Geometry & Topology

Volume 27 (2023)

Issue 7 (pages 2497–2936)

msp

GEOMETRY & TOPOLOGY
msp.org/gt

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

The subscription price for 2023 is US \$740/year for the electronic version, and \$1030/year (+\$70, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Geometry & Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Geometry & Topology (ISSN 1465-3060 printed, 1364-0380 electronic) is published 9 times per year and continuously online, by Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840. Periodical rate postage paid at Oakland, CA 94615-9651, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

GT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Geometry & Topology **27**:7 (2023) 2497–2648
DOI: 10.2140/gt.2023.27.2497
Published: 19 September 2023

The combinatorial formula for open gravitational descendents

RAN J TESSLER

Pandharipande, Solomon and Tessler (2014) defined descendent integrals on the moduli space of Riemann surfaces with boundary, and conjectured that the generating function of these integrals satisfies the open KdV equations. We prove a formula for these integrals in terms of sums of Feynman diagrams. This formula is a generalization of the combinatorial formula of Kontsevich (1992) to the open setting. In order to overcome the main challenges of the open setting, which are orientation questions and the existence of boundary and boundary conditions, new techniques are developed. These techniques, which are interesting in their own right, include a characterization of graded spin structure in terms of open and nodal Kasteleyn orientations, and a new formula for the angular form of S^{2n-1}–bundles.

Buryak and Tessler (2017) proved the conjecture of Pandharipande, Solomon and Tessler based on the work presented here.

53D45

1 Introduction

The study of the intersection theory on the moduli space of open Riemann surfaces was initiated by Pandharipande, Solomon and Tessler in [31]. The authors constructed a descendent theory in genus 0 and obtained a complete description of it. In all genera, they conjectured that the generating series of the descendent integrals satisfies the open KdV equations. This conjecture can be considered as an open analog of Witten's famous conjecture in [38].

The construction of the positive-genus analog will be carried out in joint work with Solomon [35], and is reviewed here. A physical interpretation of these constructions can be found in Dijkgraaf and Witten [15].

In this paper, after recalling the constructions of [31; 35], we prove a formula for all the descendent integrals as sums over amplitudes of special Feynman diagrams, which we call odd critical nodal ribbon graphs. With this formula one can effectively calculate all the open descendents.

Based on this formula, the conjecture of [31] is proved in Buryak and Tessler [10], and a calculation of finer invariants is performed in Alexandrov, Buryak and Tessler [2].

1.1 Witten's conjecture

1.1.1 Intersection numbers Denote by $\mathcal{M}_{g,l}$ the moduli space of compact connected Riemann surfaces with l distinct marked points. P Deligne and D Mumford [13] defined a natural compactification of it via stable curves. Given g and l, a stable curve is a compact connected complex curve with l distinct marked points and finitely many singularities, all of which are simple nodes. We require the automorphism group of the surface to be finite, and the marked points and nodes are all distinct. The moduli space of stable curves of fixed g and l is denoted by $\overline{\mathcal{M}}_{g,l}$. It is known that this space is a nonsingular complex orbifold of complex dimension $3g-3+l$. For the basic theory the reader is referred to Deligne and Mumford [13] and Harris and Morrison [17].

In his seminal paper [38], E Witten, motivated by theories of 2–dimensional quantum gravity, initiated new directions in the study of $\overline{\mathcal{M}}_{g,l}$. For each marking index i he considered the tautological line bundles

$$\mathbb{L}_i \to \overline{\mathcal{M}}_{g,l}$$

whose fiber over a point

$$[\Sigma, z_1, \ldots, z_l] \in \overline{\mathcal{M}}_{g,l}$$

is the complex cotangent space $T^*_{z_i}\Sigma$ of Σ at z_i. Let

$$\psi_i \in H^2(\overline{\mathcal{M}}_{g,l}, \mathbb{Q})$$

denote the first Chern class of $\mathbb{L}_i$, and write

$$\langle \tau_{a_1}\tau_{a_2}\cdots\tau_{a_l}\rangle_g^c := \int_{\overline{\mathcal{M}}_{g,l}} \psi_1^{a_1}\psi_2^{a_2}\cdots\psi_l^{a_l}. \tag{1}$$

The integral on the right-hand side of (1) is well defined when the stability condition

$$2g-2+l>0$$

is satisfied, all the a_i are nonnegative integers, and the dimension constraint

$$3g-3+l=\sum_i a_i$$

holds. In all other cases, $\left\langle \prod_{i=1}^{l} \tau_{a_i}\right\rangle_g^c$ is defined to be zero. The intersection products (1) are often called *descendent integrals* or *intersection numbers*.

