, a contradiction.

So, it has to be that either $(a, b) \sim (1, 1)$ or $(a, b) \sim (2, 1)$. Hence, it must be that $c = (F_i, F_{i+1})$ for some i. The argument for the case where c is an ∞ curve under Case 2 is identical. □

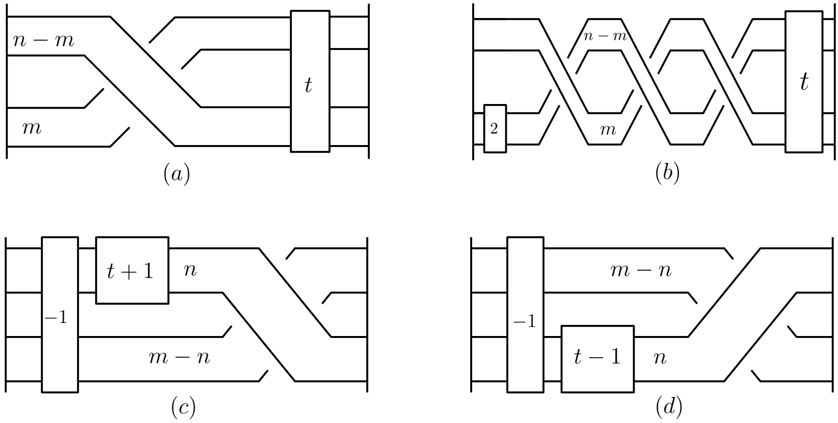

Figure 24

We end this section with a remark which was observed by the authors at the initial stages of the research and was also communicated to the authors by F. Misev.

Remark 3.6. An alternative and perhaps slightly easier way to see the existence of Fibonacci numbers for unknotted curves for the figure-eight knot is as follows: Recall that the figure-eight knot is fibered and its pseudo-Anosov monodromy $\phi : \Sigma \to \Sigma$, where Σ is the genus-one Seifert surface, induces a linear map on the first homology $H_1(\Sigma) = \mathbb{Z} \oplus \mathbb{Z}$ described by the matrix $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. By applying this matrix repeatedly to the unknotted curves (vectors) (0, 1) and $(1, 0)$ one obtains other unknotted curves that have Fibonacci numbers as their entries exactly as predicted in Proposition 3.4.

We add that this approach cannot capture the full strength of the results about the figure-eight knot: namely, showing that any unknotted curve as in Lemma 3.5 on the genus-one Seifert surface of the figure-eight knot must be a Fibonacci curve or characterizing all homologically essential curves on the Seifert surface of the figure-eight knot as in Proposition 3.3. Moreover our proof technique is by hand and works uniformly for all other twist knots we study in this paper.

3C. *Twist knot with $t > 1$: part one.* In this section we consider twist knot $K = K_t$, $t \geq 2$, and give the proof of Theorem 1.3.

Proposition 3.7. *All essential, simple closed curves on Σ_K can be characterized as the closure of one of the braids in Figure 24.*

Proof. It suffices to show all possible curves for an arbitrary m and n such that $\gcd(m, n) = 1$ are the closures of braids in Figure 24. Here too there are four cases to consider but we will analyze these in slightly different order than in the previous two sections.

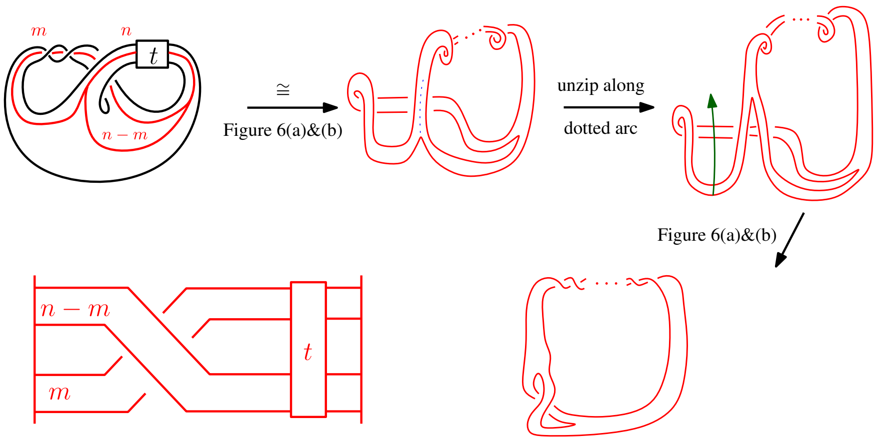

Figure 25

Case 1: (m, n) ∞ curve with $n > m > 0$. In this case the curve is the closure of a positive braid, and this is explained in Figure 25. More precisely, we start with the curve which is drawn in the top left of the figure, and after a sequence of isotopies this becomes the curve in the bottom right of the figure, which is obviously the closure of the braid in the bottom left of the figure. In particular, when $n > m \geq 1$, none of these curves will be unknotted.

Case 2: (m, n) loop curve with $n > m > 0$. In this case too the curve is the closure of a positive braid, and this is explained in Figure 26. In particular, when $n > m > 1$, none of these curves will be unknotted.

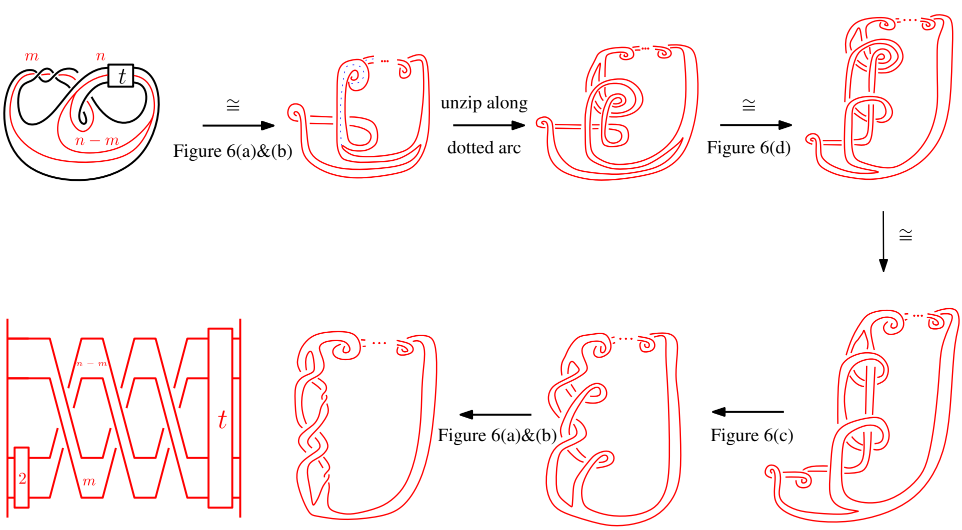

Figure 26

In the remaining two cases we will follow a slightly different way of identifying our curves as braid closures. We will see (as is evident in parts (c) and (d) of Figure 24) that the braids will not be positive or negative braids for general m, n and t values. We will then verify how under the various hypothesis listed in Theorem 1.3 these braids can be reduced to positive or negative braids.

Case 3: (m,n) ∞ curve with $m > n > 0$. We explain in Figure 27 below how the (m,n) ∞ curve with $m > n > 0$ is the closure of the braid in the bottom left of the figure. This braid is not obviously a positive or negative braid.

Case 3a: (m,n) ∞ curve with $m > n > 0$ and $m - tn > 0$. We want to show the braid in the bottom left of Figure 27 under the hypothesis that $m - tn > 0$ can be made a negative braid. We achieve this in Figure 28. More precisely, in part (a) of the figure we see the braid that we are working on. We apply the move in Figure 6(f) and some obvious simplifications to reach the braid in part (d). In part (e) of the figure we reorganize the braid: more precisely, since $m - tn > 0$ and $m - n = m - tn + (t-1)n$, we can split the piece of the braid in part (d) made of $m - n$ strands as the stack of $m - tn$ strands and a set of $t - 1$ many n strands. We then apply the move in Figure 6(f) repeatedly ($t - 1$ times) to obtain the braid in part (f). We note that the block labeled as "all negative crossings" is not important for our purposes to draw explicitly but we emphasize that each time we apply the move in Figure 6(f) it produces a full left-handed twist between a set of n strands and the rest. Next, sliding -1 full twists one by one from n strands over the block of these negative crossings we reach part (g). After further obvious simplifications and organizations in parts (h)–(j) we reach the braid in part (k), which is a negative braid.

Case 3b: (m,n) ∞ curve with $m > n > 0$ and $m - n < n$. We want to show in this case the braid in the bottom left of Figure 27 under the hypothesis that $m - n < n$ can be made a positive braid (regardless of t value). This is achieved in Figure 29.

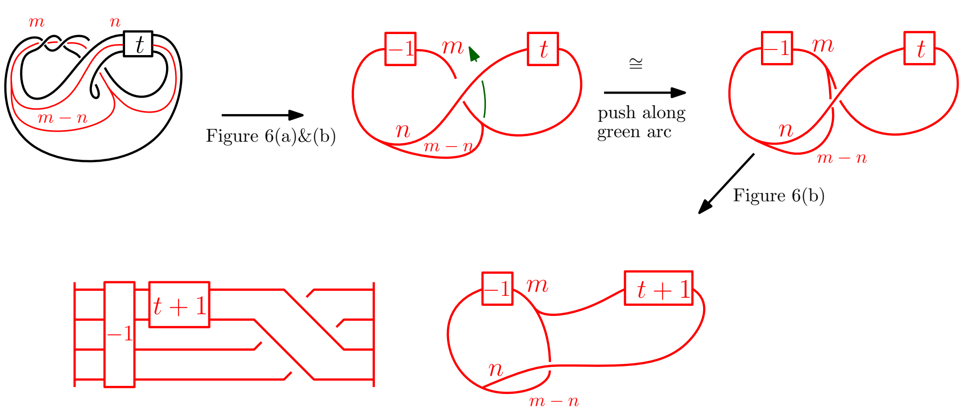

Figure 27

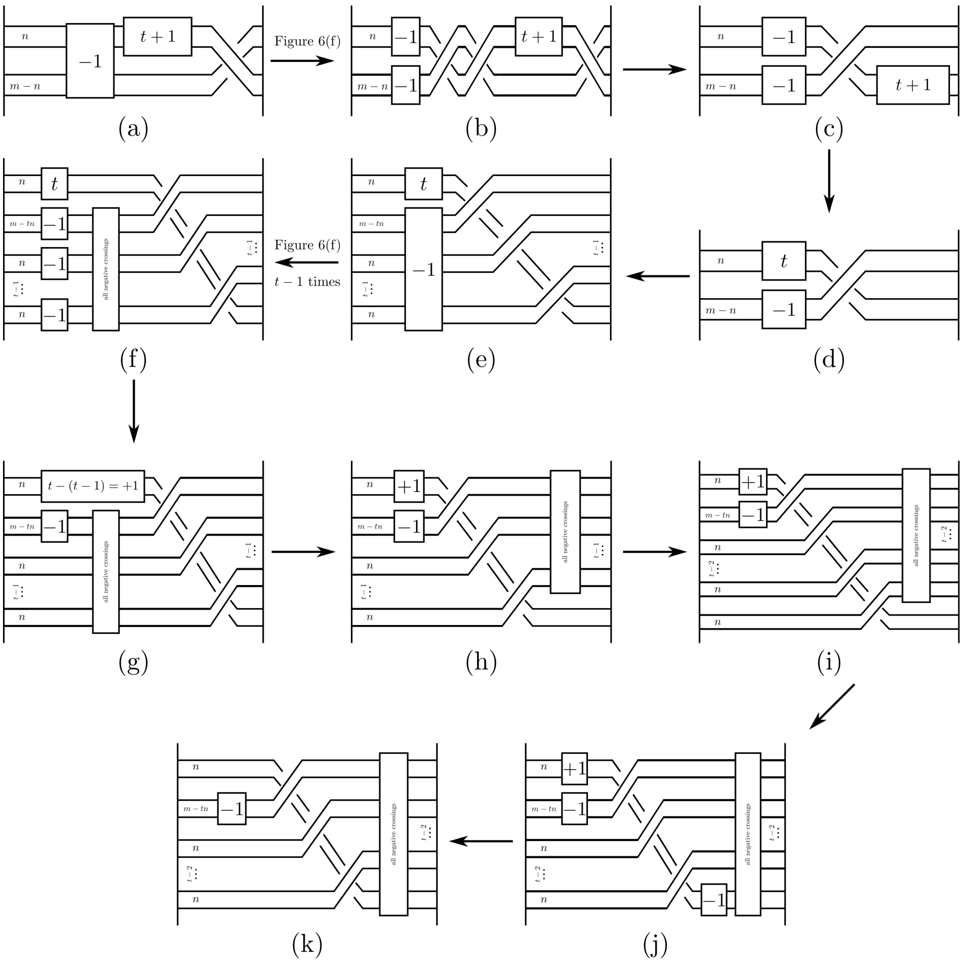

Figure 28

Case 4: (m, n) loop curve with $m > n > 0$. The arguments for this case are identical to those in Cases 3 and 3a above. The (m, n) loop curve with $m > n > 0$ is the closure of the braid that is drawn in the bottom left of Figure 30.

Case 4a: (m, n) loop curve with $m > n > 0$ and $m - tn > 0$. We show the braid, which the (m, n) ∞ curve with $m > n > 0$ is the closure of, can be made a negative braid under the hypothesis $m - tn > 0$. This follows steps very similar to those in Case 3a, which is explained through a series drawings in Figure 31.

Case 4b: (m, n) loop curve with $m > n > 0$ and $m - n < n$. Finally, we consider the (m, n) loop curve with $m > n > 0$ and $m - n < n$. Interestingly, this curve for $t > 2$ does not have to be the closure of a positive or negative braid. This will be further explored in the next section but for now we observe, through Figure 31(a)–(c), that

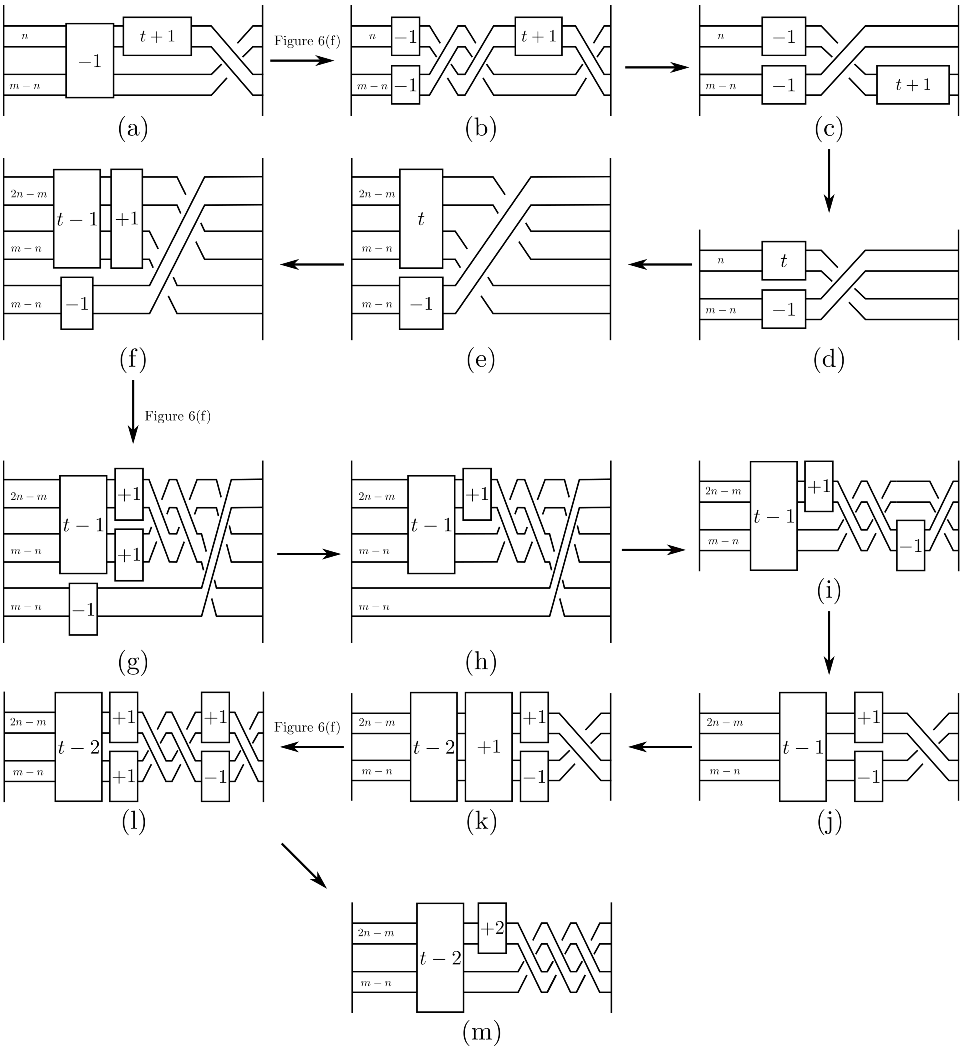

Figure 29

when $t = 2$ the curve is the closure of a negative braid: the braid in (a) in the figure is the braid from Figure 24(d). After applying the move in Figure 6(f) and simple isotopies, we obtain the braid in (c) which is clearly a negative braid when $t = 2$. □

Proof of Theorem 1.3. The proof of part (1) follows from Cases 1 and 2 above. Parts (2)(a)/(b) follows from Cases 3a/b and Case 4a above. As for part (3), observe that when $n > m$ by using Cases 1 and 2 we obtain that all homologically essential curves are the closures of positive braids. When $m > n$, we have either $m - 2n > 0$ or $m - 2n < 0$. In the former case we use Cases 3a and 4a to obtain that all homologically essential curves are the closures of negative braids. In the latter

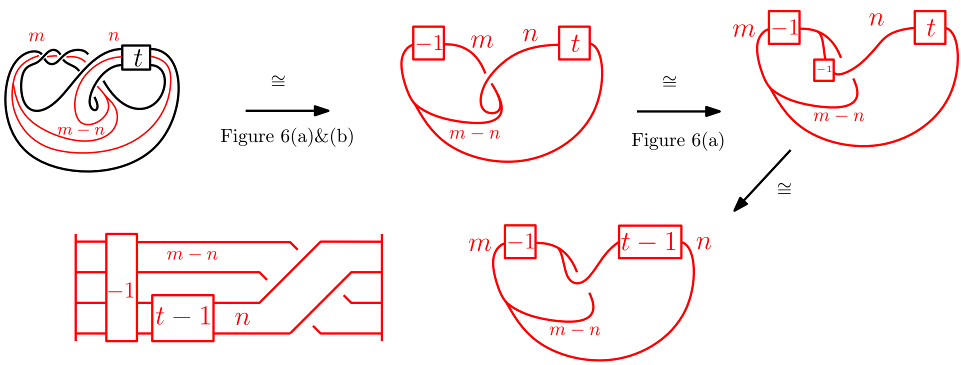

Figure 30

case, first note that $m - 2n < 0$ is equivalent to $m - n < n$, Now by Case 3b all homologically essential ∞ curves are the closures of positive braids, and by Case 4b all homologically essential loop curves are the closures of negative braids. Now by using Cromwell's result and some straightforward genus calculations we deduce that when $m > n > 1$ or $n > m \geq 1$ there are no unknotted curves among

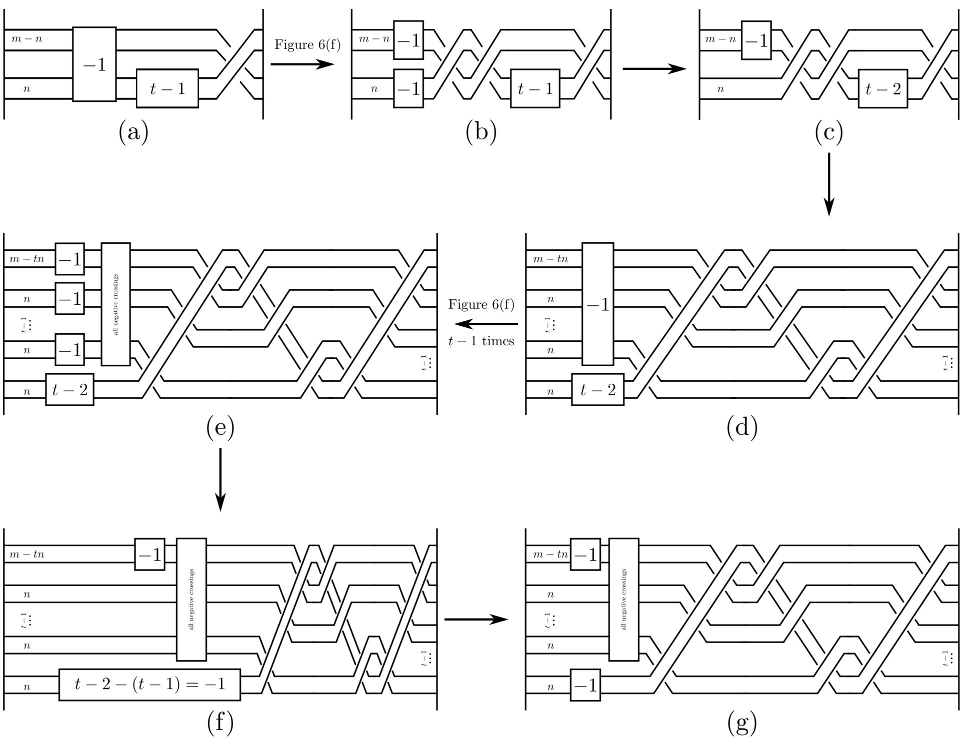

Figure 31

the (positive/negative) braid closures obtained in Cases 1–4 above. Therefore, there are exactly 5 unknotted curves among homologically essential curves on Σ_K for $K = K_t$ in Theorem 1.3. □

3D. *Twist knot with $t > 1$: part two.* In this section we consider the twist knots $K = K_t$, $t \geq 3$, and give the proof of Theorem 1.4.

Proof of Theorem 1.4. We show that the loop curve $(3, 2)$ when $t \geq 3$ is the pretzel knot $P(2t-5, -3, 2)$. This is explained in Figure 32. The braid in (a) is from Figure 24(d) with $m = 3$, $n = 2$, where we moved $(t-2)$ full right-handed twists to the top right end. We take the closure of the braid and cancel the left-handed half-twist on the top left with one of the right-handed half-twists on the top right to reach the knot in (c). In (c)–(g) we implement simple isotopies, and finally reach, in (h), the pretzel knot $P(2t-5, -3, 2)$. This knot has genus $t-1$ [Kim and Lee 2007, Corollary 2.7], and so is never unknotted as long as $t > 1$. This pretzel knot is slice exactly when $2t - 5 + (-3) = 0$, that is, when $t = 4$. The pretzel knot $P(3, -3, 3)$ is also known as 8_{20}. An interesting observation is that although $P(2t-5, -3, 2)$ for $t > 2$ is not a positive braid closure, it is a quasipositive braid closure. □

Proposition 3.8. *The (m, n) loop curve with $m - n = 1$, $n > 3$ and $t > 4$ is never slice.*

Proof. By Rudolph [1993], we have that, for a braid closure $\hat{\beta}$ when $k_+ \neq k_-$,

$$g_4(\hat{\beta}) \geq \tfrac{1}{2}(|k_+ - k_-| - n + 1),$$

where β is a braid in n strands, and $k_\pm$ is the number of positive and negative crossings in β. For quasipositive knots, equality holds, in which case, the Seifert genus is also the same as the four ball (slice) genus.

