



Counting essential surfaces in 3-manifolds

Nathan M. Dunfield¹ · Stavros Garoufalidis² ·
J. Hyam Rubinstein³

Received: 17 August 2020 / Accepted: 16 November 2021 /
Published online: 25 January 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of
Springer Nature 2022

Abstract We consider the natural problem of counting isotopy classes of essential surfaces in 3-manifolds, focusing on closed essential surfaces in a broad class of hyperbolic 3-manifolds. Our main result is that the count of (possibly disconnected) essential surfaces in terms of their Euler characteristic always has a short generating function and hence has quasi-polynomial behavior. This gives remarkably concise formulae for the number of such surfaces, as well as detailed asymptotics. We give algorithms that allow us to compute these generating functions and the underlying surfaces, and apply these to almost 60,000 manifolds, providing a wealth of data about them. We use this data to explore the delicate question of counting only connected essential surfaces and propose some conjectures. Our methods involve normal and almost normal surfaces, especially the work of Tollefson and Oertel, combined

✉ Nathan M. Dunfield
nathan@dunfield.info

Stavros Garoufalidis
stavros@mpim-bonn.mpg.de

J. Hyam Rubinstein
joachim@unimelb.edu.au

¹ Department of Mathematics, University of Illinois, 1409 W. Green Street, Urbana, IL 61801, USA

² Department of Mathematics, International Center for Mathematics, Southern University of Science and Technology, Shenzhen, China

³ School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia

with techniques pioneered by Ehrhart for counting lattice points in polyhedra with rational vertices. We also introduce a new way of testing if a normal surface in an ideal triangulation is essential that avoids cutting the manifold open along the surface; rather, we use almost normal surfaces in the original triangulation.

Contents

1	Introduction	719
1.2	Main results	720
1.5	Motivation and broader context	721
1.7	The key ideas behind Theorem 1.3	722
1.8	Making Theorem 1.3 algorithmic	724
1.9	Ideal triangulations and almost normal surfaces	724
1.10	Computations and patterns	725
1.11	The view from measured laminations	725
1.13	Understanding counts by genus	727
2	Background and conventions	727
2.1	Numbers	727
2.2	Surfaces in 3-manifolds	727
2.3	Nonorientable surfaces	728
2.5	Triangulations	729
2.6	Normal surfaces	730
2.7	Short generating functions and quasi-polynomials	731
3	Isotopy classes of essential normal surfaces	732
3.4	Dependent faces	734
3.10	Complete lw-faces in detail	735
4	Surface counts are almost quasi-polynomial	739
4.2	Counting surfaces via lattice points	740
4.9	Ideal triangulations	743
4.13	Whither nonorientable and bounded surfaces	746
5	Proof of the decision theorem	746
5.2	Graph of incompressible surfaces	747
5.4	Isotopic normal pairs	748
6	Almost normal surfaces in ideal triangulations	750
6.1	Tightening almost normal surfaces	751
6.2	Sweepouts and thin position	751
6.5	Finiteness of (almost) normal surfaces	752
6.8	Another graph of normal surfaces	754
6.12	Algorithm	755
7	Computations, examples, and patterns	757
7.1	Very large manifolds	758
7.2	Surface counts by Euler characteristic	759
7.3	Sample LW complexes	759
7.4	Isotopies of lw-surfaces	760
7.5	Barely large knots and those without meridional essential surfaces	762
7.7	Code and data	764
8	Patterns of surface counts by genus	765
8.1	Independence of $a_M(g)$ and $B_M(x)$	765
8.4	Regular genus counts and the Lambert series	768

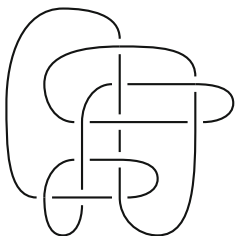
8.6 Examples 768
 8.7 Asymptotics of genus counts 770
 References 773

1 Introduction

Essential surfaces have played a central role in 3-manifold topology for at least the last 70 years, being both a key tool and a fundamental object of study. Roughly, these are compact embedded surfaces $F \subset M^3$ where $\pi_1 F \rightarrow \pi_1 M$ is injective; throughout this introduction, see Sect. 2 for precise definitions and conventions. While some compact 3-manifolds contain no essential surfaces at all (the 3-sphere, lens spaces), others contain infinitely many isotopy classes of essential surfaces of the same topological type (the 3-torus contains infinitely many essential 2-tori). However, for M that are irreducible and atoroidal (i.e. contain no essential spheres or tori), the number of essential F of a fixed topological type is always finite [38, Corollary 2.3]. For example, any hyperbolic 3-manifold is irreducible and atoroidal, and these form the main class of interest here.

A natural problem is thus to describe in a structured way the set of essential surfaces in a given 3-manifold M , in particular to list and to count them. Focusing on those F that are closed, connected, and orientable, define $a_M(g)$ to be the number of isotopy classes of essential surfaces in M of genus g . There are plenty of hyperbolic 3-manifolds where $a_M(g) = 0$ for all g , including all those that are exteriors of 2-bridge knots [36]. In contrast, for the exterior of X of the Conway knot $K11n34$, we can use Theorem 1.4 below to compute the values of $a_X(g)$ shown in Table 1, as well as further values such as $a_X(50) = 56,892$ and $a_X(100) = 444,038$.

Table 1 The first few values of $a_X(g)$ where X is the exterior of the Conway knot shown at left

	g	$a_X(g)$	g	$a_X(g)$	g	$a_X(g)$	g	$a_X(g)$
 $K11n34$	1	0	7	87	13	602	19	1993
	2	6	8	208	14	1168	20	3484
	3	9	9	220	15	1039	21	2924
	4	24	10	366	16	1498	22	4126
	5	37	11	386	17	1564	23	3989
	6	86	12	722	18	2514	24	6086

While the sequence in Table 1 is a complete mystery to us, if we broaden our perspective to include disconnected surfaces, we get a relatively simple pattern that we can describe completely. Specifically, for any M define $b_M(n)$ to be the number of isotopy classes of closed orientable essential surfaces F in M with $\chi(F) = n$. For the Conway exterior X , we show (see Fig. 3):

$$b_X(-2n) = \frac{2}{3}n^3 + \frac{9}{4}n^2 + \frac{7}{3}n + \frac{7 + (-1)^n}{8} \quad \text{for all } n \geq 1. \quad (1.1)$$

The formula for b_X would be a polynomial in n were it not for the final term which oscillates mod 2. The first main result of this paper, Theorem 1.3 below, shows that the count b_M always has this kind of almost polynomial structure for a broad class of 3-manifolds M .

1.2 Main results

We can encode a function $s: \mathbb{N} \rightarrow \mathbb{Q}$ by its *generating function* $S(x) = \sum_{n=0}^{\infty} s(n)x^n$ in the formal power series ring $\mathbb{Q}[[x]]$. We say this generating function is *short* when $S(x) = P(x)/Q(x)$ for polynomials P and Q in $\mathbb{Q}[x]$ where $Q(x)$ is a product of cyclotomic polynomials. For example, the function $s(n) = b_X(-2n)$ from (1.1) above has a short generating function, namely

$$S(x) = \frac{-x^5 + 3x^4 - 2x^3 + 2x^2 + 6x}{(x+1)(x-1)^4}.$$

Having a short generating function is equivalent to $s(n)$ being a quasi-polynomial for all but finitely many values of n , see Sect. 2.7. Quasi-polynomials first arose in Ehrhart's work on counting lattice points in polyhedra with rational vertices [24] and have many applications to enumerative combinatorics [59, Chapter 4]; curiously, they also appear in quantum topology [28, 29, 32]. We can now state:

1.3 Theorem *Suppose M is a compact orientable irreducible ∂ -irreducible atoroidal acylindrical 3-manifold that does not contain a closed nonorientable essential surface. Let $b_M(n)$ be the number of isotopy classes of closed essential surfaces F in M with $\chi(F) = n$, and $B_M(x) = \sum_{n=1}^{\infty} b_M(-2n)x^n$ be the corresponding generating function. Then $B_M(x)$ is short.*

Here, we can ensure that M has no closed nonorientable essential surfaces by requiring that $H_1(\partial M; \mathbb{F}_2) \rightarrow H_1(M; \mathbb{F}_2)$ is onto, see Proposition 2.4. Thus, Theorem 1.3 applies to the exterior of any hyperbolic knot in S^3 . We discuss possible extensions of Theorem 1.3 to nonorientable surfaces, as well as to surfaces with boundary, in Sect. 4.13.

All aspects of Theorem 1.3 can be made algorithmic, both in theory and in practice. The theoretical part is:

1.4 Theorem *There exists an algorithm that takes as input a triangulation \mathcal{T} of a manifold M as in Theorem 1.3 and computes $P(x), Q(x) \in \mathbb{Q}[x]$ such that $B_M(x) = P(x)/Q(x)$. Moreover, there is an algorithm that given $n \in \mathbb{N}$ outputs a list of normal surfaces in \mathcal{T} uniquely representing all isotopy classes of essential surfaces with $\chi = -2n$. Finally, there is an algorithm that given an essential normal surface F with $\chi(F) = -2n$ finds the isotopic surface in the preceding list.*

In Sect. 6, we refine Theorem 1.4 into a practical algorithm that uses ideal triangulations and their special properties. Then in Sect. 7, we compute $B_M(x)$ for almost 60,000 examples. It is natural to ask whether one could permit nonorientable essential surfaces in Theorem 1.3, as well as essential surfaces with boundary; we outline some of the difficulties inherent in such extensions in Sect. 4.13 below.

1.5 Motivation and broader context

From Theorem 1.3 and the discussion in Sect. 2.7, the sequence $b_M(-2n)$ grows at most polynomially in n . It is not always the case that $b_M(-2n)$ is asymptotic to cn^d : we found an example where $b_M(-2n)$ is $n/2 + 1$ for n even and 0 for n odd. However, by Lemma 2.8, we get precise asymptotics if we smooth the sequence by setting $\bar{b}_M(-2n) = \sum_{k=1}^n b_M(-2k)$:

1.6 Corollary *For each M as in Theorem 1.3, either $b_M(-2n) = 0$ for all n or there exists $d \in \mathbb{N}$ and $c > 0$ in \mathbb{Q} such that $\lim_{n \rightarrow \infty} \bar{b}_M(-2n)/n^d = c$.*

We conjecture in Sect. 1.11 below that d is the dimension of the space $\mathcal{ML}_0(M)$ of measured laminations without boundary in M , and c is the volume of a certain subset of $\mathcal{ML}_0(M)$.

As $a_M(g) \leq b_M(-2g + 2)$ for each g , we have that $a_M(g)$ also grows at most polynomially in g . In stark contrast, if we allow *immersed* surfaces, then Kahn-Markovic [43] showed that, for any closed hyperbolic 3-manifold M , the number of essential immersed surfaces of genus g grows like g^{2g} .

This distinction between counts of embedded versus immersed surfaces parallels the following story a dimension down. For a closed hyperbolic surface Y of genus g , Mirzakhani [49] showed that the number $s_Y(L)$ of embedded essential multicurves in Y of geodesic length at most L satisfies $s_Y(L) \sim n(Y)L^{6g}$ for some $n(Y) > 0$; in contrast, the number $c_Y(L)$ of primitive closed geodesics of length at most L satisfies $c_Y(L) \sim e^L/L$, see e.g. [18]. In fact, Mirzakhani proved much more: given an essential multicurve γ , the count $s_Y(L, \gamma)$ of multicurves in the mapping class group orbit of γ also satisfies

$s_Y(L, \gamma) \sim n_\gamma(Y)L^{6g}$ with $n_\gamma(Y) > 0$. In particular, this gives asymptotics for the counts of all *connected* essential curves, analogous in our setting to a_M as opposed to b_M ; we hint at how this connection might be further developed in Sect. 1.13. There are also similarities between the setting of [49] and the measured lamination perspective on Theorem 1.3 outlined in Sect. 1.11. The fact that we count surfaces by Euler characteristic, which is discrete, rather than by a continuous notion such as length or area, is what allows us get precise formulas for b_M as well as asymptotics. (More directly analogous to the surface case, one could try to count embedded essential surfaces in a closed hyperbolic 3-manifold M in terms of the area of their least area representatives. As such representatives satisfy $\pi|\chi(F)| \leq \text{Area}(F) \leq 2\pi|\chi(F)|$ by [33, Lemma 6], it is not inconceivable that there are good asymptotics here as well given Corollary 1.6.)

The algorithm of Sect. 6 relies heavily on ideal triangulations and their normal and almost-normal surfaces. Curiously, normal surfaces are also used to construct recent topological quantum invariants of 3-manifolds, specifically the 3D-index of Dimofte, Gaiotto and Gukov [21, 22]. The latter is a collection of Laurent series with integer coefficients which are defined using an ideal triangulation and depend only on the number of tetrahedra around each edge of the triangulation, as encoded in the Neumann-Zagier matrices. The 3D-index is a topological invariant of cusped hyperbolic 3-manifolds [31] that can be expressed as a generating series of generalized normal surfaces in a 1-efficient triangulation [30], a class of surfaces that includes both normal and almost normal surfaces. It would be very interesting to connect the topological invariants of Theorem 1.3 with the 3D-index.

1.7 The key ideas behind Theorem 1.3

We first explain how the perspective of branched surfaces, especially the work of Oertel [54], naturally relates the sequence $b_M(-2n)$ to counting lattice points in an expanding family of rational polyhedra; combined with Ehrhart's work [24] on the latter topic, this discussion will make Theorem 1.3 very plausible. We then sketch how Tollefson [64] reinterpreted and extended Oertel's branched surface picture in the context of normal surface theory, and how this viewpoint allows us to actually prove Theorem 1.3. For ease of exposition, we assume throughout that M is closed and contains only orientable surfaces by Proposition 2.4.

A branched surface \mathcal{B} in a 3-manifold M is the analog, one dimension up, of a train track on a surface; see [26, 54] for definitions and general background. A surface F is carried by \mathcal{B} if it is isotopic into a fibered neighborhood $N(\mathcal{B})$ of \mathcal{B} so that it is transverse to the vertical interval fibers. Such an F is determined by the nonnegative integer weights it associates to the sectors of \mathcal{B} , which are the

components of \mathcal{B} minus its singular locus. Such weights correspond to a surface if and only if they satisfy a system of homogenous linear equations that are analogous to the switch conditions for a train track. The set of all nonnegative *real* weights satisfying these equations gives a finite-sided polyhedral cone $\mathcal{ML}(\mathcal{B})$, which corresponds to measured laminations carried by \mathcal{B} . Here, each integer lattice point in $\mathcal{ML}(\mathcal{B})$ corresponds to a surface carried by \mathcal{B} . As the equations defining $\mathcal{ML}(\mathcal{B})$ have integer coefficients, each edge ray of the cone $\mathcal{ML}(\mathcal{B})$ contains a lattice point.

For M as in Theorem 1.3, by Theorem 4 of [54] there is a finite set $\mathcal{B}_1, \dots, \mathcal{B}_n$ of branched surfaces that together carry all essential surfaces in M and also carry *only* essential surfaces. Moreover, two surfaces carried by one \mathcal{B}_i are isotopic if and only if they correspond to the same lattice point in $\mathcal{ML}(\mathcal{B}_i)$. Putting aside the important issue of surfaces being carried by several of these branched surfaces, here is how to count essential surfaces carried by a fixed \mathcal{B}_i . First, there is a linear function $\bar{\chi}: \mathcal{ML}(\mathcal{B}_i) \rightarrow \mathbb{R}$ which on lattice points gives the Euler characteristic of the corresponding surface. Because M is irreducible and atoroidal, every essential surface has $\chi < 0$; as each edge ray of the cone $\mathcal{ML}(\mathcal{B}_i)$ contains a lattice point corresponding to an essential surface, we conclude that $\bar{\chi}$ is proper and nonpositive on $\mathcal{ML}(\mathcal{B}_i)$. Hence $P = \bar{\chi}^{-1}(-1)$ is a compact polytope with, it turns out, rational vertices. Thus, the contribution to $b_M(-2n)$ of surfaces carried by \mathcal{B}_i is exactly the number of lattice points in $2n \cdot P$, where the latter denotes the dilation of P by a factor of $2n$. The foundational work of Ehrhart [24] shows that this count of lattice points is quasi-polynomial.

If no surface is carried by multiple \mathcal{B}_i , the sketch just given would essentially prove Theorem 1.3 as sums of quasi-polynomials are again quasi-polynomial. However, there is no avoiding this issue in general, and we deal with it by using the work of Tollefson [64], who built on [54] to provide a concrete description of isotopy classes of essential surfaces in the context of normal surface theory. If we fix a triangulation \mathcal{T} of M , then every essential surface in M can be isotoped to be normal with respect to \mathcal{T} ; throughout, see Sect. 2.6 for definitions and general background. There can be many normal representatives of the same essential surface, so to reduce this redundancy, Tollefson focuses on those that are least weight in that they meet the 1-skeleton of \mathcal{T} in as few points as possible. We define a *lw-surface* to be a normal surface that is essential and least weight. To prove Theorem 1.3, we need to count such lw-surfaces modulo isotopy in M .

Let $\mathcal{S}_{\mathcal{T}}$ be the normal surface solution space, which is a finite rational polyhedral cone whose admissible integer points correspond to normal surfaces in \mathcal{T} , and let $\mathcal{P}_{\mathcal{T}}$ be its projectivization. A normal surface F is *carried* by a face C of $\mathcal{P}_{\mathcal{T}}$ if the projectivization of the lattice point corresponding to F is in C . An admissible face C of $\mathcal{P}_{\mathcal{T}}$ is a *lw-face* if every normal surface it carries is a

lw-surface. While it is not obvious that any lw-faces exist, Tollefson showed that every lw-surface is carried by a lw-face. To make the parallel with the previous discussion explicit, each lw-face C has a corresponding branched surface \mathcal{B}_C which carries, in the prior sense, exactly the surfaces carried by C in the current sense. The collection of all lw-faces is a complex we denote $\mathcal{LW}_{\mathcal{T}}$; see Fig. 3 for an example in the case of a triangulation of the Conway knot exterior.

Tollefson shows moreover that every lw-surface is carried by a lw-face that is *complete*: if F and G are isotopic lw-surfaces and C carries F then it also carries G . The isotopies between lw-surfaces carried by the same complete lw-face can be understood using a foliation of C by affine subspaces parallel to some fixed linear subspace W_C ; roughly, surfaces F and G carried by C are isotopic if their lattice points differ by an element of W_C . See Sect. 3 and especially Theorem 3.5 for details, including the key notion of $\text{dep}(C)$. This translates the problem of counting essential surfaces carried by a complete face to one of counting *projections* of lattice points in the cone over C after we quotient out by W_C . This is exactly the setting of recent work of Nguyen and Pak [52], which we use in Sect. 4 to complete the proof of Theorem 1.3.

1.8 Making Theorem 1.3 algorithmic

Since Haken, normal surfaces have played a key role in the study of algorithmic questions about 3-manifolds. Despite this, Tollefson in [64] did not give an algorithm for finding the lw-faces of $\mathcal{P}_{\mathcal{T}}$ nor determining their properties such as completeness. Section 5 here focuses on establishing Theorem 5.1, which gives an algorithm for computing all complete lw-faces. One important tool for this is Theorem 5.3, which shows that if F and G are isotopic lw-surfaces then there is a sequence of isotopic lw-surfaces $F = F_1, F_2, \dots, F_{n-1}, F_n = G$ with each pair (F_i, F_{i+1}) disjoint and cobounding a product region. Combined with results from Sect. 3, especially Theorem 3.3, we can strengthen the arguments behind Theorem 1.3 to prove Theorem 1.4.

1.9 Ideal triangulations and almost normal surfaces

When the 3-manifold M has nonempty boundary, the proofs of Theorems 1.3 and 1.4 use ideal triangulations rather than finite ones (see Sect. 4.9). Our computations were with M where ∂M is a single torus whose interior admits a complete hyperbolic metric of finite-volume, and we used ideal triangulations there as well, especially as they have several advantages. For example, they typically have fewer tetrahedra than finite triangulations, which speeds up normal surface computations. Most importantly, when the ideal triangulation

admits a strict angle structure, Lackenby showed [45] that the number of connected normal surfaces of a fixed genus is finite and described how they can be enumerated. In Sect. 6, we explain how to exploit this to give a *practical* version of the algorithms in Theorem 1.4. Unlike the proof of Theorem 1.4, we make heavy use of almost normal surfaces, including those with tubes, and in particular the process of tightening (also called normalizing) an almost normal surface.

The usual method for testing if a normal surface F in M is essential is to cut M open along F , triangulate the result, and then use normal surfaces to search for a compressing disk; a key difficulty with this is that the triangulation of $M \setminus F$ is usually much more complicated than the original one. Here, we introduce a completely new method for determining when F is essential that does not require cutting M open but rather uses almost normal surfaces in the original triangulation (Sect. 6.8).

Our implementation of the algorithm in Sect. 6 can be found at [23] and makes heavy use of Regina [8], SageMath [57], and Normaliz [11]. It includes code for tightening almost normal surfaces, as well as dealing with general normal surfaces with tubes, both of which have explored extensively in theory but never before in practice.

1.10 Computations and patterns

Sections 7 and 8 detail our experiments using the algorithm of Sect. 6. In particular, we applied it to more than 59,000 manifolds, including more than 4,300 where $\dim \mathcal{LW}_T > 0$. We include overall statistics about the complexes \mathcal{LW}_T , the generating functions $B_M(x)$, and the sequences $a_M(g)$ in Tables 2, 3, 4, 5 and 6 and 7, 8 and 9, as well as detailed examples of \mathcal{LW}_T in Figs. 1, 2, 3 and 4. In Sect. 8.1, we give examples showing that, perhaps surprisingly, neither of $B_M(x)$ and $a_M(g)$ determines the other.

For the more mysterious $a_M(g)$, while we are unable to find a pattern in these sequences in many cases, there are some M where we conjecture relatively simple formulae for $a_M(g)$; see Conjecture 8.2 and Table 9. In Conjecture 8.9, we posit the existence of general asymptotics for (a smoothed version of) $a_M(g)$ based on the striking plots in Figs. 6 and 7, where we computed $a_M(g)$ out to $g = 200$ in many cases.