Let t_i (for $i \geq 0$) and u be formal variables, and put

$$\gamma := \sum_{i=0}^{\infty} t_i\tau_i.$$

Let
$$F_g^c(t_0, t_1, \ldots) := \sum_{n=0}^{\infty} \frac{\langle \gamma^n \rangle_g^c}{n!}$$
be the generating function of the genus g descendent integrals (1). The bracket $\langle \gamma^n \rangle_g^c$ is defined by the monomial expansion and the multilinearity in the variables t_i. The generating series

(2) $$F^c := \sum_{g=0}^{\infty} u^{2g-2} F_g^c$$

is called the (*closed*) *free energy*. The exponent $\tau^c := \exp(F^c)$ is called the (*closed*) *partition function*.

1.1.2 KdV equations

Set
$$\langle\langle \tau_{a_1} \tau_{a_2} \cdots \tau_{a_l} \rangle\rangle^c := \frac{\partial^l F^c}{\partial t_{a_1} \partial t_{a_2} \cdots \partial t_{a_l}}.$$
Witten's conjecture [38] says that the closed partition function τ^c becomes a tau function of the KdV hierarchy after the change of variables $t_n = (2n+1)!!\, T_{2n+1}$. In particular, it implies that the closed free energy F^c satisfies the following system of partial differential equations for $n \geq 1$:
$$(2n+1)u^{-2} \langle\langle \tau_n \tau_0^2 \rangle\rangle^c = \langle\langle \tau_{n-1} \tau_0 \rangle\rangle^c \langle\langle \tau_0^3 \rangle\rangle^c + 2 \langle\langle \tau_{n-1} \tau_0^2 \rangle\rangle^c \langle\langle \tau_0^2 \rangle\rangle^c + \tfrac{1}{4} \langle\langle \tau_{n-1} \tau_0^4 \rangle\rangle^c.$$
These equations are known in mathematical physics as the KdV equations. Witten [38] proved that the intersection numbers (1) satisfy the string equation
$$\left\langle \tau_0 \prod_{i=1}^{l} \tau_{a_i} \right\rangle_g^c = \sum_{j=1}^{l} \left\langle \tau_{a_j - 1} \prod_{i \neq j} \tau_{a_i} \right\rangle_g^c \quad \text{for } 2g-2+l > 0.$$
Witten has shown that the KdV equations, together with the string equation, determine the closed free energy F^c completely. R Dijkgraaf, E Verlinde and H Verlinde [14] reformulated an alternative description to Witten's conjecture in terms of the Virasoro algebra, and they have shown that the two descriptions are equivalent.

1.2 Kontsevich's proof

Witten's conjecture was proved by M Kontsevich [25]. The proof of [25] consisted of two parts. The first part was to prove a combinatorial formula for the gravitational descendents. Let $\mathcal{R}_{g,n}$ be the set of isomorphism classes of trivalent ribbon graphs of genus g with n marked faces. Denote by $V(G)$ the set of vertices of a graph $G \in \mathcal{R}_{g,n}$. Introduce formal variables λ_i, with $i \in [n]$. For an edge $e \in E(G)$, let

$$\lambda(e) := \frac{1}{\lambda_i + \lambda_j},$$

where i and j are the numbers of faces adjacent to e. The following formula holds:

$$\sum_{a_1,\dots,a_n \geq 0} \left\langle \prod_{i=1}^{n} \tau_{a_i} \right\rangle_g^c \prod_{i=1}^{n} \frac{(2a_i - 1)!!}{\lambda_i^{2a_i+1}} = \sum_{G \in \mathcal{R}_{g,n}} \frac{2^{|E(G)|-|V(G)|}}{|\mathrm{Aut}(G)|} \prod_{e \in E(G)} \lambda(e). \tag{3}$$

The second step of Kontsevich's proof was to translate the combinatorial formula into a matrix integral. Then, by using nontrivial analytical tools and the theory of the KdV hierarchy, he was able to prove that F^c satisfies the KdV equations of Section 1.1.2. Other proofs for Witten's conjecture were given, for example, in Mirzakhani [29] and Okounkov and Pandharipande [30].

1.3 Open intersection numbers and the open KdV equations

1.3.1 Open intersection numbers In [31], R Pandharipande, J Solomon and the author constructed an intersection theory on the moduli space of stable marked disks. Let $\overline{\mathcal{M}}_{0,k,l}$ be the moduli space of stable marked disks with k boundary marked points and l internal marked points. This space carries a natural structure of a compact smooth oriented manifold with corners. One can easily define the tautological line bundles $\mathbb{L}_i$ for an internal marking i, as in the closed case.