Now for the loop curve $c = (m, n)$ as in Figure 31(c), we have that

$$k_+ = (t-2)n(n-1) \quad \text{and} \quad k_- = (m-n)(m-n-1) + 3(m-n)n.$$

Hence, when $m - n = 1$, we get that $k_- = 3n$. Notice also that for $n \geq 3$, $t \geq 4$, we have $k_+ > k_-$. Thus, for $n > 3$, $t > 4$, $m - n = 1$, we obtain $c = \hat{\beta}$ is never slice because

$$g_4(\hat{\beta} = c) \geq \tfrac{1}{2}((t-2)n(n-1) - 3n - m + 1) = n((t-2)(n-1) - 4) > 0.$$

One can manually check that the $(4, 3)$ loop curve when $t = 3$ is also not slice. □

Remark 3.9. The inequality in the proof above can also be thought as a generalization to the Seifert genus calculation formula we used for positive/negative braid closures, since for those braids when $|k_+ - k_-|$ is the number of crossings, n, the braid number, is exactly the number of Seifert circles. Thus Rudolph's inequality

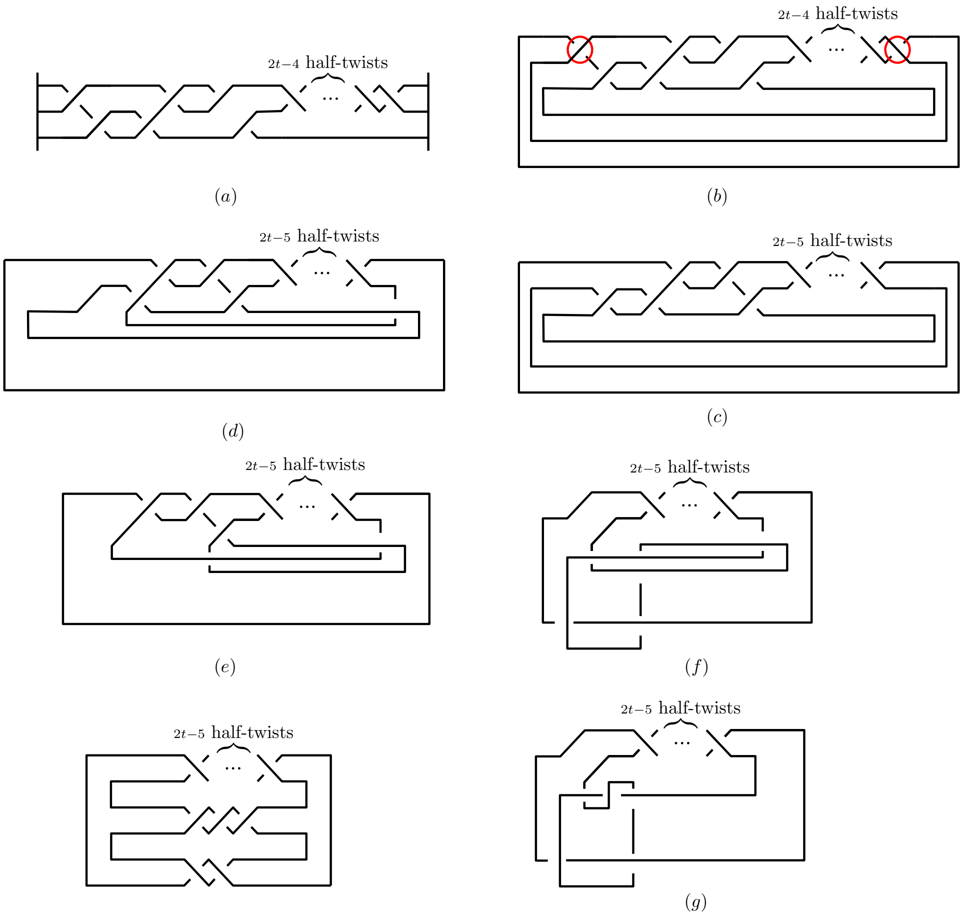

Figure 32

can also be used in the previous cases to show that there are no slice knots in the cases where we found that there are no unknotted curves.

4. Whitehead doubles

Proof of Theorem 1.6. Let $f : S^1 \times D^2 \to S^3$ denote a smooth embedding such that $f(S^1 \times \{0\}) = K$. Set $T = f(S^1 \times D^2)$. Up to isotopy, the collection of essential, simple closed, oriented curves in ∂T is parametrized by

$$\{m\mu + n\lambda \mid m, n \in \mathbb{Z} \text{ and } \gcd(m, n) = 1\},$$

where μ denotes a meridian in ∂T and λ denotes a standard longitude in ∂T coming from a Seifert surface [Rolfsen 1990; Saveliev 1999]. With this parametrization, the only curves that are null-homologous in T are $\pm\mu$ and the only curves that are

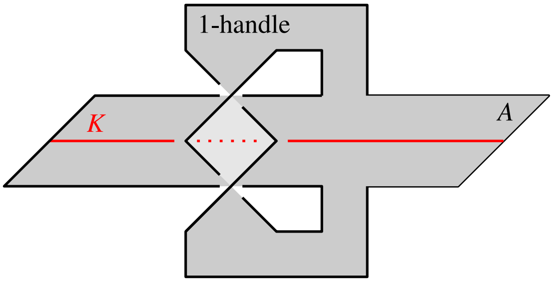

Figure 33. Standard genus-one Seifert surface F for a double of K.

null-homologous in $S^3 \setminus \mathrm{int}(T)$ are $\pm\lambda$. Of course $\pm\mu$ will bound embedded disks in T, but $\pm\lambda$ will not bound embedded disks in $S^3 \setminus \mathrm{int}(T)$ as K is a nontrivial knot. In other words, the only compressing curves for ∂T in S^3 are meridians.

Suppose now that C is a smooth, simple closed curve in the interior of T and there is a smoothly embedded 2-disk, say Δ, in S^3 such that $\partial\Delta = C$. Since C lies in the interior of T, we may assume that Δ meets ∂T transversely in a finite number of circles. Initially observe that if $\Delta \cap \partial T = \varnothing$, then we can use Δ to isotope C in the interior of T so that the result of this isotopy is a curve in the interior of T that misses a meridional disk for T. Now suppose that $\Delta \cap \partial T \neq \varnothing$. We show, in this case too, C can be isotoped to a curve that misses a meridional disk for T. To this end, let σ denote a simple closed curve in $\Delta \cap \partial T$ such that σ is innermost in Δ. That is, σ bounds a subdisk, Δ' say, in Δ, and the interior of Δ' misses ∂T. There are two cases, depending on whether or not σ is essential in ∂T. If σ is essential in ∂T, then, as has already been noted, σ must be a meridian. As such, Δ' will be a meridional disk in T, and C misses Δ'. If σ is not essential in ∂T, then σ bounds an embedded 2-disk, say D, in ∂T. It is possible that Δ meets the interior of D, but we can still cut and paste Δ along a subdisk of D to reduce the number of components in $\Delta \cap \partial T$. Repeating this process yields that if C is smoothly embedded curve in the interior of T and C is unknotted in S^3, then C can be isotoped in the interior of T so as to miss a meridional disk for T. (see [Rolfsen 1990, Theorem 9] and [Jaco 1980, page 13] for a use of similar ideas).

With all this in place, we return to discuss the Whitehead double of K. Suppose that F is a standard, genus-one Seifert surface for a double of K. See Figure 5. The surface F can be viewed as an annulus A with a 1-handle attached to it. Here K is a core circle for A, and the 1-handle is attached to A as depicted in Figure 33

Observe that F can be constructed so that it lives in the interior of T. Now, the curve C that passes once over the 1-handle and zero times around A obviously misses a meridional disk for T, and it obviously is unknotted in S^3. On the other hand, if C is any other essential simple closed curve in the interior of F, then C must go around A some positive number of times. It is not difficult, upon orienting, to see that C can be isotoped so that the strands of C going around A are coherently oriented. As such, C is homologous to some nonzero multiple of K in T. This, in turn, implies that C cannot be isotoped in T so as to miss some meridional disk for T. It follows that C cannot be an unknot in S^3. □

5. Contractible 4-manifolds and final remarks

Proof of Corollaries 1.9 and 1.10. In light of Theorem 1.7, the natural task is to determine the self-linking number s, with respect to the framing induced by the Seifert surface, for the unknotted curves found in Theorems 1.2 and 1.6. For this we use the Seifert matrix given by $S = \begin{pmatrix} -1 & -1 \\ 0 & t \end{pmatrix}$ where we use two obvious cycles — both oriented counterclockwise — in Σ_K. Recall that if $c = (m, n)$ is a loop curve, then m and n strands are endowed with the same orientation and hence the same signs. On the other hand, for an ∞ curve they will have opposite orientations and hence opposite signs. Therefore, given t, the self-linking number of a $c = (m, n)$ loop curve is $s = -m^2 - mn + n^2 t$, and the self-linking number of an (m, n) ∞ curve is $s = -m^2 + mn + n^2 t$. A quick calculation shows that the six unknotted curves in Figure 2 for $K_{-1} = T_{2,3}$ share self-linking numbers $s = -1, -3$. As we explained during the proof of Theorem 1.2 the infinitely many unknotted curves for the figure-eight knot $K_1 = 4_1$ reduce (via isotopies) to unknotted curves with $s = -1$ or $s = 1$. The five unknotted curves in Figure 4 for K_t, $t < -1$ or $t > 1$, share self-linking numbers $s = -1$, t and $t - 2$ (see [Cochran and Davis 2015]). Finally, Theorem 1.4 finds a slice but not unknotted curve, which is the curve $(3, 2)$ with $t = 4$. One can calculate from the formula above that this curve has self-linking number $s = 1$. Finally, the unique unknotted curve from Theorem 1.6 has self-linking number $s = -1$. The proofs follow as an obvious consequence of these calculations and Theorem 1.7 and its generalization in [Etnyre and Tosun 2023]. □

Next, we verify through Figure 34 that how not every essential curve on the genus-one Seifert surface of a twist knot must be the closure of a positive (or negative) braid closure. For example, we will show that an $(m, n) = (5, 2)$ ∞ curve on the Seifert surface of the twist knot K_3 as a smooth knot is the twist knot $m(5_2)$, which is known not to be a positive braid closure (e.g., via the KnotInfo database). To this end, we start with the braid as in Figure 34(a), which is the braid in Figure 27 where we substitute $m = 5$, $n = 2$ and $t = 3$. We then apply the move in Figure 6(f) to the full negative twist on 5 strands to obtain the braid in Figure 34(b). After a cancellation between a (-1) twist and a $(+4)$ twist and a small isotopy we get the braid in Figure 34(c). We apply the move in Figure 6(f) again, this time to the full negative twist on 3 strands from the bottom to obtain the braid in Figure 34(d). A small simplification gives the braid in Figure 34(e). Observe that the top strands can be eliminated — here it will be easier to think

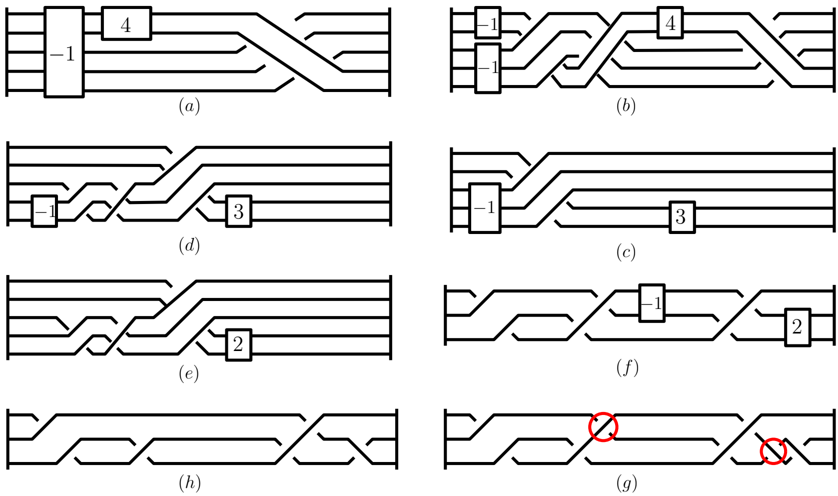

Figure 34. The knot $m(5_2)$ is an essential curve on the genus-one Seifert surface of the twist knot K_3.

of the corresponding braid closure — to get the 3-braid in Figure 34(f). A further simplification gives the braid in Figure 34(g). We can organize and simplify this braid by canceling the encircled half-crossings. This gives the braid in Figure 34(h). We claim that the closure of this braid is the knot $m(5_2)$-mirror of 5_2. One can see this by taking the closure and applying simple plane isotopies. This method is quite easy (and fun) but slightly lengthier. An alternative method is to observe that this braid has braid description $-1, -2, -2, -2, -1, 2$, which we can reorder, via braid isotopy, to be $-2, -2, -2, -1, 2, -1$. Now a quick inspection in the KnotInfo database [Livingston and Moore 2024] shows that the knot 5_2 has braid description $1, 1, 1, 2, -1, 2$. So the closure of the braid in Figure 34 is indeed $m(5_2)$. The KnotInfo database can also be used to verify the knot $m(5_2)$ is not the closure of a positive/negative braid.

Acknowledgments

We thank Audrick Pyronneau and Nicolas Fontova for helpful conversations. We thank Filip Misev for useful comments on an early draft of this paper. We are also grateful to the referee for their careful reading and many suggestions. Dey, King, and Shaw were supported in part by an NSF grant (DMS-2105525). Tosun was supported in part by grants from the NSF (CAREER DMS-2144363 and DMS-2105525) and the Simons Foundation (636841, BT and 2023 Simons Fellowship). Part of this work was carried out while Tosun was a member at the Institute for Advanced Study, and he acknowledges support from the Charles Simonyi Endowment at the Institute for Advanced Study.

References

[Casson and Gordon 1978] A. J. Casson and C. M. Gordon, "On slice knots in dimension three", pp. 39–53 in *Algebraic and geometric topology*, *II*, edited by R. J. Milgram, Proc. Sympos. Pure Math. **32**, Amer. Math. Soc., Providence, RI, 1978. MR Zbl

[Casson and Harer 1981] A. J. Casson and J. L. Harer, "Some homology lens spaces which bound rational homology balls", *Pacific J. Math.* **96**:1 (1981), 23–36. MR Zbl

[Cochran and Davis 2015] T. D. Cochran and C. W. Davis, "Counterexamples to Kauffman's conjectures on slice knots", *Adv. Math.* **274** (2015), 263–284. MR Zbl

[Cromwell 1989] P. R. Cromwell, "Homogeneous links", *J. London Math. Soc.* (2) **39**:3 (1989), 535–552. MR Zbl

[Şavk 2024] O. Şavk, "A survey of the homology cobordism group", *Bull. Amer. Math. Soc.* (*N.S.*) **61**:1 (2024), 119–157. MR Zbl

[Etnyre and Tosun 2023] J. B. Etnyre and B. Tosun, "Homology spheres bounding acyclic smooth manifolds and symplectic fillings", *Michigan Math. J.* **73**:4 (2023), 719–734. MR Zbl

[Fickle 1984] H. C. Fickle, "Knots, $\mathbb{Z}$-homology 3-spheres and contractible 4-manifolds", *Houston J. Math.* **10**:4 (1984), 467–493. MR Zbl

[Fintushel and Stern 1985] R. Fintushel and R. J. Stern, "Pseudofree orbifolds", *Ann. of Math.* (2) **122**:2 (1985), 335–364. MR Zbl

[Hirsch 1961] M. W. Hirsch, "On imbedding differentiable manifolds in euclidean space", *Ann. of Math.* (2) **73** (1961), 566–571. MR Zbl

[Jaco 1980] W. Jaco, *Lectures on three-manifold topology*, CBMS Regional Conference Series in Mathematics **43**, Amer. Math. Soc., Providence, RI, 1980. MR Zbl

[Kim and Lee 2007] D. Kim and J. Lee, "Some invariants of pretzel links", *Bull. Austral. Math. Soc.* **75**:2 (2007), 253–271. MR Zbl

[Kirby 1978] R. Kirby, "Problems in low dimensional manifold theory", pp. 273–312 in *Algebraic and geometric topology*, *II*, edited by R. J. Milgram, Proc. Sympos. Pure Math. **32**, Amer. Math. Soc., Providence, RI, 1978. MR Zbl

[Livingston and Moore 2024] C. Livingston and A. H. Moore, "KnotInfo: table of knot invariants", January 7, 2024, available at knotinfo.math.indiana.edu.

[Manolescu 2016] C. Manolescu, "Pin(2)-equivariant Seiberg–Witten Floer homology and the triangulation conjecture", *J. Amer. Math. Soc.* **29**:1 (2016), 147–176. MR Zbl

[Rohlin 1952] V. A. Rohlin, "New results in the theory of four-dimensional manifolds", *Doklady Akad. Nauk SSSR* (*N.S.*) **84** (1952), 221–224. In Russian. MR

[Rohlin 1965] V. A. Rohlin, "The embedding of non-orientable three-manifolds into five-dimensional Euclidean space", *Dokl. Akad. Nauk SSSR* **160** (1965), 549–551. In Russian. MR Zbl

[Rolfsen 1990] D. Rolfsen, *Knots and links*, revised 2nd ed., Mathematics Lecture Series **7**, Publish or Perish, 1990. MR Zbl

[Rudolph 1993] L. Rudolph, "Quasipositivity as an obstruction to sliceness", *Bull. Amer. Math. Soc.* (*N.S.*) **29**:1 (1993), 51–59. MR Zbl

[Saveliev 1999] N. Saveliev, *Lectures on the topology of* 3*-manifolds: an introduction to the Casson invariant*, de Gruyter, Berlin, 1999. MR Zbl

[Stern 1978] R. Stern, "Some more Brieskorn spheres which bound contractible manifolds", *Notices Amer. Math. Soc.* **25**:4 (1978), abstract 78T–G75.

[Stoimenow 2003] A. Stoimenow, "Positive knots, closed braids and the Jones polynomial", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **2**:2 (2003), 237–285. MR Zbl

[Tosun 2022] B. Tosun, "Stein domains in $\mathbb{C}^2$ with prescribed boundary", *Adv. Geom.* **22**:1 (2022), 9–22. MR Zbl

[Wall 1965] C. T. C. Wall, "All 3-manifolds imbed in 5-space", *Bull. Amer. Math. Soc.* **71** (1965), 564–567. MR Zbl

[Zeeman 1965] E. C. Zeeman, "Twisting spun knots", *Trans. Amer. Math. Soc.* **115** (1965), 471–495. MR Zbl

Received August 8, 2023. Revised January 8, 2024.

SUBHANKAR DEY
DURHAM UNIVERSITY
DURHAM, ENGLAND
UNITED KINGDOM
subhankar.dey@durham.ac.uk

VERONICA KING
UNIVERSITY OF TEXAS AUSTIN
AUSTIN, TX
UNITED STATES
viking0905@gmail.com

COLBY T. SHAW
SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GA
UNITED STATES
cshaw44@gatech.edu

BÜLENT TOSUN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ALABAMA
TUSCALOOSA, AL
UNITED STATES
and
SCHOOL OF MATHEMATICS
INSTITUTE FOR ADVANCED STUDY
PRINCETON, NJ
UNITED STATES
btosun@ua.edu

BRUCE TRACE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ALABAMA
TUSCALOOSA, AL
UNITED STATES
btrace@ua.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 330, No. 1, 2024
https://doi.org/10.2140/pjm.2024.330.157

ON THE GAUSS MAPS OF COMPLETE MINIMAL SURFACES IN $\mathbb{R}^n$

DINH TUAN HUYNH

Dedicated to Professor Doan The Hieu

We prove that the generalized Gauss map of a nonflat complete minimal surface immersed in $\mathbb{R}^n$ can omit a generic hypersurface D of degree at most $n^{n+2}(n+1)^{n+2}$.

1. Introduction

Let $f = (x_1, x_2, \dots, x_n) : M \to \mathbb{R}^n$ be an oriented surface immersed in $\mathbb{R}^n$. Using systems of isothermal coordinates (x, y), one can consider M as a Riemann surface. We are interested in the class of minimal surfaces, namely, those which have minimal areas for all small perturbations. It is a well-known fact that if M is minimal, then its generalized Gauss map $g : M \to \mathbb{CP}^{n-1}$, defined as

$$g(z) := [\partial f / \partial z],$$

where $z = x + iy$ is a holomorphic chart on M, is a holomorphic map.

In the particular case where $n = 3$, recalling that the classical Gauss map of M is the map sending each point $p \in M$ to the point in the unit sphere corresponding to the unit normal vector of M at p. By identifying the unit sphere with the complex projective line via the stereographic projection, one can view the classical Gauss map as a map of M into $\mathbb{CP}^1$. Osserman [18] proved that if M is a nonflat complete minimal surface immersed in $\mathbb{R}^3$, then the complement of the image of its Gauss map is of logarithmic capacity zero in $\mathbb{CP}^1$. This interesting result could be regarded as a significant improvement of the classical Bernstein's theorem. Strengthening this result, Xavier [22] proved that in this situation, the Gauss map of M can avoid at most 6 points. Sharp result was obtained by Fujimoto [10], where he proved that indeed, the Gauss map of M can avoid at most 4 points.

Passing to higher-dimensional case, first step was made by Fujimoto [9], where the intersection between the generalized Gauss maps of a complete minimal surface immersed in $\mathbb{R}^n$ and family of hyperplanes in $\mathbb{CP}^{n-1}$ was considered. Precisely, Fujimoto established that:

MSC2020: primary 53A10; secondary 32H30.
Keywords: value distribution theory, Gauss map, minimal surface, hyperbolicity.

Theorem 1.1. *If the generalized Gauss map of a nonflat complete minimal surface in $\mathbb{R}^n$ is nondegenerate, it can omit at most $q = n^2$ hyperplanes in $\mathbb{CP}^{n-1}$ in general position.*

Later, Fujimoto himself [11] decreased the number of hyperplanes in the above statement to $q = \frac{1}{2}n(n+1)$ and it turns out that this number is sharp. Ru [19] was able to remove the nondegenerate assumption of the generalized Gauss map in Fujimoto's result. Since then, by adapting tools and techniques from value distribution theory of holomorphic curves to study generalized Gauss maps, many generalizations of the above works of Fujimoto–Ru were made. Note that in these results, it is required the presence of many hypersurfaces.

In this paper, based on recent progresses towards the hyperbolicity problem [1; 2; 4; 6; 7; 8; 13; 21], we consider the case when there is only one hypersurface of high enough degree.

Theorem 1.2 (Main Theorem). *Let M be a nonflat complete minimal surface immersed in $\mathbb{R}^n$ and let $G : M \to \mathbb{CP}^{n-1}$ be its generalized Gauss map. Then G can avoid a generic hypersurface $D \subset \mathbb{CP}^{n-1}$ of degree at most*

$$d = n^{n+2}(n+1)^{n+2}.$$

2. Logarithmic jet differentials

Let X be a complex projective variety of dimension n. For a point $x \in X$, consider the holomorphic germs $(\mathbb{C}, 0) \to (X, x)$. Two such germs are said to be equivalent if they have the same Taylor expansion up to order k in some local coordinates around x. The equivalence class of an analytic germ $f : (\mathbb{C}, 0) \to (X, x)$ is called the *k-jet of f*, denoted by $j_k(f)$, which is independent of the choice of local coordinates. A k-jet $j_k(f)$ is said to be *regular* if $\mathrm{d}f(0) \neq 0$. For a given point $x \in X$, denote by $j_k(X)_x$ the vector space of all k-jets of analytic germs $(\mathbb{C}, 0) \to (X, x)$, set

$$J_k(X) := \bigcup_{x \in X} J_k(X)_x,$$

and consider the natural projection

$$\pi_k : J_k(X) \to X.$$

Then $J_k(X)$ carries the structure of a holomorphic fiber bundle over X, which is called the *k-jet bundle over X*. Note that in general, $J_k(X)$ is not a vector bundle. When $k = 1$, the 1-jet bundle $J_1(X)$ is canonically isomorphic to the tangent bundle T_X of X.

For an open subset $U \subset X$, for $\omega \in H^0(U, T_X^*)$, for a k-jet $j_k(f) \in J_k(X)|_U$, the pullback $f^*\omega$ is of the form $A(z)\,\mathrm{d}z$ for some analytic function A, where z is the global coordinate of $\mathbb{C}$. Since each derivative $A^{(j)}$ $(0 \leq j \leq k-1)$ is well defined, independent of the representation of f in the class $j_k(f)$, the analytic 1-form ω induces the holomorphic map

$$\tilde{\omega} : J_k(X)|_U \to \mathbb{C}^k, \quad j_k(f) \to \big(A(z), A(z)^{(1)}, \dots, A(z)^{(k-1)}\big). \tag{2-1}$$

Hence on an open subset U, a given local holomorphic coframe $\omega_1 \wedge \cdots \wedge \omega_n \neq 0$ yields a trivialization

$$H^0(U, J_k(X)) \to U \times (\mathbb{C}^k)^n$$

by providing the following new nk independent coordinates:

$$\sigma \to (\pi_k \circ \sigma; \tilde{\omega}_1 \circ \sigma, \dots, \tilde{\omega}_n \circ \sigma),$$

where $\tilde{\omega}_i$ are defined as in (2-1). The components $x_i^{(j)}$ $(1 \le i \le n,\ 1 \le j \le k)$ of $\tilde{\omega}_i \circ \sigma$ are called the *jet-coordinates*. In a more general setting, where ω is a section over U of the sheaf of meromorphic 1-forms, the induced map $\tilde{\omega}$ is meromorphic.