1.11 The view from measured laminations

For surfaces, a central tool for studying their topology, geometry, and dynamics is measured laminations; for example, the space $\mathcal{ML}(F)$ of all measured laminations on a surface F plays a key role in [49]. In 3-dimensions, building

on work of Morgan and Shalen [50, 51], independently Hatcher [34] and Oertel [55] studied measured laminations on 3-manifolds in detail, organizing them into a topological space $\mathcal{ML}(M)$. Note here an essential surface, with or without boundary, can be viewed as a measured lamination, and the set of all essential surfaces nearly injects into $\mathcal{ML}(M)$ (see page 6 of [34] for the caveat which involves the two nonorientable surfaces in a semifibration) with its image being a discrete set of points. While for a surface F of genus g the space $\mathcal{ML}(F)$ is just homeomorphic to \mathbb{R}^{6g-6} , for a general 3-manifold M the space $\mathcal{ML}(M)$ can be singular, being built from open strata each of which is a PL manifold. The charts on the individual strata come from branched surfaces; specifically, one uses the polyhedral cones $\mathcal{ML}(\mathcal{B}_i)$ associated with certain essential branched surfaces \mathcal{B}_i as sketched in Sect. 1.7; see [34, 55] for details.

Let $\mathcal{ML}_0(M)$ denote the subset of measured laminations that are disjoint from ∂M . The topological dimension of $\mathcal{ML}_0(M)$ is the maximum of $\dim(\mathcal{ML}(\mathcal{B}))$ for the appropriate class of essential branched surfaces \mathcal{B} without boundary. Because of the theory of Oertel [54] that underlies [64], we are highly confident that:

1.12 Conjecture *The dimension of \mathcal{ML}_0 is the maximum of $\dim C - \dim W_C + 1$ where C is an essential lw-face of \mathcal{PT} and W_C is defined in Theorem 3.5.*

If Conjecture 1.12 holds, then in Corollary 1.6 where $\bar{b}_M(-2n) \sim cn^d$ one has $d = \dim(\mathcal{ML}_0)$, thus giving an intrinsic characterization of that exponent. We further posit that the coefficient c in these asymptotics has the following natural interpretation. As mentioned, the PL structure on the strata of $\mathcal{ML}_0(M)$ comes from charts to $\mathcal{ML}(\mathcal{B}_i)$ for certain branched surfaces \mathcal{B}_i ; in particular, one gets PL coordinate change maps between (possibly empty) subsets of each pair $\mathcal{ML}(\mathcal{B}_i)$ and $\mathcal{ML}(\mathcal{B}_j)$, see Proposition 4.1 of [34]. These coordinate change maps must take lattice points to lattice points, since these correspond to the special measured laminations that come from essential surfaces. Hence the coordinate change maps should have derivatives that are in $GL_n\mathbb{Z}$ and so are (unsigned) volume preserving. This would give a well-defined measure (in the Lebesgue class) on each strata of $\mathcal{ML}_0(M)$; this is a direct analog of Thurston's notion of volume on $\mathcal{ML}(F)$ where F is a surface, which is defined in terms of the integral PL structure on $\mathcal{ML}(F)$ coming from train track charts.

Recall for any branched surface \mathcal{B}_i , there is a linear map $\bar{\chi}: \mathcal{B}_i \rightarrow \mathbb{R}$ which gives the Euler characteristic of the corresponding surface at each lattice point. These should piece together to give a PL map $\bar{\chi}: \mathcal{ML}_0(M) \rightarrow \mathbb{R}$. In the setting of Theorem 1.3, the subset $P = \bar{\chi}^{-1}([-1, 0])$ in $\mathcal{ML}_0(M)$ will be compact. We conjecture that the coefficient c is precisely $\text{vol}(P)$.

1.13 Understanding counts by genus

The key problem to overcome in understanding $a_M(g)$ is to determine, for a complete lw-face C , which lattice points in $\tilde{C} = \mathbb{R}_{\geq 0} \cdot C$ correspond to connected surfaces. Agol, Hass, and Thurston showed in [3, §4] how counting the number of connected components of a normal surface can be reframed as counting the number of orbits of a family of interval isometries acting on $\{1, 2, \dots, N\}$. Such families of interval isometries include both classical and non-classical interval exchange transformations on surfaces [27], but are considerably more general. Geometrically, a family of interval isometries can be thought of as an interval I of some length L to which finitely many bands of specified widths are attached, without any restriction on how many bands are glued to any subinterval of I . For normal surfaces, the interval I is basically an arbitrary concatenation of the edges of the ambient triangulation \mathcal{T} , and the bands correspond to families of normal arcs in the corners of each face of \mathcal{T}^2 ; see Corollary 13 of [3] for details. (For each admissible face C of $\mathcal{P}_{\mathcal{T}}$, one can also think about this in terms of the associated branched surface \mathcal{B}_C .) Thus, a general theory of the number of orbits of the integer points of such a family of isometries should allow one to develop a detailed picture for $a_M(g)$.

Currently, the best understood case is for a suitable train track τ on a surface F , where Mirzakhani [49] gives asymptotics on the portion of integer points in $\mathcal{ML}(\tau)$ that correspond to connected curves, see also [10] for a detailed discussion. (Here, one uses the total weight of a point in $\mathcal{ML}(\tau)$ as the “length” of the associated multicurve, rather than Euler characteristic in the 3-dimensional setting.) Even for simple train tracks, it seems that the counts of connected curves can be irregular in the sense of Sect. 8.4, so there is work to be done even in that setting.

2 Background and conventions

2.1 Numbers

We use \mathbb{N} to denote the nonnegative integers, i.e. $\mathbb{N} = \{0, 1, 2, \dots\}$.

2.2 Surfaces in 3-manifolds

Throughout the rest of this paper, every 3-manifold M will be compact, orientable, irreducible (every embedded sphere bounds a ball) and ∂ -irreducible (every properly embedded disk bounds a ball with some disk in ∂M). Surfaces need not be orientable, but will always be embedded in any ambient 3-manifold, and in particular be compact. Moreover, a surface F in a 3-manifold M will be assumed to be *properly* embedded with $F \cap \partial M = \partial F$, except for com-

pressing disks and ∂ -compressing disks which we define next. A *compressing disk* for a surface F in a 3-manifold M is a disk $D \subset M$ where $D \cap F = \partial D$ and ∂D does not bound a disk in F . An orientable surface F in M is *incompressible* when it has no compressing disks and is neither a sphere nor a disk. (A more general notion of incompressibility allows certain spheres and disks, but none such exist in an irreducible and ∂ -irreducible manifold.) Since M is ∂ -irreducible, any parallel copy of a component of ∂M is incompressible.

A ∂ -*compressing disk* D for a surface F in M is one where ∂D consists of an arc α in F and an arc β in ∂M , the interior of D is disjoint from $F \cup \partial M$, and α does not bound a disk in F with a segment of ∂F . An orientable surface F in M is ∂ -*incompressible* when it has no ∂ -compressing disks and is not itself a disk. A surface F in M is ∂ -*parallel* when every connected component of F is isotopic, keeping ∂F fixed, into ∂M ; when $\partial F = \emptyset$, this is equivalent to F being ambient isotopic to a union of parallel copies of components of ∂M .

An orientable surface F in M is *essential* when it is incompressible, ∂ -incompressible, and no connected component is ∂ -parallel. A 3-manifold is *atoroidal* when it does not contain an essential torus (this is sometimes called *geometrically atoroidal*). Similarly, it is *acylindrical* when it does not contain an essential annulus (also called *anannular*).

2.3 Nonorientable surfaces

For a nonorientable surface F in M , we define it to be incompressible, ∂ -incompressible, or essential when the boundary of a regular neighborhood of F has that same property. One could instead apply the above definitions directly to nonorientable surfaces, which give significantly weaker conditions in general. Sources such as [26,64] use the terms injective and ∂ -injective for what we here call incompressible and ∂ -incompressible to distinguish the possible definitions in the nonorientable case. Some corner cases are worth mentioning. First, with our conventions, a connected surface F in M is incompressible if and only if $\pi_1 F \rightarrow \pi_1 M$ is injective and F is not a sphere, a disk, or $\mathbb{R}P^2$. Also, if M is the twisted interval bundle over a nonorientable closed surface F , then F is incompressible but not essential.

In our main results, we require that M contain no closed nonorientable essential surfaces, and the following proposition provides an easily checkable sufficient condition for this to be the case:

2.4 Proposition *Suppose M is a compact orientable 3-manifold. Every closed embedded surface in M is orientable if and only if $H_2(\partial M; \mathbb{F}_2) \rightarrow H_2(M; \mathbb{F}_2)$ is onto.*

Thus a closed M contains only orientable surfaces if and only if $H_2(M; \mathbb{F}_2) = 0$. Using the long exact sequence of the pair, you can check that the homological condition in Proposition 2.4 is equivalent to $\dim H_1(M; \mathbb{F}_2) = \frac{1}{2} \dim H_1(\partial M; \mathbb{F}_2)$.

Proof of Proposition 2.4 It suffices to consider the case when M is connected. First, note that any closed surface F in M gives a class in $H_2(M; \mathbb{F}_2)$. Moreover, any c in $H_2(M; \mathbb{F}_2)$ can be represented by a closed surface F that is connected (by adding tubes between components if needed) and nonempty (by adding a sphere bounding a ball if $c = 0$). In the rest of this proof, all surfaces will be connected, nonempty, and embedded in M .

As M is orientable, any nonorientable surface F is nonseparating. Also, given a nonseparating orientable surface F we can build a nonorientable surface as follows: take an embedded arc α in M that meets F only at its endpoints and goes from one side of F to the other; attaching a tube to F along α gives the desired nonorientable surface. Thus every closed surface F in M is orientable if and only if every closed surface is separating. So we will prove that the homological hypotheses of the proposition are equivalent to every closed surface in M being separating. When M is closed, the proposition is now immediate since a closed surface F is 0 in $H_2(M; \mathbb{F}_2)$ if and only if it is separating.

To prove the proposition when M has boundary, it suffices to show that the class $[F]$ of a closed surface F is in the image of $H_2(\partial M; \mathbb{F}_2)$ if and only if F is separating. If F is separating, then F divides M into two pieces A and B and we have $[F] = [A \cap \partial M] = [B \cap \partial M]$, so $[F]$ comes from $H_2(\partial M; \mathbb{F}_2)$ as claimed. If instead F is nonseparating, let γ be a loop disjoint from ∂M that meets F in a single point; hence the homology intersection pairing $H_2(M; \mathbb{F}_2) \times H_1(M; \mathbb{F}_2) \rightarrow \mathbb{F}_2$ has $[F] \cdot [\gamma] = 1$. As any $c \in H_2(\partial M; \mathbb{F}_2)$ has $c \cdot [\gamma] = 0$, it follows that $[F]$ does not come from $H_2(\partial M; \mathbb{F}_2)$. So we have characterized which F give classes coming from $H_2(\partial M; \mathbb{F}_2)$, completing the proof. \square

2.5 Triangulations

A *triangulation* of a compact 3-manifold is a cell complex made from finitely many tetrahedra by gluing some of their 2-dimensional faces in pairs via orientation-reversing affine maps so that the link of every vertex is either a sphere or a disc. (For such face gluings, the link condition is equivalent to the complex being a 3-manifold, see e.g. [62, Prop. 3.2.7].) In particular, a triangulation is not necessarily a simplicial complex, but rather what is sometimes called a semi-simplicial, pseudo-simplicial, or singular triangulation.

An *ideal triangulation* of a compact 3-manifold with nonempty boundary is a cell complex \mathcal{T} made out of finitely many tetrahedra by gluing *all* of their

2-dimensional faces in pairs as above with no conditions on the vertex links. Here, the manifold M being triangulated is not the underlying space of \mathcal{T} but rather the subset of it gotten by removing a small regular neighborhood of each vertex. Put another way, the manifold M is what you get by gluing together *truncated* tetrahedra in the corresponding pattern. Hence M will be a compact 3-manifold with nonempty boundary, and $\mathcal{T} \setminus \mathcal{T}^0$ is homeomorphic to the interior of M , where \mathcal{T}^i denotes the i -skeleton of \mathcal{T} . See e.g. [63] for more background on ideal triangulations.

We will work with both kinds of triangulations in this paper and will sometimes refer to the first kind as *finite triangulations* for clarity.

2.6 Normal surfaces

Our conventions and notation for normal surfaces closely follow [64, §2], which the reader should consult for additional details beyond the sketch we give here. Throughout, we consider a fixed triangulation \mathcal{T} of a compact 3-manifold M , which can be either finite or ideal. However, in the ideal case, we only consider closed normal surfaces, not the spun-normal ones of [63, 67]. An *elementary disk* E in a tetrahedron Δ is a disk meeting each face of $\partial\Delta$ in either a straight line or the empty set; note ∂E is determined by $E \cap \Delta^1$, and [64, pg. 1089] gives a convention so that the interior of E is determined by $E \cap \Delta^1$ as well. A surface F in M is *normal* when it is in general position with the skeleta of \mathcal{T} and meets each tetrahedron of \mathcal{T} in elementary disks. A normal surface F is completely determined by $F \cap \mathcal{T}^1$. A *normal isotopy* of M is one that leaves every simplex in \mathcal{T} invariant. The normal isotopy classes of elementary disks in a tetrahedron Δ are called the *disk types*, of which there are seven: three kinds of triangles and four kinds of quadrilaterals (or quads for short). Fixing an ordering of the t tetrahedra in \mathcal{T} and the seven disk types, a normal surface F gives a tuple $\vec{F} \in \mathbb{N}^{7t}$ by counting the number of occurrences of each disk type; these are called the *normal coordinates* of F , or more precisely the triangle-quad normal coordinates. Note that the vector \vec{F} determines F up to normal isotopy.

The coordinates of \vec{F} satisfy a system of homogenous linear equations, called the *matching equations* in [64], one for each arc type in a face of \mathcal{T}^2 . In the vector space \mathbb{R}^{7t} , the intersection of the solutions to the matching equations with the positive orthant gives a polyhedral cone $\mathcal{S}_{\mathcal{T}}$ called the *normal solution space*. A vector $\vec{x} \in \mathcal{S}_{\mathcal{T}}$ is *admissible* when for every tetrahedron of \mathcal{T} there is at most one quad coordinate of \vec{x} that is nonzero. The points in $\mathcal{S}_{\mathcal{T}}$ corresponding to normal surfaces are precisely the admissible integral points.

A key property of a normal surface F is its *weight* $\text{wt}(F)$ which is the number of times it intersects \mathcal{T}^1 and can be viewed as its combinatorial area. This notion of weight extends to a linear function $\text{wt}: \mathbb{R}^{7t} \rightarrow \mathbb{R}$ as follows.

For an elementary disk E_i corresponding to coordinate i , each vertex of E_i is incident on an edge of T^1 ; take c_i to be the sum of the reciprocals of the valences of those edges. Defining $\text{wt}(\vec{x}) = \sum_i c_i x_i$, we have $\text{wt}(F) = \text{wt}(\vec{F})$ for every normal surface F .

The *projective solution space* \mathcal{P}_T for T is abstractly the quotient of $\mathcal{S}_T \setminus \{0\}$ modulo positive scaling. It is useful to concretely identify \mathcal{P}_T with a subset of \mathcal{S}_T , and here [64] uses the points of \mathcal{S}_T whose coordinates sum to 1. However, we instead use the convention that $\mathcal{P}_T = \{\vec{x} \in \mathcal{S}_T \mid \text{wt}(\vec{x}) = 1\}$ as this simplifies the statement of a key result of [64]. We will use $\vec{F}^* = (1/\text{wt}(\vec{F}))\vec{F}$ to denote the projectivization of \vec{F} and call it the *projective normal class* of F .

The *carrier* C_F of a normal surface F is the unique minimal face of \mathcal{P}_T containing \vec{F}^* . The faces of \mathcal{S}_T and hence \mathcal{P}_T correspond to having some of the defining inequalities $x_i \geq 0$ become equalities. Thus the carrier C_F is the face of \mathcal{P}_T cut out by the requirement that $x_i = 0$ whenever $F_i = 0$.

If normal surfaces F and G are *compatible* in the sense that they never have distinct quad types in a single tetrahedron, then they have a natural ‘‘cut and paste’’ geometric sum that is also a normal surface. This new surface is called their *normal sum* and denoted $F + G$. Its normal coordinates are $\vec{F} + \vec{G}$ and in particular the normal sum is determined up to normal isotopy by the normal isotopy classes of F and G , even though $F \cap G$ can change under normal isotopy of the surfaces individually.

2.7 Short generating functions and quasi-polynomials

Throughout this subsection, see Chapter 4 of [59] for details and further background. We can encode a function $s : \mathbb{N} \rightarrow \mathbb{Q}$ by its *generating function* $S(x) = \sum_{n=0}^\infty s(n)x^n$ in the ring $\mathbb{Q}[[x]]$ of formal power series. This generating function is *short* when $S(x) = P(x)/Q(x)$ for polynomials P and Q in $\mathbb{Q}[x]$ where Q is a product of cyclotomic polynomials. Equivalently, the generating function is short if and only if

$$S(x) = \sum_{i=1}^k \frac{c_i x^{a_i}}{(1 - x^{b_i})^{d_i}} \quad \text{for some } c_i \in \mathbb{Q} \quad \text{and} \quad a_i, b_i, d_i \in \mathbb{N}.$$

If s has a short generating function $S = P/Q$ where further $\deg P < \deg Q$, then we say that s is a *quasi-polynomial*. Equivalently, a function $s : \mathbb{N} \rightarrow \mathbb{Q}$ is quasi-polynomial if and only if there exists $L \in \mathbb{N}$ and polynomials $f_0, f_1, \dots, f_{L-1} \in \mathbb{Q}[x]$ such that $s(n) = f_k(n)$ if $n \equiv k \pmod L$, see Proposition 4.4.1 of [59]. When s has a short generating function, it is equal to a fixed quasi-polynomial except for finitely many inputs [59, Proposition 4.2.2].

We will be interested exclusively in s where $s(n) \in \mathbb{N}$ for all n . When such an s has a short generating function $S(x) = P(x)/Q(x)$, where $Q \in \mathbb{Z}[x]$ is a product of cyclotomic polynomials, then P must also be in $\mathbb{Z}[x]$; this is because $P(x) = S(x)Q(x)$ as elements of $\mathbb{Q}[[x]]$ and $S(x)Q(x)$ is a product of elements in $\mathbb{Z}[[x]]$.

We end this section with the lemma that gives Corollary 1.6 from Theorem 1.3:

2.8 Lemma *Suppose $s: \mathbb{N} \rightarrow \mathbb{Q}$ with all $s(n) \geq 0$ has a short generating function and consider $\bar{s}(n) = \sum_{k=0}^n s(k)$. Then either $s(n) = 0$ for all large n or there exists $d \in \mathbb{N}$ and $c > 0$ in \mathbb{Q} such that $\bar{s}(n) \sim cn^d$.*

Proof Since we only care about asymptotics, assume that s is a quasi-polynomial with $f_0, f_1, \dots, f_{L-1} \in \mathbb{Q}[x]$ where $s(n) = f_\ell(n)$ if $n \equiv \ell \pmod L$. Assume some $f_\ell \neq 0$ as otherwise we are done. Set $e = \max(\deg f_\ell)$, which is at least 0, and let c_ℓ be the coefficient on x^e in f_ℓ , so that $f_\ell(n) = c_\ell n^e + O(n^{e-1})$. Then as $s(n) \geq 0$ for all n it follows that $c_\ell > 0$ if $\deg f_\ell = e$ and otherwise $c_\ell = 0$; in particular, all $c_\ell \geq 0$ and $\sum_\ell c_\ell > 0$. Separating the sum in $\bar{s}(n)$ into congruence classes modulo L , we write

$$\bar{s}(n) = \sum_{\ell=0}^{L-1} \bar{s}^{(\ell)}(n) \quad \text{where} \quad \bar{s}^{(\ell)}(n) = \sum_{\substack{j=0 \\ j \equiv \ell \pmod L}}^n s(j) = \sum_{k=0}^{\lfloor (n-\ell)/L \rfloor} f_\ell(\ell + Lk) \tag{2.9}$$

Using that $(\ell + Lk)^e$ is a polynomial in k with leading term $L^e k^e$, we get $f_\ell(\ell + Lk) = c_\ell(\ell + Lk)^e + O(n^{e-1}) = c_\ell L^e k^e + O(n^{e-1})$ where $n = \ell + Lk$. Thus

$$\begin{aligned} \bar{s}^{(\ell)}(n) &= \sum_{k=0}^{\lfloor (n-\ell)/L \rfloor} (c_\ell L^e k^e + O(n^{e-1})) = c_\ell L^e \left(\sum_{k=0}^{\lfloor (n-\ell)/L \rfloor} k^e \right) + O(n^e) \\ &= \frac{c_\ell L^e}{e+1} \left[\frac{n-\ell}{L} \right]^{e+1} + O(n^e) = \frac{c_\ell}{(e+1)L} n^{e+1} + O(n^e) \end{aligned}$$

where we have used $\sum_{k=0}^m k^e = \frac{m^{e+1}}{e+1} + O(m^e)$. Set $d = e + 1$ and $c = \frac{1}{dL} \sum_\ell c_\ell > 0$, and it now follows from (2.9) that $\bar{s}(n) \sim cn^d$ as required. \square

3 Isotopy classes of essential normal surfaces

In this section, we discuss and refine Tollefson’s work on isotopy classes of incompressible surfaces from the point of view of normal surface theory. In

particular, this allows us to build a bijection between isotopy classes of such surfaces and certain equivalence classes of lattice points in a collection of rational cones. We will use this framework to prove Theorem 1.3 in Sect. 4. We begin by explaining some key facts from Tollefson [64] using the notation that we reviewed in Sect. 2.6. Throughout, we consider a compact orientable irreducible ∂ -irreducible 3-manifold M equipped with a fixed finite triangulation \mathcal{T} .

A *lw-surface* is a compact orientable incompressible ∂ -incompressible normal surface that is least weight among all normal surfaces in its isotopy class; such surfaces play a key role in [64]. (The term lw-surface is not actually used in [64] but makes its results easier to state.) A face C of $\mathcal{P}_{\mathcal{T}}$ is a *lw-face* when every orientable normal surface carried by C is a lw-surface. We use $\mathcal{LW}_{\mathcal{T}}$ to denote the set of all lw-faces of $\mathcal{P}_{\mathcal{T}}$. Clearly, $\mathcal{LW}_{\mathcal{T}}$ is a subcomplex of $\mathcal{P}_{\mathcal{T}}$. A lw-face C is *complete* if whenever it carries an orientable normal surface F it also carries every lw-surface isotopic to F . A key fact for us is:

3.1 Theorem [64, Theorem 4.5] *Every lw-surface is carried by a complete lw-face. In particular, any lw-face is contained in some complete lw-face.*

On a complete lw-face, Tollefson characterizes the various possible forms for isotopy relations among the surfaces that it carries. As we will describe, these have to be relatively simple on the interior C° of C , but proper faces of C can have different isotopy relations. Tollefson introduces the notion of a *PIC-partition* to encode all of these. We will not work with PIC-partitions directly, but reframe the underlying structure in a way more suited for the proof of Theorem 1.3. To give our structure theorem, we first need some definitions and a useful characterization of when a face of $\mathcal{LW}_{\mathcal{T}}$ is complete.