In order to define gravitational descendents, we must specify boundary conditions. The main construction in [31] is a construction of boundary conditions for $\mathbb{L}_i \to \overline{\mathcal{M}}_{0,k,l}$. In [31], vector spaces $\mathcal{S}_i = \mathcal{S}_{i,0,k,l}$ of *multisections* of $\mathbb{L}_i \to \partial\overline{\mathcal{M}}_{0,k,l}$, which satisfy the following requirements, were defined. Suppose $a_1, \dots, a_l$ are nonnegative integers with $2\sum_i a_i = \dim_{\mathbb{R}} \overline{\mathcal{M}}_{0,k,l} = k + 2l - 3$. Then:

(a) For any generic choice of multisections $s_{ij} \in \mathcal{S}_i$ for $1 \leq j \leq a_i$, the multisection

$$s = \bigoplus_{\substack{i \in [l] \\ 1 \leq j \leq a_i}} s_{ij}$$

vanishes nowhere on $\partial\overline{\mathcal{M}}_{0,k,l}$.

(b) For any two such choices s and s' we have

$$\int_{\overline{\mathcal{M}}_{0,k,l}} e(E, s) = \int_{\overline{\mathcal{M}}_{0,k,l}} e(E, s'),$$

where $E = \bigoplus_i \mathbb{L}_i^{a_i}$ and $e(E, s)$ is the relative Euler class.

The multisections s_{ij}, as above, are called *canonical*. With this construction the open gravitational descendents in genus 0 are defined by

$$\langle \tau_{a_1}\tau_{a_2}\cdots\tau_{a_l}\sigma^k\rangle_0^o := 2^{-\frac{1}{2}(k-1)}\int_{\overline{\mathcal{M}}_{0,k,l}} e(E,s), \tag{4}$$

where E is as above and s is canonical.

In a forthcoming paper [35], J Solomon and the author define a generalization for all genera. Suppose g, k and l are such that

$$2g-2+k+2l>0 \quad \text{with } 2\mid g+k-1. \tag{5}$$

In [35] a moduli space $\overline{\mathcal{M}}_{g,k,l}$, which classifies stable surfaces with boundaries and some extra structure, is constructed; see Section 2.3 for a precise definition. The moduli space $\overline{\mathcal{M}}_{g,k,l}$ is a smooth oriented compact orbifold with corners, of real dimension

$$3g-3+k+2l. \tag{6}$$

Note that naively, without adding an extra structure, the moduli of real stable curves of positive genus is nonorientable.

Again, on $\overline{\mathcal{M}}_{g,k,l}$ one defines vector spaces $\mathcal{S}_i = \mathcal{S}_{i,g,k,l}$ for $i \in [l]$, for which the genus g analogs of requirements (a) and (b) from above hold. Write

$$\langle \tau_{a_1}\tau_{a_2}\cdots\tau_{a_l}\sigma^k\rangle_g^o := 2^{-\frac{1}{2}(g+k-1)}\int_{\overline{\mathcal{M}}_{g,k,l}} e(E,s) \tag{7}$$

for the corresponding higher-genus descendents. Introduce one more formal variable s. The *open free energy* is the generating function

$$F^o(s,t_0,t_1,\dots;u) := \sum_{g=0}^{\infty} u^{g-1}\sum_{l=0}^{\infty}\frac{\langle \gamma^l\delta^k\rangle_g^o}{n!\,k!}, \tag{8}$$

where $\gamma := \sum_{i\ge 0} t_i\tau_i$ and $\delta := s\sigma$, and again we use the monomial expansion and the multilinearity in the variables t_i and s.

The descriptions of $\overline{\mathcal{M}}_{g,k,l}$ and its construction, and of the boundary conditions and their construction, are given in Section 2. Throughout this article we shall write $\langle\cdots\rangle$ for $\langle\cdots\rangle_g^o$, as closed descendents will not be considered, and the genus can be read from the numbers $k, l, a_1, \dots, a_l$.