Now, suppose that $D \subset X$ is a normal crossing divisor on X. This means that at each point $x \in X$, there exist some local coordinates $z_1, \dots, z_\ell, z_{\ell+1}, \dots, z_n$ $(\ell = \ell(x))$ centered at x in which D is defined by

$$D = \{z_1 \cdots z_\ell = 0\}.$$

Following Iitaka [14], the *logarithmic cotangent bundle of X along D*, denoted by $T_X^*(\log D)$, corresponds to the locally free sheaf generated by

$$\frac{\mathrm{d}z_1}{z_1}, \dots, \frac{\mathrm{d}z_\ell}{z_\ell}, z_{\ell+1}, \dots, z_n$$

in the above local coordinates around x.

A holomorphic section $s \in H^0(U, J_k(X))$ over an open subset $U \subset X$ is said to be a *logarithmic k-jet field* if $\tilde{\omega} \circ s$ are analytic for all sections $\omega \in H^0(U', T_X^*(\log D))$, for all open subsets $U' \subset U$, where $\tilde{\omega}$ are induced maps defined as in (2-1). Such logarithmic k-jet fields define a subsheaf of $J_k(X)$, and this subsheaf is itself a sheaf of sections of a holomorphic fiber bundle over X, called the *logarithmic k-jet bundle over X along D*, denoted by $J_k(X, -\log D)$ (see [16]).

The group $\mathbb{C}^*$ admits a natural fiberwise action defined as follows. For local coordinates

$$z_1, \dots, z_\ell, z_{\ell+1}, \dots, z_n \quad (\ell = \ell(x))$$

centered at x in which $D = \{z_1 \dots z_\ell = 0\}$, for any logarithmic k-jet field along D represented by some germ $f = (f_1, \dots, f_n)$, if $\varphi_\lambda(z) = \lambda z$ is the homothety with ratio $\lambda \in \mathbb{C}^*$, the action is given by

$$\begin{cases} (\log(f_i \circ \varphi_\lambda))^{(j)} = \lambda^j (\log f_i)^{(j)} \circ \varphi_\lambda & (1 \le i \le \ell), \\ (f_i \circ \varphi_\lambda)^{(j)} = \lambda^j f_i^{(j)} \circ \varphi_\lambda & (\ell + 1 \le i \le n). \end{cases}$$

A *logarithmic jet differential of* ***order*** k *and* ***degree*** m at a point $x \in X$ is a polynomial $Q(f^{(1)}, \dots, f^{(k)})$ on the fiber over x of $J_k(X, -\log D)$ enjoying weighted homogeneity:

$$Q(j_k(f \circ \varphi_\lambda)) = \lambda^m Q(j_k(f)) \quad (\lambda \in \mathbb{C}^*).$$

Consider the symbols

$$\mathrm{d}^j \log z_i \quad (1 \le j \le k,\ 1 \le i \le \ell) \qquad \text{and} \qquad \mathrm{d}^j z_i \quad (1 \le j \le k,\ \ell + 1 \le i \le n).$$

Set the weight of $\mathrm{d}^j \log z_i$ or $\mathrm{d}^j z_i$ to be j. Then a logarithmic jet differential of order k and weight k along D at x is a weighted homogeneous polynomial of degree m whose variables are these symbols. Denote by $E^{GG}_{k,m} T^*_X(\log D)_x$ be the vector space spanned by such polynomials and set

$$E^{GG}_{k,m} T^*_X(\log D) := \bigcup_{x \in X} E^{GG}_{k,m} T^*_X(\log D)_x.$$

By Faà di Bruno's formula [3; 15], one can check that $E^{GG}_{k,m} T^*_X(\log D)$ carries the structure of a vector bundle over X, called *logarithmic Green–Griffiths vector bundle* [12]. A global section of $E^{GG}_{k,m} T^*_X(\log D)$ is called a *logarithmic jet differential* of order k and weight m along D. Locally, a logarithmic jet differential form can be written as

$$\sum_{\substack{\alpha_1, \dots, \alpha_k \in \mathbb{N}^n \\ |\alpha_1| + 2|\alpha_2| + \cdots + k|\alpha_k| = m}} A_{\alpha_1, \dots, \alpha_k} \left(\prod_{i=1}^{\ell} (\mathrm{d} \log z_i)^{\alpha_{1,i}} \prod_{i=\ell+1}^{n} (\mathrm{d} z_i)^{\alpha_{1,i}} \right) \cdots \left(\prod_{i=1}^{\ell} (\mathrm{d}^k \log z_i)^{\alpha_{k,i}} \prod_{i=\ell+1}^{n} (\mathrm{d}^k z_i)^{\alpha_{k,i}} \right), \tag{2-2}$$

where

$$\alpha_\lambda = (\alpha_{\lambda,1}, \dots, \alpha_{\lambda,n}) \in \mathbb{N}^n \quad (1 \le \lambda \le k)$$

are multiindices of length

$$|\alpha_\lambda| = \sum_{1 \le i \le n} \alpha_{\lambda,i},$$

and where $A_{\alpha_1, \dots, \alpha_k}$ are locally defined holomorphic functions.

Assigning the weight s for $(\mathrm{d}^s z_i)/z_i$, then one can rewritten $\mathrm{d}^j \log z_i$ as an isobaric polynomial of weight j of variables $(\mathrm{d}^s z_i)/z_i$ $(1 \le s \le j)$ with integer coefficients, namely

$$\mathrm{d}^j \log z_i = \sum_{\substack{\beta = (\beta_1, \dots, \beta_j) \in \mathbb{N}^j \\ \beta_1 + 2\beta_2 + \cdots + j\beta_j = j}} b_{j\beta} \left(\frac{\mathrm{d} z_i}{z_i} \right)^{\beta_1} \cdots \left(\frac{\mathrm{d}^j z_i}{z_i} \right)^{\beta_j},$$

where $b_{j\beta} \in \mathbb{Z}$. Conversely, one can also express $(\mathrm{d}^j z_i)/z_i$ as an isobaric polynomial of weight j of variables $\mathrm{d}^s \log z_i$ $(1 \le s \le j)$ with integer coefficients [2]. Thus one can also use the following trivialization of logarithmic jet differentials:

$$\sum_{\substack{\beta_1,\dots,\beta_k\in\mathbb{N}^n\\ |\beta_1|+2|\beta_2|+\cdots+k|\beta_k|=m}} B_{\beta_1,\dots,\beta_k}\biggl(\prod_{i=1}^{\ell}\Bigl(\frac{\mathrm{d}z_i}{z_i}\Bigr)^{\beta_{1,i}}\prod_{i=\ell+1}^{n}(\mathrm{d}z_i)^{\beta_{1,i}}\biggr) \cdots\biggl(\prod_{i=1}^{\ell}\Bigl(\frac{\mathrm{d}^k z_i}{z_i}\Bigr)^{\beta_{k,i}}\prod_{i=\ell+1}^{n}(\mathrm{d}^k z_i)^{\beta_{k,i}}\biggr), \tag{2-3}$$

where

$$\beta_\lambda=(\beta_{\lambda,1},\dots,\beta_{\lambda,n})\in\mathbb{N}^n \quad (1\le\lambda\le k)$$

are multiindices of length

$$|\beta_\lambda|=\sum_{1\le i\le n}\beta_{\lambda,i},$$

and where $B_{\beta_1,\dots,\beta_k}$ are locally defined holomorphic functions.

Demailly [5] refined the Green–Griffiths' theory and considered the subbundle $E_{k,m}T_X^*(\log D)$ of $E_{k,m}^{GG}T_X^*(\log D)$, whose sections are logarithmic jet differentials that are invariant under arbitrary reparametrization of the source $\mathbb{C}$. Let

$$(X,D,V)$$

be a *log-direct manifold*, i.e., a triple consisting of a projective manifold X, a simple normal crossing divisor D on X and a holomorphic subbundle V of the logarithmic tangent bundle $T_X(-\log D)$. Starting with a log-direct manifold $(X_0,D_0,V_0):=(X,D,T_X(-\log D))$, one then defines $X_1:=\mathbb{P}(V_0)$ together with the natural projection $\pi_1:X_1\to X_0$. Setting $D_1:=\pi_1^*D_0$, so that π_1 becomes a log-morphism, and defines the subbundle $V_1\subset T_{X_1}(-\log D_1)$ as

$$V_{1,(x,[v])}:=\bigl\{\xi\in T_{X_1,(x,[v])}(-\log D_1):\pi_*\xi\in\mathbb{C}\cdot v\bigr\},$$

one obtains the log-direct manifold (X_1,D_1,V_1) from the initial one. Any germ of a holomorphic map $f:(\mathbb{C},0)\to(X\setminus D,x)$ can be lifted to $f^{[1]}:\mathbb{C}\to X_1\setminus D_1$. Inductively, one can construct on $X=X_0$ the *Demailly–Semple tower*:

$$(X_k,D_k,V_k)\to\cdots\to(X_1,D_1,V_1)\to(X_0,D_0,V_0),$$

together with the projections $\pi_k:X_k\to X_0$. Denote by $\mathcal{O}_{X_k}(1)$ the tautological line bundle on X_k. Then the direct image $(\pi_k)_*\mathcal{O}_{X_k}(m)$ of $\mathcal{O}_{X_k}(m)=\mathcal{O}_{X_k}(1)^{\otimes m}$, denoted by $E_{k,m}T_X^*(\log D)$, is a locally free subsheaf of $E_{k,m}^{GG}T_X^*(\log D)$ generated by all polynomial operators in the derivatives up to order k, which are furthermore invariant under any change of parametrization $(\mathbb{C},0)\to(\mathbb{C},0)$. From the construction, one can immediately check that:

Theorem 2.1 (direct image formula). *For any ample line bundle $\mathcal{A}$ on X, one has*

$$H^0(X,E_{k,m}T_X^*(\log D)\otimes\mathcal{A}^{-1})\cong H^0(X_k,\mathcal{O}_{X_k}(m)\otimes\pi_k^*\mathcal{A}^{-1}). \tag{2-4}$$

The bundles $E_{k,m}^{GG}\,T_X^*(\log D)$, $E_{k,m}T_X^*(\log D)$ are fundamental tools in studying the degeneracy of holomorphic curves into $\mathbb{C}\setminus D$. By the fundamental vanishing theorem of entire curves [5; 21], for any ample line bundle $\mathcal{A}$ on X, a nontrivial global section of $E_{k,m}^{GG}\,T_X^*(\log D)\otimes\mathcal{A}^{-1}$ gives a corresponding algebraic differential equation that all entire holomorphic function $f:\mathbb{C}\to X\setminus D$ must satisfy. The existence of these sections was proved recently [6; 15], provided that the order of jet is high enough. However, despite many efforts, the problem of controlling the base locus of these bundles can be only handled under the condition that the degree of D must be very large compared with the dimension of the variety [1; 2; 7; 8; 21].

Now we consider the case where D is a generic hypersurface of degree d in $\mathbb{CP}^n$. To guarantee the existence of logarithmic jet differentials along D, we consider the order jet $k=n+1$ and put

$$k'=\tfrac{1}{2}k(k+1),\qquad \delta=(k+1)\,n+k.$$

Fixing two positive integers $\epsilon>0$ and $r>\delta^{k-1}k(\epsilon+k\delta)$. For a smooth hypersurface D, denote by $Y_k(D)$ the log-Demailly–Semple k-jet tower associated to $(\mathbb{CP}^n,\,D,\,T_{\mathbb{CP}^n}(-\log D))$. For a line bundle L on $\mathcal{O}_{Y_k(D)}$, denote by $\mathrm{Bs}(\mathcal{O}_{Y_k(D)}L)$ the base locus of the line bundle L. We will employ the following key result in [2].

Proposition 2.2 [2, Corollary 4.5]. *There exist $\beta,\tilde\beta\in\mathbb{N}$ such that for any $\alpha\geq 0$ and for any generic hypersurface $D\in|\mathcal{O}_{\mathbb{CP}^n(1)}^{\epsilon+(r+k)\,\delta}|$, one has*

$$\mathrm{Bs}\big(\mathcal{O}_{Y_k(D)}(\beta+\alpha\,\delta^{k-1}k')\otimes\pi_{0,k}^*\,\mathcal{O}_{\mathbb{CP}^n(1)}^{\tilde\beta+\alpha(\delta^{k-1}k(\epsilon+k\delta)-r)}\big)\subset Y_k(D)^{\mathrm{sing}}\cup\pi_{0,k}^{-1}(D).$$

Using this result, Brotbek–Deng confirmed the logarithmic Kobayashi conjecture in the case where the degree of D is large enough. We extract from their proof that:

Theorem 2.3. *Let $D\subset\mathbb{P}^n(\mathbb{C})$ be a generic smooth hypersurface in $\mathbb{P}^n(\mathbb{C})$ having large enough degree*

$$d\geq(n+1)^{n+3}(n+2)^{n+3}.$$

Let $f:\Delta\to\mathbb{CP}^n$ be a nonconstant holomorphic disk. If $f(\Delta)\not\subset D$, then for jet order $k=n+1$, there exist some weighted degree m, vanishing order $\widetilde m$ with $\widetilde m>2m$ and some global logarithmic jet differential

$$\mathscr{P}\in H^0\big(\mathbb{CP}^n,\,E_{k,m}^{GG}\,T_{\mathbb{CP}^n}^*(\log D)\otimes\mathcal{O}_{\mathbb{CP}^n}(1)^{-\widetilde m}\big)$$

such that

$$\mathscr{P}(j_k(f))\not\equiv 0. \tag{2-5}$$

Proof. We follow the arguments in [2, Corollary 4.9], with a slightly modification to get higher vanishing order. First, putting

$$r_0=2\delta^{k-1}k'+\delta^{k-1}(\delta+1)^2=\delta^{k-1}(\delta+1)(\delta+2).$$

Since

$$k(k+\delta-1+k\delta) < (\delta+1)^2,$$

any integer number $d \geq (r_0+k)\,\delta + 2\delta$ can be written as

$$d = \epsilon + (r+k)\,\delta,$$

where $k \leq \epsilon \leq k+\delta-1$ and $r > 2\delta^{k-1}k' + \delta^{k-1}k(\epsilon+k\delta)$. For such d, since

$$\lim_{\alpha\to\infty} \frac{\beta+\alpha\,\delta^{k-1}k'}{-\tilde{\beta}-\alpha(\delta^{k-1}k(\epsilon+k\delta)-r)} = \frac{\alpha\,\delta^{k-1}k'}{r-\delta^{k-1}k(\epsilon+k\delta)} < \tfrac{1}{2},$$

using Proposition 2.2, for $\alpha \gg 1$ large enough, there exists some global logarithmic jet differential

$$\mathscr{P} \in H^0\big(\mathbb{CP}^n, E^{GG}_{k,m}\, T^*_{\mathbb{CP}^n}(\log D) \otimes \mathcal{O}_{\mathbb{CP}^n}(1)^{-\widetilde{m}}\big)$$

satisfying (2-5) with $m = \beta + \alpha\,\delta^{k-1}k'$, $\widetilde{m} = -\tilde{\beta} - \alpha(\delta^{k-1}k(\epsilon+k\delta)-r)$ and $\widetilde{m} > 2m$. Hence it remains to giving a lower bound for $(r_0+k)\,\delta+2\delta$. This could be done by a straightforward computation:

$$(r_0+k)\,\delta + 2\delta = \big(\delta^{k-1}(\delta+1)(\delta+2)+k+2\big)\,\delta < (n+1)^{n+3}(n+2)^{n+3}. \quad \square$$

3. Value distribution theory for holomorphic maps from unit disc into projective spaces

Let $E = \sum_i \alpha_i\, a_i$ be a divisor on the unit disc Δ where $\alpha_i \geq 0$, $a_i \in \Delta$ and let $k \in \mathbb{N} \cup \{\infty\}$. For each $0 < t < 1$, denote by Δ_t the disk $\{z \in \mathbb{C},\, |z| < t\}$. Summing the k-truncated degrees of the divisor on disks by

$$n^{[k]}(t, E) := \sum_{a_i \in \Delta_t} \min\{k, \alpha_i\} \quad (0 < t < 1),$$

the *truncated counting function* at level k of E is then defined by taking the logarithmic average

$$N^{[k]}(r, E) := \int_0^r \frac{n^{[k]}(t, E)}{t}\, \mathrm{d}t \quad (0 < r < 1).$$

When $k = \infty$, we write $n(t, E)$, $N(r, E)$ instead of $n^{[\infty]}(t, E)$, $N^{[\infty]}(r, E)$. Let $f : \Delta \to \mathbb{CP}^n$ be an entire curve having a reduced representation $f = [f_0 : \cdots : f_n]$ in the homogeneous coordinates $[z_0 : \cdots : z_n]$ of $\mathbb{CP}^n$. Let $D = \{Q = 0\}$ be a divisor in $\mathbb{CP}^n$ defined by a homogeneous polynomial $Q \in \mathbb{C}[z_0, \ldots, z_n]$ of degree $d \geq 1$. If $f(\Delta) \not\subset D$, we define the *truncated counting function* of f with respect to D as

$$N^{[k]}_f(r, D) := N^{[k]}(r, (Q \circ f)_0),$$

where $(Q \circ f)_0$ denotes the zero divisor of $Q \circ f$.

The *proximity function* of f for the divisor D is defined as

$$m_f(r, D) := \int_0^{2\pi} \log \frac{\|f(re^{i\theta})\|^d \|Q\|}{|Q(f)(re^{i\theta})|} \frac{d\theta}{2\pi},$$

where $\|Q\|$ is the maximum absolute value of the coefficients of Q and

$$\|f(z)\| = \max\{|f_0(z)|, \dots, |f_n(z)|\}.$$

Since $|Q(f)| \leq \|Q\| \cdot \|f\|^d$, one has $m_f(r, D) \geq 0$.

Lastly, the *Cartan order function* of f is defined by

$$T_f(r) := \frac{1}{2\pi} \int_0^{2\pi} \log \|f(re^{i\theta})\| \, d\theta.$$

With the above notations, the Nevanlinna theory consists of two fundamental theorems (for comprehensive presentations, see [17; 20]).

Theorem 3.1 (First Main Theorem). *Let $f : \Delta \to \mathbb{P}^n(\mathbb{C})$ be a holomorphic curve and let D be a hypersurface of degree d in $\mathbb{CP}^n$ such that $f(\Delta) \not\subset D$. Then for every $r > 1$, the following holds*:

$$m_f(r, D) + N_f(r, D) = d\, T_f(r) + O(1),$$

whence

$$N_f(r, D) \leq d\, T_f(r) + O(1). \tag{3-1}$$

On the other side, in the harder part, so-called Second Main Theorem, one tries to bound the order function from above by some sum of certain counting functions. Such types of results were given in several situations, and most of them were relied on the following key estimate.

Theorem 3.2 (logarithmic derivative lemma). *Let g be a nonconstant meromorphic function on the unit disc and let $k \geq 1$ be a positive integer number. Then for any $0 < r < 1$, the following estimate holds*:

$$m_{g^{(k)}/g}(r) := m_{g^{(k)}/g}(r, \infty) = O\left(\log \frac{1}{1-r}\right) + O(\log T_g(r)) \qquad \|,$$

where the notation $\|$ means that the above estimate holds true for all $0 < r < 1$ outside a subset $E \subset (0, 1)$ with

$$\int_E \frac{dr}{1-r} < \infty.$$

4. An application of the logarithmic derivative lemma

It is a well-known fact that the growth of the order function of an entire holomorphic curve could be used to determine its rationality. Replacing the source of the curve by the unit disc Δ, one has:

Definition 4.1. A holomorphic map $f : \Delta \to \mathbb{CP}^n$ is said to be transcendental if

$$\limsup_{r\to 1} \frac{T_f(r)}{\log \frac{1}{1-r}} = \infty.$$

Theorem 4.2. *Let $f : \Delta \to \mathbb{CP}^n$ be a holomorphic map and $D \subset \mathbb{CP}^n$ be a generic hypersurface having large enough degree*:

$$d \geq (n+1)^{n+3}(n+2)^{n+3}.$$

If f avoids D, then it is not transcendental.

Proof. Employing the logarithmic jet differentials supplied by Theorem 2.3, following the arguments as in [13] and using the logarithmic derivative lemma for meromorphic functions on unit disc, one gets

$$T_f(r) \leq N_f^{[1]}(r, D) + O\left(\log \frac{1}{1-r}\right) + O(\log T_f(r)) = O\left(\frac{1}{1-r}\right) + O(\log T_f(r)) \quad \|,$$

whence concludes the proof. □

We will also need the following results due to Fujimoto [9].

Proposition 4.3 [9, Proposition 2.5]. *Let φ be a nowhere zero holomorphic function on Δ which is not transcendental. Then, for each positive integer number λ, the following estimate holds*:

$$\int_0^{2\pi} \left| \frac{d^{\lambda-1}}{dz^{\lambda-1}} \left(\frac{\varphi'}{\varphi} \right) (re^{i\theta}) \right| d\theta \leq \frac{\text{Const.}}{(1-r)^{\lambda}} \log \frac{1}{1-r} \quad (0 < r < 1).$$

Corollary 4.4 [9, Lemma 3.4]. *Let $\varphi_1, \ldots, \varphi_n$ be nowhere zero holomorphic functions on Δ which are not transcendental. Then, for any n-tuple of positive integer numbers $(\lambda_1, \ldots, \lambda_n)$ and for any positive real number t with $tn < 1$, the following estimate holds*:

$$\int_0^{2\pi} \left| \prod_{j=1}^{n} \left(\frac{\varphi_j'}{\varphi_j} \right)^{(\lambda_j - 1)} (re^{i\theta}) \right|^t d\theta \leq \frac{\text{Const.}}{(1-r)^s} \left(\log \frac{1}{1-r} \right)^s \quad (0 < r < 1),$$

where $s = t\left(\sum_{j=1}^n \lambda_j\right)$.

5. Proof of the Main result

Proposition 5.1. *Let $D \subset \mathbb{CP}^n$ be a generic smooth hypersurface of degree d and let $f : \Delta \to \mathbb{CP}^n \setminus D$ be a nondegenerate holomorphic curve. Suppose that there exists a global logarithmic jet differential*

$$\mathscr{P} \in H^0\big(\mathbb{P}^n(\mathbb{C}),\, E^{GG}_{n,m}\, T^*_{\mathbb{CP}^n}(\log D) \otimes \mathcal{O}_{\mathbb{CP}^n}(1)^{-\widetilde{m}}\big)$$

such that

$$\mathscr{P}(j_n(f)) \not\equiv 0. \tag{5-1}$$

Then, there exists a positive constant K such that

$$\int_0^{2\pi} \big|\mathscr{P}(j_n(f))(re^{i\theta})\big|^{2/\widetilde{m}} \, \|f(re^{i\theta})\|^2 \, d\theta \leq \frac{K}{(1-r)^{2m/\widetilde{m}}} \left(\log \frac{1}{1-r}\right)^{2m/\widetilde{m}}$$

for $(0 < r < 1)$.