If F is an orientable surface in M and m a positive integer, a disjoint union of m parallel copies of F is called a *multiple* of F and denoted mF . When F is normal, we always take mF to be a normal surface whose normal coordinates are $m\vec{F}$. To mirror what happens algebraically in the normal case, for a nonorientable surface F one defines $2F$ as the boundary G of a regular neighborhood of F and then mF as either $\frac{m}{2}G$ or $F \cup \frac{m-1}{2}G$ depending on the parity of m . Surfaces F and G are *projectively isotopic* when they have multiples that are isotopic. We say two normal surfaces are *projectively normally isotopic* if they have multiples that are normally isotopic. Note here that the admissible rational points of $\mathcal{P}_{\mathcal{T}}$ correspond exactly to projective normal isotopy classes of normal surfaces.

3.2 Remark Our definitions of least-weight and completeness for a face C differ from those in [64] in that we only look at *orientable* normal surfaces F carried by C whereas [64] allows nonorientable F . However, it is easy to see our definitions are equivalent to the originals. For example, if C is a lw-face with our definition and F is a nonorientable surface carried by C , then $2F$

is a lw-surface and hence F itself is incompressible and ∂ -incompressible. Moreover, if G is any normal surface isotopic to F then $2G$ is isotopic to $2F$ and hence $\text{wt}(2G) \geq \text{wt}(2F)$ which implies $\text{wt}(G) \geq \text{wt}(F)$; thus F is least weight among all normal surfaces in its isotopy class. The equivalence of the two definitions of completeness is similar, using that if C carries $2G$ then it carries G .

Important for us throughout this paper is that whether a face is (complete) least-weight is determined by any one surface carried by its interior:

3.3 Theorem *Suppose F is an orientable normal surface carried by the interior of a face C of $\mathcal{P}_{\mathcal{T}}$. Then the following are equivalent:*

- (a) C is a lw-face.
- (b) F is a lw-surface.
- (c) Every connected component of F is a lw-surface.

If C is a lw-face, the following are equivalent:

- (d) C is complete.
- (e) C carries every lw-surface isotopic to F .
- (f) C carries every lw-surface isotopic to a connected component of F .

We will prove Theorem 3.3 below in Sect. 3.10.

3.4 Dependent faces

A face D of a lw-face C is C -dependent if there exists a surface carried by D that is projectively isotopic to one carried by C° ; otherwise, the face D is C -independent. The collection of C -independent faces of C clearly forms a subcomplex \mathcal{D} of ∂C and we define $\text{dep}(C)$ to be $C \setminus \bigcup_{D \in \mathcal{D}} D$. Note that if D is a C -dependent face of C , then $D^\circ \subset \text{dep}(C)$ since if any $\vec{x} \in D^\circ$ was in a C -independent face E then D would be a face of E , contradicting that D is C -dependent. As any point of C is in the interior of some face, we see that $\text{dep}(C)$ is also the union of D° over all C -dependent faces D of C .

Tollefson completely characterized the isotopy relations among the surfaces carried by each $\text{dep}(C)$. We rework this as:

3.5 Theorem *For each face C of $\mathcal{LW}_{\mathcal{T}}$ there is a rational linear subspace W_C such that the following holds. Any two surfaces F and G carried by $\text{dep}(C)$ are projectively isotopic if and only if $\vec{F}^* - \vec{G}^*$ is in W_C . Moreover, if F and G are orientable then they are isotopic if and only if $\vec{F} - \vec{G}$ is in W_C . Also,*

$$\text{dep}(C) = \{ \vec{x} \in C \mid \vec{x} + W_C \text{ meets } C^\circ \} \tag{3.6}$$

so that in particular any F carried by $\text{dep}(C)$ is projectively isotopic to one carried by C° . Finally, the subspace W_C is contained in $\ker(\text{wt})$ and given

any F carried by C° there exist surfaces $\vec{F}_1, \dots, \vec{F}_k$ projectively isotopic to F and carried by C° such that the $\vec{F}^* - \vec{F}_i^*$ span W_C .

The example in Sect. 7.4 may help you understand the statement of Theorem 3.5.

For the practical algorithms in Sect. 6, we will need the following additional properties of W_C :

3.7 Corollary *Suppose C is a face of \mathcal{LW}_T . If surfaces F and G carried by C are projectively isotopic, then $\vec{F}^* - \vec{G}^* \in W_C$. Also, if D is a face of C , then $W_D \subset W_C$. Finally, if F is any orientable surface carried by C° , then W_C is spanned by all $\vec{G} - \vec{H}$ where G is a connected component of F and H is isotopic to G and carried by C .*

Combined with Theorem 3.5, the next result will be key to proving Theorem 1.3:

3.8 Theorem *The complex \mathcal{LW}_T is the disjoint union of the $\text{dep}(C)$ as C ranges over the complete lw-faces of \mathcal{P}_T . Moreover, if C and C' are distinct complete lw-faces, then no surface carried by $\text{dep}(C)$ is projectively isotopic to one carried by $\text{dep}(C')$. Consequently, for an orientable incompressible ∂ -incompressible surface F , there is a unique complete lw-face C such that $\text{dep}(C)$ carries a surface (non-projectively) isotopic to F .*

3.9 Remark We defined $\mathcal{P}_T = \{\vec{x} \in \mathcal{S}_T \mid \text{wt}(\vec{x}) = 1\}$ rather than $\mathcal{P}'_T = \{\vec{x} \in \mathcal{S}_T \mid \sum x_i = 1\}$ in order to state Theorem 3.5 in the above form. Tollefson uses \mathcal{P}'_T , and ends up with a partition of $\text{dep}(C)$ along a family of typically *nonparallel* affine subspaces whereas our partition is along *parallel* affine subspaces. While both \mathcal{P}_T and \mathcal{P}'_T are projectivizations of \mathcal{S}_T , the map that identifies them is not affine but rather projective and so this is not a contradiction.

3.10 Complete lw-faces in detail

We begin with the proof of Theorem 3.3 as it is needed to prove Theorem 3.8.

Proof of Theorem 3.3 First, recall that faces of \mathcal{P}_T are defined by setting a subset of the normal coordinates to 0. Consequently, a normal surface G is carried by C if and only if every connected component of it is carried by C . More generally, if K and L are compatible normal surfaces, then C carries $K + L$ if and only if it carries K and L individually.

A normal surface F is carried by C° if and only if the carrier of F is equal to C ; the equivalence of (a) and (b) is thus Theorem 4.2 of [64]. From the definition it is clear that (c) implies (b), so to complete the proof of the first part of the theorem we will show (a) implies (c). This holds because if F' is a

component of F then, as noted above, C carries F' and thus F' is a lw-surface as C is a lw-face.

For the second part, by definition (d) implies (e), and (d) implies (f) since every component of F is also carried by C . Since C carries a surface if and only if it carries all of its components, we see that (f) easily gives (e). So it remains to prove (e) implies (d). So suppose C is a least-weight face of $\mathcal{P}_{\mathcal{T}}$ such that every lw-surface projectively isotopic to F is carried by C . We must show that C is complete, so suppose K is a lw-surface carried by C and L is a lw-surface isotopic to K . As $\vec{F}^* \in C^\circ$, we can pick a lw-surface E with $\vec{E}^* \in C^\circ$ and \vec{F}^* in the interior of the line segment joining \vec{K}^* to \vec{E}^* . Then there are positive integers $\{m, a, b\}$ with $mF = aK + bE$. Applying Corollary 4.3 of [64] with $G = aK$, $G' = aL$, and $H = H' = bE$, we conclude that aL and bE are compatible and that $aL + bE$ is isotopic to mF . By hypothesis, as $aL + bE$ is projectively isotopic to F , it is carried by C . Then aL is carried by C and hence L is carried by C as well as $a > 0$. Hence C is complete as claimed. \square

We now turn to the proof of Theorem 3.5 for which we will need:

3.11 Lemma *Suppose C is a compact convex polyhedron in \mathbb{R}^n and W a subspace of \mathbb{R}^n . Set $\text{dep}(C, W) = \{\vec{x} \in C \mid \vec{x} + W \text{ meets } C^\circ\}$. If D is a face of C , then the intersection $D \cap \text{dep}(C, W)$ is either empty or contains D° .*

Proof Passing to a subspace if necessary, we assume that $\dim C = n$ and hence C° is open in \mathbb{R}^n . There are finitely many nonzero linear functionals ℓ_i on \mathbb{R}^n , say indexed by a set I , and $\alpha_i \in \mathbb{R}$, such that $C = \{\vec{x} \in \mathbb{R}^n \mid \ell_i(\vec{x}) \geq \alpha_i \text{ for all } i \in I\}$. Then $C^\circ = \{\vec{x} \in \mathbb{R}^n \mid \ell_i(\vec{x}) > \alpha_i \text{ for all } i \in I\}$. For a face D of C , define I_D to be the indices in I where $\ell_i(\vec{x}) = \alpha_i$ on all of D .

Now assume $D \cap \text{dep}(C, W)$ is nonempty, and pick $\vec{x} \in D$ and $\vec{w} \in W$ with $\vec{x} + \vec{w}$ in C° . For any $i \in I$, we have $\ell_i(\vec{x} + \vec{w}) > \alpha_i$, which for those $i \in I_D$ implies $\ell_i(\vec{w}) > 0$ since $\ell_i(\vec{x}) = \alpha_i$. Given \vec{y} in $D^\circ = \{\vec{x} \in D \mid \ell_i(\vec{x}) > \alpha_i \text{ for all } i \notin I_D\}$, we need to show that it is in $\text{dep}(C, W)$. For $\epsilon > 0$, consider $\vec{v} = \vec{y} + \epsilon\vec{w}$. For $i \notin I_D$, we have $\ell_i(\vec{v}) = \ell_i(\vec{y}) + \epsilon\ell_i(\vec{w})$; since $\ell_i(\vec{y}) > \alpha_i$, we can thus make $\ell_i(\vec{v}) > \alpha_i$ as well by choosing ϵ small enough. On the other hand, for $i \in I_D$ we have $\ell_i(\vec{v}) = \alpha_i + \epsilon\ell_i(\vec{w}) > \alpha_i$ for any positive ϵ as $\ell_i(\vec{w}) > 0$ for such i . Thus \vec{v} is in C° for small ϵ and so $\vec{y} \in \text{dep}(C, W)$ as needed. \square

Proof of Theorem 3.5 This result is essentially a reframing of Theorem 5.5 of [64] on the existence of a PIC-partition for C , but to see this one must use a number of details from the proof of that theorem. Hence we will simply prove Theorem 3.5 directly relying on results earlier in that paper. Suppose F is any lw-surface carried by C° . Let V_F be the subspace of \mathbb{R}^{7t} spanned by all \vec{G} where G is carried by C and projectively isotopic to F . Using that V_F is finite-dimensional, we can find mutually isotopic lw-surfaces F_1, \dots, F_k carried by

C , each projectively isotopic to F with $\vec{F}_1^* = \vec{F}^*$, such that $\{\vec{F}_1, \dots, \vec{F}_k\}$ is a basis for V_F . Then Theorem 5.3 of [64] implies that every normal surface carried by $A_F = C \cap V_F$ is projectively isotopic to F .

Consider the affine subspace $X = \{\vec{x} \in \mathbb{R}^{7t} \mid \text{wt}(\vec{x}) = 1\}$ and note $\mathcal{P}_{\mathcal{T}} = \mathcal{S}_{\mathcal{T}} \cap X$ where $\mathcal{S}_{\mathcal{T}}$ is the normal solution space. Define $X_F = X \cap V_F$ which is also the smallest affine subspace containing $\{\vec{F}_1^*, \dots, \vec{F}_k^*\}$. Note here that A_F is also $C \cap X_F$. If \vec{G} is another orientable normal surface carried by C° , then either $A_F = A_G$ or $A_F \cap A_G = \emptyset$ depending on whether or not F and G are projectively isotopic. When A_F and A_G are disjoint, we claim that X_F and X_G are still parallel; formally, the tangent space to an affine subspace $Y \subset \mathbb{R}^{7t}$ is $TY = \{\vec{y}_1 - \vec{y}_2 \mid \vec{y}_1, \vec{y}_2 \in Y\}$ and we will show $TX_F = TX_G$.

As $\vec{G}^* \in C^\circ$, we can find a lw-surface H with $\vec{H}^* \in C^\circ$ and \vec{G}^* on the interior of the line segment between \vec{F}_1^* and \vec{H}^* . Hence there are positive integers $\{m, a, b\}$ such that $m\vec{G}^* = a\vec{F}_1^* + b\vec{H}^*$. Set $G_i = aF_i + bH$. By Corollary 4.3 of [64], all the G_i are isotopic to $G_1 = mG$ and hence lie in V_G . The F_i are isotopic lw-surfaces and so have the same weight, and consequently so do the G_i and hence

$$\vec{F}_i^* - \vec{F}_j^* = \frac{\text{wt}(G_1)}{a \cdot \text{wt}(F_1)} (\vec{G}_i^* - \vec{G}_j^*) \quad \text{for all } i, j.$$

In particular, we have $TX_F \subset TX_G$. Reversing the roles of F and G shows $TX_G \subset TX_F$ as claimed.

Now set $W_C = TX_F$ for any orientable F carried by C° . Note that W_C is spanned by the $\vec{F}_i - \vec{F}_j = \text{wt}(F_1)(\vec{F}_i^* - \vec{F}_j^*)$ from above, and hence by the $\vec{F}^* - \vec{F}_i^*$ since $\vec{F}_1^* = \vec{F}^*$. As F is arbitrary, this verifies the claims in the last sentence of the statement of the theorem since in addition each $\vec{F}_i - \vec{F}_j$ is in $\ker(\text{wt})$.

We extend our notion of A_F to all $\vec{y} \in C^\circ$ by setting $A_{\vec{y}} = (\vec{y} + W_C) \cap C$. Let $\tilde{A} = \bigcup_{\vec{y} \in C^\circ} A_{\vec{y}}$. We will show:

3.12 Claim $\tilde{A} = \text{dep}(C)$.

Before proving the claim, let us show that Theorem 3.5 follows from it. First, the Eq. (3.6) holds since \tilde{A} is also $\{\vec{x} \in C \mid \vec{x} + W_C \text{ meets } C^\circ\}$. Second, the claim that surfaces F and G carried by \tilde{A} are projectively isotopic if and only if $\vec{F} - \vec{G} \in W_C$ follows because \tilde{A} is partitioned by the $A_{\vec{y}}$ which for rational \vec{y} correspond exactly to projective isotopy classes of surfaces carried by C . Finally, if F and G are orientable surfaces carried by $\text{dep}(C)$, we need to show they are isotopic if and only if $\vec{F} - \vec{G} \in W_C$. If they are isotopic, they must have the same weight and so $\vec{F} - \vec{G}$ is a multiple of $\vec{F}^* - \vec{G}^*$, and the latter must be in W_C as F and G are projectively isotopic. Conversely, if $\vec{F} - \vec{G} \in W_C$, the surfaces are projectively isotopic and as $W_C \subset \ker(\text{wt})$ it

follows $\text{wt}(F) = \text{wt}(G)$. As F and G are orientable, least weight, of the same weight, and projectively isotopic, they are actually isotopic as needed.

To prove Claim 3.12, note both sets contain C° , so for each face D of ∂C we will check that both sets agree on D° . By Lemma 3.11, either $\tilde{A} \cap D$ is empty or it contains D° . If the former, then no surface carried by D can be projectively isotopic to one carried by C° , and so D is C -independent and hence $\text{dep}(C) \cap D = \emptyset$ as well. If the latter, then any rational point in D° gives a surface projectively isotopic to one carried by C° ; hence, D is C -dependent. As noted in Sect. 3.4, this implies $D^\circ \subset \text{dep}(C)$. This proves the claim and hence the theorem. \square

Proof of Corollary 3.7 First, suppose F and G are projectively isotopic and carried by C . Since this does not change $\vec{F}^* - \vec{G}^*$, we will assume F and G are orientable and actually isotopic. Let H be an orientable surface carried by C° . Then by Corollary 4.3 of [64], we have $H + F$ and $H + G$ are isotopic, and hence by Theorem 3.5 above we have $\overrightarrow{(H + F)} - \overrightarrow{(H + G)} = \vec{F} - \vec{G} = \text{wt}(F)(\vec{F}^* - \vec{G}^*)$ is in W_C as needed.

Second, if D is a face of C , then by the last part of Theorem 3.5 we have W_D is spanned by certain $\vec{F}^* - \vec{F}_i^*$ where F and F_i are carried by D . By what we just showed, all of these are in W_C as well, proving $W_D \subset W_C$.

Finally, fix an orientable surface F carried by C° and define Z to be the span of all $\vec{G} - \vec{H}$ where G is a connected component of F and H is isotopic to G and carried by C . We need to show $W_C = Z$. By the first part of this corollary, we know $Z \subset W_C$. From Theorem 3.5, there are surfaces F_1, \dots, F_k projectively isotopic to F where the $\vec{F}^* - \vec{F}_i^*$ span W_C . We can moreover arrange that each F_i is isotopic to F so that the $\vec{F} - \vec{F}_i$ span W_C . To see $\vec{F} - \vec{F}_i$ is in Z , let G_1, \dots, G_n be the connected components of F . Under an isotopy between F and F_i , let G'_j be the connected component of F_i corresponding to G_j . Then $\vec{F} - \vec{F}_i = \sum (\vec{G}_j - \vec{G}'_j)$ which is in Z , giving $W_C = Z$ and completing the proof of the corollary. \square

Turning now to the proof of Theorem 3.8, we begin with a lemma:

3.13 Lemma *Suppose C is a complete lw-face. A maximal C -independent face D of C is also complete.*

Proof Pick a lw-surface F carried by D° . By Theorem 3.3, it suffices to show that given a lw-surface G isotopic to F then G is carried by D . By completeness of C , we know G is carried by C . By Theorem 5.3 of [64], every normal surface carried by the segment $L = [\vec{F}^*, \vec{G}^*]$ in C is projectively isotopic to F . Let E be the minimal face of C containing L ; since L is just a segment, it meets E° . We cannot have E be C as then F is projectively isotopic to some surface in C° , violating that D is C -independent. For the same reason, the face E cannot

be C -dependent as then $E^\circ \subset \text{dep}(C)$ and hence by Theorem 3.5 any surface carried by E° is projectively isotopic to one carried by C° . Thus E must be C -independent and we know that it contains \vec{F}^* which is an interior point of the maximal C -independent face D ; consequently, we must have $E = D$ and so G is carried by D . Thus D is complete as claimed. \square

Proof of Theorem 3.8 We start with:

3.14 Claim Any lw-surface F is carried by $\text{dep}(C)$ for some complete lw-face C .

By Theorem 3.1, the surface F is carried by some complete lw-face. It is immediate from the definition that the intersection of two complete lw-faces is again complete, so there exists a minimal complete lw-face C carrying F . Let D be the face of C containing F in its interior. If D is C -dependent, then $D^\circ \subset \text{dep}(C)$ and so $F \in \text{dep}(C)$ as desired. So assume D is C -independent. Let E be a maximal C -independent face of C containing D , and note $E \neq C$ as C is C -dependent. By Lemma 3.13, the face E is complete and so we have found a smaller complete face containing F than C , a contradiction. So we have proven Claim 3.14.

To prove Theorem 3.8 it remains to show:

3.15 Claim If C_1 and C_2 are distinct (but perhaps not disjoint) complete lw-faces, then no surface carried by $\text{dep}(C_1)$ is projectively isotopic to one carried by $\text{dep}(C_2)$. In particular, the sets $\text{dep}(C_1)$ and $\text{dep}(C_2)$ are disjoint.

Suppose not and that F_1 and F_2 are projectively isotopic normal surfaces carried by $\text{dep}(C_1)$ and $\text{dep}(C_2)$ respectively. By Theorem 3.5, we can further assume each F_i is carried by C_i° . Replacing them with multiples if necessary, we can assume that they are actually isotopic. By Corollary 4.6 of [64], it follows that both F_1 and F_2 must be carried by $C_1 \cap C_2$; as each F_i is carried by C_i° , we must have $C_1 = C_2$, a contradiction. This proves Claim 3.15 and hence the theorem. \square

4 Surface counts are almost quasi-polynomial

The first of this section’s two main results is:

4.1 Theorem Suppose M is a closed irreducible atoroidal 3-manifold that contains no nonorientable essential surfaces and C is a complete lw-face of \mathcal{P}_T . Let $b_C(n)$ be the number of isotopy classes of closed essential surfaces F carried by $\text{dep}(C)$ with $\chi(F) = n$, and let $B_C(x) = \sum_{n=1}^\infty b_C(-2n)x^n$ be the corresponding generating function. Then $B_C(x)$ is short.

The other main result of this section is Theorem 4.12, which is the analog of Theorem 4.1 when M has boundary. Combining Theorem 4.1 with the results from the last section, we can now give:

Proof of Theorem 1.3 when M is closed As M is closed, a (necessarily closed) surface in M is essential exactly when it is incompressible. Therefore, by Theorem 3.8, each isotopy class of essential surface is carried by $\text{dep}(C)$ for a unique complete lw-face C . As sums of short generating functions are also short, Theorem 1.3 now follows from Theorem 4.1. \square

4.2 Counting surfaces via lattice points

We now turn to the proof of Theorem 4.1, so let M be a closed irreducible atoroidal 3-manifold with triangulation \mathcal{T} . From now on, fix a complete lw-face C of $\mathcal{P}_{\mathcal{T}}$ and consider the cone $\tilde{C} \subset \mathcal{S}_{\mathcal{T}}$, that is $\mathbb{R}_{\geq 0} \cdot C = \{t\vec{x} \mid t \in \mathbb{R}_{\geq 0}, \vec{x} \in C\}$. Recall that $\text{dep}(C) \subset C$ is the complement of its C -independent faces, and define $\text{dep}(\tilde{C}) = (\mathbb{R}_{\geq 0} \cdot \text{dep}(C)) \setminus \{\vec{0}\}$. Now $b_C(n)$ in Theorem 4.1 is the number of isotopy classes of normal surfaces F with $\vec{F} \in \text{dep}(\tilde{C})$ and $\chi(F) = n$.

As motivation, let us start with the easy case when the subspace W_C from Theorem 3.5 is zero. Then $\text{dep}(\tilde{C}) = \tilde{C}^\circ$, and two surfaces F and G carried by $\text{dep}(\tilde{C})$ are isotopic if and only if $\vec{F} = \vec{G}$. In our triangle-quad coordinates, there is a linear function $\chi : \mathbb{R}^{7t} \rightarrow \mathbb{R}$ such that the Euler characteristic of a normal surface F is given by $\chi(\vec{F})$, see [42, Algorithm 9.1]. Every normal surface F carried by C is incompressible and hence $\chi(F) \leq -1$ as M is closed, irreducible, and atoroidal. Thus $\chi < 0$ on every vertex of C which implies χ is proper on \tilde{C} and so the set $X = \{\vec{x} \in \tilde{C} \mid \chi(x) = -1\}$ is a compact polyhedron. Now $b_C(-n)$ is simply the size of the set $(nX^\circ) \cap \mathbb{Z}^{7t}$, and counting lattice points in dilations of a compact polyhedron has been studied extensively starting with the work of Ehrhart in the 1960s. In particular, Theorem 4.6.26 of [59], whose proof uses Ehrhart-Macdonald reciprocity, tells us that the generating function $B_C(x)$ is short, proving Theorem 4.1 when $W_C = 0$.

When W_C is nonzero, to count isotopy classes of surfaces we need to identify lattice points in $\text{dep}(\tilde{C})$ that differ by an element of W_C . We do so in the following way. Let V be the linear subspace of \mathbb{R}^{7t} spanned by all vectors in C , and let W be W_C . Define $V(\mathbb{Z}) = V \cap \mathbb{Z}^{7t}$ and $W(\mathbb{Z}) = W \cap \mathbb{Z}^{7t}$. Using Smith normal form, we can find a complementary rational subspace $L \subset V$ to W such that the lattice $V(\mathbb{Z})$ is the direct sum of $W(\mathbb{Z})$ and $L(\mathbb{Z}) = L \cap \mathbb{Z}^{7t}$. Let $T : V \rightarrow L$ be the projection operator associated with the decomposition $V = W \oplus L$. We can now turn our question of counting isotopy classes of surfaces into one about counting certain lattice points in $L(\mathbb{Z})$:

4.3 Lemma *The set $T(\text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t})$ is in bijection with isotopy classes of normal surfaces carried by $\text{dep}(C)$.*

Proof Normal surfaces carried by $\text{dep}(C)$ correspond to lattice points in $\text{dep}(\tilde{C})$. As W is the kernel of T , the claim is equivalent to saying that if

F and G are normal surfaces with \vec{F} and \vec{G} in $\text{dep}(\tilde{C})$, then F is isotopic to G if and only if $\vec{F} - \vec{G} \in W$. As we are assuming that all incompressible surfaces in M are orientable, this follows immediately from Theorem 3.5. \square

To prove Theorem 4.1, we will need a tool for counting points in sets such as $T(\text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t})$. Recently, Nguyen and Pak [52], building on [19], established exactly the result we need here. To apply [52], we need the linear map T to be *integral* in the sense that its matrix with respect to any \mathbb{Z} -bases of $V(\mathbb{Z})$ and $L(\mathbb{Z})$ has integer entries, but that is clear from its definition. Since we want to count by Euler characteristic, we first study $\chi : V \rightarrow \mathbb{R}$:

4.4 Lemma *The restriction $\chi : V(\mathbb{Z}) \rightarrow \mathbb{R}$ is integral and, since M is closed, irreducible, and atoroidal, the function χ is proper on \tilde{C} and negative on $\tilde{C} \setminus \{\vec{0}\}$.*

Proof For $\chi|_V$, note that \tilde{C} has nonempty interior as a subset of V and contains open balls of arbitrary size. Hence, given any $\vec{v} \in V(\mathbb{Z})$, we can find $\vec{x}, \vec{y} \in \tilde{C}(\mathbb{Z})$ with $\vec{v} = \vec{x} - \vec{y}$. There are normal surfaces F and G with $\vec{F} = \vec{x}$ and $\vec{G} = \vec{y}$, and so $\chi(\vec{v}) = \chi(F) - \chi(G)$ is in \mathbb{Z} as needed to show $\chi|_V$ is integral.

For $\chi|_{\tilde{C}}$, every normal surface F carried by C is incompressible and hence $\chi(F) \leq -1$ as M is closed, irreducible, and atoroidal. Thus $\chi < 0$ on every vertex of C which implies it is proper on \tilde{C} and negative on $\tilde{C} \setminus \{\vec{0}\}$. \square

Now we combine χ and T as follows. Define $\bar{T} : V \rightarrow L \oplus \mathbb{R}$ by $\bar{T}(\vec{x}) = (T(\vec{x}), -\chi(\vec{x}))$, which is integral as both its component functions are, and we have:

4.5 Lemma *The set $\bar{T}(\text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t})$ is in bijection with isotopy classes of normal surfaces carried by $\text{dep}(C)$.*

Proof By Lemma 4.3, it suffices to show that projecting away the second factor of $L \oplus \mathbb{R}$ gives a bijection between $\bar{T}(\text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t})$ and $T(\text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t})$. This projection is clearly onto, so this reduces to showing that for normal surfaces F and G with \vec{F} and \vec{G} in $\text{dep}(C)$ and $T(\vec{F}) = T(\vec{G})$ then $-\chi(\vec{F}) = -\chi(\vec{G})$. But the latter holds since $\vec{F} - \vec{G} \in W$ implies the surfaces F and G must be isotopic by Theorem 3.5 and thus homeomorphic. \square

To apply [52], we will need one more property about T and \bar{T} :

4.6 Lemma *There is a lattice $L'(\mathbb{Z})$ containing $L(\mathbb{Z})$ which has a \mathbb{Z} -basis such that $T(\tilde{C}) \subset \mathbb{R}_{\geq 0}^d$ under the induced identification of L with \mathbb{R}^d . Moreover, the same holds for \bar{T} .*

Proof The claim for \bar{T} follows immediately from that for T since $-\chi(\tilde{C}) = [0, \infty)$ by Lemma 4.4. So now we consider only T .

Fourier-Motzkin elimination tells us that the image $T(\tilde{C})$ is again a polyhedral cone. We first show that $T(\tilde{C})$ is a *pointed cone*, that is, one that does

not contain a line. By Theorem 3.5, we know $W \subset \ker(\text{wt})$. Therefore, the map $\text{wt}: V \rightarrow \mathbb{R}$ factors through the projection $T: V \rightarrow L$. Hence all of $T(\tilde{C}) \setminus \{\vec{0}\}$ is strictly to the positive side of the hyperplane $(\text{wt}|_L)^{-1}(0)$ and so $T(\tilde{C})$ is a pointed cone.

We will now find a basis for $L(\mathbb{Q})$ as a \mathbb{Q} -vector space with the property that $T(\tilde{C})$ lies in the positive orthant; this suffices to prove the lemma as we can scale the basis elements by $a > 0$ in \mathbb{Q} so that the lattice they generate contains $L(\mathbb{Z})$. Let $\{\vec{v}_i\}$ denote the vertices of C . Note that if we can find a basis $\{\ell_j\}$ of $L(\mathbb{Q})^* = \text{Hom}(L(\mathbb{Q}), \mathbb{Q})$ where $\ell_j(\vec{v}_i) > 0$ for all i and j , then the algebraically dual basis \vec{e}_k of $L(\mathbb{Q})$, that is, the one where $\ell_j(\vec{e}_k) = \delta_{jk}$, is the basis we seek. Fix any basis $\{\beta_j\}$ of $L(\mathbb{Q})^*$ where $\beta_1 = \text{wt}$ and for $\epsilon \in \mathbb{Q}^\times$ consider the new basis $\{\ell_j\}$ where $\ell_1 = \beta_1$ and all other $\ell_j = \beta_1 + \epsilon\beta_j$. Since we showed above that $\beta_1(\vec{v}_i) = \text{wt}(\vec{v}_i) > 0$ for each i , for small enough ϵ we have $\ell_j(\vec{v}_i) > 0$ for all i and j as needed to prove the lemma. \square

Next, we introduce the language needed to state the conclusion of [52]. A set A of points in \mathbb{N}^n has an associated generating function:

$$f_A(\mathbf{t}) = \sum_{\vec{a} \in A} \mathbf{t}^{\vec{a}} \quad \text{in } \mathbb{Z}[[t_1, \dots, t_n]]$$

where $\mathbf{t}^{\vec{a}} = t_1^{a_1} \dots t_n^{a_n}$ for $\vec{a} = (a_1, \dots, a_n)$.

We say that A has a *short generating function* when there are $c_i \in \mathbb{Q}$ and $\vec{a}_i, \vec{b}_{ij} \in \mathbb{Z}^n$ such that:

$$f_A(\mathbf{t}) = \sum_{i=1}^N \frac{c_i \mathbf{t}^{\vec{a}_i}}{(1 - \mathbf{t}^{\vec{b}_{i1}}) \dots (1 - \mathbf{t}^{\vec{b}_{in}})}. \tag{4.7}$$

These multivariable short generating series were introduced by Barvinok and play a key role in polynomial time algorithms for counting lattice points in convex polyhedra [7].

Using the lattice $L'(\mathbb{Z}) \oplus \mathbb{Z} \subset L \oplus \mathbb{R}$, where $L'(\mathbb{Z})$ is from Lemma 4.6, we henceforth view $\bar{T}(\text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t})$ as subset of \mathbb{N}^{d+1} . The key to Theorem 4.1 is:

4.8 Lemma *The set $\bar{T}(\text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t})$ has a short generating function.*

Proof We will construct a rational polyhedron $Q \subset \tilde{C}$ such that

$$Q \cap \mathbb{Z}^{7t} = \text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t}.$$

By Lemma 4.6, we have $\bar{T}(Q) \subset \mathbb{R}_{\geq 0}^{d+1}$ using the basis of $L'(\mathbb{Z})$ given there. Additionally, the projection \bar{T} is integral with respect to the lattices $V(\mathbb{Z})$ and

$L'(\mathbb{Z})$ since $L'(\mathbb{Z}) \supset L(\mathbb{Z})$. Therefore, Theorem 1.1 of Nguyen–Pak [52] will apply and give that the generating function for $\bar{T}(Q \cap \mathbb{Z}^{7t})$ is short, proving the lemma.

Now $\text{dep}(\tilde{C})$ is simply \tilde{C} with some closed faces removed, and we can use the following standard trick to construct Q . For a face D of C its *active variables* are

$$I_D = \{i \in [1, 2, \dots, 7t] \mid x_i = 0 \text{ on } D \text{ but } x_i > 0 \text{ somewhere on } C\}.$$

Thus D is the subset of C cut out by $x_i = 0$ for $i \in I_D$, or equivalently the locus where $\sum_{i \in I_D} x_i = 0$ since each $x_i \geq 0$ on C . Then $\text{dep}(\tilde{C})$ consists of those $\vec{x} \in V$ where:

- (a) all $x_i \geq 0$,
- (b) for each C -independent face D one has $\sum_{i \in I_D} x_i > 0$,
- (c) and finally $\sum_{i=1}^{7t} x_i > 0$ as the origin is not in $\text{dep}(\tilde{C})$.

If we define Q to be those $\vec{x} \in V$ where all $x_i \geq 0$, where for each C -independent face D one has $\sum_{i \in I_D} x_i \geq 1$, and finally where $\sum_{i=1}^{7t} x_i \geq 1$, then we have $Q \cap \mathbb{Z}^{7t} = \text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t}$ as needed. □

Proof of Theorem 4.1 Let $f(\mathbf{t})$ be the generating function for $\bar{T}(\text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t})$. The variable t_{d+1} in $f(\mathbf{t})$ corresponds to $-\chi$ and by Lemma 4.4 the function χ is proper on \tilde{C} ; thus, there are only finitely many terms of $f(\mathbf{t})$ with any given power of t_{d+1} . Hence $g(t) = f(1, \dots, 1, t)$ is a well-defined element of $\mathbb{Z}[[t]]$, and indeed by Lemma 4.3 it is the generating function $B_C(x)$ we seek with x replaced by t^2 .

Since $f(\mathbf{t})$ is short by Lemma 4.8, it remains to use this to see that $g(t)$ is also short. Provided no denominator in (4.7) has a factor of $(1 - t_1^{a_1} t_2^{a_2} t_3^{a_3} \dots t_d^{a_d})$, that is, has no t_{d+1} -term, then this is immediate. To handle the general case, we will use results from [68], noting that our notion of a generating function being short is equivalent to rationality in the sense of Definition 1.4 of [68]. First, set $S = \bar{T}(\text{dep}(\tilde{C}) \cap \mathbb{Z}^{7t})$. As the generating function $f(\mathbf{t})$ of S is short, Theorem 1.5 of [68] gives that S is a Presburger set, that is, there is a Presburger formula F which tests points in \mathbb{N}^{d+1} for membership in S . Writing points in \mathbb{N}^{d+1} as (\vec{c}, p) with $\vec{c} \in \mathbb{N}^d$ and $p \in \mathbb{N}$, we see from Definition 1.6 of [68] that $g(t)$ is the generating function of the Presburger counting function $p \mapsto \#\{\vec{c} \in \mathbb{N}^d \mid F(\vec{c}, p)\}$. Therefore, by Theorem 1.10 of [68], specifically $A \Rightarrow C$, the generating function $g(t)$ is short. □

4.9 Ideal triangulations

Our proof of Theorem 1.3 when ∂M is nonempty will use ideal triangulations instead of finite ones. The theory of *closed* normal surfaces in ideal triangula-

tions is nearly identical to that of normal surfaces in finite triangulations; after all, normal surfaces stay away from the vertices of the ambient triangulation, which is the only place where the topology differs between the two cases. Indeed, we claim that all the results of [64] hold for closed normal surfaces in ideal triangulations without any changes to the proofs. Manifolds with boundary are allowed in [64], so switching from finite to ideal triangulations can be viewed as using a slightly different type of finite cellulation as the background to do normal surface theory, specifically a cellulation by *truncated* tetrahedra. The combinatorics of closed normal surfaces in truncated tetrahedra is almost indistinguishable from standard normal surface theory in a finite triangulation, and hence the proofs in [64] work as written in our new context. Consequently, the results of Sect. 3 also hold for closed surfaces in ideal triangulations.

From now on, suppose M is a compact irreducible ∂ -irreducible 3-manifold with nonempty boundary and \mathcal{T} an ideal triangulation of M , which exists by e.g. [41, Proposition 3]. The *vertex link* H_v of a vertex $v \in \mathcal{T}^0$ is the normal surface consisting of one triangle in each tetrahedron corner where the vertex in that corner corresponds to v . The vertex link H_v should be viewed as a parallel copy of the corresponding boundary component of M . An ideal triangulation \mathcal{T} is ∂ -efficient when the only connected normal surfaces that are boundary parallel are the vertex links. For example, if \mathcal{T} has a positive angle structure then it is ∂ -efficient by [44, Proposition 4.4]. Provided M is acylindrical, then any minimal ideal triangulation is ∂ -efficient by [41, Theorem 4], so such triangulations always exist for the manifolds we consider in Theorem 1.3.

Our goal now is to weed out the inessential incompressible surfaces, i.e. those with a ∂ -parallel component, from our counts. When ∂M includes a torus, this is not just an aesthetic preference but a requirement since there are infinitely many isotopy classes of (disconnected) incompressible surfaces with the same Euler characteristic.

4.10 Lemma *Suppose \mathcal{T} is a ∂ -efficient ideal triangulation of a 3-manifold M that contains no nonorientable closed incompressible surfaces. Let C be a lw -face of $\mathcal{P}_{\mathcal{T}}$. If C carries no vertex link then every normal surface carried by C is essential. If C carries some vertex link then no normal surface carried by $\text{dep}(C)$ is essential.*

Proof Let $I_C \subset \{1, 2, \dots, 7t\}$ be the indices of the coordinates on \mathbb{R}^{7t} which vanish on all of C . Then $C = \{\vec{x} \in \mathcal{P}_{\mathcal{T}} \mid x_i = 0 \text{ for all } i \in I_C\}$ and $C^\circ = \{\vec{x} \in C \mid x_i > 0 \text{ for all } i \notin I_C\}$.

First, suppose C carries an inessential normal surface F . As \mathcal{T} is ∂ -efficient, the surface F is the disjoint union of a normal surface G (possibly empty) and some vertex link H_v ; in particular $F = G + H_v$. From the above description of C , it is clear that as $G + H_v$ is carried by C , both G and H_v are also carried by C . This proves the first claim.

Second, suppose C carries some vertex link H_v . Since H_v is carried by C , for each index i corresponding to a triangle in H_v we have $i \notin I_C$. Hence, for any normal surface F carried by C° , each triangle that appears in H_v also has positive weight in F . Consequently, the surface F has a component which is normally isotopic to some H_v and in particular is inessential. More broadly, suppose G is a normal surface carried by $\text{dep}(C)$. As C is least-weight, by Theorem 3.5, the surface G is projectively isotopic to some F carried by C° , and by the previous argument the latter has a component which is H_v . By the hypotheses on M , both G and F are orientable since they are incompressible; as they are projectively isotopic, it follows that G also has a component isotopic to H_v . In particular, the surface G is inessential. This proves the second claim. \square

We call a lw-face C *essential* if every normal surface carried by C is essential. By Lemma 4.10, a lw-face C is either essential or every normal surface carried by $\text{dep}(C)$ is inessential. Hence Theorem 3.8 gives:

4.11 Theorem *Suppose \mathcal{T} is a ∂ -efficient ideal triangulation of a 3-manifold M that contains no nonorientable closed incompressible surfaces. For each orientable essential surface F there exists a unique complete essential lw-face C such that $\text{dep}(C)$ carries a surface (non-projectively) isotopic to F .*

We can now prove the analog of Theorem 4.1 for manifolds with boundary:

4.12 Theorem *Suppose M is an irreducible ∂ -irreducible atoroidal acylindrical 3-manifold with $\partial M \neq \emptyset$ that contains no nonorientable essential surfaces. Suppose \mathcal{T} is a ∂ -efficient ideal triangulation of M and C is a complete essential lw-face of $\mathcal{P}_{\mathcal{T}}$. The generating function $B_C(x)$ corresponding to the counts of isotopy classes of closed essential surfaces carried by $\text{dep}(C)$ is short.*

Proof As C carries only essential surfaces, we have $\chi : \tilde{\mathcal{C}} \rightarrow \mathbb{R}$ is proper since all essential surfaces have $\chi \leq -1$. This gives the analog of Lemma 4.4 in our setting, and the proof of the theorem is now identical to that of Theorem 4.1. \square

We now complete the proof of the first main theorem of this paper:

Proof of Theorem 1.3 when M has boundary Take \mathcal{T} to be a minimal ideal triangulation of M , which is ∂ -efficient by [41, Theorem 4] since M is acylindrical. By Theorem 4.11, every isotopy class of essential surface is carried by $\text{dep}(C)$ for a unique complete essential lw-face C . As sums of short generating functions are also short, Theorem 1.3 now follows from Theorem 4.12. \square

4.13 Whither nonorientable and bounded surfaces

We would of course like to remove the hypothesis in Theorem 1.3 that M contains no closed nonorientable essential surfaces, and also broaden the count to allow essential surfaces with boundary. We now outline some of the difficulties inherent in such extensions.

For nonorientable closed essential surfaces, what prevents us from just including them in the count is that while $\mathcal{LW}_{\mathcal{T}}$ can carry nonorientable essential surfaces, it need not carry *all* of them. (This distinction is also present in the branched surface perspective of [54].) The issue is that you can have a nonorientable essential normal surface F which is least weight in its isotopy class but where its double $2F$, while normal and essential, may not be least weight. This does not happen for an orientable F since the double is just two parallel copies of F . One could sidestep this issue by just counting orientable surfaces, but picking those out of each lw-face seems tricky for the following reason. Note that F is orientable if and only if $2F$ has twice the number of connected components as F . As we see from the a_M versus b_M discussion in Sects. 1 and 8, counting components is subtle. Consequently, we suspect there are examples where the count of orientable surfaces does not have a short generating function.

A key obstruction to counting surfaces with boundary is actually the issue of orientability. Unlike in the closed case with Proposition 2.4, there is no homological condition we can impose that a priori eliminates the possibility of nonorientable essential surfaces with boundary. For example, the exterior of a knot in S^3 can contain such nonorientable surfaces (e.g. many checkerboard surfaces for alternating knots). As [64] and our Sect. 3 do allow orientable surfaces with boundary, we are hopeful that if nonorientable closed surfaces can be dealt with, then counting bounded surfaces will also be possible.

5 Proof of the decision theorem

This section is devoted to proving Theorem 1.4, which says that there are algorithms for finding the generating function in Theorem 1.3 as well as enumerating representatives of the isotopy class of essential surfaces and determining which isotopy class a given surface belongs to. In this section, we do not worry about the efficiency of these algorithms, merely their existence; the actual method used to compute the examples in Sect. 7 uses some of the ideas here but in the modified form of Sect. 6 which is specific to when ∂M is a nonempty union of tori.

Throughout this section, let M be a compact orientable irreducible ∂ -irreducible 3-manifold with a fixed triangulation \mathcal{T} , which is a finite triangulation when M is closed or an ideal triangulation otherwise. With the exception

of the proof of Theorem 1.4 itself at the very end, in this section we do *not* require that M is acylindrical or atoroidal, nor that \mathcal{T} is ∂ -efficient; moreover, the manifold M may contain nonorientable closed essential surfaces. The algorithms in Theorem 1.4 follow the approach of the proof of Theorem 1.3 closely, so the first thing we will need is:

5.1 Theorem *There exists an algorithm for computing the collection $\mathcal{CLW}_{\mathcal{T}}$ of complete lw-faces of $\mathcal{P}_{\mathcal{T}}$.*

Given a normal surface F in \mathcal{T} , one can algorithmically determine whether or not it is incompressible and ∂ -incompressible; indeed, this is essentially Haken's original application of normal surface theory, see e.g. [42, Algorithm 9.6]. The tricky part of computing $\mathcal{CLW}_{\mathcal{T}}$ is figuring out the isotopy relationships between different such normal surfaces.