Proof. Let s be the canonical section of the ample line bundle $\mathcal{E} := \mathcal{O}_{\mathbb{CP}^n}(1)$. Since $\mathscr{P}$ vanishes on $\mathcal{E}$ with vanishing order $\widetilde{m}$, in any local chat U_α of $\mathbb{CP}^n$, one can represent $\mathscr{P}s^{\widetilde{m}}$ as an isobaric polynomial $\mathscr{P}^\alpha_s$ of weight m of variables

$$\frac{d^\lambda\, u^\alpha_{j,\lambda}}{u^\alpha_{j,\lambda}} \quad (1 \leq \lambda \leq k,\ 1 \leq j \leq n),$$

with local holomorphic coefficients, where $u_{j,\lambda}$ are rational functions on $\mathbb{CP}^n$. Consequently, we get that

$$|\mathscr{P}(j_k(f))| \cdot \|f\|^{\widetilde{m}} \leq \sum_\alpha \left| \mathscr{P}^\alpha_s \left(\frac{d^\lambda (u^\alpha_{j,\lambda} \circ f)}{u^\alpha_{j,\lambda} \circ f} \right) \right|.$$

Since $0 < \frac{2}{\widetilde{m}} < 1$, using the elementary inequality

$$(x_1 + \cdots + x_r)^{2/\widetilde{m}} < x_1^{2/\widetilde{m}} + \cdots + x_r^{2/\widetilde{m}} \quad (x_i > 0),$$

the above estimate yields

$$\big|\mathscr{P}(j_n(f))(re^{i\theta})\big|^{2/\widetilde{m}} \, \|f(re^{i\theta})\|^2 < \sum_\alpha \left| \mathscr{P}^\alpha_s \left(\frac{d^\lambda (u^\alpha_{j,\lambda} \circ f)}{u^\alpha_{j,\lambda} \circ f} \right) \right|^{2/\widetilde{m}}.$$

Hence it suffices to prove

$$\int_0^{2\pi} \left| \mathscr{P}_s^{2/\widetilde{m}} \left(\frac{d^\lambda (u^\alpha_{j,\lambda} \circ f)}{u^\alpha_{j,\lambda} \circ f} \right) \right|^{2m/\widetilde{m}} d\theta \leq \frac{\text{Const.}}{(1-r)^{2m/\widetilde{m}}} \left(\log \frac{1}{1-r}\right)^{2m/\widetilde{m}} \quad (0 < r < 1).$$

By assumption, f avoids D, hence it is not transcendental by Theorem 4.2. Since each function $u^\alpha_{j,\lambda}$ is rational, it follows that $u^\alpha_{j,\lambda} \circ f$ is also not transcendental.

Now, observing that each term

$$\frac{d^{\lambda}(u^{\alpha}_{j,\lambda}\circ f)}{u^{\alpha}_{j,\lambda}\circ f}$$

can be represented as a polynomial $\mathscr{P}^{\alpha}_{j,\lambda}$ of variables

$$\frac{(u^{\alpha}_{j,\lambda}\circ f)'}{u^{\alpha}_{j,\lambda}\circ f},\ldots,\left(\frac{(u^{\alpha}_{j,\lambda}\circ f)'}{u^{\alpha}_{j,\lambda}\circ f}\right)^{\lambda-1},$$

which is isobaric of weight λ, using Corollary 4.4, one gets the desired result. □

We will also need the following result of Yau [23] in the sequence.

Theorem 5.2 [23]. *Let M be a complete Riemann manifold equipped with a volume form $d\sigma$. Let h be a nonnegative and nonconstant smooth function on M such that $\Delta \log h = 0$ almost everywhere. Then $\int_M h^p\,d\sigma = \infty$ for any $p > 0$.*

Now we enter the details of the proof of the Main Theorem. Let f be the conjugate of G, which is a holomorphic map. It suffices to prove that f is constant. Suppose on the contrary that this is not the case. Let $\pi : \widetilde{M} \to M$ be the universal covering of M. Then $\widetilde{M}$ is also considered as a minimal surface in $\mathbb{R}^n$. Hence without lost of generality, we may assume $M = \widetilde{M}$. Since there is no compact minimal surface in $\mathbb{R}^n$, it follows that M is biholomorphic to either $\mathbb{C}$ or Δ. Thus we may assume $M = \mathbb{C}$ or $M = \Delta$. The first case was excluded by recent work towards Kobayashi's conjecture (see [2]). Hence it suffices to work in the case where $M = \Delta$. The area form of the metric on M induced from the flat metric on $\mathbb{R}^n$ is given by

$$d\sigma = 2\|f\|^2\,du\wedge dv.$$

Let $\mathscr{P}$ be a global logarithmic jet differential supplied by Theorem 2.3. Then it is clear that $h = |\mathscr{P}(j_k(f)| \not\equiv 0$ and $\Delta \log h = 0$ for any z out side the zero set of h. Since Δ is complete, simply connected and of nonpositive curvature, it has the infinite area with respect to the metric induced from $\mathbb{R}^n$. Using Theorem 5.2, one obtains that

$$\int_{\Delta} h^{2/\widetilde{m}}\,d\sigma = \infty. \tag{5-2}$$

On the other hand, using Proposition 5.1, one has

$$\begin{aligned}\int_{\Delta} h^{2/\widetilde{m}}\,d\sigma &= 2\int_{\Delta} h^{2/\widetilde{m}}\,\|f\|^2\,du\,dv\\ &= 2\int_0^1 r\,dr\left(\int_0^{2\pi} h(re^{i\theta})^{2/\widetilde{m}}\,\|f(re^{i\theta})\|^2\,d\theta\right)\\ &\le K\int_0^1 \frac{r}{(1-r)^{2m/\widetilde{m}}}\left(\log\frac{1}{1-r}\right)^{2m/\widetilde{m}} dr.\end{aligned}$$

The last integral in the above estimate is finite since $2m < \widetilde{m}$. This contradicts (5-2). Therefore, the map f must be constant, whence concludes the proof of the Main Theorem.

6. Some discussions

Theorem 1.1 can be recovered via the above jet method. Indeed, according to Siu [21], the Wronskian can be employed to build a suitable logarithmic jet differentials. Precisely, let us consider the inhomogeneous coordinates $x_1, x_2, \dots, x_n$ of $\mathbb{CP}^n$. Let $\{H_i\}_{1\le i\le q}$ be the family of hyperplanes in general position in $\mathbb{CP}^n$. For each $1 \le i \le q$, denote by F_i the linear form of variables $x_1, \dots, x_n$ defining the hyperplane H_i. Put

$$\omega = \frac{\mathrm{Wron}(dx_1, \dots, dx_n)}{F_1 \dots F_q},$$

where Wron denotes the Wronskian. The point is that by the assumption of general position, at any point $x = (x_1, \dots, x_n)$, there exists a set $I = \{i_1, \dots, i_n\}$ having cardinality n such that F_j are nowhere zero in a neighborhood U of x for all $j \notin I$. Locally on U, one can write ω as

$$\omega = \mathrm{Const.}\,\frac{\mathrm{Wron}(d\log F_{i_1}(x), \dots, d\log F_{i_n}(x))}{\prod_{j\notin I} F_j(x)},$$

and hence, ω gives rise to a logarithmic jet differentials along the divisor $\sum_{i=1}^q H_i$. The denominator $F_1 \dots F_q$ in ω gives the vanishing order q at the infinity hyperplane, hence direct computation yields immediately that ω is of weight $m = \frac{1}{2}n(n+1)$ and vanishes on the infinity hyperplane with the vanishing order $\widetilde{m} = q - (n+1)$.

Finally, in view of the result of Fujimoto–Ru, one can expect that the optimal degree bound in the statement of our Main Theorem should be $\frac{1}{2}n(n+1)$.

Conjecture 6.1. *Let M be a nonflat complete minimal surface in $\mathbb{R}^n$ and let $G : M \to \mathbb{CP}^{n-1}$ be its generalized Gauss map. Then G could avoid a generic hypersurface $D \subset \mathbb{CP}^{n-1}$ of degree at most*

$$d = \tfrac{1}{2}n(n+1).$$

Acknowledgements

This work is supported by the Vietnam Ministry of Education and Training under the grant number B2024-DHH-14. I would like to thank Prof. Doan The Hieu for his encouragements. I want to thank Song-Yan Xie for helpful suggestions which improved the exposition. I warmly thank the referee for the careful reading of the first version of the manuscript and for the comments that helped me clarify the presentation.

References

[1] G. Bérczi, "Towards the Green–Griffiths–Lang conjecture via equivariant localisation", *Proc. Lond. Math. Soc.* (3) **118**:5 (2019), 1057–1083. MR Zbl

[2] D. Brotbek and Y. Deng, "Kobayashi hyperbolicity of the complements of general hypersurfaces of high degree", *Geom. Funct. Anal.* **29**:3 (2019), 690–750. MR Zbl

[3] G. M. Constantine and T. H. Savits, "A multivariate Faà di Bruno formula with applications", *Trans. Amer. Math. Soc.* **348**:2 (1996), 503–520. MR Zbl

[4] L. Darondeau, "On the logarithmic Green–Griffiths conjecture", *Int. Math. Res. Not.* **2016**:6 (2016), 1871–1923. MR Zbl

[5] J.-P. Demailly, "Algebraic criteria for Kobayashi hyperbolic projective varieties and jet differentials", pp. 285–360 in *Algebraic geometry* (Santa Cruz, 1995), Proc. Sympos. Pure Math. **62(2)**, Amer. Math. Soc., Providence, RI, 1997. MR Zbl

[6] J.-P. Demailly, "Hyperbolic algebraic varieties and holomorphic differential equations", *Acta Math. Vietnam.* **37**:4 (2012), 441–512. MR Zbl

[7] J.-P. Demailly, "Recent results on the Kobayashi and Green–Griffiths–Lang conjectures", *Jpn. J. Math.* **15**:1 (2020), 1–120. MR Zbl

[8] S. Diverio, J. Merker, and E. Rousseau, "Effective algebraic degeneracy", *Invent. Math.* **180**:1 (2010), 161–223. MR Zbl

[9] H. Fujimoto, "On the Gauss map of a complete minimal surface in $\mathbf{R}^m$", *J. Math. Soc. Japan* **35**:2 (1983), 279–288. MR Zbl

[10] H. Fujimoto, "On the number of exceptional values of the Gauss maps of minimal surfaces", *J. Math. Soc. Japan* **40**:2 (1988), 235–247. MR Zbl

[11] H. Fujimoto, "Modified defect relations for the Gauss map of minimal surfaces, II", *J. Differential Geom.* **31**:2 (1990), 365–385. MR Zbl

[12] M. Green and P. Griffiths, "Two applications of algebraic geometry to entire holomorphic mappings", pp. 41–74 in *The Chern Symposium* 1979: *Proc. Internat. Sympos.* (Berkeley, Calif., 1979), Springer, New York, 1980. MR Zbl

[13] D. T. Huynh, D.-V. Vu, and S.-Y. Xie, "Entire holomorphic curves into projective spaces intersecting a generic hypersurface of high degree", *Ann. Inst. Fourier* (*Grenoble*) **69**:2 (2019), 653–671. MR Zbl

[14] S. Iitaka, *Algebraic geometry*: *an introduction to birational geometry of algebraic varieties*, Graduate Texts in Mathematics **76**, Springer, New York, 1982. MR Zbl

[15] J. Merker, "Algebraic differential equations for entire holomorphic curves in projective hypersurfaces of general type: optimal lower degree bound", pp. 41–142 in *Geometry and analysis on manifolds*, Progr. Math. **308**, Birkhäuser, Cham, 2015. MR Zbl

[16] J. Noguchi, "Logarithmic jet spaces and extensions of de Franchis' theorem", pp. 227–249 in *Contributions to several complex variables*, Aspects Math. **E9**, Friedr. Vieweg, Braunschweig, 1986. MR Zbl

[17] J. Noguchi and J. Winkelmann, *Nevanlinna theory in several complex variables and Diophantine approximation*, Fundamental Principles of Mathematical Sciences **350**, Springer, Tokyo, 2014. MR Zbl

[18] R. Osserman, "Global properties of minimal surfaces in E^3 and E^n", *Ann. of Math.* (2) **80** (1964), 340–364. MR Zbl

[19] M. Ru, "On the Gauss map of minimal surfaces immersed in $\mathbf{R}^n$", *J. Differential Geom.* **34**:2 (1991), 411–423. MR Zbl

[20] M. Ru, *Nevanlinna theory and its relation to Diophantine approximation*, 2nd ed., World Scientific Publishing, Hackensack, NJ, 2021. MR Zbl

[21] Y.-T. Siu, "Hyperbolicity of generic high-degree hypersurfaces in complex projective space", *Invent. Math.* **202**:3 (2015), 1069–1166. MR Zbl

[22] F. Xavier, "The Gauss map of a complete nonflat minimal surface cannot omit 7 points of the sphere", *Ann. of Math.* (2) **113**:1 (1981), 211–214. MR Zbl

[23] S. T. Yau, "Some function-theoretic properties of complete Riemannian manifold and their applications to geometry", *Indiana Univ. Math. J.* **25**:7 (1976), 659–670. MR Zbl

Received January 13, 2024. Revised March 14, 2024.

DINH TUAN HUYNH
DEPARTMENT OF MATHEMATICS
HUE UNIVERSITY OF EDUCATION
HUE UNIVERSITY
HUE CITY
VIETNAM

dinhtuanhuynh@hueuni.edu.vn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 330, No. 1, 2024
https://doi.org/10.2140/pjm.2024.330.171

EXPLICIT BOUNDS ON TORSION OF CM ABELIAN VARIETIES OVER p-ADIC FIELDS WITH VALUES IN LUBIN–TATE EXTENSIONS

YOSHIYASU OZEKI

Let K and k be p-adic fields. Let L be the composite field of K and a certain Lubin–Tate extension over k (including the case where $L = K(\mu_{p^\infty})$). We show that there exists an explicitly described constant C, depending only on K, k and an integer $g \geq 1$, which satisfies the following property: if $A_{/K}$ is a g-dimensional CM abelian variety, then the order of the p-primary torsion subgroup of $A(L)$ is bounded by C. We also give a similar bound in the case where $L = K(\sqrt[p^\infty]{K})$. Applying our results, we study bounds of orders of torsion subgroups of some CM abelian varieties over number fields with values in full cyclotomic fields.

1. Introduction

Let p be a prime number and K a p-adic field (that is, a finite extension of $\mathbb{Q}_p$). It is a theorem of Mattuck [1955] that, for a g-dimensional abelian variety A over K and a finite extension L/K, the Mordell–Weil group $A(L)$ is isomorphic to the direct sum of $\mathbb{Z}_p^{\oplus g\cdot[L:\mathbb{Q}_p]}$ and a finite group. We study some properties of the torsion subgroup $A(L)_{\mathrm{tor}}$ of $A(L)$. Clark and Xarles [2008] gave an explicit upper bound of the order of $A(L)_{\mathrm{tor}}$ of $A(L)$ in terms of p, g and some numerical invariants of L if A has anisotropic reduction (here, we say that A has anisotropic reduction if its Néron special fiber does not contain a copy of $\mathbb{G}_m$). This includes the case where A has potential good reduction. We consider the case where L/K is *of infinite degree*. There are some situations in which the torsion part $A(L)_{\mathrm{tor}}$ is finite. Suppose that A has potential good reduction. It is a theorem of Imai [1975] that $A(K(\mu_{p^\infty}))_{\mathrm{tor}}$ is finite. Here, $K(\mu_{p^\infty})$ is the extension field of K obtained by adjoining all p-power roots of unity. Moreover, Kubo and Taguchi [2013] showed that $A(K(\sqrt[p^\infty]{K}))_{\mathrm{tor}}$ is also finite, where $K(\sqrt[p^\infty]{K})$ is the extension field of K obtained by adjoining all p-power roots of all elements of K. The author showed in [Ozeki 2024] that there exists a "uniform" bound of the order of $A(K(\sqrt[p^\infty]{K}))_{\mathrm{tor}}$ under the assumption that A has complex multiplication. (Here we say that A has complex multiplication

MSC2020: 11G10.
Keywords: abelian varieties, Lubin–Tate extensions.

if there exists a ring homomorphism $F \to \mathbb{Q} \otimes_{\mathbb{Z}} \mathrm{End}_{\overline{K}}(A)$ for some algebraic number field F of degree $2g$.)

The main purpose of this paper is to give explicit upper bounds of the orders of $A(K(\mu_{p^\infty}))_{\mathrm{tor}}$ and $A(K(\sqrt[p^\infty]{K}))_{\mathrm{tor}}$ for abelian varieties A/K with complex multiplication. For this, we should note that to give an upper bound of the order of the prime-to-p part of $A(K(\mu_{p^\infty}))_{\mathrm{tor}}$ is not so difficult. In fact, the reduction map gives an injection from the prime-to-p part of the group which we want to study into certain rational points of the reduction $\bar{A}$ of A (if A has good reduction), and the order of the target is bounded by the Weil bound. Hence the essential obstruction for our purpose appears in a study of the p-part $A(K(\mu_{p^\infty}))[p^\infty]$ of $A(K(\mu_{p^\infty}))_{\mathrm{tor}}$.

Let us state our main results. For a p-adic field k and a uniformizer π of k, we denote by k_π/k the Lubin–Tate extension associated with π (that is, k_π is the extension field of k obtained by adjoining all π-power torsion points of the Lubin–Tate formal group associated with π; see [Yoshida 2008] for more details). For example, we have $k_\pi = \mathbb{Q}_p(\mu_{p^\infty})$ if $k = \mathbb{Q}_p$ and $\pi = p$. We set $d_L := [L : \mathbb{Q}_p]$ for any p-adic field L. For any integer $n > 0$, we set

$$\Phi(n) := \mathrm{Max}\{m \in \mathbb{Z}_{>0} \mid \varphi(m) \text{ divides } 2n\},$$
$$H(n) := \gcd\{\#\mathrm{GSp}_{2n}(\mathbb{Z}/N\mathbb{Z}) \mid N \geq 3\}.$$

Here, φ is Euler's totient function. There are some upper bounds related with $H(n)$ and $\Phi(n)$ (see Section 5). It is a theorem of Silverberg [1992, Corollary 3.3] that we have $H(n) < 2(9n)^{2n}$ for any $n > 0$. It follows from elementary arguments that we have $\Phi(n) < 6n\sqrt[3]{n}$ for $n > 1$. Furthermore, a lower bound (5-3) of φ proved by Rosser and Schoenfeld [1962] gives $\Phi(n) < 4n \log\log n$ for $n > 3^{3^9}$.

Theorem 1.1 (a special case of Theorem 3.1). *Let $g > 0$ be a positive integer. Let k be a p-adic field with residue cardinality q_k and π a uniformizer of k. Assume the following conditions*:

(i) $q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi)$ *is a root of unity*;[1] *and*

(ii) d_k *is prime to* $(2g)!$.

Denote by $0 < \mu < p$ the minimum integer such that $(q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi))^\mu = 1$. For any g-dimensional abelian variety A over a p-adic field K with complex multiplication,

$$A(Kk_\pi)[p^\infty] \subset A[p^C],$$

where

$$C := 2g^2 \cdot (2g)! \cdot \Phi(g)H(g) \cdot \mu \cdot d_{Kk} + 12g^2 - 18g + 10.$$

In particular,

$$\#A(Kk_\pi)[p^\infty] \leq p^{2gC}.$$

[1]This condition is equivalent to saying that some finite extension of k_π contains $\mathbb{Q}_p(\mu_{p^\infty})$ (see [Ozeki 2020, Lemma 2.7(2)]).

As an immediate consequence of the theorem above, we obtain a result for cyclotomic extensions; see Corollary 3.7. Furthermore, the method of our proof of Theorem 1.1 can be applied to the field $K(\sqrt[p^\infty]{K})$ discussed in Kubo and Taguchi, which gives a refinement of the main theorem of [Ozeki 2024].

Theorem 1.2. *Let $g > 0$ be a positive integer. For any g-dimensional abelian variety A over a p-adic field K with complex multiplication, we have*

$$A(K(\sqrt[p^\infty]{K}))[p^\infty] \subset A[p^C],$$

where

$$C := 2g^2 \cdot (2g)! \cdot p^{1+v_p(2)} \cdot (\Phi(g)H(g))^2 \cdot p^{v_p(d_K)} d_K + 12g^2 - 18g + 10.$$

(*Here*, v_p *is the p-adic valuation normalized by* $v_p(p) = 1$.) *In particular*,

$$\#A(K(\sqrt[p^\infty]{K}))[p^\infty] \le p^{2gC}.$$

We can consider some further topics. For example, we do not know what will happen if we remove the CM assumption from the above theorems. Our proofs in this paper deeply depend on the theory of locally algebraic representations, which can be adapted only for abelian representations. This is the main reason why we cannot remove the CM assumption from our arguments. To overcome this obstruction, it seems to be helpful for us to study the case of (not necessarily CM) elliptic curves. We will study this case in future work. We are also interested in giving the list of the groups that appear as $A(Kk_\pi)[p^\infty]$ or $A(K(\sqrt[p^\infty]{K}))[p^\infty]$. However, this should be quite difficult; the author does not know such classification results even for $A(K)[p^\infty]$.

Combining the cyclotomic case of Theorem 1.1 and Ribet's arguments given in [Katz and Lang 1981], we can obtain a result on a bound of the order of the torsion subgroup of some CM abelian variety defined over a number field with values in full cyclotomic fields. (Here, a number field is a finite extension of $\mathbb{Q}$.)

Theorem 1.3. *Let $g > 0$ be an integer. Let K be a number field of degree d, and denote by h the narrow class number of K. Let $K(\mu_\infty)$ be the field obtained by adjoining to K all roots of unity. Let A be a g-dimensional abelian variety over K with complex multiplication which has good reduction everywhere. Then*

$$A(K(\mu_\infty))_{\mathrm{tor}} \subset A[N],$$

where

$$N := \Bigl(\prod_p p\Bigr)^{2g^2\cdot(2g)!\cdot\Phi(g)H(g)\cdot dh+12g^2-18g+10}.$$

Here, p *ranges over the prime numbers such that either* $p \le (1+\sqrt{2}^{dh})^{2g}$ *or p is ramified in* K.

We should note that Chou [2019] gave the complete list of the groups that appear as $A(\mathbb{Q}(\mu_\infty))_{\mathrm{tor}}$ as A ranges over all elliptic curves defined over $\mathbb{Q}$. For CM elliptic curves A over a number field K, more precise observations for the order of $A(K(\mu_\infty))_{\mathrm{tor}}$ than ours are studied in [Chou et al. 2021].

Notation. For any perfect field F, we denote by G_F the absolute Galois group of F. In this paper, a p-adic field is a finite extension of $\mathbb{Q}_p$. If F is an algebraic extension of $\mathbb{Q}_p$, we denote by $\mathcal{O}_F$ the ring of integers of F. We also denote by F^{ab} the maximal abelian extension of F (in a fixed algebraic closure of F). We put $d_F = [F : \mathbb{Q}_p]$ if F is a p-adic field. For a finite extension F'/F, we denote by $e_{F'/F}$ and $f_{F'/F}$ the ramification index of F'/F and the extension degree of the residue field extension of F'/F, respectively. We set $e_F := e_{F/\mathbb{Q}_p}$ and $f_F := f_{F/\mathbb{Q}_p}$, and also set $q_F := p^{f_F}$. Finally, we denote by Γ_F the set of $\mathbb{Q}_p$-algebra embeddings of F into a (fixed) algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$.

2. Evaluations of some p-adic valuations for characters

Fix an algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$. Throughout this section, we assume that all p-adic fields are subfields of $\overline{\mathbb{Q}}_p$. Denote by v_p the p-adic valuation normalized by $v_p(p) = 1$. For any continuous character ψ of G_K, we often regard ψ as a character of $\mathrm{Gal}(K^{\mathrm{ab}}/K)$. Denote by Art_K the local Artin map $K^\times \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$. We set $\psi_K := \psi \circ \mathrm{Art}_K$. Denote by $\hat{K}^\times$ the profinite completion of $K^\times$. Note that the local Artin map induces a topological isomorphism $\mathrm{Art}_K : \hat{K}^\times \xrightarrow{\sim} \mathrm{Gal}(K^{\mathrm{ab}}/K)$. For a uniformizer π_K of K, denote by $\chi_{\pi_K} : G_K \to \mathcal{O}_K^\times$ the Lubin–Tate character associated with π_K (see [Serre 1989, Chapter III, A4]). By definition, the character χ_{π_K} is characterized by $\chi_{\pi_K,K}(\pi_K) = 1$ and $\chi_{\pi_K,K}(x) = x^{-1}$ for any $x \in \mathcal{O}_K^\times$. Let π be a uniformizer of k and denote by k_π the Lubin–Tate extension of k associated with π. The field corresponding to the kernel of the Lubin–Tate character $\chi_\pi : G_k \to \mathcal{O}_k^\times$ is k_π, and k_π is a totally ramified abelian extension of k.