5.2 Graph of incompressible surfaces

We use the following framework for understanding isotopies among normal surfaces. Let $\mathcal{G}_{\mathcal{T}}$ be the graph whose vertices are connected incompressible closed normal surfaces, more precisely the normal isotopy classes of such surfaces, and where there is an edge between surfaces F and G exactly when F and G can be normally isotoped to be disjoint and cobound a product region. Given an integer w , we use $\mathcal{G}_{\mathcal{T}}^{\leq w}$ to denote the subgraph whose vertices are all surfaces $F \in \mathcal{G}_{\mathcal{T}}$ of weight at most w . Since there are only finitely many normal surfaces of bounded weight, each $\mathcal{G}_{\mathcal{T}}^{\leq w}$ is finite. Moreover, given w , the graph $\mathcal{G}_{\mathcal{T}}^{\leq w}$ can be algorithmically constructed as follows. First, the vertices of $\mathcal{G}_{\mathcal{T}}^{\leq w}$ can be found by enumerating all connected normal surfaces of weight at most w and then testing each for incompressibility. Second, for each pair of surfaces F and G in $\mathcal{G}_{\mathcal{T}}^{\leq w}$, one can test if they can be normally isotoped to be disjoint using Algorithm 9.5 of [42]; specifically, this can be done if and only if F and G are compatible and the normal sum $F + G$ consists of two connected components where one is normally isotopic to F and the other to G . When they can be made disjoint in this way, there is a unique way to do so up to normal isotopy. Finally, for each pair of surfaces F and G that can be normally isotoped apart, we test all components of M cut along $F \cup G$ for being products using Algorithm 9.7 of [42]. This completes the algorithm for constructing $\mathcal{G}_{\mathcal{T}}^{\leq w}$.

Any two surfaces in the same connected component of $\mathcal{G}_{\mathcal{T}}$ are of course isotopic. It turns out the converse is true as well, in the following strong form:

5.3 Theorem *If F and G are two vertices of $\mathcal{G}_{\mathcal{T}}$ that are isotopic, then they are joined by a path in $\mathcal{G}_{\mathcal{T}}$ passing only through vertices H with $\text{wt}(H) \leq$*

$\max(\text{wt}(F), \text{wt}(G))$. In particular, the isotopy classes of surfaces in any $\mathcal{G}_{\mathcal{T}}^{\leq w}$ correspond precisely to the connected components of $\tilde{\mathcal{G}}_{\mathcal{T}}^{\leq w}$.

We prove this theorem in Sect. 5.4 below, but we first use it to derive Theorem 5.1.

Proof of Theorem 5.1 First, compute the polytope $\mathcal{P}_{\mathcal{T}}$ from the normal surface equations. For each face C of $\mathcal{P}_{\mathcal{T}}$, fix a normal surface F_C that is carried by its interior. Compute the graph $\mathcal{G}_{\mathcal{T}}^{\leq w}$ where w is the maximum weight of any F_C . The incompressible F_C are those that are vertices of $\mathcal{G}_{\mathcal{T}}^{\leq w}$, and, by Theorem 5.3, we know exactly which F_C are least-weight. When F_C is least-weight, we also know every other least-weight surface isotopic to it. Applying Theorem 3.3 now identifies exactly the faces C that are in $\mathcal{CLW}_{\mathcal{T}}$. □

5.4 Isotopic normal pairs

Throughout, let M be a compact orientable irreducible 3-manifold with a fixed triangulation \mathcal{T} as in the previous section. An *isotopic normal pair* (F, G) is an isotopic pair of closed incompressible normal surfaces F and G that meet transversely in the sense of [64, Page 1091]. Define the complexity of such a pair by

$$c(F, G) = (\max(\text{wt}(F), \text{wt}(G)), \min(\text{wt}(F), \text{wt}(G)), \#(F \cap G)),$$

where $\#(F \cap G)$ denotes the number of connected components of $F \cap G$. We will compare complexities lexicographically.

If (F, G) is an isotopic normal pair where F and G are disjoint, then by Lemma 5.3 of [66] the surfaces F and G are parallel, i.e. cobound a region homeomorphic to $F \times I$. In this case, the pair (F, G) gives rise to an edge of $\mathcal{G}_{\mathcal{T}}$. The key result of this subsection is:

5.5 Theorem *If (F, G) is an isotopic normal pair with $F \cap G \neq \emptyset$ then, after possibly interchanging F and G , there exists a normal surface F' that is isotopic to F and disjoint from it that meets G transversely with $c(F', G) < c(F, G)$.*

Given any (F, G) isotopic normal pair with $w = \max(\text{wt}(F), \text{wt}(G))$, we can apply Theorem 5.5 repeatedly until we arrive at a pair (F'', G'') where F'' and G'' are disjoint. This proves Theorem 5.3 above, since each application of Theorem 5.5 gives an edge in $\mathcal{G}_{\mathcal{T}}^{\leq w}$ and there is also an edge from F'' to G'' .

Suppose \tilde{F} and \tilde{G} are subsurfaces of F and G respectively, with $\partial\tilde{F} = \partial\tilde{G}$. Here is one way to make precise the notation that \tilde{F} and \tilde{G} are “parallel rel boundary”. Given a compact surface H , define $P(H)$ as the quotient of $H \times I$ where for each $h \in \partial H$ the set $\{h\} \times I$ has been collapsed to a

point. A *product region* between \tilde{F} and \tilde{G} is an embedding $f : P(H) \rightarrow M$ where $f(H \times \{0\}) = \tilde{F}$ and $f(H \times \{1\}) = \tilde{G}$; here we do not insist that $P = f(P(H))$ meets $F \cup G$ only in $\tilde{F} \cup \tilde{G}$. The first step in proving Theorem 5.5 is to show:

5.6 Lemma *Suppose (F, G) is an isotopic normal pair with $F \cap G \neq \emptyset$. After possibly interchanging F and G , there exist subsurfaces $\tilde{F} \subset F$ and $\tilde{G} \subset G$ where $\tilde{G} \cap F = \partial\tilde{G}$ and $\partial\tilde{F} = \partial\tilde{G}$ and $\text{wt}(\tilde{G}) \leq \text{wt}(\tilde{F})$ with \tilde{G} and \tilde{F} bounding a product region P with $P \cap F = \tilde{F}$.*

Proof Suppose all components of $F \cap G$ are essential in both F and G . Then by Proposition 5.4 of [66], there exist subsurfaces $\tilde{F} \subset F$ and $\tilde{G} \subset G$ with $\tilde{F} \cap G = \partial\tilde{F} = \partial\tilde{G} = F \cap \tilde{G}$ where $\tilde{F} \cup \tilde{G}$ bounds a product region P where $P \cap F = \tilde{F}$ and $P \cap G = \tilde{G}$. Relabeling, we can arrange that $\text{wt}(\tilde{G}) \leq \text{wt}(\tilde{F})$ to prove the lemma in this case.

Suppose instead some component of $F \cap G$ is inessential in one of F or G . Among all disks contained in one of F or G bounded by a component of $F \cap G$, let D be one of least weight, which exists since the weight (i.e., the number of intersection points with the 1-skeleton of \mathcal{T}) of any subsurface is a nonnegative integer. By passing to an innermost component, we can assume D meets $F \cap G$ only along ∂D . After relabeling, we can assume this D is contained in G and then set $\tilde{G} = D$. As F is incompressible, the curve $\partial\tilde{G}$ must bound a disk \tilde{F} in F . Together the disks $\tilde{F} \cup \tilde{G}$ form a sphere which must bound a ball as M is irreducible, and hence \tilde{F} and \tilde{G} bound the required product region P . By our initial choice of \tilde{G} , we must have $\text{wt}(\tilde{G}) \leq \text{wt}(\tilde{F})$ as desired. □

Proof of Theorem 5.5 Let \tilde{F} and \tilde{G} be given by Lemma 5.6. Set $F_0 = (F \setminus \tilde{F}) \cup \tilde{G}$ which is isotopic to F via the product region P . Move F_0 slightly so that it is disjoint from F , and notice that $\text{wt}(F_0) = \text{wt}(F) - \text{wt}(\tilde{F}) + \text{wt}(\tilde{G}) \leq \text{wt}(F)$. If F' is a normalization of the incompressible surface F_0 , we have $\text{wt}(F') \leq \text{wt}(F_0)$; by the barrier theory [40, Theorem 3.2(1)], the surface F' is disjoint from F , and we can perturb F' slightly to be transverse to G .

If $\text{wt}(F') < \text{wt}(F)$ we now have our desired (F', G) as $c(F', G) < c(F, G)$ where the two complexities differ in one of the first two components. If instead $\text{wt}(F') = \text{wt}(F)$, then we must have $\text{wt}(F_0) = \text{wt}(F)$. This means that F_0 is normally isotopic to F' as all normalization moves that change the normal isotopy class strictly reduce the weight. Then $\#(F' \cap G) = \#(F_0 \cap G) = \#(F \cap G) - \#\partial\tilde{F} < \#(F \cap G)$ and hence $c(F', G) < c(F, G)$ with the complexities differing only in the last component. □

We turn now to the proof of Theorem 1.4, so now the manifold M is atoroidal, acylindrical, and does not contain a nonorientable essential surface.

Proof of Theorem 1.4 As input, we are given a triangulation \mathcal{T} of M which will be finite if M is closed or could be finite or ideal if M has boundary. If M has boundary and we are given a finite triangulation, convert it to an ideal one using the procedure described in the proof of [47, Theorem 1.1.13], which is relevant as per [41, Proposition 3]. When M has boundary, apply the algorithm of [39, Theorem 4.7] so that the ideal triangulation \mathcal{T} we are working with is ∂ -efficient.

Start by computing $\mathcal{CLW}_{\mathcal{T}}$ via Theorem 5.1. As in the proof of Theorem 1.3, the claim that we can compute the overall generating function algorithmically follows if we can implement Theorem 4.1 or Theorem 4.12 as appropriate for a particular face C of $\mathcal{CLW}_{\mathcal{T}}$. (In the case when M has boundary, by Lemma 4.10 we can skip any C which carries a vertex link, which is easy to test.) From the proofs of those theorems, we need algorithms for two things: finding the subspace W_C from Theorem 3.5 and applying Theorem 1.1 of [52]. The latter is provided by [52] itself, so we focus on the former.

To compute W_C , first let F_0 be any normal surface carried by the interior of C and compute $\mathcal{G}_{\mathcal{T}}^{\leq \text{wt}(F_0)}$. Now apply Theorem 5.3 to find all least-weight surfaces F_1, \dots, F_k that are isotopic to F . As C is complete, all the F_i are carried by C and hence are mutually compatible. By Lemma 3.5, we have that W_C is spanned by the $\vec{F}_0 - \vec{F}_k$, giving us the needed description of W_C .

The second claim of the theorem, that we can give unique normal representatives of the isotopy classes of incompressible surfaces with $\chi = -2n$, is easy by looking at the lattice points in the sublevel sets of χ on $\tilde{C} \subset \mathcal{S}_{\mathcal{T}}$ for each face C of $\mathcal{CLW}_{\mathcal{T}}$ and modding out by W_C . The final claim, that we can determine the isotopy class of a given incompressible surface F , can be done using this list and $\mathcal{G}_{\mathcal{T}}^{\leq \text{wt}(F)}$ because of Theorem 5.3. \square

6 Almost normal surfaces in ideal triangulations

This section discusses the algorithm used to implement Theorem 1.4 for the computations in Sect. 7. The key difference compared to the proof of Theorem 1.4 in Sect. 5 is that we use almost normal surfaces, rather than normal ones, to determine which normal surfaces are incompressible and to find isotopies between them.

In this section, we study manifolds with boundary a union of tori using ideal triangulations admitting a *partially flat angle structure* in the sense of [45]. Since we are restricting to M with $\chi(\partial M) = 0$, we require the angles at each ideal vertex to sum to exactly π rather than at most π as in (i) on page 916 of [45]. Such triangulations impose restrictions on the topology of the underlying manifold M . The only connected closed normal surfaces in \mathcal{T} with $\chi \geq 0$ are vertex links, and M is irreducible, ∂ -irreducible, atoroidal, and acylindrical [45, Theorem 2.2]; in particular, the interior of M admits a

finite-volume complete hyperbolic metric, and \mathcal{T} is a ∂ -efficient triangulation of M . A general algorithm for finding such a \mathcal{T} in this setting is given in [45, §2] and in practice one easily finds a \mathcal{T} admitting the stronger notion of a *strict angle structure* from [35].

6.1 Tightening almost normal surfaces

An *almost normal surface* is a surface S in \mathcal{T} built from the same elementary discs as normal surfaces except for exactly one piece, which is either an almost normal octagon or made by joining two elementary discs in the same tetrahedron by an unknotted tube. Given a transverse orientation of an orientable almost normal surface A in \mathcal{T} , we can “destabilize” the exceptional piece in that direction and then perform normalization moves. This process, called *tightening* the surface A , moves it in only one direction and terminates in a normal surface which we denote $T_+(A)$, see [58, Chapter 4] for details. While the sequence of normalization moves is not unique, the tightened surface $T_+(A)$ is well-defined: together A and $T_+(A)$ bound the *canonical compression body* of A defined in [58, §4.1] which we denote $V_+(A)$. Here, the compression body $V_+(A)$ is built from $A \times I$ by adding 2- and 3-handles, so that A is the minus boundary of $V_+(A)$ and $T_+(A)$ is the plus boundary. In particular, we have $\chi(A) \leq \chi(T_+(A))$ with equality if and only if $V_+(A)$ is just $A \times I$. Moreover, since \mathcal{T} contains no normal 2-spheres, every component of $T_+(A)$ has genus at least 1. We will use $T_-(A)$ to denote the tightening of A in the opposite transverse direction with $V_-(A)$ the corresponding canonical compression body.

As per [58], the tightening process can be followed by tracking just the intersection of each surface with the 2-skeleton of \mathcal{T} . Thus it amounts to looking at a family of arcs in \mathcal{T}^2 , which need not all be normal, and then doing a sequence of bigon moves across edges of \mathcal{T}^1 until one is left only with normal arcs. It is thus straightforward to implement once one creates an appropriate data structure to do the bookkeeping, though our code is the first time this has been done.

6.2 Sweepouts and thin position

The notions of sweepouts [56] and Gabai’s thin position [61] can independently be used to prove the existence of almost normal surfaces in many situations. We will need the following two such results, which are quite standard.

6.3 Lemma *Suppose N and N' are normal surfaces cobounding a product region V . Then there are disjoint surfaces $N = N_0, A_1, N_1, A_2, \dots, N_{n-1}$,*

$A_n, N_n = N'$ in V with the N_k normal and the A_k almost normals such that $T_-(A_k) = N_{k-1}$ and $T_+(A_k) = N_k$.

Proof Since \mathcal{T} is ideal and the surfaces N and N' are closed, the product region V between them contains no vertices of \mathcal{T} . The usual sweepout or thin position argument for a product, see e.g. [58, Theorem 6.2.2], gives the needed sequence of surfaces. \square

6.4 Lemma *Suppose V is a nonproduct compression body in M where both ∂_-V and ∂_+V are normal surfaces in \mathcal{T} . Then there exists an almost normal surface $A \subset V$ such that $T_-(A)$ and A are parallel inside V to ∂_-V and $T_+(A)$ is a (proper) compression of ∂_-V .*

Proof As \mathcal{T} is ideal, there are no vertices of \mathcal{T} inside V , and a push off of ∂_-V into V is a strongly irreducible Heegaard surface for V . By a slight strengthening of [56,60] in the same manner as [45, Theorem 4.2], we can find an almost normal surface A in V that is parallel to ∂_-V . Transversely orient A away from ∂_-V . Since A is incompressible in the negative direction, we have $T_-(A)$ is parallel to ∂_-V . We are done if $T_+(A)$ is a compression of A . Otherwise, the region between ∂_-V and $T_+(A)$ is a product, and we repeat the argument on the compression body bounded by the normal surfaces $T_+(A)$ and ∂_+V . As there is a bound on the number of disjoint normal surfaces in \mathcal{T} , none of which are normally isotopic, this will terminate and so produce the surface we seek. \square

6.5 Finiteness of (almost) normal surfaces

For $g \geq 2$, we define $\mathcal{N}_{\mathcal{T}}^g$ to be the set of connected normal surfaces of genus g in \mathcal{T} , up to normal isotopy. Correspondingly, the set of such almost normal surfaces is $\mathcal{A}_{\mathcal{T}}^g$, again up to normal isotopy. A key result for us is Theorem 4.3 of [45]

6.6 Theorem [45] *When \mathcal{T} has a partially flat angle structure, both $\mathcal{N}_{\mathcal{T}}^g$ and $\mathcal{A}_{\mathcal{T}}^g$ are finite and algorithmically computable.*

We will sketch the proof of Theorem 6.6 as it outlines the algorithm for finding $\mathcal{N}_{\mathcal{T}}^g$ and $\mathcal{A}_{\mathcal{T}}^g$, which is an important component of the overall algorithm given in Sect. 6.12. This discussion is most natural in the setting of the *quadrilateral coordinates* for normal surfaces introduced in [65] rather than the standard triangle-quad coordinates we've used so far. We now describe the basics of quad coordinates, referring to [15] for details. As the name suggests, in these coordinates a normal surface F is recorded by just the $3t$ weights on the quadrilateral discs, where t is the number of tetrahedra of \mathcal{T} . It turns out this determines F up to any vertex-linking components H_v as in Sect. 4.9. There

are still linear equations, one for each edge of \mathcal{T} , characterizing the admissible vectors in \mathbb{N}^{3t} that give normal surfaces; we use $\mathcal{S}'_{\mathcal{T}}$ and $\mathcal{P}'_{\mathcal{T}}$ to denote the corresponding linear solution space and its intersection with the positive orthant. (The relationship between the vertices of $\mathcal{P}_{\mathcal{T}}$ and $\mathcal{P}'_{\mathcal{T}}$ is described in detail in [15].) Given an admissible vector $\vec{v} \in \mathbb{N}^{3t}$ carried by $\mathcal{P}'_{\mathcal{T}}$, we take the associated normal surface to be the one with those quad weights and no vertex-linking components; following [15], we call such surfaces *canonical*. Two things to keep in mind about quad coordinates:

- (a) In standard coordinates, adding vector representatives corresponds to the geometric Haken sum. In quad coordinates, adding vector representatives corresponds to geometric Haken sum *followed by removing all copies of the vertex links*. Hence the total weight is additive in standard coordinates but only subadditive in quad coordinates. Correspondingly, the total weight of a surface is only piecewise linear in quad coordinates.
- (b) Because we have an angle structure on \mathcal{T} , the Euler characteristic function is linear in quad coordinates. (In contrast, it is only piecewise linear in quad coordinates for finite triangulations.) Specifically, consider the linear function $\chi : \mathbb{R}^{3t} \rightarrow \mathbb{R}$ defined as follows. Consider the basis vector e_i corresponding to a quad Q in a tetrahedron σ . We set

$$\chi(e_i) = -1 + \frac{\theta_1 + \theta_2 + \theta_3 + \theta_4}{2\pi},$$

where the θ_k are the angles assigned to the four edges of σ that Q meets. Then $\chi(F) = \chi(\bar{F})$ by [44, Proposition 4.3]. Moreover, for each face C of $\mathcal{P}'_{\mathcal{T}}$ that carries only admissible vectors, the function $\chi : (\mathbb{R}_+ \cdot C) \rightarrow \mathbb{R}$ is in fact proper, nonpositive, and zero only at the origin [45, Theorem 2.1]; when the angle structure is strict, this is immediate for all of $\mathcal{P}'_{\mathcal{T}}$ since each $\chi(e_i) < 0$.

Turning to almost normal surfaces, those with octagons can be described in terms of lattice points in certain polytopes, and we will use the quad-octagon coordinates of [16], as opposed to the standard quad-octagon-tri coordinates, to record them. Almost normal surfaces with tubes will be encoded by a *normal* surface together with the pair of adjacent normal discs that the tube runs between. With these preliminaries in hand, we can now give:

Proof of Theorem 6.6 First, consider the case of $\mathcal{N}^g_{\mathcal{T}}$ which is contained in the preimage $\chi^{-1}(2 - 2g)$ for the map $\chi : \mathbb{R}^{3t} \rightarrow \mathbb{R}$ defined in (b) above. Since χ is proper on the cone over each admissible face of $\mathcal{P}'_{\mathcal{T}}$, there are only finitely many lattice points in $\chi^{-1}(2 - 2g)$ corresponding to surfaces. These can be enumerated and tested for whether the surfaces are connected, giving us exactly $\mathcal{N}^g_{\mathcal{T}}$.

For \mathcal{A}_T^g , we consider the cases of octagons and tubes separately. For octagons, the map χ is again proper on the relevant polytope, and so this case works out the same as \mathcal{N}_T^g . For tubes, one first enumerates all normal surfaces (not necessarily connected) with $\chi = 4 - 2g$. For each such surface N , one considers all possible tubes and selects those that produce a connected surface, i.e. an element of \mathcal{A}_T^g . \square

6.7 Remark A single normal surface F can give rise to many different almost normal surfaces with tubes, where we are considering almost normal surfaces up to normal isotopy. To keep the computation manageable, we considered non-normal isotopies of tubes for a fixed normal surface F . That is, for an almost normal surface A made by adding a tube to F , we can “slide” the attaching points of the tube through one of the faces of the tetrahedron that contains it to get another almost normal surface built on the same F . In our actual computations, we considered such surfaces up to this equivalence. It is not hard to show that two tubed surfaces that are equivalent in this sense have the same canonical compression body and hence the same tightenings.

6.8 Another graph of normal surfaces

We now turn \mathcal{N}_T^g into a graph by adding edges as follows. For each $A \in \mathcal{A}_T^g$, we pick a transverse orientation arbitrarily and consider its two tightenings $T_{\pm}(A)$:

- (a) If both $T_{\pm}(A)$ are homeomorphic to A , we add an (undirected) edge joining $T_-(A)$ and $T_+(A)$. In this situation, both $V_{\pm}(A)$ are products and so $T_-(A)$ and $T_+(A)$ are isotopic.
- (b) If $T_+(A)$ is homeomorphic to A but $T_-(A)$ is not, mark the vertex $T_+(A)$ in \mathcal{N}_T^g as compressible. Also do the same with the roles of $T_+(A)$ and $T_-(A)$ reversed.