Proposition 2.1. *Let $\psi_1, \dots, \psi_n : G_K \to \overline{\mathbb{Q}}_p^\times$ be continuous characters. Then*

$$\mathrm{Min}\left\{ \sum_{i=1}^{n} v_p(\psi_i(\sigma) - 1) \,\middle|\, \sigma \in G_{Kk_\pi} \right\} \leq \mathrm{Min}\left\{ \sum_{i=1}^{n} v_p(\psi_{i,Kk}(\omega) - 1) \,\middle|\, \omega \in \mathrm{Nr}_{Kk/k}^{-1}(\pi^{f_{Kk/k}\mathbb{Z}}) \right\}.$$

Proof. This is [Ozeki 2024, Proposition 2.1] but we include a proof here for completeness. Let M be the maximal unramified extension of k contained in Kk. The group $\mathrm{Art}_k^{-1}(\mathrm{Gal}(k^{\mathrm{ab}}/M))$ contains $\mathrm{Art}_k^{-1}(\mathrm{Gal}(k^{\mathrm{ab}}/k^{\mathrm{ur}})) = \mathcal{O}_k^\times$. Furthermore, $\mathrm{Art}_k^{-1}(\mathrm{Gal}(k^{\mathrm{ab}}/M))$ is a subgroup of $\hat{k}^\times = \pi^{\hat{\mathbb{Z}}} \times \mathcal{O}_k^\times$ of index $[M : k] = f_{Kk/k}$. Thus it holds that $\mathrm{Art}_k^{-1}(\mathrm{Gal}(k^{\mathrm{ab}}/M)) = \pi^{f_{Kk/k}\hat{\mathbb{Z}}} \times \mathcal{O}_k^\times$. Since $\mathrm{Art}_k^{-1}(\mathrm{Gal}(k^{\mathrm{ab}}/k_\pi)) = \pi^{\hat{\mathbb{Z}}}$, we obtain that $\mathrm{Art}_k^{-1}(\mathrm{Gal}(k^{\mathrm{ab}}/Mk_\pi)) = \pi^{f_{Kk/k}\hat{\mathbb{Z}}}$. If we denote by $\mathrm{Res}_{Kk/k}$ the natural restriction map from $\mathrm{Gal}((Kk)^{\mathrm{ab}}/Kk)$ to $\mathrm{Gal}(k^{\mathrm{ab}}/k)$, it is not difficult to check that $\mathrm{Res}_{Kk/k}^{-1}(\mathrm{Gal}(k^{\mathrm{ab}}/Mk_\pi)) = \mathrm{Gal}((Kk)^{\mathrm{ab}}/Kk_\pi)$, and therefore we find that $\mathrm{Art}_{Kk}^{-1}\big(\mathrm{Gal}((Kk)^{\mathrm{ab}}/Kk_\pi)\big) = \mathrm{Nr}_{Kk/k}^{-1}(\pi^{f_{Kk/k}\hat{\mathbb{Z}}})$. Now the lemma follows from

$$\mathrm{Min}\left\{ \sum_{i=1}^{n} v_p(\psi_i(\sigma)-1) \;\middle|\; \sigma \in G_{Kk_\pi} \right\} = \mathrm{Min}\left\{ \sum_{i=1}^{n} v_p(\psi_{i,Kk} \circ \mathrm{Art}_{Kk}^{-1}(\sigma)-1) \;\middle|\; \sigma \in \mathrm{Gal}((Kk)^{\mathrm{ab}}/Kk_\pi) \right\}. \qquad \square$$

We often use p-adic Hodge theory, which plays an important role in this paper. For the basic notion of p-adic Hodge theory, it is helpful for the reader to refer to [Fontaine 1994a; 1994b]. Let B_{cris} be the Fontaine's p-adic period ring and set $D_{\mathrm{cris}}^K(V) := (B_{\mathrm{cris}} \otimes_{\mathbb{Q}_p} V)^{G_K}$ for any $\mathbb{Q}_p$-representation V of G_K. Let us denote by K_0 the maximal unramified subextension of $K/\mathbb{Q}_p$ and denote by φ_{K_0} the arithmetic Frobenius map of K_0, that is, the (unique) lift of the p-th power map on the residue field of K_0. Since $B_{\mathrm{cris}}^{G_K} = K_0$, $D_{\mathrm{cris}}^K(V)$ is a K_0-vector space. Moreover, $D_{\mathrm{cris}}^K(V)$ is a filtered φ-module over K; it is of finite dimension over K_0, it is equipped with a bijective φ_{K_0}-semilinear Frobenius operator φ and it is equipped with a decreasing exhaustive and separated filtration on $D_{\mathrm{cris}}^K(V) \otimes_{K_0} K$. We say that V is crystalline if the equality $\dim_{\mathbb{Q}_p} V = \dim_{K_0} D_{\mathrm{cris}}^K(V)$ holds. Let M be a finite extension of $\mathbb{Q}_p$ and $\psi : G_K \to M^\times$ a continuous character. We denote by $M(\psi)$ the $\mathbb{Q}_p$-representation of G_K underlying a 1-dimensional M-vector space endowed with an M-linear action by G_K via ψ. We say that ψ is crystalline if $M(\psi)$ is crystalline. On the other hand, we denote by $\underline{K}^\times$ the Weil restriction $\mathrm{Res}_{K/\mathbb{Q}_p}(\mathbb{G}_m)$. This is an algebraic torus such that, for a $\mathbb{Q}_p$-algebra R, the R-valued points $\underline{K}^\times(R)$ of $\underline{K}^\times$ is $\mathbb{G}_m(R \otimes_{\mathbb{Q}_p} K)$.

Proposition 2.2. *Let $\psi : G_K \to M^\times$ be a continuous character.*

(1) *$M(\psi)$ is crystalline if and only if there exists a (necessarily unique) $\mathbb{Q}_p$-homomorphism $\psi_{\mathrm{alg}} : \underline{K}^\times \to \underline{M}^\times$ such that ψ_K and ψ_{alg} (on $\mathbb{Q}_p$-points) coincide on $\mathcal{O}_K^\times$ ($\subset \underline{K}^\times(\mathbb{Q}_p)$).*

(2) *Assume $M(\psi)$ is crystalline, and let ψ_{alg} be as in (1). (Note $M(\psi^{-1})$ is also crystalline.) The filtered φ-module $D_{\mathrm{cris}}^K(M(\psi^{-1})) = (B_{\mathrm{cris}} \otimes_{\mathbb{Q}_p} M(\psi^{-1}))^{G_K}$ over K is free of rank 1 over $K_0 \otimes_{\mathbb{Q}_p} M$, and its K_0-linear endomorphism φ^{f_K} is given by the action of the product $\psi_K(\pi_K) \cdot \psi_{\mathrm{alg}}^{-1}(\pi_K) \in M^\times$. Here, π_K is any uniformizer of K.*

Proof. This is Proposition B.4 of [Conrad 2011]. $\square$

Let $\psi : G_K \to M^\times$ be a crystalline character. For any $\sigma \in \Gamma_M$, let $\chi_{\sigma M} : I_{\sigma M} \to \sigma M^\times$ be the restriction to the inertia $I_{\sigma M}$ of the Lubin–Tate character associated with any choice of uniformizer of σM (it depends on the choice of a uniformizer of σM, but its restriction to the inertia subgroup does not). Assume that K contains the Galois closure of $M/\mathbb{Q}_p$. Then

$$\psi = \prod_{\sigma\in\Gamma_M} \sigma^{-1}\circ \chi_{\sigma M}^{h_\sigma}$$

on the inertia I_K for some integer h_σ. Equivalently, the character ψ_{alg} on $\mathbb{Q}_p$-points coincides with $\prod_{\sigma\in\Gamma_M}\sigma^{-1}\circ \mathrm{Nr}_{K/\sigma M}^{-h_\sigma}$. Note that $\{h_\sigma \mid \sigma\in\Gamma_M\}$ is the set of Hodge–Tate weights of $M(\psi)$, that is, $C\otimes_{\mathbb{Q}_p} M(\psi)\simeq \bigoplus_{\sigma\in\Gamma_M} C(h_\sigma)$, where C is the completion of $\overline{\mathbb{Q}}_p$.

For integers d, h and a p-adic field M, we define a constant $C(d, M, h)$ by

$$C(d,M,h) := v_p(d/d_M)+h+\frac{1}{2}d_M\Big(d_M+v_p(e_M)-\frac{1}{e_M}+v_p(2)(d_M-1)\Big). \tag{2-1}$$

Theorem 2.3. *Let $\psi_1,\dots,\psi_n : G_K\to M^\times$ be crystalline characters and $h\ge 0$ an integer. Assume that M is a Galois extension of $\mathbb{Q}_p$ and K contains M. Suppose that, for each i, we have*

$$\psi_i = \prod_{\sigma\in\Gamma_M}\sigma^{-1}\circ\chi_M^{h_{i,\sigma}}$$

on the inertia I_K; thus $\{h_{i,\sigma}\mid\sigma\in\Gamma_M\}$ is the set of Hodge–Tate weights of $M(\psi_i)$. We assume the following conditions:

(i) *$\{h_{i,\sigma}\mid\sigma\in\Gamma_M\}$ contains at least two different integers for each i. (In particular, we have $M\neq\mathbb{Q}_p$.)*

(ii) *$\mathrm{Min}\{v_p(h_{i,\sigma}-h_{i,\tau})\mid\sigma,\tau\in\Gamma_M\}\le h$ for each i.*

Then:

(1) *There exists an element $\hat{\omega}\in\ker \mathrm{Nr}_{M/\mathbb{Q}_p}$ such that for every $1\le i\le n$,*

$$1+v_p(2)\le v_p(\psi_{i,K}(\hat{\omega})^{-1}-1)\le \delta_{(i)}+C(d_K,M,h). \tag{2-2}$$

Here,

$$\delta_{(i)} := \begin{cases} 0 & \text{if } i=1,2,\\ 2i-5 & \text{if } i\ge 3.\end{cases}$$

(2) *Let $\hat{\omega}$ be as in (1). For any $x\in K^\times$, there exists an integer $0\le s(x)\le n$ such that for every $1\le i\le n$,*

$$v_p(\psi_{i,K}(x\hat{\omega}^{p^{s(x)}})^{-1}-1)\le n+\delta_{(i)}+C(d_K,M,h). \tag{2-3}$$

Proof. Take an element $x\in\mathcal{O}_M$ such that $\mathcal{O}_M=\mathbb{Z}_p[x]$. We set $p':=p$ or $p':=4$ if $p\neq 2$ or $p=2$, respectively, and put $x'=p'x$. Set $m_{r,\sigma}^{\tau}:=d_{K/M}(h_{r,\tau\sigma}-h_{r,\sigma})$

for $1 \le r \le n$ and $\sigma, \tau \in \Gamma_M$. We also set

$$y^\tau_{r,\ell} := \sum_{\sigma \in \Gamma_M} m^\tau_{r,\sigma} (\sigma^{-1} x')^{\ell-1}$$

for $1 \le \ell \le d_M$. (Note that $y^\tau_{r,1} = 0$.) Set

$$\omega_\ell := \exp((x')^{\ell-1}) \quad \text{and} \quad \omega^\tau_\ell := \frac{\tau \omega_\ell}{\omega_\ell}$$

for any $1 \le \ell \le d_M$ and $\tau \in \Gamma_M$. Here, exp denotes the p-adic exponential map (see [Neukirch 1999, Chapter II, Proposition 5.5]). By construction, $\omega^\tau_\ell \in \ker \mathrm{Nr}_{M/\mathbb{Q}_p}$.

Lemma 2.4. $$\exp(y^\tau_{r,\ell}) = \psi_{r,K}(\omega^\tau_\ell)^{-1}.$$

Proof. We see

$$\psi_{r,K}(\omega_\ell)^{-1} = \prod_{\sigma \in \Gamma_M} \sigma^{-1} \circ \mathrm{Nr}_{K/M}(\omega_\ell)^{h_{r,\sigma}} = \biggl(\prod_{\sigma \in \Gamma_M} \sigma^{-1} \omega_\ell^{h_{r,\sigma}} \biggr)^{d_{K/M}}.$$

We also have $\psi_{r,K}(\tau\omega_\ell)^{-1} = \bigl(\prod_{\sigma \in \Gamma_M} \sigma^{-1} \tau \omega_\ell^{h_{r,\sigma}}\bigr)^{d_{K/M}} = \bigl(\prod_{\sigma \in \Gamma_M} \sigma^{-1} \omega_\ell^{h_{r,\tau\sigma}}\bigr)^{d_{K/M}}$. Thus we have

$$\psi_{r,K}(\omega^\tau_\ell)^{-1} = \biggl(\prod_{\sigma \in \Gamma_M} \sigma^{-1} \omega_\ell^{h_{r,\tau\sigma} - h_{r,\sigma}} \biggr)^{d_{K/M}} = \prod_{\sigma \in \Gamma_M} \sigma^{-1} \omega_\ell^{m^\tau_{r,\sigma}}.$$

On the other hand, we have

$$\begin{aligned} \exp(y^\tau_{r,\ell}) = \exp\biggl(\sum_{\sigma \in \Gamma_M} m^\tau_{r,\sigma} (\sigma^{-1} x')^{\ell-1} \biggr) &= \prod_{\sigma \in \Gamma_M} \exp((\sigma^{-1} x')^{\ell-1})^{m^\tau_{r,\sigma}} \\ &= \prod_{\sigma \in \Gamma_M} \sigma^{-1} \omega_\ell^{m^\tau_{r,\sigma}}. \end{aligned}$$ □

We furthermore need the following evaluation.

Lemma 2.5. *For each $1 \le r \le n$, there exist $\tau_r \in \Gamma_M$ and an integer $2 \le \ell_r \le d_M$ such that*

$$v_p(y^{\tau_r}_{r,\ell_r}) \le C(d_K, M, h).$$

Proof. Fix r. By assumption (i), there exist $\tau_1, \tau_2 \in \Gamma_M$ such that $h_{r,\tau_1} \ne h_{r,\tau_2}$. Choose τ_1 and τ_2 so that $v_p(h_{r,\tau_1} - h_{r,\tau_2}) = \mathrm{Min}\{v_p(h_{r,\sigma} - h_{r,\tau}) \mid \sigma, \tau \in \Gamma_M\}$, and set $\tau := \tau_2 \tau_1^{-1} \in \Gamma_M$. We write $\Gamma_M = \{\tau_1, \tau_2, \ldots, \tau_{d_M}\}$. Note that $m^\tau_{r,\tau_1} = d_{K/M}(h_{r,\tau_2} - h_{r,\tau_1})$ is not zero. We denote by $X \in M_d(\mathcal{O}_M)$ the matrix whose (i, j)-component is $(\tau_i^{-1} x')^{j-1}$. Then we have

$$(y^\tau_{r,1} \cdots y^\tau_{r,d_M}) = (m^\tau_{r,\tau_1} \cdots m^\tau_{r,\tau_{d_M}}) X \tag{2-4}$$

and

$$\det X = \prod_{1\le i<j\le d_M} (\tau_j^{-1}x' - \tau_i^{-1}x') = (p')^{\frac{1}{2}d_M(d_M-1)} \prod_{1\le i<j\le d_M} (\tau_j^{-1}x - \tau_i^{-1}x).$$

We also have

$$\begin{aligned} v_p\Big(\prod_{1\le i<j\le d_M} (\tau_j^{-1}x - \tau_i^{-1}x)\Big) &= \sum_{1\le i<j\le d_M} v_p(\tau_j^{-1}x - \tau_i^{-1}x) \\ &= \frac{1}{2}\sum_{1\le i,j\le d_M, i\ne j} v_p(\tau_j^{-1}x - \tau_i^{-1}x) \\ &= \frac{1}{2}d_M v_p(\mathcal{D}_{M/\mathbb{Q}_p}) \le \frac{1}{2}d_M\Big(1+v_p(e_M)-\frac{1}{e_M}\Big). \end{aligned}$$

(see [Serre 1979, Chapter 3, Section 6, Proposition 13]), where $\mathcal{D}_{M/\mathbb{Q}_p}$ is the different ideal of $M/\mathbb{Q}_p$. We find

$$v_p(\det X) \le \frac{1}{2}d_M\Big(d_M + v_p(e_M) - \frac{1}{e_M} + v_p(2)(d_M-1)\Big). \tag{2-5}$$

By (2-4), we have $m^\tau_{r,\tau_1} \det X = \sum_{\ell=1}^{d_M} y^\tau_{r,\ell}x_\ell$ for some $x_\ell \in \mathcal{O}_M$, which gives the fact that there exists an integer $\ell_r = \ell$ with the property that $v_p(y^\tau_{r,\ell}) \le v_p(m^\tau_{r,\tau_1} \det X)$. By (2-5), we have

$$v_p(y^\tau_{r,\ell}) \le v_p(d_{K/M}) + v_p(h_{r,\tau_1} - h_{r,\tau_2}) + v_p(\det X) \le C(d_K, M, h),$$

as desired. We remark that ℓ is not equal to 1 since $y^\tau_{r,1}$ is zero. □

Now we return to the proof of Theorem 2.3. Take τ_r and ℓ_r as in Lemma 2.5 with the additional condition that

$$v_p(y^{\tau_r}_{r,\ell_r}) = \mathrm{Min}\{v_p(y^\tau_{r,\ell}) \mid \tau \in \Gamma_M, 2 \le \ell \le d_M\}. \tag{2-6}$$

Here we consider an element $\hat{\omega} \in \ker \mathrm{Nr}_{M/\mathbb{Q}_p}$ which is of the form $\hat{\omega} = \prod_{r=1}^n (\omega^{\tau_r}_{\ell_r})^{s_r}$, where s_r is defined inductively by the following:

$$(s_1, s_2) = \begin{cases} (0,1) & \text{if } v_p(y^{\tau_1}_{1,\ell_1}) = v_p(y^{\tau_2}_{1,\ell_2}), \\ (1,0) & \text{if } v_p(y^{\tau_1}_{1,\ell_1}) \ne v_p(y^{\tau_2}_{1,\ell_2}) \text{ and } v_p(y^{\tau_1}_{2,\ell_1}) = v_p(y^{\tau_2}_{2,\ell_2}), \\ (1,1) & \text{if } v_p(y^{\tau_1}_{1,\ell_1}) \ne v_p(y^{\tau_2}_{1,\ell_2}) \text{ and } v_p(y^{\tau_1}_{2,\ell_1}) \ne v_p(y^{\tau_2}_{2,\ell_2}). \end{cases}$$

$$s_3 = \begin{cases} p & \text{if } v_p(s_1 y^{\tau_1}_{3,\ell_1} + s_2 y^{\tau_2}_{3,\ell_2}) \ne v_p(p y^{\tau_3}_{3,\ell_3}), \\ p^2 & \text{if } v_p(s_1 y^{\tau_1}_{3,\ell_1} + s_2 y^{\tau_2}_{3,\ell_2}) = v_p(p y^{\tau_3}_{3,\ell_3}). \end{cases}$$

For $r \ge 4$,

$$s_r = \begin{cases} p s_{r-1} & \text{if } v_p\big(\sum_{j=1}^{r-1} s_j y^{\tau_j}_{r,\ell_j}\big) \ne v_p(p s_{r-1} y^{\tau_r}_{r,\ell_r}), \\ p^2 s_{r-1} & \text{if } v_p\big(\sum_{j=1}^{r-1} s_j y^{\tau_j}_{r,\ell_j}\big) = v_p(p s_{r-1} y^{\tau_r}_{r,\ell_r}). \end{cases}$$

We claim that we have

$$1+v_p(2) \le v_p\left(\sum_{r=1}^{n} s_r y_{i,\ell_r}^{\tau_r}\right) \le \delta_{(i)} + C(d_K, M, h)$$

for any i, where $\delta_{(i)}$ is as in the statement (1). The inequality $1+v_p(2) \le v_p\big(\sum_{r=1}^{n} s_r y_{i,\ell_r}^{\tau_r}\big)$ is clear since we always have $1+v_p(2) \le v_p(y_{i,\ell}^{\tau})$ by definition of $y_{i,\ell}^{\tau}$. We show $v_p\big(\sum_{r=1}^{n} s_r y_{i,\ell_r}^{\tau_r}\big) \le \delta_{(i)} + C(d_K, M, h)$ by induction on i.

- Suppose either $i=1$ or $i=2$. By (2-6) and the inequality $0 < v_p(s_r)$ for $r \ge 3$, it is not difficult to check $v_p\big(\sum_{r=1}^{n} s_r y_{i,\ell_r}^{\tau_r}\big) = v_p(y_{i,\ell_i}^{\tau_i})$. Furthermore, we have $v_p(y_{i,\ell_i}^{\tau_i}) \le C(d_K, M, h) = \delta_{(i)} + C(d_K, M, h)$ by Lemma 2.5.
- Suppose $i \ge 3$. By definition of s_i we have $v_p\big(\sum_{r=1}^{i-1} s_r y_{i,\ell_r}^{\tau_r}\big) \ne v_p(s_i y_{i,\ell_i}^{\tau_i})$. We also have $v_p\big(\sum_{r=i}^{n} s_r y_{i,\ell_r}^{\tau_r}\big) = v_p(s_i y_{i,\ell_i}^{\tau_i})$ since $v_p(s_i y_{i,\ell_i}^{\tau_i}) < v_p(s_r y_{i,\ell_r}^{\tau_r})$ for $i < r$. Hence, it follows from Lemma 2.5 that we have

$$\begin{aligned} v_p\left(\sum_{r=1}^{n} s_r y_{i,\ell_r}^{\tau_r}\right) &= \mathrm{Min}\left\{ v_p\left(\sum_{r=1}^{i-1} s_r y_{i,\ell_r}^{\tau_r}\right), v_p(s_i y_{i,\ell_i}^{\tau_i})\right\} \\ &\le v_p(p s_{i-1} y_{i,\ell_i}^{\tau_i}) \le 1 + v_p(s_{i-1}) + C(d_K, M, h) \end{aligned}$$

  if $i \ge 4$. Since we have $v_p(s_{i-1}) \le 2(i-3)$ if $i \ge 4$, the claim for $i \ge 4$ follows. The claim for $i=3$ follows by a similar manner; we have $v_p(\sum_{r=1}^{n} s_r y_{3,\ell_r}^{\tau_r}) \le v_p(p y_{3,\ell_3}^{\tau_3}) \le 1 + C(d_K, M, h) = \delta_{(3)} + C(d_K, M, h)$.

By construction of $\hat{\omega}$ and Lemma 2.4, we see

$$\psi_{i,K}(\hat{\omega})^{-1} = \prod_{r=1}^{n} \psi_{i,K}(\omega_{\ell_r}^{\tau_r})^{-s_r} = \prod_{r=1}^{n} \exp(s_r y_{i,\ell_r}^{\tau_r}) = \exp\left(\sum_{r=1}^{n} s_r y_{i,\ell_r}^{\tau_r}\right).$$

Thus we find $v_p(\psi_{i,K}(\hat{\omega})^{-1} - 1) = v_p(\sum_{r=1}^{n} s_r y_{i,\ell_r}^{\tau_r})$. Therefore, the claim above gives Theorem 2.3(1).

To show Theorem 2.3(2), we set $m_i := \psi_{i,K}(x)^{-1} - 1$ and $\theta_i^{(s)} = \psi_{i,K}(\hat{\omega}^{p^s})^{-1} - 1$ for any $s \ge 0$. It follows from the condition $v_p(\psi_{i,K}(\hat{\omega})^{-1} - 1) \ge 1 + v_p(2)$ that the equality $v_p(\theta_i^{(s)}) = s + v_p(\theta_i^{(0)})$ holds. For each $1 \le i \le n$, there exists at most only one integer $s \ge 0$ so that $v_p(m_i) = v_p(\theta_i^{(s)})$ since $\{v_p(\theta_i^{(s)})\}_s$ is strictly increasing. Hence, there exists an integer $0 \le s(x) \le n$ with the property that $v_p(m_i) \ne v_p(\theta_i^{(s(x))})$ for every $1 \le i \le n$ (by the pigeonhole principle). With this choice of $s(x)$, we obtain $v_p(\psi_{i,K}(x\hat{\omega}^{p^{s(x)}})^{-1} - 1) = v_p(m_i + \theta_i^{(s(x))} + m_i \theta_i^{(s(x))}) \le v_p(\theta_i^{(n)}) = n + v_p(\theta_i^{(0)})$. This finishes the proof of (2). □

3. Proof of main theorems

The main purpose of this section is to show Theorems 1.1 and 1.2. For Theorem 1.1, we show a slightly refined statement as follows.