The main result of this subsection is:

6.9 Theorem *Isotopy classes of closed essential surfaces in M of genus g are in bijection with the connected components of \mathcal{N}_T^g where no surface is marked as compressible.*

Given how the edges in \mathcal{N}_T^g were defined, to prove Theorem 6.9, it suffices to show the following two lemmas:

6.10 Lemma *If $N \in \mathcal{N}_T^g$ is essential and isotopic to $N' \in \mathcal{N}_T^g$ then there is a path joining them in \mathcal{N}_T^g .*

Proof Let $w = \max(\text{wt}(N), \text{wt}(N'))$ and consider the graph $\mathcal{G}_T^{\leq w}$ from Sect. 5.2. Temporarily viewing N and N' as vertices of $\mathcal{G}_T^{\leq w}$, since they

are isotopic surfaces, Theorem 5.3 gives a sequence of normal surfaces $N = N_0, N_1, \dots, N_n = N'$ where N_k and N_{k+1} can be normally isotoped to be disjoint and cobound a product region P_k . Applying Lemma 6.3 to P_k gives a path in $\mathcal{N}_{\mathcal{T}}^g$ joining N_k to N_{k+1} . Concatenating these paths together gives a path in $\mathcal{N}_{\mathcal{T}}^g$ joining N to N' as needed. \square

6.11 Lemma *If $N \in \mathcal{N}_{\mathcal{T}}^g$ is compressible, it can be joined by a path in $\mathcal{N}_{\mathcal{T}}^g$ to a surface N' that is marked as compressible.*

Proof We first show there exists a nontrivial compression body V in M with $\partial_- V = N$ and $\partial_+ V$ a normal surface. Splitting M open along N and using the characteristic compression body of [12, §2], we can find a nontrivial compression body $V \subset M$ with $\partial_- V = N$ and $\partial_+ V$ is incompressible in the complement of N . Since N is normal, barrier theory [40, Theorem 3.2(1)] tells us that we can normalize $\partial_+ V$ in the complement of N via an isotopy, giving us the desired compression body.

By Lemma 6.4, there is an almost normal surface A in V so that $T_-(A)$ and A are parallel to N and $T_+(A)$ is a compression of A . Set $N' = T_-(A)$, which is marked as compressible because of the surface A . As N and N' are parallel, by Lemma 6.3 they are joined by a path in $\mathcal{N}_{\mathcal{T}}^g$, completing the proof of the lemma. \square

6.12 Algorithm

The input for the algorithm is an ideal triangulation \mathcal{T} with a partially flat angle structure of a manifold M where $H_2(\partial M; \mathbb{F}_2) \rightarrow H_2(M; \mathbb{F}_2)$ is onto so M contains no nonorientable surfaces by Proposition 2.4. (Here, the angles are given as rational multiples of π ; the set of partially flat angle structures form a convex polytope with rational vertices, so this is not a real restriction.) The output is the list $\{(C, W_C)\}$ of complete essential lw-faces C of $\mathcal{LW}_{\mathcal{T}} \subset \mathcal{P}_{\mathcal{T}}$ together with the corresponding subspaces W_C . Before starting, recall that a *vertex surface* is a normal surface F where \bar{F} is a primitive lattice point on the ray corresponding to an admissible vertex of $\mathcal{P}_{\mathcal{T}}$. Note that every vertex surface is connected, and that each vertex-linking torus H_v for $v \in \mathcal{T}^0$ is a vertex surface [15, Corollary 4.4].

- (1) Enumerate all vertex surfaces for the normal surface equations for \mathcal{T} in standard triangle-quad coordinates via [15, Algorithm 5.17]. Then use Algorithm 3.2 of [17] to find all admissible faces of $\mathcal{P}_{\mathcal{T}}$. Set g_0 to be the maximum genus of any vertex surface.
- (2) For each g with $2 \leq g \leq g_0$, enumerate the finite sets $\mathcal{N}_{\mathcal{T}}^g$ and $\mathcal{A}_{\mathcal{T}}^g$ as described in the proof of Theorem 6.6. Apply the tightening procedure of Sect. 6.1 to each surface A in $\mathcal{A}_{\mathcal{T}}^g$ to compute $T_-(A)$ and $T_+(A)$. As

detailed in Sect. 6.8, this information makes $\mathcal{N}_{\mathcal{T}}^g$ into a graph where certain vertices are labeled compressible.

We then compute a complete list of lw-surfaces of genus g from the graph $\mathcal{N}_{\mathcal{T}}^g$ using Theorem 6.9 as follows: for each connected component of $\mathcal{N}_{\mathcal{T}}^g$ where no surface was marked as compressible, compute the weight of each surface and then take all those of minimal weight for that component. This also computes all isotopy relations among the lw-surfaces of genus g . Because of the angle structure on M , there are no essential tori in M and the only nonessential lw-surfaces are the vertex links. Thus we now have a complete list of all essential lw-surfaces of genus at most g_0 .

- (3) From the list of admissible faces of $\mathcal{P}_{\mathcal{T}}$ which were computed in Step 1, select those where all vertices are among the essential lw-surfaces enumerated in Step 2. For each such C , select a surface F_C carried by its interior, e.g. take F_C to be the sum of the vertex surfaces of C . Use Algorithm 9.4 of [42] to decompose F_C into its connected components which are again normal surfaces. If any component of F_C has genus greater than g_0 , replace g_0 with the maximum genus of any component of F_C and re-run Step 2. By Theorem 3.3, the face C is least-weight if and only if every connected component of F_C is a lw-surface. So we now determine whether or not C is least-weight by using the list of lw-surfaces of genus at most g_0 from Step 2. Finally, if C is least-weight, it is essential by Lemma 4.10 and the observation that if C carried a vertex link H_v , then H_v would have to be one of the vertex surfaces of C . We now have a complete list of all essential lw-faces of $\mathcal{P}_{\mathcal{T}}$.
- (4) Now we determine which essential lw-faces are complete. Given such a face C , let F_C be the preferred surface in its interior. By Step 2, we know all lw-surfaces isotopic to a connected component of F_C . By Theorem 3.3, the face C is complete if and only if all these other surfaces are carried by C .
- (5) It remains to determine the subspace W_C for each complete essential lw-face C . Let G_1, \dots, G_k be all lw-surfaces isotopic to a connected component of F_C , all of which will be carried by C as it is complete. By the last part of Corollary 3.7, the vectors $\vec{G}_i - \vec{G}_j$ span W_C . This concludes the algorithm.

6.13 Remark It is clearly to our advantage to keep g_0 as small as possible, which suggests several performance improvements. For example, in Step 3 it pays to search the interior of C for an F_C whose components have the least genus. More elaborately, say that a normal surface N has an *obvious compression* when there is a chain of quads forming an annulus around a thin edge. In our setting, such surfaces cannot be essential, and we can discard in Step 1 any vertex surfaces with obvious compressions before setting g_0 . Because the notion of obvious compression can be framed as an admissibility

criteria on the faces of $\mathcal{P}_{\mathcal{T}}$ that is compatible with [17, Algorithm 3.2], it is not hard to check that this does not affect the correctness of the answer.

6.14 Remark The above algorithm in particular determines whether or not M contains a closed essential surface. It would be very interesting to study the practical efficiency of this algorithm as compared to the more traditional approach of [9, 13] involving testing for incompressibility by cutting M open along candidate surfaces. While we did implement Steps 1–3 of our algorithm for the computations in Sect. 7, we used a high-level but slow programming language and did not optimize the code extensively. Consequently, we did not have a good basis for making this comparison.

7 Computations, examples, and patterns

In this section, we describe the results of computing $\mathcal{LW}_{\mathcal{T}}$ for some 59,096 manifolds with torus boundary. These manifolds were drawn from two censuses. The first was the 44,692 orientable hyperbolic 3-manifolds that have ideal triangulations with at most 9 tetrahedra where ∂M is a single torus and $H_1(M; \mathbb{F}_2) = 0$ [14]. The second was the 14,656 hyperbolic knots in S^3 with at most 15 crossings whose exteriors have ideal triangulations with at most 17 ideal tetrahedra [37]. These two censuses have little overlap, with only 216 manifolds common to both.

Using the default triangulation for each manifold provided by SnapPy [20], we used Algorithm 6.12 to try to compute $\mathcal{LW}_{\mathcal{T}}$. We succeeded except for 36 triangulations where the computation ran out of time or memory. Combined, the computations took about 8 CPU-months, with the running time for a single manifold having a mean of 5.85 min and a maximum of 3 days. The median time was 0.8 seconds for the cusped census and 1.3 min for the knot exteriors.

With the initial triangulations, some 182 manifolds had distinct essential lw-surfaces that were isotopic. To avoid computing the generating function $B_M(x)$ in the general case where one is taking the quotient by the subspaces W_C , we replaced 168 of these triangulations with others where there were no such isotopies. Except for Sect. 7.4, we will unfairly lump the remaining 14 manifolds with distinct isotopic essential lw-surfaces in with the 36 whose computations timed out, and restrict our analysis to the other 59,082. A summary of these manifolds is given in Table 2, where we use the following terminology. Recall a 3-manifold M is *large* when it contains a closed essential surface and *small* otherwise. We call a large manifold M *barely large* when every closed essential surface is a multiple of a finite collection of such surfaces; otherwise M will be *very large*. In the language of Sect. 1.11, the terms small, barely large, and very large correspond, respectively, to $\mathcal{ML}_0(M) = \emptyset$, $\dim(\mathcal{ML}_0(M)) = 1$, and $\dim(\mathcal{ML}_0(M)) > 1$.

Table 2 Summary of the manifolds where we tried to compute $\mathcal{LW}_{\mathcal{T}}$. There are 216 manifolds common to both samples, which is why the last row is not the sum of the previous two. The 14 knot exteriors where there was a non-normal isotopy of lw-surfaces are all large; they are likely very large, but we did not check this

Sample	Count	Small	Barely large	Very large	Isotopy of lw	Failed
Cusped census	44,692	38,358	6046	288	0	0
Knot exteriors	14,656	10,554	3	4049	14	36
Combined	59,132	48,703	6049	4330	14	36

Table 3 Statistics about the complexes $\mathcal{LW}_{\mathcal{T}}$ for the 4330 very large manifolds, broken down by $\dim \mathcal{LW}_{\mathcal{T}}$. The properties recorded are: the number of connected components (comps), the number of vertices (verts), the number of maximal faces (max faces), and the largest number of vertices in any face (face size). For each numerical property, we give the mean in the μ column as well as the min-max interval in the range column

Dim	Count	Comps		Verts		Max faces		Face size	
		μ	Range	μ	Range	μ	Range	μ	Range
1	1697	1.1	[1, 4]	2.8	[2, 14]	1.8	[1, 10]	2	[2, 2]
2	1810	1.1	[1, 2]	5.6	[3, 16]	3.3	[1, 13]	3.1	[3, 4]
3	606	1.2	[1, 3]	10.6	[6, 21]	7.8	[1, 26]	4.8	[4, 7]
4	205	1.0	[1, 2]	16.4	[8, 44]	11.3	[2, 48]	7.7	[5, 12]
5	12	1.0	[1, 1]	19.3	[16, 21]	13.6	[7, 18]	10.8	[10, 12]

It is natural to ask what is the smallest volume of a hyperbolic manifold that is barely or very large. In our sample, the smallest manifold that is barely large is $m137$, which has volume $V_{oct} \approx 3.663862376$, and the smallest knot exterior that is barely large is that of $K15n153789$, which has volume about 9.077985047. Similarly, the smallest manifold we found that is very large is $s783$ which has volume about 5.333489566, and the smallest such knot exterior is that of $K10n10 = 10_{153}$, which has volume about 7.374343889.

7.1 Very large manifolds

For the 4,330 manifolds where $\dim \mathcal{LW}_{\mathcal{T}} \geq 1$, the complexes $\mathcal{LW}_{\mathcal{T}}$ run the gamut from a single edge (for 760 manifolds) up to monsters like $\mathcal{LW}_{\mathcal{T}}$ for $K13n3838$ which is connected with 44 vertices and 48 maximal faces all of dimension 4, where each maximal face has between 5 and 9 vertices. Basic statistics about the topology and combinatorics of the $\mathcal{LW}_{\mathcal{T}}$ are given in Table 3. All but 178 of these complexes are *pure*, that is, every maximal face has the same dimension; the exceptions are 140 cases where each component

Table 4 Statistics about the 88 distinct generating functions $B_M(x)$ for the 4330 very large manifolds, broken down by $\dim \mathcal{LW}_T$. Here, each $B_M(x)$ has rational form $P(x)/Q(x)$ for some $P, Q \in \mathbb{Z}[x]$ with $\deg P = \deg Q$. The properties recorded are: the number of distinct $B_M(x)$ (count), the values of $\deg P$ (degree), the observed periods of $B_M(x)$ (periods), and the range of the ℓ_1 -norm of the combined coefficients of the polynomials P and Q (ℓ_1 -norm)

Dim	Count	Degree	Periods	ℓ_1 -norm
1	35	2, 3, 4, 6, 8	1, 2, 3, 6	[7, 53]
2	18	3, 6, 7	1, 2, 3	[16, 38]
3	24	4, 5, 6, 7, 8	1, 2	[26, 71]
4	9	5, 6, 8	1, 2	[38, 94]
5	2	7, 8	2	[78, 88]

of \mathcal{LW}_T is pure but there are components of differing dimensions, and 38 cases where \mathcal{LW}_T is connected and impure.

While the combinatorics of some of these complexes is quite elaborate, the underlying topology of all \mathcal{LW}_T in our sample is simple in that every connected component is actually contractible. Moreover, for a component Y of dimension d , each $(d - 1)$ -face is glued to at most two d -faces; consequently, all components of dimension 1 are homeomorphic to intervals rather than more general trees. Here, contractibility was checked as follows. First, each \mathcal{LW}_T was converted to a simplicial complex (some 3,603 of the \mathcal{LW}_T are in fact simplicial, for the rest a barycentric subdivision of the polyhedral complex was used). We then checked that every component had vanishing reduced homology and trivial fundamental group using [57], which implies contractibility.

7.2 Surface counts by Euler characteristic

For each of the 4,330 very large manifolds, we computed the generating function $B_M(x)$ from Theorem 1.3 starting from \mathcal{LW}_T by using Normaliz [11]. This resulted in only 88 distinct generating functions whose properties are summarized in Table 4 and examples of which are given in Tables 5 and 6.

7.3 Sample LW complexes

We next give several examples of \mathcal{LW}_T for specific triangulations of knot exteriors. To start off, Fig. 1 gives an example of a simple \mathcal{LW}_T which is unusual in having components of different dimensions. Then Fig. 2 describes one of the most complicated 2-dimensional examples we found. Figure 3 shows the fairly complicated 3-dimensional complex coming from the Conway knot

Table 5 Nine of the most common $B_M(x)$, which together account for 3284 (75.8%) of the 4330 very large manifolds. The properties recorded are the dimension of \mathcal{LW}_T (dim), the rational form $P(x)/Q(x)$ of $B_M(x)$, the period of $B_M(x)$ (per), the ℓ_1 -norm of the combined coefficients of P and Q , an example manifold with this $B_M(x)$ (sample M), and the number of manifolds with this $B_M(x)$ (count)

Dim	$B_M(x)$	Per	ℓ_1	Sample M	Count
1	$\frac{-x^2 + 2x}{(x - 1)^2}$	1	7	$K10n10$	1009
1	$\frac{-x^4 + 2x^2}{(x - 1)^2(x + 1)^2}$	2	7	$K14n11913$	259
2	$\frac{-x^3 + 3x^2 - 4x}{(x - 1)^3}$	1	16	$t12766$	1459
2	$\frac{-x^6 + 3x^4 - 6x^2}{(x - 1)^3(x + 1)^3}$	2	18	$K15n93515$	82
3	$\frac{-x^4 + 4x^3 - 5x^2 + 6x}{(x - 1)^4}$	1	32	$K12n605$	219
3	$\frac{-x^5 + 3x^4 - 2x^3 + 2x^2 + 6x}{(x - 1)^4(x + 1)}$	2	26	$K11n34$	139
4	$\frac{-x^6 + 4x^5 - 5x^4 - 2x^3 - 2x^2 - 8x}{(x - 1)^5(x + 1)}$	2	42	$K14n1808$	62
4	$\frac{-x^5 + 5x^4 - 10x^3 + 10x^2 - 8x}{(x - 1)^5}$	1	66	$K12n214$	44
5	$\frac{-x^8 + 4x^7 - 3x^6 - 4x^5 + 14x^4 + 2x^3 + 14x^2 + 10x}{(x - 1)^6(x + 1)^2}$	2	88	$K15n15582$	11

$K11n34$. In dimension four, we were only able to visualize one of the very simplest examples in Fig. 4, and for dimension five we simply gave up.

All 38 examples where \mathcal{LW}_T is connected and impure have dimension 4 or 5; those of dimension 4 also have maximal faces of dimension 2 and those of dimension 5 also have maximal faces of dimension 3. One of the simplest such is $K13n857$ where \mathcal{LW}_T consists of seven 4-simplices plus two triangles, where the triangles are glued together to form a square, and then one edge of that square is glued to the main mass of 4-simplices.

7.4 Isotopies of lw-surfaces

An example of a non-normal isotopy of lw-surfaces occurs in the 13-tetrahedra triangulation:

$$T = nvLAAvAPQkcdfgfhkmjlmklmwcadtfaaoaedrg$$

Table 6 Six of the most complicated $B_M(x)$ in our sample. The properties recorded are the dimension of \mathcal{LW}_T (dim), the rational form $P(x)/Q(x)$ of $B_M(x)$, the period of $B_M(x)$ (per), the ℓ_1 -norm of the combined coefficients of P and Q (ℓ_1), and a manifold with this generating function (M)

dim	$B_M(x)$	per	ℓ_1	M
1	$\frac{-2x^8 - 4x^7 - 2x^6 + 6x^5 + 13x^4 + 8x^3 + 2x^2}{(x-1)^2(x+1)^2(x^2+x+1)^2}$	6	53	$K15n138922$
2	$\frac{-2x^6 + 5x^4 - 4x^3 - 15x^2 - 4x}{(x-1)^3(x+1)^3}$	2	38	$K15n27228$
2	$\frac{-2x^7 + 2x^6 - x^5 + x^4 - 9x^3 - 5x^2 - 4x}{(x-1)^3(x^2+x+1)^2}$	3	32	$K15n86383$
3	$\frac{-3x^8 + 13x^6 + 2x^5 - 14x^4 - 4x^3 + 17x^2 + 2x}{(x-1)^4(x+1)^4}$	2	71	$K15n139871$
4	$\frac{-2x^8 + 4x^7 + 4x^6 - 14x^5 - 12x^4 - 6x^3 - 22x^2 - 8x}{(x-1)^5(x+1)^3}$	2	94	$K13n1795$
5	$\frac{-x^7 + 5x^6 - 9x^5 + 5x^4 + 8x^3 + 10x}{(x-1)^6(x+1)}$	2	78	$K13n2458$

of the exterior of $K13n585$. To determine \mathcal{LW}_T , we enumerated normal and almost normal surfaces down to $\chi = -8$. In this range, there are 138 connected normal surfaces, 261 connected almost normal surfaces with octagons, and 603 almost normal surfaces with tubes. By tightening the almost normal surfaces, we found there are 11 connected essential lw-surfaces with $\chi \geq -8$ with four non-normal isotopies among them. Figure 5 shows the complex \mathcal{LW}_T which consists of an edge $B = [N_{12}, N_{23}]$ and a triangle $C = [N_{23}, N_4, N_7]$. For the face C it is the surface N_{116} that plays the role of F_C in steps (3–5) of Algorithm 6.12, and the isotopies $N_{115} \sim N_{116} \sim N_{118}$ are what determine the subspace W_C . Here, the subspace W_E for $E = [N_4, N_7]$ is the same as W_C . In general, even if a face E of C is parallel to W_C , it could be that W_E is a proper subspace of W_C ; see the example at the start of Section 5 of [64] for more on this important phenomenon. Notice also that $\text{dep}(C)$ is the complement of $\{N_{23}\} \cup E$ and that the surfaces N_{70} and N_{71} are *projectively isotopic* to a surface carried by the interior of C , but not isotopic to such a surface.

Because $W_E = W_C$, every isotopy class of essential lw-surface is uniquely represented by a surface carried by $B \cup [N_{23}, N_4]$. This allows us to easily compute that $B_M(x) = (-x^2 + 2x)/(x - 1)^2$. This is also what one gets from the triangulation:

$$S = nvLALAwAQkedffgijikmlmlmfvaetcaangcbn$$

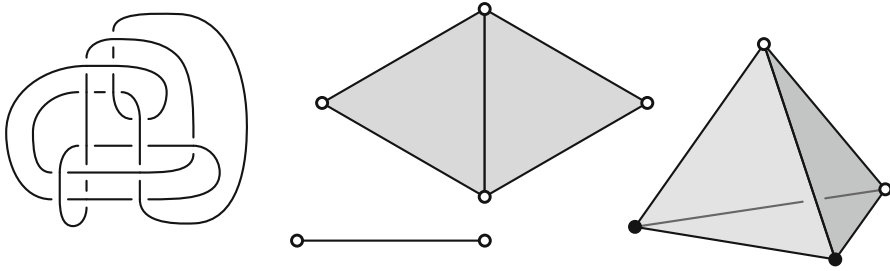


Fig. 1 For the knot $K15n51747$ shown at left, at right is the complex $\mathcal{LW}_{\mathcal{T}}$ for a triangulation of its exterior with 17 ideal tetrahedra. This example is unusual in that there are components of different dimensions. The vertex surfaces are either genus 2 (solid vertices) or genus 3 (open vertices). Here $B_M(x) = (-3x^7 + 3x^6 + 9x^5 - 9x^4 - 9x^3 + 9x^2 + 2x)/((x - 1)^4(x + 1)^3)$

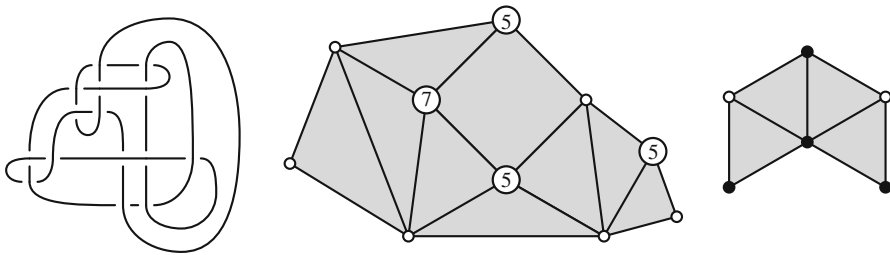


Fig. 2 For the knot $K15n18579$ shown at left, at right is the complex $\mathcal{LW}_{\mathcal{T}}$ for a triangulation of its exterior with 17 ideal tetrahedra. It is one of the most complicated examples in our sample with $\dim \mathcal{LW}_{\mathcal{T}} = 2$; note that one face is a square rather than a triangle. The vertex surfaces are genus 2 (solid vertices), genus 3 (open vertices), or genus 5 or 7 as labeled. Here $B_M(x) = (-2x^6 + 5x^4 - 4x^3 - 15x^2 - 4x)/((x - 1)^3(x + 1)^3)$

where $\mathcal{LW}_{\mathcal{S}}$ is two edges sharing a common vertex and there are no isotopies between essential lw-surfaces.

7.5 Barely large knots and those without meridional essential surfaces

A striking contrast in Table 2 is that there are more than 6,000 barely large manifolds in the cusped census yet only three such knot exteriors. Many constructions of closed essential surfaces in knot exteriors come from tubing essential surfaces with meridional boundary, and there are classes of knots where all closed essential surfaces are of this form, including Montesinos knots [53], alternating knots [48], and their generalizations [1, 2]. A connected meridional surface F in M can be tubed along ∂M in two distinct ways, resulting in a pair of disjoint surfaces; hence if both tubings are essential then M is very large. However, barely large knot exteriors do exist: Baker identified an infinite family of barely large knots with a single incompressible genus 2 surface in [6, §4.7.1]. Additionally, Adams-Reid [5] and Eudave-Muñoz [25]

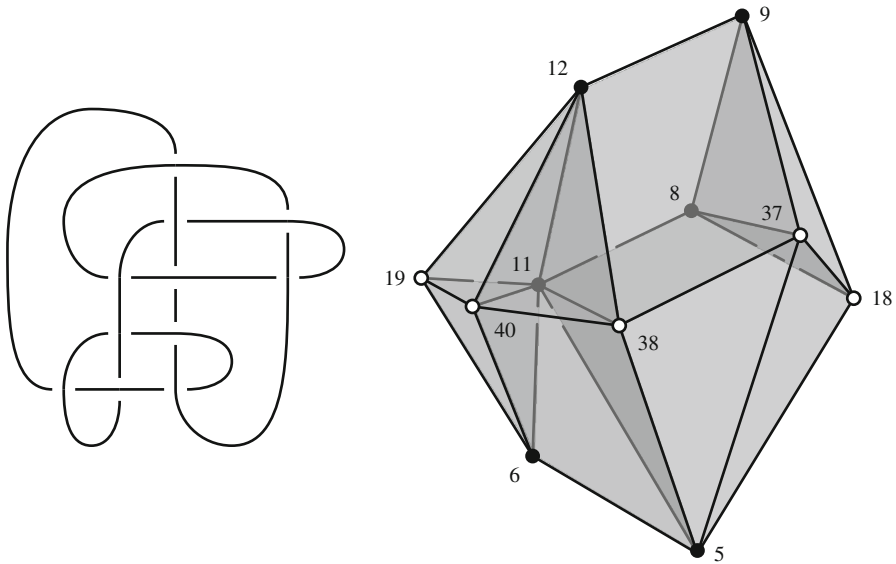


Fig. 3 The complex $\mathcal{LW}_{\mathcal{T}}$ for a triangulation of the exterior of the Conway knot $K 11n34$. The 11 vertex surfaces are $N_5, N_6, N_8, N_9, N_{11}, N_{12}, N_{18}, N_{19}, N_{37}, N_{38}, N_{40}$, where the notation follows [23]. The first six surfaces have genus 2 (dark vertices above) and the rest genus 3 (white vertices above). There are seven 3-dimensional faces: four tetrahedra, a pyramid with quadrilateral base ($N_5, N_6, N_{40}, N_{38}, N_{11}$), a triangular prism ($N_{11}, N_{12}, N_{38}, N_8, N_9, N_{37}$), and the one in the lower right whose faces are four triangles and two quadrilaterals ($N_{11}, N_8, N_{37}, N_{38}$ and N_{11}, N_8, N_{18}, N_5). Here $B_M(x) = (-x^5 + 3x^4 - 2x^3 + 2x^2 + 6x)/((x + 1)(x - 1)^4)$

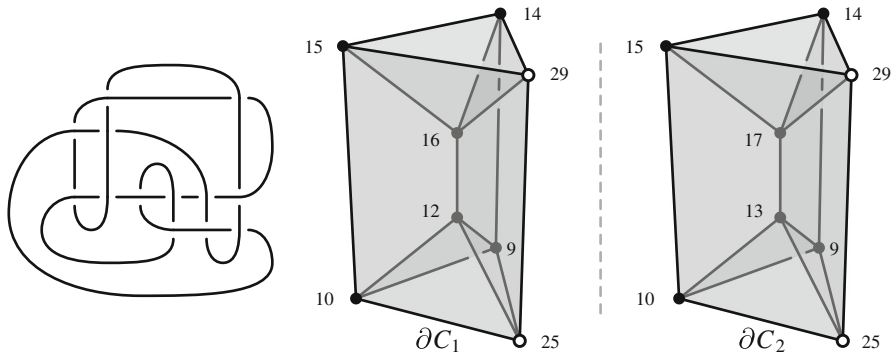


Fig. 4 For a 15-tetrahedra triangulation of the exterior of the knot $K 13n1019$ shown at left, the complex $\mathcal{LW}_{\mathcal{T}}$ consists of two 4-dimensional faces C_1 and C_2 with the same combinatorics that are glued together along a single 3-dimensional face. The boundary of each C_i is depicted above via an identification of ∂C_i with S^3 ; hence, in each case there is an additional face of ∂C_i on the outside, namely a triangular prism whose vertices are $N_{10}, N_{25}, N_9, N_{15}, N_{29}, N_{14}$. It is these two outside faces that are identified to form $\mathcal{LW}_{\mathcal{T}}$. As usual, solid and open vertices correspond to surfaces of genus 2 and 3 respectively, and the numbering follows [23]. Here $B_M(x) = (-x^5 + 5x^4 - 10x^3 + 10x^2 - 8x)/(x - 1)^5$

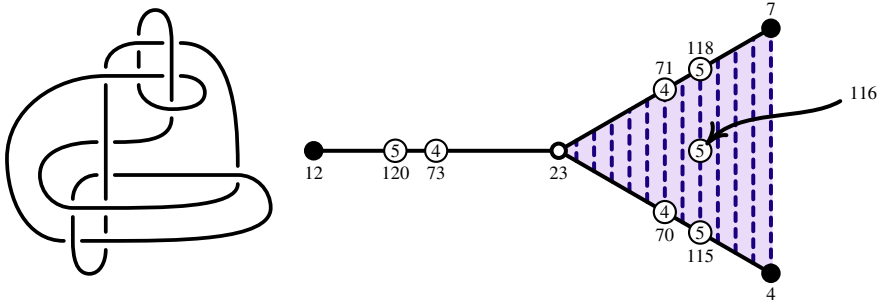


Fig. 5 For this triangulation of the exterior of $K13n585$, there are 11 connected lw-surfaces down to $\chi = -8$: three of genus 2 (N_4, N_7, N_{12}), one of genus 3 (N_{23}), three of genus 4 ($N_{70} = N_4 + N_{23}$ and $N_{71} = N_7 + N_{23}$ and $N_{73} = N_{12} + N_{23}$), and four of genus 5 ($N_{115} = 2N_4 + N_{23}$ and $N_{116} = N_4 + N_7 + N_{23}$ and $N_{117} = 2N_7 + N_{23}$ and $N_{120} = 2N_{12} + N_{23}$). The complete list of isotopies between them is: $N_4 \sim N_7, N_{70} \sim N_{71}$, and $N_{115} \sim N_{116} \sim N_{118}$. The complex \mathcal{LWT} consists of the edge $B = [N_{12}, N_{23}]$ and the triangle $C = [N_{23}, N_4, N_7]$ shown above. Here $W_B = 0$, but for the faces C and $E = [N_4, N_7]$, the subspaces W_C and W_E are both 1-dimensional; indeed, $W_C = W_E$ with the induced decomposition of C into projective isotopy classes indicated by the dashed vertical lines

gave examples of closed essential surfaces that cannot come from a meridional tubing construction. Still, the following appears to be new:

7.6 Theorem *There exists a knot in S^3 , namely $K15n153789$, whose exterior is large (indeed, barely large) and where the meridian is not the boundary slope of any essential surface.*

Here, the knot $K15n153789$ is one of the three examples of barely large knots we found; its exterior contains a unique essential surface, which has genus 2. We checked the boundary slope condition by noting that there are no spinnormal surfaces with meridional boundary slope in the triangulation:

$$kLLLzPQkccfegjihijlnahwdavhpk_bBaB$$

of its exterior.

7.7 Code and data

Complete data and the code used to compute it are available at [23]. Regina [8] was used as the underlying engine for triangulations and normal surface computations, including enumeration of vertex and fundamental (almost) normal surfaces, and Normaliz [11] was used for computing $B_M(x)$, with the whole computation taking place inside SageMath [57] using the Python wrappings of these libraries. The code for dealing with almost normal surfaces with tubes, tightening almost normal surfaces (with either tubes or octagons), and implementing Algorithm 6.12 was all completely new.

To help validate our code, we started with a sample of 6,510 of the manifolds from Table 2 and generated 5 random triangulations of each. Then the complete algorithm was run on all 32,550 triangulations and the output compared to ensure that each triangulation gave the same surface counts and other associated data. This technique proved extremely effective at finding bugs in the code (and, if we are being honest, our thinking), including subtle ones that only manifest themselves in corner cases. Additionally, we compared our data to the lists of which knots are small/large from [9]; on the common set of 1,764 knots, our data matched theirs exactly.

8 Patterns of surface counts by genus

We now return to the question of counting *connected* essential surfaces in a given 3-manifold in terms of their genus. As we know how to count all essential surfaces by Euler characteristic given the complex \mathcal{LW}_T , we approach this by identifying the connected surfaces in that larger count. This problem has an arithmetic flavor, and is related to counting primitive lattice points, as well as to the work of Mirzakhani discussed in Sect. 1.5.

Let $a_M(g)$ denote the number of isotopy classes of connected essential surfaces of genus g . For each of the 4,330 very large examples in Table 2, we computed the first 20 values of $a_M(g)$ starting from \mathcal{LW}_T as follows. Let g be fixed. For each face C of \mathcal{LW}_T , let $\tilde{C} = \mathbb{R}_{\geq 0} \cdot C \subset \mathcal{S}_T$ be the corresponding cone. We used Normaliz [11] to find all lattice points carried by the interior of the rational polytope $\{\vec{x} \in \tilde{C} \mid \chi(\vec{x}) = 2 - 2g\}$. For each corresponding normal surface, we checked connectivity using Algorithm 9.4 of [42]. As $W_C = 0$ for all these examples, the number of such lattice points corresponding to connected surfaces is the contribution of the interior C° to $a_M(g)$.

We found 94 distinct patterns for $(a_M(2), \dots, a_M(21))$. Table 7 lists the most common patterns and Table 8 gives the most complicated. For one manifold exhibiting each pattern, we computed additional values of $a_M(g)$, nearly always up to at least $g = 50$ and in more than 40 cases up to $g = 200$. This data is available at [23], where the largest single value is $a_M(51) = 3,072,351$ for the exterior M of $K15n33595$.

From now on, we work with $n = g - 1$ rather than g as the index for the count of connected surfaces, and so define $\tilde{a}_M(n) = a_M(n + 1)$; this simplifies the arithmetic, for example giving $\tilde{a}_M(n) \leq b_M(-2n)$ rather than $a_M(g) \leq b_M(2 - 2g)$.

8.1 Independence of $a_M(g)$ and $B_M(x)$

We next give four examples showing that neither $\tilde{a}_M(n)$ nor $B_M(x)$ determines the other. We start with two manifolds with the same $B_M(x)$ but different

Table 7 The ten most common patterns of $a_M(g)$ for $2 \leq g \leq 21$, which together account for 3629 (83.8%) of the 4330 very large manifolds. A sample manifold for each pattern is given in the second column, and the final column is the number of times the pattern appears

$a_M(g)$	M	Count
4, 2, 4, 4, 8, 4, 12, 8, 12, 8, 20, 8, 24, 12, 16, 16, 32, 12, 36, 16	$t09753$	1473
2, 1, 2, 2, 4, 2, 6, 4, 6, 4, 10, 4, 12, 6, 8, 8, 16, 6, 18, 8	$t12198$	918
0, 2, 0, 1, 0, 2, 0, 2, 0, 4, 0, 2, 0, 6, 0, 4, 0, 6, 0, 4	$K14n11913$	259
6, 4, 8, 8, 16, 8, 24, 16, 24, 16, 40, 16, 48, 24, 32, 32, 64, 24, 72, 32	$K12n605$	219
8, 4, 8, 8, 16, 8, 24, 16, 24, 16, 40, 16, 48, 24, 32, 32, 64, 24, 72, 32	$K11n73$	169
0, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0	$K14n13645$	148
6, 9, 24, 37, 86, 87, 208, 220, 366, 386, 722, 602, 1168, 1039, 1498, 1564, 2514, 1993, 3484, 2924	$K11n34$	139
6, 7, 18, 29, 64, 73, 156, 177, 290, 321, 550, 521, 896, 865, 1236, 1297, 1950, 1731, 2714, 2499	$K11n42$	131
2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0	$o937085$	91
0, 6, 0, 5, 0, 12, 0, 16, 0, 31, 0, 28, 0, 58, 0, 53, 0, 82, 0, 79	$K15n93515$	82

Table 8 The eight of the most complicated patterns of $a_M(g)$ for $2 \leq g \leq 21$. These all come from examples where $\dim \mathcal{LW}_{\mathcal{T}} \geq 4$

$a_M(g)$	M
8, 14, 46, 89, 224, 305, 674, 905, 1536, 1955, 3326, 3771, 6150, 7019, 9850, 11611, 16714, 17767, 25490, 27415	K12n214
8, 16, 54, 98, 264, 318, 806, 984, 1794, 2098, 3994, 4074, 7368, 7632, 11552, 12976, 20114, 19396, 30670, 30550	K12n210
12, 21, 61, 109, 261, 320, 721, 880, 1480, 1762, 3094, 3115, 5429, 5666, 8019, 9086, 13596, 13059, 20062, 19841	K13n3763
10, 25, 71, 140, 352, 473, 1058, 1386, 2389, 2939, 5152, 5585, 9422, 10311, 14887, 17057, 25304, 25573, 38238, 39603	K15n15582
12, 16, 51, 99, 235, 345, 711, 999, 1649, 2209, 3551, 4319, 6593, 7919, 10971, 13231, 18275, 20555, 28063, 31485	K15n15220
8, 18, 57, 110, 270, 356, 785, 1013, 1737, 2092, 3667, 3942, 6614, 7134, 10397, 11710, 17426, 17422, 26131, 26891	K15n23198
12, 34, 110, 216, 532, 708, 1558, 2018, 3462, 4176, 7314, 7876, 13204, 14256, 20778, 23404, 34820, 34832, 52226, 53766	K13n3838
12, 30, 109, 231, 549, 861, 1737, 2511, 4059, 5643, 8859, 10941, 16623, 20229, 27303, 33729, 46215, 52455, 71079, 80271	K15n33595
10, 21, 73, 143, 385, 513, 1224, 1605, 2870, 3542, 6409, 7010, 12051, 13231, 19463, 22436, 33614, 34307, 51700, 53862	K13n2458

$\tilde{a}_M(n)$. Let A and B be the exteriors of the knots $K14n22185$ and $K13n586$ respectively, and we use \mathcal{T} and \mathcal{S} to denote their standard triangulations. Both $\mathcal{LW}_{\mathcal{T}}$ and $\mathcal{LW}_{\mathcal{S}}$ consist of a single edge C whose vertices correspond to genus 2 surfaces F and G . Moreover, the lattice points in the cone over C are simply $u\vec{F} + v\vec{G}$ for $u, v \in \mathbb{N}$. Thus, the surfaces with $\chi = -2n$ are the lattice points in \mathbb{N}^2 on the line $x + y = n$, which gives $b_M(-2n) = n + 1$ and hence $B_M(x) = (-x^2 + 2x)/(x - 1)^2$.

For \mathcal{T} , the surfaces F and G can be made disjoint after a normal isotopy, and hence every normal surface carried by C is a disjoint union of parallel copies of F and G . Thus the only connected essential surfaces in A are F and G , giving $\tilde{a}_A(1) = 2$ and $\tilde{a}_A(n) = 0$ for $n > 1$. In contrast, we find that the first 30 values of $\tilde{a}_B(n)$ are:

2, 1, 2, 2, 4, 2, 6, 4, 6, 4, 10, 4, 12, 6, 8, 8, 16, 6, 18, 8, 12, 10, 22, 8, 20,
12, 18, 12, 28, 8

Now, if $u\vec{F} + v\vec{G}$ is connected then $\gcd(u, v) = 1$. The above data is consistent with the converse being true, or equivalently $\tilde{a}_B(n)$ is exactly the number of primitive lattice points in \mathbb{N}^2 on the line $x + y = n$, which is the Euler totient function $\phi(n)$ when $n > 1$. This pattern continues for all $n \leq 500$, so we may safely posit:

8.2 Conjecture *For the exterior B of the knot $K13n586$, one has $\tilde{a}_B(n) = \phi(n)$ for all $n > 1$.*

Since $\tilde{a}_B(1) = 2$, equivalently the conjecture is that $\tilde{a}_B(n) = \epsilon(n) + \phi(n)$ for all $n \geq 1$ where $\epsilon(n)$ is 1 when $n = 1$ and 0 otherwise. This count of primitive lattice points can be related to the corresponding count of all lattice points via the Möbius inversion formula, making Conjecture 8.2 equivalent to:

$$\tilde{a}_B(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) (d + 1) \tag{8.3}$$

for all $n \geq 1$, where μ is the Möbius function. *Note added in proof:* Lee [46] has recently proved Conjecture 8.2 by analyzing the number of connected components of $u\vec{F} + v\vec{G}$ using the method of [3].

A pair with the same $\tilde{a}_M(n)$ but different $B_M(x)$ are the census manifolds $X = v3394$ and $Y = o9_{43058}$. Both have exactly four connected essential surfaces, all of genus two, but

$$B_X(x) = \frac{-2x^2 + 4x}{(x - 1)^2} \quad \text{and} \quad B_Y(x) = \frac{-x^2 + 4x}{(x - 1)^2}.$$

For the standard triangulations \mathcal{X} and \mathcal{Y} of X and Y , the complexes $\mathcal{LW}_{\mathcal{X}}$ and $\mathcal{LW}_{\mathcal{Y}}$ are quite different: the first consists of an edge and a disjoint vertex, but $\mathcal{LW}_{\mathcal{Y}}$ consists of two edges sharing a common vertex. All vertex surfaces are connected, and so the vertices of $\mathcal{LW}_{\mathcal{X}}$ and $\mathcal{LW}_{\mathcal{Y}}$ correspond to three of the four essential genus 2 surfaces; in both cases, the fourth is hiding as a fundamental surface in the interior of an edge.

8.4 Regular genus counts and the Lambert series

For the manifold $B = K13n586$ in Sect. 8.1, while the count $\tilde{a}_B(n)$ does not have a short generating function, from (8.3) we see its Möbius transform $p(n) = \sum_{d|n} \tilde{a}_B(d)$ is a polynomial, specifically $p(n) = n + 1$. This motivates our next definition. Recall that Dirichlet convolution on arithmetic functions $f, g: \mathbb{Z}_{\geq 1} \rightarrow \mathbb{C}$ is defined by $(f * g)(n) = \sum_{d|n} f(n/d)g(d)$. We say that $\tilde{a}_M(n)$ is *regular* if $1 * \tilde{a}_M$ has a short generating series. Equivalently, if we set $p_M = 1 * \tilde{a}_M$, regularity is equivalent to $p_M(n)$ being a quasi-polynomial for all large n . Thus, when $\tilde{a}_M(n)$ is regular, we can use Möbius inversion $\tilde{a}_M(n) = (\mu * p_M)(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) p_M(d)$ to compute \tilde{a}_M from the simpler p_M , as we did in (8.3). In the language of generating functions, the count \tilde{a}_M is regular if and only if its Lambert series

$$L_{A_M}(x) = \sum_{n=1}^{\infty} \tilde{a}_M(n) \frac{x^n}{1 - x^n} \tag{8.5}$$

is short, since the coefficients of this series are precisely $1 * \tilde{a}_M$.

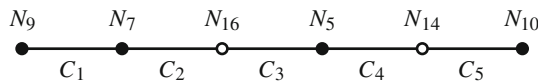
8.6 Examples

Of the 94 observed patterns for $\tilde{a}_M(n)$ in our sample, we conjecture that exactly 54 of them are regular, including 7 of the 10 manifolds in Table 7, with the exceptions being $K11n34$, $K11n42$, and $K15n93515$. Examples of our conjectured formulae for $L_{A_M}(x)$ are given in Table 9. In contrast, we believe all the examples in Table 8 are irregular.

One example where the count appears irregular, though still highly structured, is $M = o9_{41176}$. Specifically we conjecture that $\tilde{a}_M(n)$ is equal to $f(n) = \phi(n) + 1$ for $n \geq 2$ (here $\tilde{a}_M(1) = 5$). While f is quite simple, we have $1 * f = n + \sigma_0$ where $\sigma_0(n)$ is the number of divisors of n , and $\sigma_0(n)$ does not have a short generating function. For the standard triangulation \mathcal{T} of M , the complex $\mathcal{LW}_{\mathcal{T}}$ is:

Table 9 Eight examples of our conjectured Lambert series $LA_M(x)$ for manifolds where $\tilde{a}_M(n)$ appears regular. The first five are from Table 7 and the last three are among the most complicated we found

$LA_M(x)$	Per	ℓ_1	M
$\frac{-2x^2 + 4x}{(x - 1)^2}$	1	10	t09753
$\frac{-x^2 + 2x}{(x - 1)^2}$	1	7	t12198
$\frac{-x^4 + 2x^2}{(x - 1)^2(x + 1)^2}$	2	7	K14n11913
$\frac{-2x^2 + 6x}{(x - 1)^2}$	1	12	K12n605
$\frac{-4x^2 + 8x}{(x - 1)^2}$	1	16	K11n73
$\frac{-4x^6 + 2x^5 + 16x^4 + 4x^3 - 14x^2 - 6x}{(x - 1)^3(x + 1)^3}$	2	54	K15n67261
$\frac{-2x^8 - 4x^7 - 2x^6 + 4x^5 + 9x^4 + 6x^3 + 2x^2}{(x - 1)^2(x + 1)^2(x^2 + x + 1)^2}$	6	45	K15n129923
$\frac{-2x^8 - 4x^7 - 2x^6 + 6x^5 + 13x^4 + 8x^3 + 2x^2}{(x - 1)^2(x + 1)^2(x^2 + x + 1)^2}$	6	53	K15n138922



using the conventions of [23]. Here the vertex surfaces N_{14} and N_{16} have genus 3 and the others have genus 2. We conjecture that the faces contribute to \tilde{a}_M as follows:

- (a) The interior of C_1 carries a single connected surface N_8 which has genus 2. Here $N_7 + N_9 = 2N_8$.
- (b) The interior of C_2 carries no connected surfaces as N_7 and N_{16} are disjoint.
- (c) The interior of C_3 carries a unique surface of genus g for each $g \geq 4$, namely $(g - 3)N_5 + N_{16}$. It is this face that contributes the $+1$ to $\tilde{a}_M(n)$.
- (d) The connected surfaces carried by C_4 are exactly $uN_5 + vN_{14}$ for $u, v > 0$ and $\text{gcd}(u, v) = 1$. The situation is the same for C_5 , with N_5 replaced with N_{10} . Together, these faces contribute the $\phi(n)$ to $\tilde{a}_M(n)$.