Theorem 3.1. *Let $g > 0$ be a positive integer. Let k be a p-adic field with residue cardinality q_k and π a uniformizer of k. Put $p' = p$ or $p' = 4$ if $p \neq 2$ or $p = 2$, respectively. Let $\mu \geq 1$ be the smallest integer*[2] *so that*

$$(q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi))^{\mu} \equiv 1 \mod p'.$$

Assume the following conditions:[3]

(i) $v_p\big((q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi))^{\mu} - 1\big) > g \cdot (2g)! \cdot \Phi(g)H(g) \cdot \mu \cdot d_{Kk/k} f_k$, *and*

(ii) d_k *is prime to* $(2g)!$.

Then, for any g-dimensional abelian variety A over a p-adic field K with complex multiplication, we have

$$A(Kk_{\pi})[p^{\infty}] \subset A[p^C],$$

where

$$C := 2g^2 \cdot (2g)! \cdot \Phi(g)H(g) \cdot \mu \cdot d_{Kk} + 12g^2 - 18g + 10.$$

In particular,

$$\#A(Kk_{\pi})[p^{\infty}] \leq p^{2gC}.$$

Our proofs of Theorems 3.1 and 1.2 proceed by similar methods. As in the previous section, we fix an algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$ and suppose that K is a subfield of $\overline{\mathbb{Q}}_p$. In this section, we often use the following technical constants:

$$L_g(m) := [\log_p(1 + p^{\frac{1}{2}m})^{2g}],$$
$$C(m, M, h) := v_p\Big(\frac{m}{d_M}\Big) + h + \frac{1}{2}d_M\Big(d_M + v_p(e_M) - \frac{1}{e_M} + v_p(2)(d_M - 1)\Big).$$

Here, $m \geq 1$ and $h \geq 0$ are integers and M is a p-adic field.

Remark 3.2. (1) $mg \leq L_g(m) < g(m + 1 + v_p(2))$ for any prime p and $m \geq 1$, and $L_g(m) < g(m + 1)$ if $(p, m) \neq (2, 1), (2, 2)$.

(2) Moreover,[4]

$$L_g(m) = mg \quad \text{for } m \geq 8g.$$

This can be checked as follows: It suffices to show $(1 + p^{\frac{1}{2}m})^{2g} < p^{mg+1}$ for $m \geq 8g$. This inequality is equivalent to $(1 + p^{-\frac{1}{2}m})^{2g} < p$. Thus it is enough to show $(1 + 2^{-\frac{1}{2}m_0})^{2g} < 2$ where $m_0 := 8g$. By the inequalities $2g < 2^{2g}$ and $\binom{2g}{r} < 2^{2g}$ for $0 \leq r \leq 2g$, we find, as desired,

$$(1+2^{-\frac{1}{2}m_0})^{2g} = 1+\sum_{r=1}^{2g}\binom{2g}{r}\Big(\frac{1}{2}\Big)^{\frac{1}{2}rm_0} < 1+2g\cdot 2^{2g}\Big(\frac{1}{2}\Big)^{\frac{1}{2}m_0} < 1+\Big(\frac{1}{2}\Big)^{\frac{1}{2}m_0-4g} = 2.$$

[2] If $q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi)$ is a root of unity, the constant μ here coincides with the μ in Theorem 1.1.

[3] Condition (i) depends on the choice of K. However, the author hopes that this condition can be replaced with one that does not depend on K, as in Theorem 1.1(i).

[4] The value $8g$ here is "rough" but it is enough for our proofs.

Special cases. We consider Theorem 3.1 under some additional hypothesis. In this section, we show:

Proposition 3.3. *Let the situation be as in Theorem 3.1 except assuming not* (i) *but*

(i)′ $v_p\big((q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi))^{\mu}-1\big) > L_g((2g)!\cdot\mu\cdot d_{Kk/k}f_k)$.

Moreover, *we assume that A has good reduction over K and all the endomorphisms of A are defined over K. Put*

$$C_g(K,k) = v_p(d_{Kk}) + \tfrac{1}{2}(2g)!\big((2g)! + v_p((2g)!) + v_p(2)((2g)!-1)\big),$$
$$\Delta_g(K,k) = \mathrm{Max}\big\{C_g(K,k),\, L_g((2g)!\cdot\mu\cdot d_{Kk/k}f_k)\big\}.$$

Then

$$A(Kk_\pi)[p^\infty] \subset A[p^C],$$

where

$$C := 2g\Delta_g(K,k) + 12g^2 - 18g + 10.$$

Proof. Put $T = T_p(A)$ and $V = V_p(A)$ for brevity. Let $\rho : G_K \to \mathrm{GL}_{\mathbb{Z}_p}(T)$ be the continuous homomorphism obtained by the G_K-action on T. Fix an isomorphism $\iota : T \xrightarrow{\sim} \mathbb{Z}_p^{\oplus 2g}$ of $\mathbb{Z}_p$-modules. We have an isomorphism $\hat{\iota} : \mathrm{GL}_{\mathbb{Z}_p}(T) \simeq \mathrm{GL}_{2g}(\mathbb{Z}_p)$ relative to ι. We abuse notation by writing ρ for the composite map $G_K \to \mathrm{GL}_{\mathbb{Z}_p}(T) \simeq \mathrm{GL}_{2g}(\mathbb{Z}_p)$ of ρ and $\hat{\iota}$. Now let $P \in T$ and denote by $\overline{P}$ the image of P in T/p^nT. By definition, we have $\iota(\sigma P) = \rho(\sigma)\iota(P)$ for $\sigma \in G_K$. Suppose that $\overline{P} \in (T/p^nT)^{G_{Kk_\pi}}$. This implies $\sigma P - P \in p^nT$ for any $\sigma \in G_{Kk_\pi}$. This is equivalent to saying that $(\rho(\sigma) - E)\iota(P) \in p^n\mathbb{Z}_p^{\oplus 2g}$, and this in particular implies $\det(\rho(\sigma) - E)\iota(P) \in p^n\mathbb{Z}_p^{\oplus 2g}$ for any $\sigma \in G_{G_{Kk_\pi}}$. Thus $\det(\rho(\sigma) - E)P \in p^nT$ for any $\sigma \in G_{Kk_\pi}$. Put

$$c = \mathrm{Min}\{v_p(\det(\rho(\sigma) - E)) \mid \sigma \in G_{Kk_\pi}\}.$$

Then we see $P \in p^{n-c}T$ (if c is finite and $n > c$) and this shows $(T/p^nT)^{G_{Kk_\pi}} \subset p^{n-c}T/p^nT$. This implies an inequality

$$A(Kk_\pi)[p^\infty] \subset A[p^c] \tag{3-1}$$

if c is finite.

On the other hand, we recall that A has complex multiplication and all the endomorphisms of A are defined over K. Thus there exists an injective ring homomorphism from a number field F of degree $2g$ into $\mathbb{Q} \otimes_{\mathbb{Z}} \mathrm{End}_K(A)$. By [Serre and Tate 1968, Theorem 5(i)], we know that V is a free $F \otimes_{\mathbb{Q}} \mathbb{Q}_p$-module of rank one and the G_K-action on V commutes with $F \otimes_{\mathbb{Q}} \mathbb{Q}_p$-action. Let $\prod_{i=1}^n F_i$ denote the decomposition of $F \otimes_{\mathbb{Q}} \mathbb{Q}_p$ into a finite product of p-adic fields. This induces a decomposition $V \simeq \bigoplus_{i=1}^n V_i$ of $\mathbb{Q}_p[G_K]$-modules. Each V_i is equipped with a structure of one-dimensional F_i-modules and the G_K-action on V_i commutes with the F_i-action. Let $\rho_i : G_K \to \mathrm{GL}_{\mathbb{Q}_p}(V_i)$ be the homomorphism obtained by the G_K-action on V_i. Since ρ_i is abelian, it follows that $(V_i \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p)^{\mathrm{ss}} \simeq \bigoplus_{j=1}^{d_{F_i}} \overline{\mathbb{Q}}_p(\psi_{i,j})$ for some continuous characters $\psi_{i,j} : G_K \to \overline{\mathbb{Q}}_p^{\times}$. Here, the superscript "ss" stands for the semisimplification. As is well known, $\psi_{i,j}$ satisfies the following properties (since the G_K-action on V_i is given by a character $G_K \to F_i^{\times}$):

(a) $\psi_{i,1}, \dots, \psi_{i,d_{F_i}}$ are $\mathbb{Q}_p$-conjugate with each other, that is, $\psi_{i,k} = \tau_{k\ell} \circ \psi_{i,\ell}$ for some $\tau_{k\ell} \in G_{\mathbb{Q}_p}$.

(b) $\psi_{i,1}, \dots, \psi_{i,d_{F_i}}$ have values in a p-adic field M_i (in the fixed algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$) which is $\mathbb{Q}_p$-isomorphic to the Galois closure of $F_i/\mathbb{Q}_p$ (in an algebraic closure of F_i). We remark that d_{M_i} divides $d_{F_i}!$.

In particular,

$$v_p(\det \rho_i(\sigma) - E) = d_{F_i} v_p(\psi_i(\sigma) - 1),$$

where $\psi_i := \psi_{i,1}$. Let M be the composite field of $M_1, \dots, M_n$, and we regard $\psi_1, \dots, \psi_n$ as characters of G_K with values in $M^{\times}$, that is, $\psi_i : G_K \to M^{\times}$. The field M is a Galois extension of $\mathbb{Q}_p$ in $\overline{\mathbb{Q}}_p$ and d_M divides $d_{F_1}!\, d_{F_2}! \cdots d_{F_n}!$. Since $\sum_{i=1}^{n} d_{F_i} = 2g$, we find

$$d_M \mid (2g)!. \tag{3-2}$$

(Here, we recall that the product of n natural numbers is divisible by $n!$ for any natural number n.) In particular, we have $M \cap k = \mathbb{Q}_p$ since d_k is prime to $(2g)!$, and then we obtain

$$\ker \mathrm{Nr}_{M/\mathbb{Q}_p} \subset \ker \mathrm{Nr}_{Mk/k} \subset \ker \mathrm{Nr}_{K_M k/k}.$$

Here, K_M is the composite KM of K and M. It follows from Proposition 2.1 that

$$\begin{aligned} c &\le \mathrm{Min}\{v_p(\det(\rho(\sigma) - E)) \mid \sigma \in G_{K_M k_\pi}\} \\ &= \mathrm{Min}\left\{ \sum_{i=1}^{n} d_{F_i} v_p(\psi_i(\sigma) - 1) \;\middle|\; \sigma \in G_{K_M k_\pi} \right\} \\ &\le \mathrm{Min}\left\{ \sum_{i=1}^{n} d_{F_i} v_p(\psi_{i,K_M k}(\pi\omega)^{-1} - 1) \;\middle|\; \omega \in \ker \mathrm{Nr}_{K_M k/k} \right\} \\ &\le \mathrm{Min}\left\{ \sum_{i=1}^{n} d_{F_i} v_p(\psi_{i,K_M k}(\pi\omega)^{-1} - 1) \;\middle|\; \omega \in \ker \mathrm{Nr}_{M/\mathbb{Q}_p} \right\} \\ &\le \mathrm{Min}\left\{ \sum_{i=1}^{n} d_{F_i} v_p(\psi^{\mu}_{i,K_M k}(\pi\omega)^{-1} - 1) \;\middle|\; \omega \in \ker \mathrm{Nr}_{M/\mathbb{Q}_p} \right\}. \end{aligned} \tag{3-3}$$

Here, μ is the integer appeared in the statement of Theorem 3.1. Note that ψ_i is a crystalline character since A has good reduction over K (see [Fontaine 1982, Section 6]; see also [Coleman and Iovita 1999, Theorem 1]). By rearranging the numbering of subscripts, we may suppose the following situation for some $0 \le r \le n$.

(I) For $1 \le i \le r$, the set of the Hodge–Tate weights of $M(\psi_i)$ is $\{0, 1\}$.

(II) For $r < i \le n$, the set of the Hodge–Tate weights of $M(\psi_i)$ is either $\{1\}$ or $\{0\}$.

Lemma 3.4. *For $r < i \le n$ and any $\omega \in \ker \mathrm{Nr}_{M/\mathbb{Q}_p}$, we have*

$$v_p(\psi^{\mu}_{i,K_M k}(\pi\omega)^{-1} - 1) \le L_g((2g)! \cdot d_{Kk/k} f_k \cdot \mu).$$

Proof. In this proof we set $L := K_M k$. We know that the morphism $\psi_{i,\mathrm{alg}} : \underline{L}^\times \to \underline{M}^\times$ corresponding to $\psi_i|_{G_L}$ is trivial or $\mathrm{Nr}^{-1}_{L/\mathbb{Q}_p}$ on $\mathbb{Q}_p$-points. This in particular gives $\psi_{i,L}(\omega) = 1$. Since $\pi_L^{e_{L/k}}\pi^{-1}$ is a p-adic unit for any uniformizer π_L of L, we find

$$\psi_{i,L}(\pi\omega)^{-1} = \psi_{i,L}(\pi)^{-1} = \psi_{i,L}(\pi_L^{-e_{L/k}} \cdot \pi_L^{e_{L/k}}\pi^{-1}) = \alpha_i^{-e_{L/k}} \cdot \psi_{i,\mathrm{alg}}(\pi)^{-1},$$

where $\alpha_i := \psi_{i,L}(\pi_L)\psi_{i,\mathrm{alg}}(\pi_L)^{-1}$. Denote by L' the unramified extension of L of degree $\mu e_{L/k}$.

(I) Suppose that the set of the Hodge–Tate weights of $M(\psi_i)$ is $\{0\}$. In this case, $\psi_{i,\mathrm{alg}}$ is trivial and thus we have $\psi^{\mu}_{i,L}(\pi\omega)^{-1} = \alpha_i^{-\mu e_{L/k}}$. It follows from Lemma 9 of [Ozeki 2024] that $\psi^{\mu}_{i,L}(\pi\omega)^{-1}$ is a unit root of the characteristic polynomial $f(T)$ of the geometric Frobenius endomorphism of $\bar{A}_{/\mathbb{F}_{L'}}$. Since $f(1) = \#\bar{A}(\mathbb{F}_{q_{L'}})$, we see $v_p(\psi^{\mu}_{i,L}(\pi\omega)^{-1} - 1) \le v_p(\#\bar{A}(\mathbb{F}_{q_{L'}})) \le [\log_p \#\bar{A}(\mathbb{F}_{q_{L'}})]$. It follows from the Weil bound that $v_p(\psi^{\mu}_{i,L}(\pi\omega)^{-1} - 1) \le L_g(f_{L'})$. Since we have $f_{L'} = \mu e_{L/k} f_L = d_{L/Kk} \cdot \mu \cdot d_{Kk/k} f_k \le (2g)! \cdot \mu \cdot d_{Kk/k} f_k$. we obtain the desired inequality.

(II) Suppose that the set of the Hodge–Tate weights of $M(\psi_i)$ is $\{1\}$. In this case $\psi_{i,\mathrm{alg}}$ is $\mathrm{Nr}^{-1}_{L/\mathbb{Q}_p}$ on $\mathbb{Q}_p$-points. If we set $\beta := q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi)$, we find

$$\begin{aligned}\psi^{\mu}_{i,L}(\pi\omega)^{-1} - 1 &= (\alpha_i^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi)^{f_{L/k}})^{\mu e_{L/k}} - 1 \\ &= ((\alpha_i^{-1}q_L)^{\mu e_{L/k}} - 1)\beta^{\mu d_{L/k}} + (\beta^{\mu d_{L/k}} - 1).\end{aligned}$$

It again follows from Lemma 9 of [Ozeki 2024] that $(\alpha_i^{-1}q_L)^{\mu e_{L/k}}$ is a unit root of the characteristic polynomial $f^\vee(T)$ of the geometric Frobenius endomorphism of $\overline{A^\vee}_{/\mathbb{F}_{L'}}$. Since $f^\vee(1) = \#\overline{A^\vee}(\mathbb{F}_{q_{L'}})$, the same argument as in (I) shows that $v_p((\alpha_i^{-1}q_L)^{\mu e_{L/k}} - 1) \le L_g(f_{L'}) \le L_g((2g)! \cdot \mu \cdot d_{Kk/k} f_k)$. In particular, we have $v_p(\beta^{\mu d_{L/k}} - 1) > v_p((\alpha_i^{-1}q_L)^{\mu e_{L/k}} - 1)$ by the assumption (i)′. Since β is a p-adic unit, we obtain $v_p(\psi^{\mu}_{i,L}(\pi\omega)^{-1} - 1) = v_p((\alpha_i^{-1}q_L)^{\mu e_{L/k}} - 1) \le L_g((2g)! \cdot \mu \cdot d_{Kk/k} f_k)$, as desired. □

By (3-3) and the lemma, in the case where $r = 0$, we have

$$c \le \sum_{i=1}^{n} d_{F_i} L_g((2g)! \cdot \mu \cdot d_{Kk/k} f_k) = 2g L_g((2g)! \cdot \mu \cdot d_{Kk/k} f_k). \tag{3-4}$$

In the rest of the proof, we assume $r > 0$. By (3-3) and the lemma again, we have

$$c \le \mathrm{Min}\left\{ \sum_{i=1}^{r} d_{F_i} v_p(\psi^{\mu}_{i,K_M k}(\pi\omega)^{-1} - 1) \,\middle|\, \omega \in \ker \mathrm{Nr}_{M/\mathbb{Q}_p} \right\} + L_g((2g)! \cdot \mu \cdot d_{Kk/k} f_k) \sum_{i=r+1}^{n} d_{F_i}.$$

Here we remark that $v_p(\mu) = 0$ and the Hodge–Tate weights of ψ^{μ}_i for each $1 \le i \le r$ consist of 0 and μ. Hence, applying Theorem 2.3 to the set of characters $\psi^{\mu}_1, \dots, \psi^{\mu}_r : G_{K_M k} \to M^{\times}$, an element $x = \pi$ and $h = 0$, there exists an element $\hat{\omega} \in \ker \mathrm{Nr}_{M/\mathbb{Q}_p}$ and an integer $0 \le s = s(\pi) \le r$ as in the theorem. Then

$$\begin{aligned} c &\le \sum_{i=1}^{r} d_{F_i} v_p(\psi^{\mu}_{i,K_M k}(\pi\hat{\omega}^{p^s})^{-1} - 1) + L_g((2g)! \cdot \mu \cdot d_{Kk/k} f_k) \sum_{i=r+1}^{n} d_{F_i} \\ &\le \sum_{i=1}^{r} d_{F_i}(r + \delta_{(i)} + C(d_{K_M k}, M, 0)) + L_g((2g)! \cdot \mu \cdot d_{Kk/k} f_k) \sum_{i=r+1}^{n} d_{F_i} \\ &\le 2g\Delta_0 + \sum_{i=1}^{r} d_{F_i}(r + \delta_{(i)}), \end{aligned}$$

where $\Delta_0 := \mathrm{Max}\{C(d_{K_M k}, M, 0), L_g((2g)! \cdot \mu \cdot d_{Kk/k} f_k)\}$. Since d_M divides $(2g)!$, we also have

$$C(d_{K_M k}, M, 0) < v_p(d_{Kk}) + \tfrac{1}{2}(2g)!\big((2g)! + v_p((2g)!) + v_p(2)((2g)! - 1)\big).$$

Thus, for the constant $\Delta_g(K, k)$ defined in the statement of the proposition, we obtain $\Delta_0 \le \Delta_g(K, k)$ and $c \le 2g\Delta_g(K, k) + \sum_{i=1}^{r} d_{F_i}(r + \delta_{(i)})$.

- If $r \le 2$, we have $\sum_{i=1}^{r} d_{F_i}(r + \delta_{(i)}) = \sum_{i=1}^{r} d_{F_i} r \le r \cdot 2g \le 4g$.
- If $r > 2$, we have $\sum_{i=1}^{r} d_{F_i}(r+\delta_{(i)}) = r\sum_{i=1}^{r} d_{F_i} + \sum_{i=3}^{r} d_{F_i}\delta_{(i)} \le n\sum_{i=1}^{n} d_{F_i} + \sum_{i=3}^{n} d_{F_i}(2n-5) \le n\cdot 2g + (2n-5)(\sum_{i=1}^{n} d_{F_i} - 2) \le 2g\cdot 2g + (4g-5)\cdot(2g-2) = 12g^2 - 18g + 10$.

Therefore, for any $r > 0$, we find

$$c \le 2g\Delta_g(K, k) + 12g^2 - 18g + 10.$$

Note that this inequality holds also for the case $r = 0$ by (3-4). Now the proposition follows from (3-1). □

General cases. We show Theorems 3.1 and 1.2. For this, we need the following observations given by Serre and Tate [1968] and Silverberg [1992].

Theorem 3.5. *Let A be a g-dimensional abelian variety over K.*

(1) *Put* $m = 3$ *or* $m = 4$ *if* $p \neq 3$ *or* $p = 3$, *respectively. Then* A *has semistable reduction over* $K(A[m])$ *and all the endomorphisms of* A *are defined over* $K(A[m])$.

(2) *Let* L *be the intersection of the fields* $K(A[N])$ *for all integers* $N > 2$. *Then all the endomorphisms of* A *are defined over* L *and* $[L : K]$ *divides* $H(g)$.

(3) *Assume* A *has potential good reduction. Let* $\rho_{A,\ell} : G_K \to \mathrm{GL}_{\mathbb{Z}_p}(T_\ell(A))$ *be the continuous homomorphism defined by the* G_K*-action on the Tate module* $T_\ell(A)$ *for any prime* ℓ.

(i) *For any prime* ℓ *not equal to* p, *let* H_ℓ *be the kernel of the restriction of* $\rho_{A,\ell}$ *to* I_K. *Then* H_ℓ *is an open subgroup of* I_K, *which is independent of the choice of* ℓ. *Moreover, if we set* $c := [I_K : H_\ell]$, *then there exists a finite totally ramified extension* L/K *of degree* c *such that* A *has good reduction over* L.

(ii) *If* A *has complex multiplication and all the endomorphisms of* A *are defined over* K, *then the constant* c *above satisfies* $c \leq \Phi(g)$.

(4) *Assume* A *has complex multiplication. Then there exists a finite extension* L/K *of degree at most* $\Phi(g)H(g)$ *such that* A *has good reduction over* L *and all the endomorphisms of* A *are defined over* L.

Proof. Item (1) follows from [Silverberg 1992, Theorem 4.1] and Raynaud's criterion of semistable reduction [SGA 7_{I} 1972, Proposition 4.7]. Item (2) is [Silverberg 1992, Theorem 4.1], and (4) is an immediate consequence of (2) and (3) since A must have potential good reduction under the condition that A has complex multiplication. The assertions in (3) are consequences of results given in Sections 2 and 4 of [Serre and Tate 1968] but some of them are not directly mentioned in loc. cit. Thus we give a proof here, just in case. The first statement related to H_ℓ in (3)(i) is [Serre and Tate 1968, Section 2, Theorem 2, p. 496]. The group H is a closed normal subgroup of G_K, which is also open in I_K. Let Γ be the closure of the subgroup of G_K generated by any choice of a lift of the q_K-th Frobenius element in $G_{\mathbb{F}_{q_K}}$. The projection $G_K \to G_{\mathbb{F}_{q_K}}$ gives an isomorphism of Γ onto $G_{\mathbb{F}_{q_K}}$; in particular, G_K is the semidirect product of Γ and I_K. Let K_Γ/K be the field extension (of infinite degree) corresponding to $\Gamma \subset G_K$, and let M/K^{ur} be the finite extension corresponding to $H := H_\ell \subset I_K$. Note that A has good reduction over M. Now we set $L := K_\Gamma \cap M$. Then L/K is totally ramified since so is K_Γ/K. Furthermore, it is immediate to check $H\Gamma \cap I_K = H$; this shows $LK^{\mathrm{ur}} = M$. Hence we obtain that A has good reduction over L and $[L : K] = [M : K^{\mathrm{ur}}] = c$. This shows (3)(i). Next we show (3)(ii). By assumptions on A, there exists a number field F of degree $2g$ which is a subalgebra of $\mathbb{Q} \otimes_{\mathbb{Z}} \mathrm{End}_K(A)$. It follows from [Serre and Tate 1968, Theorem 5(i)] that $V_\ell(A)$ has a structure of free $(F \otimes_{\mathbb{Q}} \mathbb{Q}_\ell)$-module of rank one and the G_K-action on $V_\ell(A)$ commutes with $F \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$. Thus we may consider $\rho_{A,\ell}$ as a character $G_K \to (F \otimes_{\mathbb{Q}} \mathbb{Q}_\ell)^\times$. Moreover, the image of this character restricted to I_K has values in the group $\mu(F)$ of roots of unity contained in F by [Serre and Tate 1968, Section 4, Theorem 6, p. 503]. Thus we obtain the fact that c divides the order m of $\mu(F)$. On the other hand, since μ_m is a subset of F, we have $\varphi(m) \mid 2g$. Therefore, we obtain $c \le m \le \Phi(g)$, as desired. □

Now we are ready to show our main theorems. First we show Theorem 3.1.

Proof of Theorem 3.1. Let A be as in the theorem. Since A has complex multiplication, it follows from Theorem 3.5(4) that there exists a finite extension L/K such that $d_{L/K} \le \Phi(g)H(g)$, A has good reduction over L, and all the endomorphisms of A are defined over L. In addition, we have

$$\begin{aligned} v_p\big((q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi))^{\mu}-1\big) &> g\cdot(2g)!\cdot\Phi(g)H(g)\cdot\mu\cdot d_{Kk/k}f_k \\ &= L_g\big((2g)!\cdot\Phi(g)H(g)\cdot\mu\cdot d_{Kk/k}f_k\big) \\ &\ge L_g((2g)!\cdot\mu\cdot d_{Lk/k}f_k) \end{aligned}$$

by assumption (i) and Remark 3.2(2). So we can apply Proposition 3.3 to A/L; we have

$$A(Lk_{\pi})[p^{\infty}] \subset A[p^{C'}],$$

where $C' = 2g\Delta_g(L,k) + 12g^2 - 18g + 10$. Here,

$$\begin{aligned} C_g(L,k) &= v_p(d_{Lk}) + \tfrac{1}{2}(2g)!\big((2g)! + v_p((2g)!) + v_p(2)((2g)!-1)\big), \\ \Delta_g(L,k) &= \mathrm{Max}\big\{C_g(L,k),\ L_g((2g)!\cdot\mu\cdot d_{Lk/k}f_k)\big\}. \end{aligned}$$

Note that we have $v_p(d_{Lk}) < d_{Lk} \le \Phi(g)H(g)\cdot d_{Kk}$ and $L_g((2g)!\cdot\mu\cdot d_{Lk/k}f_k) \le g\cdot(2g)!\cdot\Phi(g)H(g)\cdot\mu\cdot d_{Kk}$. Therefore, it suffices to show

$$\begin{aligned} \Phi(g)H(g)\cdot d_{Kk} + \tfrac{1}{2}(2g)!\big((2g)! + v_p((2g)!) + v_p(2)((2g)!-1)\big) \\ < g\cdot(2g)!\cdot\Phi(g)H(g)\cdot\mu\cdot d_{Kk} \end{aligned}$$

for the proof but this is clear. □

Remark 3.6. In the proof of Theorem 3.1, we referred to the field extension L/K of Theorem 3.5(4) and the upper bound $\Phi(g)H(g)$ of $[L:K]$. By Theorem 3.5(1), we may refer to the field $K(A[m])$ instead of the above L. Since we have a natural embedding from $\mathrm{Gal}(K(A[m])/K)$ into $\mathrm{GL}(A[m]) \simeq \mathrm{GL}_{2g}(\mathbb{Z}/m\mathbb{Z})$, we obtain a bound for the extension degree of $K(A[m])/K$; we have $[K(A[m]):K] \le G(g)$, where

$$G(n) := \begin{cases} \#\mathrm{GL}_{2n}(\mathbb{Z}/3\mathbb{Z}) = \prod_{i=0}^{2n-1}(3^{2n}-3^i) & \text{if } p \ne 3, \\ \#\mathrm{GL}_{2n}(\mathbb{Z}/4\mathbb{Z}) = 2^{4n^2}\prod_{i=0}^{2n-1}(2^{2n}-2^i) & \text{if } p = 3 \end{cases}$$

for $n > 0$. Note that we have $G(n) < m^{4n^2}$. It is not difficult to check the inequalities $\Phi(1)H(1) > G(1)$ and $\Phi(g)H(g) < G(g)$ for $g > 1$ (see Section 5). Hence, only in the case $g = 1$ of elliptic curves, we can obtain smaller bound than that given in Theorem 3.1 by replacing $\Phi(g)H(g)$ with $G(1)$.

Applying Theorem 1.1 with $k = \mathbb{Q}_p$ and $\pi = p$, we immediately obtain the following.

Corollary 3.7. *Let A be a g-dimensional abelian variety over a p-adic field K with complex multiplication. Then we have*

$$A(K(\mu_{p^\infty}))[p^\infty] \subset A[p^C],$$

where

$$C := 2g^2 \cdot (2g)! \cdot \Phi(g)H(g) \cdot d_K + 12g^2 - 18g + 10.$$

In particular,

$$\#A(K(\mu_{p^\infty}))[p^\infty] \le p^{2gC}.$$

Next we show Theorem 1.2.

Proof of Theorem 1.2. We follow essentially the same argument as for Theorem 3.1. Put $\hat{K} = K(\sqrt[p^\infty]{K})$.

Step 1: First we consider the case where A has good reduction over K and all the endomorphisms of A are defined over K. Put $\nu = v_p(d_K) + 1 + v_p(2)$ and

$$C_g(K) = v_p(d_K) + \nu + \tfrac{1}{2}(2g)!\big((2g)! + v_p((2g)!) + v_p(2)((2g)! - 1)\big),$$
$$\Delta_g(K) = \mathrm{Max}\big\{C_g(K),\, L_g((2g)! \cdot p^\nu \cdot d_K)\big\}.$$

Following the proof of Proposition 3.3, we show

$$A(\hat{K})[p^\infty] \subset A[p^{C'}], \tag{3-5}$$

where $C' := 2g\Delta_g(K) + 12g^2 - 18g + 10$. Let $\rho : G_K \to \mathrm{GL}_{\mathbb{Z}_p}(T_p(A)) \simeq \mathrm{GL}_{2g}(\mathbb{Z}_p)$, $M/\mathbb{Q}_p$ and $\psi_1, \ldots, \psi_n : G_K \to M^\times$ be as in the proof of Proposition 3.3. If we denote by $\hat{K}_{\mathrm{ab}}$ the maximal abelian extension of K contained in $\hat{K}$, all the points of $A(\hat{K})[p^\infty]$ are in fact defined over $\hat{K}_{\mathrm{ab}}$ since ρ is abelian. Thus, setting $c := \mathrm{Min}\{v_p(\det(\rho(\sigma) - E)) \mid \sigma \in G_{\hat{K}_{\mathrm{ab}}}\}$, we find

$$A(\hat{K})[p^\infty] = A(\hat{K}_{\mathrm{ab}})[p^\infty] \subset A[p^c] \tag{3-6}$$

if c is finite (see arguments just above (3-1)). On the other hand, we set $G := \mathrm{Gal}(\hat{K}/K)$ and $H := \mathrm{Gal}(\hat{K}/K(\mu_{p^\infty}))$. Let $\chi_p : G_K \to \mathbb{Z}_p^\times$ be the p-adic cyclotomic character. Since we have $\sigma\tau\sigma^{-1} = \tau^{\chi_p(\sigma)}$ for any $\sigma \in G$ and $\tau \in H$, we see $(G, G) \supset (G, H) \supset H^{\chi_p(\sigma)-1}$. Hence we have a natural surjection

$$H/H^{\chi_p(\sigma)-1} \twoheadrightarrow H/\overline{(G, G)} = \mathrm{Gal}(\hat{K}_{\mathrm{ab}}/K(\mu_{p^\infty})) \quad \text{for any } \sigma \in G. \tag{3-7}$$

Lemma 3.8. $\chi_p(\sigma_0) - 1 = p^\nu$ *for some* $\sigma_0 \in G$.

Proof. We set

$$K' := \begin{cases} K(\mu_p) & \text{if } p \neq 2, \\ K(\mu_4) & \text{if } p = 2. \end{cases}$$

If we denote by p^ℓ the order of the set of p-power roots of unity in K', we see $K' \cap \mathbb{Q}_p(\mu_{p^\infty}) = \mathbb{Q}_p(\mu_{p^\ell})$ and thus $\chi_p(G_{K'}) = 1 + p^\ell\mathbb{Z}_p$. Furthermore, since $[\mathbb{Q}_p(\mu_{p^\ell}) : \mathbb{Q}_p]$ divides $[K' : K][K : \mathbb{Q}_p]$, we see $p^{\ell-1-v_p(2)} \mid d_K$. Hence we obtain $\chi_p(G_{K'}) \supset 1 + p^\nu\mathbb{Z}_p$ and the lemma follows. □

By the lemma above and (3-7), we see that $\mathrm{Gal}(\hat{K}_{\mathrm{ab}}/K(\mu_{p^\infty}))$ is of exponent p^ν, that is, $\sigma \in G_{K(\mu_{p^\infty})}$ implies $\sigma^{p^\nu} \in G_{\hat{K}_{\mathrm{ab}}}$. This shows $c \leq \mathrm{Min}\{v_p(\det(\rho(\sigma)^{p^\nu} - E)) \mid \sigma \in G_{K(\mu_{p^\infty})}\}$. Mimicking the arguments for inequalities (3-3), we find

$$c \leq \mathrm{Min}\left\{ \sum_{i=1}^{n} d_{F_i} v_p(\psi_{i,K_M}^{p^\nu}(\pi\omega)^{-1} - 1) \;\middle|\; \omega \in \ker \mathrm{Nr}_{M/\mathbb{Q}_p} \right\}.$$

Now the inequality (3-6) follows by completely the same method as the proof of Proposition 3.3 (with replacing the pair (k, μ) there with $(\mathbb{Q}_p, p^\nu)$).

Step 2: Next we consider the general case. Since A has complex multiplication, it follows from Theorem 3.5(4) that there exists a finite extension L/K such that $d_{L/K} \leq \Phi(g)H(g)$, A has good reduction over L and all the endomorphisms of A are defined over L. Thus we can apply the result of Step 1 to A/L; we have

$$A(\hat{K})[p^\infty] \subset A(\hat{L})[p^\infty] \subset A[p^{C''}],$$

where $C'' := 2g\Delta_g(L) + 12g^2 - 18g + 10$. We find

$$\begin{aligned} L_g\big((2g)! \cdot p^{v_p(d_L)+1+v_p(2)} \cdot d_L\big) &= L_g\big((2g)! \cdot p^{1+v_p(2)} \cdot p^{v_p(d_{L/K})} d_{L/K} \cdot p^{v_p(d_K)} d_K\big) \\ &\leq L_g\big((2g)! \cdot p^{1+v_p(2)} \cdot (d_{L/K})^2 \cdot p^{v_p(d_K)} d_K\big) \\ &\leq g \cdot (2g)! \cdot p^{1+v_p(2)} \cdot (\Phi(g)H(g))^2 \cdot p^{v_p(d_K)} d_K. \end{aligned}$$

(For the last equality, see Remark 3.2(2).) Now Theorem 1.2 immediately follows by $\Delta_g(L) \leq g \cdot (2g)! \cdot p^{1+v_p(2)} \cdot (\Phi(g)H(g))^2 \cdot p^{v_p(d_K)} d_K$. □

One of the keys for our arguments above is a theory of locally algebraic representations. Thus our method essentially works also for abelian varieties A with the property that the G_K-action on the semisimplification of $V_p(A) \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p$ is abelian. For example, this is the case where A has good ordinary reduction.

Proposition 3.9. *Let $g > 0$ be a positive integer. Let K and k be p-adic fields. Let π be a uniformizer of k. Assume that $q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi)$ is a root of unity; we denote by $0 < \mu < p$ the minimum integer such that $(q_k^{-1}\mathrm{Nr}_{k/\mathbb{Q}_p}(\pi))^\mu = 1$. Then, for* *any g-dimensional abelian variety A over K with good ordinary reduction, we have*

$$A(Kk_\pi)[p^\infty] \subset A[p^{2gL_g(\mu d_{Kk/k} f_k)}].$$

In particular,

$$\#A(Kk_\pi)[p^\infty] \le p^{4g^2 L_g(\mu d_{Kk/k} f_k)} < p^{4g^3(\mu d_{Kk/k} f_k + 1 + v_p(2))}.$$

Proof. Put $V = V_p(A)$, $T = T_p(A)$ and $c = \mathrm{Min}\{v_p(\det(\rho(\sigma) - E)) \mid \sigma \in G_{Kk_\pi}\}$. By the same argument as the beginning of the proof of Proposition 3.3, we obtain

$$A(Kk_\pi)[p^\infty] \subset A[p^c] \tag{3-8}$$

if c is finite. Since A has good ordinary reduction, we have an exact sequence $0 \to V_1 \to V \to V_2 \to 0$ of $\mathbb{Q}_p[G_K]$-modules with the following properties:

(i) $V_1 \simeq W \otimes_{\mathbb{Q}_p} \mathbb{Q}_p(1)$ for some unramified representation W of G_K, and

(ii) V_2 is unramified.

Hence, taking a p-adic field M large enough, we have $(V \otimes_{\mathbb{Q}_p} M)^{\mathrm{ss}} \simeq \bigoplus_{i=1}^{2g} M(\psi_i)$ for some continuous crystalline characters $\psi_i : G_K \to M^\times$. Furthermore, for every i, the set of the Hodge–Tate weights of $M(\psi_i)$ is either $\{1\}$ or $\{0\}$. By Proposition 2.1, we have $c \le \sum_{i=1}^{2g} v_p(\psi_{i,Kk}^\mu(\pi)^{-1} - 1)$. Let K' be the unramified extension of Kk of degree $\mu e_{Kk/k}$. By a similar method of the proof of Lemma 3.4, we find that $\psi_{i,Kk}^\mu(\pi)^{-1}$ is a unit root of the characteristic polynomial $f(T)$ of the geometric Frobenius endomorphism of $\bar{A}_{/\mathbb{F}_{K'}}$; otherwise, $\psi_{i,Kk}^\mu(\pi)^{-1}$ is a unit root of the characteristic polynomial $f^\vee(T)$ of the geometric Frobenius endomorphism of $\overline{A^\vee}_{/\mathbb{F}_{K'}}$. We know $f(1) = \#\bar{A}(\mathbb{F}_{q_{K'}})$ and $f^\vee(1) = \#\overline{A^\vee}(\mathbb{F}_{q_{K'}})$, and their p-adic valuations are bounded by $L_g(f_{K'})$ by the Weil bound. Since we have $f_{K'} = f_{K'/Kk} f_{Kk} = \mu d_{Kk/k} f_k$, we obtain $c \le \sum_{i=1}^{2g} v_p(\psi_{i,Kk}^\mu(\pi)^{-1} - 1) \le 2gL_g(\mu d_{Kk/k} f_k)$. Now the result follows from (3-8). □

4. Abelian varieties over number fields

In this section, we suppose that K is a number field. The goal of this section is to give a proof of Theorem 1.3. The theorem is an immediate consequence of the following proposition.

Proposition 4.1. *Let g, K, d and h be as in Theorem 1.3.*

(1) *Let A be a g-dimensional abelian variety over K with semistable reduction everywhere. Let p_0 be the smallest prime number such that A has good reduction at some finite place of K above p_0. Then $A(K(\mu_\infty))[p]$ is zero if $p > (1 + \sqrt{p_0}^{dh})^{2g}$, p is unramified in K, and A has good reduction at some finite place of K above p.*

(2) *Let A be a g-dimensional abelian variety over K with complex multiplication which has good reduction everywhere. Then, for any prime p, we have*

$$A(K(\mu_\infty))[p^\infty] \subset A[p^C],$$

where $C := 2g^2 \cdot (2g)! \cdot \Phi(g)H(g) \cdot dh + 12g^2 - 18g + 10.$

Proof. Let A be a g-dimensional abelian variety over K with semistable reduction everywhere. Let K' be the maximal extension of K contained in $K(\mu_\infty)$ which is unramified at all finite places of K. Note that K' is a finite abelian extension of K. In particular, it follows from class field theory that $[K' : K]$ is a divisor of the narrow class number h of K. If we denote by L_p the maximal extension of K contained in $K(\mu_\infty)$ which is unramified at all places except for places dividing p and the infinite places, then it is shown in [Katz and Lang 1981, Appendix, Lemma] that $L_p = K'(\mu_{p^\infty})$.

(1) Here we mainly follow Ribet's arguments in [Katz and Lang 1981]. We suppose that p is prime to $2p_0$ and also suppose that p is unramified in K. Assume that $A(K(\mu_\infty))[p] \neq O$. We claim that there exists a g-dimensional abelian variety A' over K' which is K'-isogenous to A such that $A'(K')[p] \neq O$. We denote by G and H the absolute Galois groups of K' and $K(\mu_\infty)$, respectively. The assumption $A(K(\mu_\infty))[p] \neq O$ is equivalent to the assumption $A[p]^H \neq O$. Let W be a simple G-submodule of $A[p]^H$. Ribet showed in the proof of Theorem 2 of [Katz and Lang 1981] that, since A has semistable reduction everywhere over K', W is one-dimensional over $\mathbb{F}_p$ and the action of G on W factors through $\mathrm{Gal}(K'(\mu_p)/K')$. Since p is unramified at K', we find that the G-action on W is given by $\bar{\chi}_p^n$ for some $0 \le n \le p-1$, where $\bar{\chi}_p$ is the mod p cyclotomic character. Moreover, since A has good reduction at some finite place of K' above p ($\neq 2$) and p is unramified in K', it follows from the classification of Tate and Oort [1970, pp. 15–16] that n is equal to 0 or 1. Thus W is isomorphic to $\mathbb{F}_p$ or $\mathbb{F}_p(1)$. If we are in the former case, we have $A'(K')[p] \neq O$ for $A' := A$. Suppose that we are in the latter case. Then there exists a surjection $A^\vee[p] \to \mathbb{F}_p$ of G-modules. If we denote by C the kernel of this surjection, then the G-action on $A^\vee[p]$ preserves C. This implies that $A' := A^\vee/C$ is an abelian variety defined over K' and we find that there exists a trivial G-submodule of $A'[p]$ of order p. Thus we have $A'(K')[p] \neq O$. This finishes the proof of the claim.

Now we take a prime $\mathfrak{p}_0'$ of K' above p_0 such that A has good reduction at $\mathfrak{p}_0'$. Since A' above is K'-isogenous to A, we know that A' has good reduction at $\mathfrak{p}_0'$ by [Serre and Tate 1968, Section 1, Corollary 2]. If we denote by $K'_{\mathfrak{p}_0'}$ the completion of K' at $\mathfrak{p}_0'$ and also denote by $\mathbb{F}_{\mathfrak{p}_0'}$ the residue field of $K'_{\mathfrak{p}_0'}$, then reduction modulo $\mathfrak{p}_0'$ gives an injective homomorphism

$$A'(K')[p] \subset A'(K'_{\mathfrak{p}_0'})[p] \hookrightarrow \bar{A}'(\mathbb{F}_{\mathfrak{p}_0'}).$$

We recall that $A'(K')[p] \neq O$. Since the order of $\mathbb{F}_{\mathfrak{p}'_0}$ is bounded by p_0^{dh}, it follows from the Weil bound that $p < (1+\sqrt{p_0}^{dh})^{2g}$. This finishes the proof.

(2) Let A be an abelian variety as in the statement. Since A has good reduction everywhere over K, it follows from the Néron–Ogg–Shafarevich criterion that the G_K-action on $A[p^\infty]$ is unramified outside p. This gives the fact that the G_K-action on $A(K(\mu_{p^\infty}))[p^\infty]$ factors through $\mathrm{Gal}(L_p/K) = \mathrm{Gal}(K'(\mu_{p^\infty})/K)$. Thus

$$A(K(\mu_\infty))[p^\infty] = A(K'(\mu_{p^\infty}))[p^\infty].$$

Since we have $[K':\mathbb{Q}] \leq dh$, the result follows from Corollary 3.7. □

5. Bounds on $\Phi(n)$ and $H(n)$

We recall the definitions of $\Phi(n)$ and $H(n)$:

$$\Phi(n) := \mathrm{Max}\{m \in \mathbb{Z}_{>0} \mid \varphi(m) \text{ divides } 2n\},$$
$$H(n) := \gcd\{\#\mathrm{GSp}_{2n}(\mathbb{Z}/N\mathbb{Z}) \mid N \geq 3\}.$$

Here, φ is the Euler's totient function. The values of $\Phi(n)$, $H(n)$ (and $G(n)$ for $p \neq 3$; see Remark 3.6) for small n are given in Tables 1–3. In this section, we study some upper bounds of Φ and H.

The function H. For the function H, we refer to results of [Silverberg 1992, Sections 3 and 4]. The exact formula for $H(n)$ is as follows:

$$H(n) = \frac{1}{2^{n-1}} \prod_q q^{r(q)},$$

where the product is over primes $q \leq 2n+1$,

$$r(2) = [n] + \sum_{j=0}^{\infty} \left[\frac{2n}{2^j}\right] \quad \text{and} \quad r(q) = \sum_{j=0}^{\infty} \left[\frac{2n}{q^j(q-1)}\right] \quad \text{if } q \text{ is odd.}$$

Moreover, we have:

Theorem 5.1 [Silverberg 1992, Corollary 3.3]. *We have*

$$H(n) < 2(9n)^{2n}$$

for any $n > 0$.

The function Φ. Next we consider the function Φ. At first, we remark that $\Phi(n)$ must be even since $\varphi(x) = \varphi(2x)$ if x is odd. Furthermore, $\Phi(n)$ is not a power of 2. (In fact, we have $\varphi(2^r) = \varphi(2^{r-1}\cdot 3)$ if $r \geq 2$.) Thus it holds that

$$\Phi(n) = \mathrm{Max}\big\{m \in \mathbb{Z}_{>0} \mid \varphi(m) \text{ divides } 2n, \text{ and } m = 2^r x, \text{ where } r \geq 1 \text{ and } x \geq 3 \text{ is odd}\big\}. \tag{5-1}$$

We show some elementary formulas.

Proposition 5.2. (1) $\Phi(1)=6$ *and* $6 \le \Phi(n) < 6n\sqrt[3]{n}$ *for* $n>1$.