The manifold $N = t12071$ is similar in that $\tilde{a}_N(n)$ is irregular but $\tilde{a}_N(n) - 4$ appears regular.

In our sample, a simple example where \tilde{a}_M appears irregular and where we cannot glean any other structure is $W = o9_{42517}$. The first 50 values of $\tilde{a}_W(n)$ are:

6, 4, 10, 14, 26, 26, 52, 46, 76, 76, 118, 96, 172, 136, 194, 196, 286, 212,
 354, 274, 388, 360, 506, 378, 604, 490, 634, 574, 820, 568, 948, 728, 946,
 846, 1122, 864, 1356, 1040, 1316, 1146, 1644, 1140, 1800, 1392, 1716,
 1570, 2136, 1506, 2332, 1752

and [23] has all values to $n = 200$. With its usual triangulation \mathcal{W} , the complex $\mathcal{LW}_{\mathcal{W}}$ is a triangle whose vertex surfaces N_3, N_9 , and N_{11} all have genus 2. Here the edge $[N_3, N_9]$ appears to carry a single connected surface in its interior, which has genus 2, and the same for $[N_9, N_{11}]$. The remaining edge $[N_3, N_{11}]$ appears to contribute $2\phi(n)$ to $\tilde{a}_W(n)$ for $n > 1$, and the interior of the triangle contributes the mysterious:

0, 2, 6, 10, 18, 22, 40, 38, 64, 68, 98, 88, 148, 124, 178, 180, 254, 200, 318,
 258, 364, 340

8.7 Asymptotics of genus counts

We now explore the asymptotics of the sequences $\tilde{a}_M(n)$. Since $\tilde{a}_M(n) \leq b_M(-2n)$ and the latter grows polynomially, it is natural to ask whether $\tilde{a}_M(n)$ does so as well. Even in the regular case, the sequence $\tilde{a}_M(n)$ depends arithmetically on the divisors of n , so it is better to study the smoothed sequence:

$$\bar{a}_M(n) = \sum_{k \leq n} \tilde{a}_M(k). \tag{8.8}$$

Of the 94 observed patterns for $\tilde{a}_M(n)$, there are 14 where $\tilde{a}_M(n) = 0$ for all large n and 4 where we were only able to compute up to $\tilde{a}_M(20)$; we consider only the remaining 76. The plots in Figs. 6 and 7 together suggest:

8.9 Conjecture *Suppose M as in Theorem 1.3. Then either $\tilde{a}_M(n) = 0$ for all large n or there exists $s \in \mathbb{N}$ such that $\lim_{n \rightarrow \infty} \bar{a}_M(n)/n^s$ exists and is positive.*

In fact, Conjecture 8.9 holds whenever $\tilde{a}_M(n)$ is regular and the corresponding quasi-polynomial has constant leading term as we now show; this includes all the conjecturally regular examples in our sample.

8.10 Lemma *Suppose $\tilde{a}_M(n)$ is regular and the corresponding $p_M(n)$ has constant leading term, with $p_M(n) = c_r n^r + O(n^{r-1})$ for some $r \geq 1$ and positive $c_r \in \mathbb{Q}$. Then*

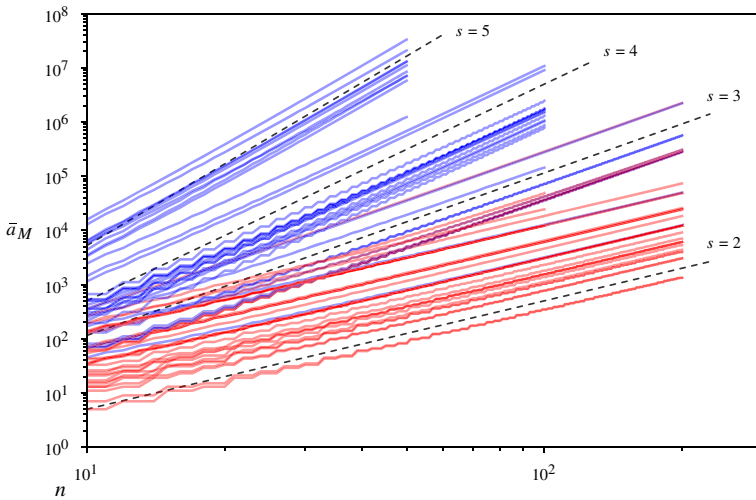


Fig. 6 This log–log plot shows the sequence $\{\bar{a}_M(n)\}$ for 76 manifolds, up to $n = 50, 100,$ or 200 depending. Those coming from conjecturally regular $\bar{a}_M(n)$ are in red whereas the likely irregular ones are in blue. The dotted lines plot $c_s n^s$ for the indicated s and some choice of c_s . Each of $\bar{a}_M(n)$ appears nearly parallel to one of these lines, consistent with $\bar{a}_M(n)$ being asymptotic to cn^s as $n \rightarrow \infty$ for some integer s and $c_s > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{n^{r+1}} \bar{a}_M(n) = \frac{c_r}{r+1} \frac{1}{\zeta(r+1)}, \tag{8.11}$$

where $\zeta(s)$ is the Riemann ζ -function.

Proof As $\sum_{k=1}^m k^s = \frac{1}{s+1} m^{s+1} + O(m^s)$, we see $\sum_{k \leq m} p_M(k) = \frac{c_r}{r+1} m^{r+1} + O(m^r)$. Now

$$\begin{aligned} \bar{a}_M(n) &= \sum_{\ell \leq n} \sum_{d|\ell} \mu(d) p_M(\ell/d) = \sum_{d \cdot k \leq n} \mu(d) p_M(k) = \sum_{d \leq n} \left(\mu(d) \sum_{k \leq \lfloor n/d \rfloor} p_M(k) \right) \\ &= \sum_{d \leq n} \mu(d) \left(\frac{c_r}{r+1} \left\lfloor \frac{n}{d} \right\rfloor^{r+1} + o\left(\left\lfloor \frac{n}{d} \right\rfloor^r\right) \right) = \sum_{d \leq n} \mu(d) \left(\frac{c_r}{r+1} \frac{n^{r+1}}{d^{r+1}} + o\left(\frac{n^r}{d^r}\right) \right), \end{aligned}$$

where we have used that $\lfloor n/d \rfloor = n/d + O(1)$ and hence by the binomial theorem $\lfloor n/d \rfloor^{r+1} = n^{r+1}/d^{r+1} + O(n^r/d^r)$. Thus

$$\frac{\bar{a}_M(n)}{n^{r+1}} = \sum_{d \leq n} \mu(d) \left(\frac{c_r}{r+1} \frac{1}{d^{r+1}} + o\left(\frac{1}{nd^r}\right) \right) = \left(\frac{c_r}{r+1} \sum_{d \leq n} \frac{\mu(d)}{d^{r+1}} \right) + o\left(\frac{\log(n)}{n}\right),$$

where we have used in the last step that $\sum_{d \leq n} 1/d^r \leq \sum_{d \leq n} 1/d \approx \log(n)$. Now as $r \geq 1$, we have $\sum_{d \leq n} \mu(d)/d^{r+1} = 1/\zeta(r+1) + o(n)$ by [4, Section 11.4], and so the lemma follows. \square

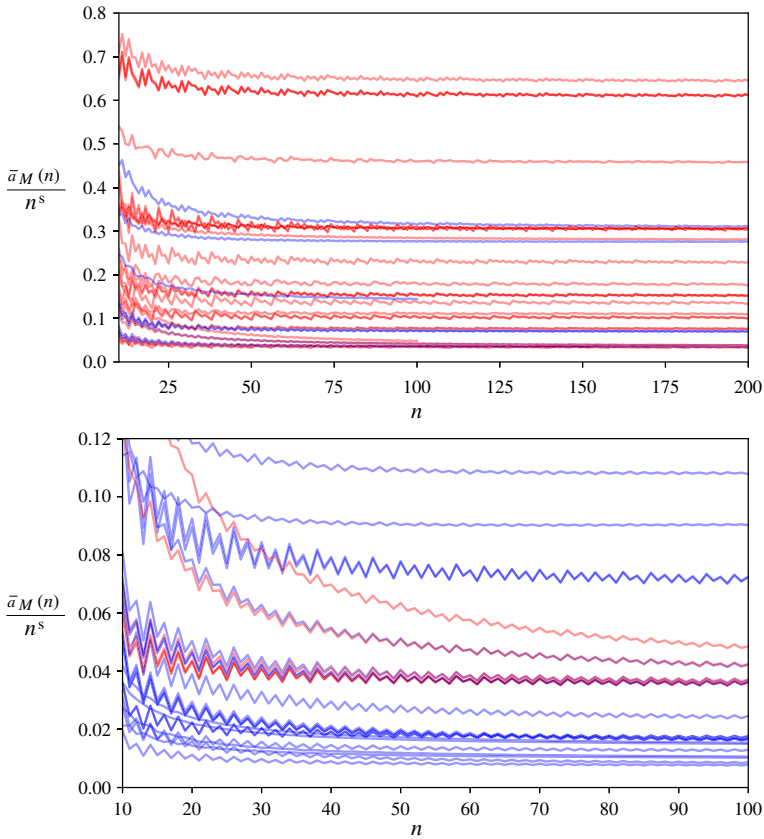


Fig. 7 Using the predicted asymptotic exponent s for each sequence $\bar{a}_M(n)$ from Fig. 6, we plot $\bar{a}_M(n)/n^s$ to test Conjecture 8.9. Here, the top plot shows those where $s = 2, 3$ and the bottom where $s = 4, 5$; again, red and blue correspond to (conjecturally) regular versus irregular sequences. For better readability, 10 sequences that lie above the given vertical scales are omitted, 9 from the top plot (all but one regular) and 1 from the bottom; these look very similar to the 66 sequences shown

8.12 Remark If one allows $r = 0$, then (8.11) still holds if you interpret the righthand side as 0, seeing that ζ has a pole at 1; we leave the details in this case to the reader.

Acknowledgements We thank Alexander Barvinok and Josephine Yu for discussions about counting lattice points that were crucial to the proof of Theorem 1.3, as well as Craig Hodgson for helpful discussions on several related projects. We also thank the referee for their very careful reading of this paper and detailed comments. Dunfield was partially supported by U.S. National Science Foundation grant DMS-1811156, and Rubinstein partially supported by Australian Research Council grant DP160104502.

References

1. Adams, C.C., Brock, J.F., Bugbee, J., Comar, T.D., Faigin, K.A., Huston, A.M., Joseph, A.M., Pesikoff, D.: Almost alternating links. *Topol. Appl.* **46**(2), 151–165 (1992)
2. Adams, C.C.: Toroidally alternating knots and links. *Topology* **33**(2), 353–369 (1994)
3. Agol, I., Hass, J., Thurston, W.: The computational complexity of knot genus and spanning area. *Trans. Am. Math. Soc.* **358**(9), 3821–3850 (2006)
4. Apostol, T.: *Introduction to Analytic Number Theory*. Springer, New York (1976). Undergraduate Texts in Mathematics
5. Adams, C.C., Reid, A.W.: Quasi-Fuchsian surfaces in hyperbolic knot complements. *J. Austral. Math. Soc. Ser. A* **55**(1), 116–131 (1993)
6. Baker, K.L.: *Knots on Once-Punctured Torus Fibers*. ProQuest LLC, Ann Arbor (2004). Thesis (Ph.D.)—The University of Texas at Austin
7. Barvinok, A.I.: A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed. *Math. Oper. Res.* **19**(4), 769–779 (1994)
8. Burton, B.A., Budney, R., Pettersson, W., et al.: *Regina: Software for low-dimensional topology*. <http://regina-normal.github.io/> (1999–2020)
9. Burton, B.A., Coward, A., Tillmann, S.: Computing closed essential surfaces in knot complements. In: *Computational Geometry (SoCG'13)*, pp. 405–413. ACM, New York (2013)
10. Bell, M.C.: Experimental statistics for Mirzakhani's theorem (**preprint**) (2019)
11. Bruns, W., Ichim, B., Römer, T., Sieg, R., Söger, C.: *Normaliz algorithms for rational cones and affine monoids (version 3.7)*. <https://www.normaliz.uni-osnabrueck.de>
12. Bonahon, F.: Cobordism of automorphisms of surfaces. *Ann. Sci. École Norm. Sup.* (4) **16**(2), 237–270 (1983)
13. Burton, B.A., Tillmann, S.: Computing closed essential surfaces in 3-manifolds 36 pages (**preprint**) (2018). [arXiv:1812.11686](https://arxiv.org/abs/1812.11686)
14. Burton, B.A.: The cusped hyperbolic census is complete, 32 pages (**preprint**) (2014). [arXiv:1405.2695](https://arxiv.org/abs/1405.2695)
15. Burton, B.A.: Converting between quadrilateral and standard solution sets in normal surface theory. *Algebr. Geom. Topol.* **9**(4), 2121–2174 (2009)
16. Burton, B.: Quadrilateral-octagon coordinates for almost normal surfaces. *Exp. Math.* **19**(3), 285–315 (2010)
17. Burton, B.A.: Enumerating fundamental normal surfaces: algorithms, experiments and invariants. In: *Proceedings of the Meeting on Algorithm Engineering & Experiments*, pp. 112–124, USA (2014). Society for Industrial and Applied Mathematics
18. Buser, P.: *Geometry and Spectra of Compact Riemann Surfaces*. Progress in Mathematics, vol. 106. Birkhäuser, Boston (1992)
19. Barvinok, A., Woods, K.: Short rational generating functions for lattice point problems. *J. Am. Math. Soc.* **16**(4), 957–979 (2003)
20. Culler, M., Dunfield, N.M., Goerner, M., Weeks, J.R.: SnapPy, a computer program for studying the topology and geometry of 3-manifolds, version 2.7.1. <http://snappy.computop.org>
21. Dimofte, T., Gaiotto, D., Gukov, S.: 3-manifolds and 3d indices. *Adv. Theor. Math. Phys.* **17**(5), 975–1076 (2013)
22. Dimofte, T., Gaiotto, D., Gukov, S.: Gauge theories labelled by three-manifolds. *Comm. Math. Phys.* **325**(2), 367–419 (2014)
23. Dunfield, N., Garoufalidis, S., Rubinstein, J. H.: Code and data to accompany this paper, Harvard Dataverse (2020). <https://doi.org/10.7910/DVN/FZIHMB>
24. Ehrhart, E.: Sur les polyèdres homothétiques bordés à n dimensions. *C. R. Acad. Sci. Paris* **254**, 988–990 (1962)
25. Eudave-Muñoz, M.: Incompressible surfaces and $(1, 1)$ -knots. *J. Knot Theory Ramific.* **15**(7), 935–948 (2006)

26. Floyd, W., Oertel, U.: Incompressible surfaces via branched surfaces. *Topology* **23**(1), 117–125 (1984)
27. Gadre, V.S.: Dynamics of non-classical interval exchanges. *Ergod. Theory Dyn. Syst.* **32**(6), 1930–1971 (2012)
28. Garoufalidis, S.: The degree of a q -holonomic sequence is a quadratic quasi-polynomial. *Electron. J. Combin.* **18**(2), 4–23 (2011)
29. Garoufalidis, S.: The Jones slopes of a knot. *Quantum Topol.* **2**(1), 43–69 (2011)
30. Garoufalidis, S., Hodgson, C.D., Hoffman, N.R., Hyam Rubinstein, J.: The 3D-index and normal surfaces. III. *J. Math.* **60**(1), 289–352 (2016)
31. Garoufalidis, S., Hodgson, C.D., Hyam Rubinstein, J., Segerman, H.: 1-efficient triangulations and the index of a cusped hyperbolic 3-manifold. *Geom. Topol.* **19**(5), 2619–2689 (2015)
32. Garoufalidis, S., Lê, T.T.Q.: The colored Jones function is q -holonomic. *Geom. Topol.* **9**, 1253–1293 (2005)
33. Hass, J.: Acylindrical surfaces in 3-manifolds. *Mich. Math. J.* **42**(2), 357–365 (1995)
34. Hatcher, A.: Measured Lamination Spaces for 3-manifolds, 36 pages (**preprint**) (1999). <https://pi.math.cornell.edu/~hatcher/Papers/ML.pdf>
35. Hodgson, C., Rubinstein, J.H., Segerman, H.: Triangulations of hyperbolic 3-manifolds admitting strict angle structures. *J. Topol.* **5**(4), 887–908 (2012)
36. Hatcher, A., Thurston, W.: Incompressible surfaces in 2-bridge knot complements. *Invent. Math.* **79**(2), 225–246 (1985)
37. Hoste, J., Thistlethwaite, M., Weeks, J.: The first 1,701,936 knots. *Math. Intell.* **20**(4), 33–48 (1998)
38. Jaco, W., Oertel, U.: An algorithm to decide if a 3-manifold is a Haken manifold. *Topology* **23**(2), 195–209 (1984)
39. Jaco, W., Rubinstein, H.: Annular-efficient triangulations of 3-manifolds, 21 pages (**preprint**) (2011). [arXiv:1108.2936](https://arxiv.org/abs/1108.2936)
40. Jaco, W., Hyam Rubinstein, J.: 0-efficient triangulations of 3-manifolds. *J. Differ. Geom.* **65**(1), 61–168 (2003)
41. Jaco, W., Rubinstein, H., Spreer, J., Tillmann, S.: On minimal ideal triangulations of cusped hyperbolic 3-manifolds. *J. Topol.* **13**(1), 308–342 (2020)
42. Jaco, W., Tollefson, J.L.: Algorithms for the complete decomposition of a closed 3-manifold. III. *J. Math.* **39**(3), 358–406 (1995)
43. Kahn, J., Marković, V.: Counting essential surfaces in a closed hyperbolic three-manifold. *Geom. Topol.* **16**(1), 601–624 (2012)
44. Lackenby, M.: Word hyperbolic Dehn surgery. *Invent. Math.* **140**(2), 243–282 (2000)
45. Lackenby, M.: An algorithm to determine the Heegaard genus of simple 3-manifolds with nonempty boundary. *Algebr. Geom. Topol.* **8**(2), 911–934 (2008)
46. Lee, C.: Essential surfaces in the exterior of $K13n586$, 13 pages (**preprint**) (2021). [arXiv:2110.10231](https://arxiv.org/abs/2110.10231)
47. Matveev, S.: Algorithmic Topology and Classification of 3-Manifolds, volume 9 of Algorithms and Computation in Mathematics, 2nd edn. Springer, Berlin (2007)
48. Menasco, W.: Closed incompressible surfaces in alternating knot and link complements. *Topology* **23**(1), 37–44 (1984)
49. Mirzakhani, M.: Growth of the number of simple closed geodesics on hyperbolic surfaces. *Ann. Math. (2)* **168**(1), 97–125 (2008)
50. Morgan, J.W., Shalen, P.B.: Valuations, trees, and degenerations of hyperbolic structures. I. *Ann. Math. (2)* **120**(3), 401–476 (1984)
51. Morgan, J.W., Shalen, P.B.: Degenerations of hyperbolic structures. II. Measured laminations in 3-manifolds. *Ann. Math. (2)* **127**(2), 403–456 (1988)
52. Nguyen, D., Pak, I.: Enumerating projections of integer points in unbounded polyhedra. *SIAM J. Discrete Math.* **32**(2), 986–1002 (2018)

53. Oertel, U.: Closed incompressible surfaces in complements of star links. *Pac. J. Math.* **111**(1), 209–230 (1984)
54. Oertel, U.: Incompressible branched surfaces. *Invent. Math.* **76**(3), 385–410 (1984)
55. Oertel, U.: Measured laminations in 3-manifolds. *Trans. Am. Math. Soc.* **305**(2), 531–573 (1988)
56. Rubinstein, J.H.: Polyhedral minimal surfaces, Heegaard splittings and decision problems for 3-dimensional manifolds. In: *Geometric Topology* (Athens, GA, 1993), volume 2 of *AMS/IP Studies in Advanced Mathematics*, pp. 1–20. American Mathematical Society, Providence (1997)
57. Stein, W., et al.: SageMath, the Sage Mathematics Software System (Version 8.9) (2019)
58. Schleimer, S.: Almost Normal Heegaard Splittings. ProQuest LLC, Ann Arbor (2001). Thesis (Ph.D.)—University of California, Berkeley
59. Stanley, R.P.: *Enumerative Combinatorics*. Vol. 1, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge (1997). With a foreword by Gian-Carlo Rota, Corrected reprint of the 1986 original
60. Stocking, M.: Almost normal surfaces in 3-manifolds. *Trans. Am. Math. Soc.* **352**(1), 171–207 (2000)
61. Thompson, A.: Thin position and the recognition problem for S^3 . *Math. Res. Lett.* **1**(5), 613–630 (1994)
62. Thurston, W.P.: *Three-Dimensional Geometry and Topology*. Vol. 1, volume 35 of *Princeton Mathematical Series*. Princeton University Press, Princeton (1997). Edited by Silvio Levy
63. Tillmann, S.: Normal surfaces in topologically finite 3-manifolds. *Enseign. Math. (2)* **54**(3–4), 329–380 (2008)
64. Tollefson, J.: Isotopy classes of incompressible surfaces in irreducible 3-manifolds. *Osaka J. Math.* **32**(4), 1087–1111 (1995)
65. Tollefson, J.: Normal surface Q -theory. *Pac. J. Math.* **183**(2), 359–374 (1998)
66. Waldhausen, F.: On irreducible 3-manifolds which are sufficiently large. *Ann. Math.* **2**(87), 56–88 (1968)
67. Walsh, G.S.: Incompressible surfaces and spinnormal form. *Geom. Dedicata* **151**, 221–231 (2011)
68. Woods, K.: Presburger arithmetic, rational generating functions, and quasi-polynomials. *J. Symb. Log.* **80**(2), 433–449 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.