(2) *Put* $t=v_2(n)+2$ *and let* $p_1=2<p_2<\cdots<p_t$ *be the first t prime numbers. Then*

$$\Phi(n) \le 2n \prod_{i=1}^{t} \frac{p_i}{p_i-1}.$$

In particular, $\Phi(n) \le 6n$ *if n is odd.*

(3) *If* $n>3$ *is an odd prime*, *we have*[5]

$$\Phi(n)=\begin{cases} 6 & \text{if } 2n+1 \text{ is not prime},\\ 4n+2 & \text{if } 2n+1 \text{ is prime}. \end{cases}$$

Proof. To check $\Phi(1)=6$ is an easy exercise. Since $\varphi(6)=2 \mid 2n$, we have $\Phi(n)\ge 6$ for any n. Suppose that $n>1$. We take an even integer $m>0$ of the form $2^r x$, where $r\ge 1$ and $x\ge 3$ is odd, such that $\varphi(m)\mid 2n$. Let $m=2^r\prod_{i=1}^{s} q_i^{e_i}$ be the prime factorization of m with $r, s, e_1, \dots, e_s \ge 1$. Since $\varphi(m)=2^{r-1}\prod_{i=1}^{s} q_i^{e_i-1}(q_i-1)$ and $\varphi(m)\mid 2n$, we have $v_2(2n)\ge r-1+s$ and thus

$$r+s \le v_2(n)+2. \tag{5-2}$$

Then we find

$$2n \ge \varphi(m) = m\Big(1-\frac{1}{2}\Big)\prod_{i=1}^{s}\Big(1-\frac{1}{q_i}\Big) \ge m\prod_{i=1}^{s+1}\Big(1-\frac{1}{p_i}\Big) \ge m\prod_{i=1}^{t}\Big(1-\frac{1}{p_i}\Big).$$

This shows (2). Furthermore, we have

$$\begin{aligned}\Phi(n) &\le 2n\prod_{i=1}^{t}\frac{p_i}{p_i-1} = 6n\prod_{i=3}^{t}\frac{p_i}{p_i-1} \le 6n\Big(\frac{5}{5-1}\Big)^{v_2(n)}\\ &\le 6n\cdot\Big(\frac{5}{4}\Big)^{\log_2(n)} < 6n\cdot 2^{\frac{1}{3}\log_2(n)}.\end{aligned}$$

Thus we obtain (1). Let us show (3). From now on we assume that $n>3$ is an odd prime. Assume that $m\ne 6$. Since n is odd, it follows from (5-2) that the prime factorization of m is of the form $m=2q^e$ for some odd prime q. Then $\frac{1}{2}\varphi(m)=q^{e-1}\frac{1}{2}(q-1)$ divides n. Since $n>3$ is a prime and $m\ne 6$, we find $e=1$ and $\frac{1}{2}(q-1)=n$. This implies $2n+1$ must be prime and $m=4n+2$. Now the result follows. ∎

[5] A prime number p is called a *Sophie Germain prime* if $2p+1$ is also prime. It is not known whether there exist infinitely many Sophie Germain primes or not. On the other hand, there exist infinitely many primes which are not Sophie Germain primes. In fact, every prime number p with $p\equiv 1 \bmod 3$ is not a Sophie Germain prime.

n	$\Phi(n)$	n	$\Phi(n)$	n	$\Phi(n)$	n	$\Phi(n)$
1	$2^1 \cdot 3^1$	31	$2^1 \cdot 3^1$	61	$2^1 \cdot 3^1$	91	$2^1 \cdot 3^1$
2	$2^2 \cdot 3^1$	32	$2^4 \cdot 3^1 \cdot 5^1$	62	$2^2 \cdot 3^1$	92	$2^2 \cdot 3^1 \cdot 47^1$
3	$2^1 \cdot 3^2$	33	$2^1 \cdot 67^1$	63	$2^1 \cdot 127^1$	93	$2^1 \cdot 3^2$
4	$2^1 \cdot 3^1 \cdot 5^1$	34	$2^2 \cdot 3^1$	64	$2^1 \cdot 3^1 \cdot 5^1 \cdot 17^1$	94	$2^2 \cdot 3^1$
5	$2^1 \cdot 11^1$	35	$2^1 \cdot 71^1$	65	$2^1 \cdot 131^1$	95	$2^1 \cdot 191^1$
6	$2^1 \cdot 3^1 \cdot 7^1$	36	$2^1 \cdot 3^3 \cdot 5^1$	66	$2^1 \cdot 3^2 \cdot 23^1$	96	$2^3 \cdot 3^1 \cdot 5^1 \cdot 7^1$
7	$2^1 \cdot 3^1$	37	$2^1 \cdot 3^1$	67	$2^1 \cdot 3^1$	97	$2^1 \cdot 3^1$
8	$2^2 \cdot 3^1 \cdot 5^1$	38	$2^2 \cdot 3^1$	68	$2^1 \cdot 137^1$	98	$2^1 \cdot 197^1$
9	$2^1 \cdot 3^3$	39	$2^1 \cdot 79^1$	69	$2^1 \cdot 139^1$	99	$2^1 \cdot 199^1$
10	$2^1 \cdot 3^1 \cdot 11^1$	40	$2^1 \cdot 3^1 \cdot 5^1 \cdot 11^1$	70	$2^1 \cdot 3^1 \cdot 71^1$	100	$2^1 \cdot 3^1 \cdot 5^3$
11	$2^1 \cdot 23^1$	41	$2^1 \cdot 83^1$	71	$2^1 \cdot 3^1$	101	$2^1 \cdot 3^1$
12	$2^1 \cdot 3^2 \cdot 5^1$	42	$2^1 \cdot 3^1 \cdot 7^2$	72	$2^1 \cdot 3^2 \cdot 5^1 \cdot 7^1$	102	$2^1 \cdot 3^1 \cdot 103^1$
13	$2^1 \cdot 3^1$	43	$2^1 \cdot 3^1$	73	$2^1 \cdot 3^1$	103	$2^1 \cdot 3^1$
14	$2^1 \cdot 29^1$	44	$2^2 \cdot 3^1 \cdot 23^1$	74	$2^1 \cdot 149^1$	104	$2^2 \cdot 3^1 \cdot 53^1$
15	$2^1 \cdot 31^1$	45	$2^1 \cdot 31^1$	75	$2^1 \cdot 151^1$	105	$2^1 \cdot 211^1$
16	$2^3 \cdot 3^1 \cdot 5^1$	46	$2^1 \cdot 3^1 \cdot 47^1$	76	$2^1 \cdot 3^1 \cdot 5^1$	106	$2^1 \cdot 3^1 \cdot 107^1$
17	$2^1 \cdot 3^1$	47	$2^1 \cdot 3^1$	77	$2^1 \cdot 23^1$	107	$2^1 \cdot 3^1$
18	$2^1 \cdot 3^2 \cdot 7^1$	48	$2^2 \cdot 3^1 \cdot 5^1 \cdot 7^1$	78	$2^1 \cdot 3^1 \cdot 79^1$	108	$2^1 \cdot 3^4 \cdot 5^1$
19	$2^1 \cdot 3^1$	49	$2^1 \cdot 3^1$	79	$2^1 \cdot 3^1$	109	$2^1 \cdot 3^1$
20	$2^1 \cdot 3^1 \cdot 5^2$	50	$2^1 \cdot 5^3$	80	$2^2 \cdot 3^1 \cdot 5^1 \cdot 11^1$	110	$2^1 \cdot 3^1 \cdot 11^2$
21	$2^1 \cdot 7^2$	51	$2^1 \cdot 103^1$	81	$2^1 \cdot 3^5$	111	$2^1 \cdot 223^1$
22	$2^1 \cdot 3^1 \cdot 23^1$	52	$2^1 \cdot 3^1 \cdot 53^1$	82	$2^1 \cdot 3^1 \cdot 83^1$	112	$2^1 \cdot 3^1 \cdot 5^1 \cdot 29^1$
23	$2^1 \cdot 47^1$	53	$2^1 \cdot 107^1$	83	$2^1 \cdot 167^1$	113	$2^1 \cdot 227^1$
24	$2^1 \cdot 3^1 \cdot 5^1 \cdot 7^1$	54	$2^1 \cdot 3^3 \cdot 7^1$	84	$2^2 \cdot 3^1 \cdot 7^2$	114	$2^1 \cdot 229^1$
25	$2^1 \cdot 11^1$	55	$2^1 \cdot 11^2$	85	$2^1 \cdot 11^1$	115	$2^1 \cdot 47^1$
26	$2^1 \cdot 53^1$	56	$2^2 \cdot 3^1 \cdot 29^1$	86	$2^1 \cdot 173^1$	116	$2^2 \cdot 3^1 \cdot 59^1$
27	$2^1 \cdot 3^4$	57	$2^1 \cdot 3^2$	87	$2^1 \cdot 59^1$	117	$2^1 \cdot 79^1$
28	$2^1 \cdot 3^1 \cdot 29^1$	58	$2^1 \cdot 3^1 \cdot 59^1$	88	$2^1 \cdot 3^1 \cdot 5^1 \cdot 23^1$	118	$2^2 \cdot 3^1$
29	$2^1 \cdot 59^1$	59	$2^1 \cdot 3^1$	89	$2^1 \cdot 179^1$	119	$2^1 \cdot 239^1$
30	$2^1 \cdot 3^2 \cdot 11^1$	60	$2^1 \cdot 3^1 \cdot 7^1 \cdot 11^1$	90	$2^1 \cdot 3^3 \cdot 11^1$	120	$2^1 \cdot 3^1 \cdot 5^2 \cdot 7^1$

Table 1. $\Phi(n)$.

Let us consider an upper bound of Φ by using an "analytic" lower bound function of φ given by Rosser and Schoenfeld. If we denote by γ Euler's constant,[6] it is shown in [Rosser and Schoenfeld 1962, Theorem 15] that[7]

$$\varphi(m) > \frac{m}{e^{\gamma}\log\log m + \frac{3}{\log\log m}} \tag{5-3}$$

for $m \geq 3$. We set

$$\Psi(n) := \mathrm{Max}\{m \in \mathbb{Z}_{>0} \mid \varphi(m) \leq 2n\}.$$

We clearly have $\Phi(n) \leq \Psi(n)$ for all $n > 0$.

Proposition 5.3. *For any real number $C > 2e^{\gamma}$, we have*

$$\Psi(n) < Cn\log\log n$$

for any n large enough.

Proof. The result should be well known as a consequence of Mertens' theorem:

$$\liminf_{n} \frac{\varphi(n)\log\log n}{n} = e^{-\gamma}.$$

Using (5-3) instead of Mertens' theorem, we can obtain a slightly refined statement (see Remark 5.4). So, for later use, we write down a proof with using (5-3). Put $f(x) = C\log\log x$. Take any integer $N > 0$ satisfying the following: for all $x > N$,

(i) $f(x) > \frac{1}{x}e^{e^2}$, and

(ii) $f(x) > 2e^{\gamma}\bigl(\log\log(xf(x)) + 1\bigr)$.

(The assumption $C > 2e^{\gamma}$ asserts the existence of such N.) Take any integer $n > N$. It suffices to show n satisfies the desired inequality. Assume there exists an integer m such that both $\varphi(m) \leq 2n$ and $m \geq nf(n)$ hold. Since $e^{\gamma} > 3/(\log\log x)$ for $x > e^{e^2}$ and $m\ (\geq nf(n)) > e^{e^2}$, we find

$$\frac{1}{e^{\gamma}}\cdot\frac{m}{\log\log m + 1} < \frac{m}{e^{\gamma}\log\log m + \frac{3}{\log\log m}} < \varphi(m) \leq 2n$$

by (5-3). Also, $nf(n)/(\log\log(nf(n)) + 1) \leq m/(\log\log m + 1)$ since the function $x/(\log\log x + 1)$ is strictly increasing for $x > e$ and $m \geq nf(n)\ (> e^{e^2}) > e$. Hence

$$\frac{1}{e^{\gamma}}\cdot\frac{nf(n)}{\log\log(nf(n)) + 1} < 2n,$$

which gives $f(n) < 2e^{\gamma}(\log\log(nf(n)) + 1)$. This contradicts condition (ii). We conclude that if $\varphi(m) \leq 2n$, then $m < nf(n)$. This implies that $\Psi(n) < nf(n) = Cn\log\log n$. ☐

[6] $\gamma = \int_1^{\infty}\bigl(\frac{1}{[x]} - \frac{1}{x}\bigr)\,dx = 0.57721\ldots$. Note also $e^{\gamma} = 1.78107\ldots$.

[7] More precisely, that theorem states $\varphi(m) > m/(e^{\gamma}\log\log m + 5/(2\log\log m))$ for $m \geq 3$ except when m is the product of the first nine primes, $m = 223092870 = 2\cdot 3\cdot 5\cdot\cdot 7\cdot 11\cdot 13\cdot 17\cdot 19\cdot 23$.

n	$H(n)$
1	$2^4 \cdot 3^1$
2	$2^8 \cdot 3^2 \cdot 5^1$
3	$2^{11} \cdot 3^4 \cdot 5^1 \cdot 7^1$
4	$2^{16} \cdot 3^5 \cdot 5^2 \cdot 7^1$
5	$2^{19} \cdot 3^6 \cdot 5^2 \cdot 7^1 \cdot 11^1$
6	$2^{23} \cdot 3^8 \cdot 5^3 \cdot 7^2 \cdot 11^1 \cdot 13^1$
7	$2^{26} \cdot 3^9 \cdot 5^3 \cdot 7^2 \cdot 11^1 \cdot 13^1$
8	$2^{32} \cdot 3^{10} \cdot 5^4 \cdot 7^2 \cdot 11^1 \cdot 13^1 \cdot 17^1$
9	$2^{35} \cdot 3^{13} \cdot 5^4 \cdot 7^3 \cdot 11^1 \cdot 13^1 \cdot 17^1 \cdot 19^1$
10	$2^{39} \cdot 3^{14} \cdot 5^6 \cdot 7^3 \cdot 11^2 \cdot 13^1 \cdot 17^1 \cdot 19^1$
11	$2^{42} \cdot 3^{15} \cdot 5^6 \cdot 7^3 \cdot 11^2 \cdot 13^1 \cdot 17^1 \cdot 19^1 \cdot 23^1$
12	$2^{47} \cdot 3^{17} \cdot 5^7 \cdot 7^4 \cdot 11^2 \cdot 13^2 \cdot 17^1 \cdot 19^1 \cdot 23^1$
13	$2^{50} \cdot 3^{18} \cdot 5^7 \cdot 7^4 \cdot 11^2 \cdot 13^2 \cdot 17^1 \cdot 19^1 \cdot 23^1$
14	$2^{54} \cdot 3^{19} \cdot 5^8 \cdot 7^4 \cdot 11^2 \cdot 13^2 \cdot 17^1 \cdot 19^1 \cdot 23^1 \cdot 29^1$
15	$2^{57} \cdot 3^{21} \cdot 5^8 \cdot 7^5 \cdot 11^3 \cdot 13^2 \cdot 17^1 \cdot 19^1 \cdot 23^1 \cdot 29^1 \cdot 31^1$
16	$2^{64} \cdot 3^{22} \cdot 5^9 \cdot 7^5 \cdot 11^3 \cdot 13^2 \cdot 17^2 \cdot 19^1 \cdot 23^1 \cdot 29^1 \cdot 31^1$
17	$2^{67} \cdot 3^{23} \cdot 5^9 \cdot 7^5 \cdot 11^3 \cdot 13^2 \cdot 17^2 \cdot 19^1 \cdot 23^1 \cdot 29^1 \cdot 31^1$
18	$2^{71} \cdot 3^{26} \cdot 5^{10} \cdot 7^6 \cdot 11^3 \cdot 13^3 \cdot 17^2 \cdot 19^2 \cdot 23^1 \cdot 29^1 \cdot 31^1 \cdot 37^1$
19	$2^{74} \cdot 3^{27} \cdot 5^{10} \cdot 7^6 \cdot 11^3 \cdot 13^3 \cdot 17^2 \cdot 19^2 \cdot 23^1 \cdot 29^1 \cdot 31^1 \cdot 37^1$
20	$2^{79} \cdot 3^{28} \cdot 5^{12} \cdot 7^6 \cdot 11^4 \cdot 13^3 \cdot 17^2 \cdot 19^2 \cdot 23^1 \cdot 29^1 \cdot 31^1 \cdot 37^1 \cdot 41^1$
21	$2^{82} \cdot 3^{30} \cdot 5^{12} \cdot 7^8 \cdot 11^4 \cdot 13^3 \cdot 17^2 \cdot 19^2 \cdot 23^1 \cdot 29^1 \cdot 31^1 \cdot 37^1 \cdot 41^1 \cdot 43^1$
22	$2^{86} \cdot 3^{31} \cdot 5^{13} \cdot 7^8 \cdot 11^4 \cdot 13^3 \cdot 17^2 \cdot 19^2 \cdot 23^2 \cdot 29^1 \cdot 31^1 \cdot 37^1 \cdot 41^1 \cdot 43^1$
23	$2^{89} \cdot 3^{32} \cdot 5^{13} \cdot 7^8 \cdot 11^4 \cdot 13^3 \cdot 17^2 \cdot 19^2 \cdot 23^2 \cdot 29^1 \cdot 31^1 \cdot 37^1 \cdot 41^1 \cdot 43^1 \cdot 47^1$
24	$2^{95} \cdot 3^{34} \cdot 5^{14} \cdot 7^9 \cdot 11^4 \cdot 13^4 \cdot 17^3 \cdot 19^2 \cdot 23^2 \cdot 29^1 \cdot 31^1 \cdot 37^1 \cdot 41^1 \cdot 43^1 \cdot 47^1$
25	$2^{98} \cdot 3^{35} \cdot 5^{14} \cdot 7^9 \cdot 11^5 \cdot 13^4 \cdot 17^3 \cdot 19^2 \cdot 23^2 \cdot 29^1 \cdot 31^1 \cdot 37^1 \cdot 41^1 \cdot 43^1 \cdot 47^1$

Table 2. $H(n)$.

Remark 5.4. For a given C, we can modify the phrase "for any n large enough" in the statement of Proposition 5.3. For example, let us consider the case $C = 4$. By studying (i) and (ii) in the above proof more carefully, we can show

$$\Psi(n) < 4n \log\log n \quad \text{for any } n > e^{(1.001e)^9}.$$

n	$G(n)$
1	$2^4 \cdot 3^1$
2	$2^9 \cdot 3^6 \cdot 5^1 \cdot 13^1$
3	$2^{13} \cdot 3^{15} \cdot 5^1 \cdot 7^1 \cdot 11^2 \cdot 13^2$
4	$2^{19} \cdot 3^{28} \cdot 5^2 \cdot 7^1 \cdot 11^2 \cdot 13^2 \cdot 41^1 \cdot 1093^1$
5	$2^{23} \cdot 3^{45} \cdot 5^2 \cdot 7^1 \cdot 11^4 \cdot 13^3 \cdot 41^1 \cdot 61^1 \cdot 757^1 \cdot 1093^1$
6	$2^{28} \cdot 3^{66} \cdot 5^3 \cdot 7^2 \cdot 11^4 \cdot 13^4 \cdot 23^1 \cdot 41^1 \cdot 61^1 \cdot 73^1 \cdot 757^1 \cdot 1093^1 \cdot 3851^1$
7	$2^{32} \cdot 3^{91} \cdot 5^3 \cdot 7^2 \cdot 11^4 \cdot 13^4 \cdot 23^1 \cdot 41^1 \cdot 61^1 \cdot 73^1 \cdot 547^1 \cdot 757^1 \cdot 1093^2 \cdot 3851^1 \cdot 797161^1$

Table 3. $G(n)$ (for $p \neq 3$).

Here we check the above inequality. Condition (ii) is equivalent to

$$(\log x)^{C/(2e^{\gamma})-1} > e\left(1+\frac{\log(C\log\log x)}{\log x}\right).$$

We assume $x > e^{e^9}$. Since $C/(2e^{\gamma})-1 > \frac{4}{3.6}-1=\frac{1}{9}$ and $\log(C\log\log x)/\log x < 0.001$, inequality (ii) holds if $(\log x)^{\frac{1}{9}} > 1.001e$, that is, $x > e^{(1.001e)^9}$. Note that (i) clearly holds for such x.

Acknowledgements

The author would like to thank Yuichiro Taguchi for useful discussions about the proof of our main results and Manabu Yoshida for advice on an earlier draft. Thanks also are due to Takaichi Fujiwara for helpful advice on the functions discussed in Section 5. This work is supported by JSPS KAKENHI grant number JP19K03433.

References

[Chou 2019] M. Chou, "Torsion of rational elliptic curves over the maximal abelian extension of $\mathbb{Q}$", *Pacific J. Math.* **302**:2 (2019), 481–509. MR Zbl

[Chou et al. 2021] M. Chou, P. L. Clark, and M. Milosevic, "Acyclotomy of torsion in the CM case", *Ramanujan J.* **55**:3 (2021), 1015–1037. MR Zbl

[Clark and Xarles 2008] P. L. Clark and X. Xarles, "Local bounds for torsion points on abelian varieties", *Canad. J. Math.* **60**:3 (2008), 532–555. MR Zbl

[Coleman and Iovita 1999] R. Coleman and A. Iovita, "The Frobenius and monodromy operators for curves and abelian varieties", *Duke Math. J.* **97**:1 (1999), 171–215. MR Zbl

[Conrad 2011] B. Conrad, "Lifting global representations with local properties", preprint, 2011, available at http://math.stanford.edu/~conrad/papers/locchar.

[Fontaine 1982] J.-M. Fontaine, "Sur certains types de représentations p-adiques du groupe de Galois d'un corps local: construction d'un anneau de Barsotti–Tate", *Ann. of Math.* (2) **115**:3 (1982), 529–577. MR Zbl

[Fontaine 1994a] J.-M. Fontaine, "Le corps des périodes p-adiques", pp. 59–111 in *Périodes p-adiques* (Bures-sur-Yvette, France, 1988), edited by J.-M. Fontaine, Astérisque **223**, Soc. Math. France, Paris, 1994. MR Zbl

[Fontaine 1994b] J.-M. Fontaine, "Représentations p-adiques semi-stables", pp. 113–184 in *Périodes p-adiques* (Bures-sur-Yvette, France, 1988), edited by J.-M. Fontaine, Astérisque **223**, Soc. Math. France, Paris, 1994. MR Zbl

[Imai 1975] H. Imai, "A remark on the rational points of abelian varieties with values in cyclotomic Z_p-extensions", *Proc. Japan Acad.* **51** (1975), 12–16. MR Zbl

[Katz and Lang 1981] N. M. Katz and S. Lang, "Finiteness theorems in geometric classfield theory", *Enseign. Math.* (2) **27**:3-4 (1981), 285–319. MR Zbl

[Kubo and Taguchi 2013] Y. Kubo and Y. Taguchi, "A generalization of a theorem of Imai and its applications to Iwasawa theory", *Math. Z.* **275**:3-4 (2013), 1181–1195. MR Zbl

[Mattuck 1955] A. Mattuck, "Abelian varieties over p-adic ground fields", *Ann. of Math.* (2) **62** (1955), 92–119. MR Zbl

[Neukirch 1999] J. Neukirch, *Algebraic number theory*, Grundl. Math. Wissen. **322**, Springer, 1999. MR Zbl

[Ozeki 2020] Y. Ozeki, "Torsion of abelian varieties and Lubin–Tate extensions", *J. Number Theory* **207** (2020), 282–293. MR Zbl

[Ozeki 2024] Y. Ozeki, "Bounds on torsion of CM abelian varieties over a p-adic field with values in a field of p-power roots", *New York J. Math.* **30** (2024), 422–435. MR Zbl

[Rosser and Schoenfeld 1962] J. B. Rosser and L. Schoenfeld, "Approximate formulas for some functions of prime numbers", *Illinois J. Math.* **6** (1962), 64–94. MR Zbl

[Serre 1979] J.-P. Serre, *Local fields*, Grad. Texts in Math. **67**, Springer, 1979. MR Zbl

[Serre 1989] J.-P. Serre, *Abelian l-adic representations and elliptic curves*, 2nd ed., Addison-Wesley, Redwood City, CA, 1989. MR Zbl

[Serre and Tate 1968] J.-P. Serre and J. Tate, "Good reduction of abelian varieties", *Ann. of Math.* (2) **88** (1968), 492–517. MR Zbl

[SGA 7_{I} 1972] A. Grothendieck, "Modèles de Néron et monodromie", pp. 313–523 in *Groupes de monodromie en géométrie algébrique*, *I* (Séminaire de Géométrie Algébrique du Bois Marie 1967–1969), edited by A. Grothendieck, Lecture Notes in Math. **288**, Springer, 1972. MR Zbl

[Silverberg 1992] A. Silverberg, "Fields of definition for homomorphisms of abelian varieties", *J. Pure Appl. Algebra* **77**:3 (1992), 253–262. MR Zbl

[Tate and Oort 1970] J. Tate and F. Oort, "Group schemes of prime order", *Ann. Sci. École Norm. Sup.* (4) **3** (1970), 1–21. MR Zbl

[Yoshida 2008] T. Yoshida, "Local class field theory via Lubin–Tate theory", *Ann. Fac. Sci. Toulouse Math.* (6) **17**:2 (2008), 411–438. MR Zbl

Received October 23, 2023. Revised May 16, 2024.

YOSHIYASU OZEKI
FACULTY OF SCIENCE
KANAGAWA UNIVERSITY
KANAGAWA-KU, YOKOHAMA
JAPAN
ozeki@kanagawa-u.ac.jp

Guidelines for Authors

Authors may submit articles at msp.org/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 330 No. 1 May 2024