1.3.2 Open KdV The initial condition

$$F^o|_{t_{\geq 1}=0} = u^{-1}\frac{s^3}{6} + u^{-1}t_0 s \tag{9}$$

follows easily from the definitions [31]. In [31] the authors conjectured the equations

$$\frac{\partial F^o}{\partial t_0} = \sum_{i=0}^{\infty} t_{i+1}\frac{\partial F^o}{\partial t_i} + u^{-1}s, \tag{10}$$

$$\frac{\partial F^o}{\partial t_1} = \sum_{i=0}^{\infty} \frac{2i+1}{3} t_i \frac{\partial F^o}{\partial t_i} + \frac{2}{3}s\frac{\partial F^o}{\partial s} + \frac{1}{2}. \tag{11}$$

They were called the *open string* and the *open dilaton* equation, correspondingly. These equations were geometrically proved in [31] for $g=0$, and for all genera in [35].

Put

$$\langle\langle \tau_{a_1}\tau_{a_2}\cdots\tau_{a_l}\sigma^k\rangle\rangle^o := \frac{\partial^{l+k}F^o}{\partial t_{a_1}\partial t_{a_2}\cdots\partial t_{a_l}\partial s^k}.$$

The main conjecture in [31] was:

Conjecture 1 (open KdV conjecture) The system of equations

$$\begin{aligned}(2n+1)u^{-1}\langle\langle\tau_n\rangle\rangle^o &\\ = u\langle\langle\tau_{n-1}\tau_0\rangle\rangle^c\langle\langle\tau_0\rangle\rangle^o &- \tfrac{1}{2}u\langle\langle\tau_{n-1}\tau_0^2\rangle\rangle^c + 2\langle\langle\tau_{n-1}\rangle\rangle^o\langle\langle\sigma\rangle\rangle^o + 2\langle\langle\tau_{n-1}\sigma\rangle\rangle^o,\end{aligned} \tag{12}$$

with $n \geq 1$, is satisfied.

In [31], the equations (12) were called the *open KdV equations*. It is easy to see that F^o is fully determined by the open KdV equations (12), the initial condition (9) and the closed free energy F^c. They also made a Virasoro-type conjecture, which also fully describes the open descendents. Both conjectures were proved in [31] for $g=0$. In [5], Buryak proved the equivalence of the two conjectures. Based on the work presented here, the conjecture was proven for all genus in [10]; see Section 1.5 below for more details.

1.4 The open combinatorial formula

Here and below the genus of a Riemann surface with boundary Σ, smooth or nodal, is defined as the usual genus of the doubled surface obtained from gluing two copies of Σ along the common boundary $\partial\Sigma$.

Definition 1.1 Let g,k,l be nonnegative integers which satisfy conditions (5), and let $\mathcal{B},\mathcal{I}$ be sets with $|\mathcal{B}|=k$, $|\mathcal{I}|=l$. Let $(\Sigma,\{x_i\}_{i\in\mathcal{B}},\{z_i\}_{i\in\mathcal{I}})$ be a genus g surface with boundary, whose set of boundary markings is $\mathcal{B}$, and set of internal markings is $\mathcal{I}$. A $(g,\mathcal{B},\mathcal{I})$*–smooth trivalent ribbon graph* is an embedding $\iota\colon G\to\Sigma$ of a connected graph G into $(\Sigma,\{x_i\}_{i\in\mathcal{B}},\{z_i\}_{i\in\mathcal{I}})$, such that:

(a) $\{x_i\}_{i\in\mathcal{B}}\subseteq\iota(V(G))$, where $V(G)$ is the set of vertices of G. We henceforth consider $\{x_i\}$ as vertices.

(b) The degree of every x_i is 2.

(c) The degree of any vertex $v\in V(G)\setminus\{x_i\}_{i\in\mathcal{B}}$ is 3.

(d) $\partial\Sigma\subseteq\iota(G)$.

(e) If $l\geq 1$, then
$$\Sigma\setminus\iota(G)=\coprod_{i\in\mathcal{I}}D_i,$$
where each D_i is a topological open disk, with $z_i\in D_i$. We call the disk D_i the face marked i.

(f) If $l=0$, then $\iota(G)=\partial\Sigma$ and $k=3$. Such a component is called *trivalent ghost*.

The genus $g(G)$ of the graph G is the genus of Σ. The number of the boundary components of G or Σ is denoted by $b(G)$, and $V^I(G)$ stands for the set of internal vertices. Denote by $B(G)$ the set of boundary marked points $\{x_i\}_{i\in\mathcal{B}}$, and by $I(G)\simeq\mathcal{I}$ the set of faces.

Definition 1.2 An *odd critical nodal ribbon graph* is $G=\left(\coprod_i G_i\right)/N$, where:

(a) The $\iota_i\colon G_i\to\Sigma_i$ are smooth trivalent ribbon graphs.

(b) $N\subset\left(\bigcup_i V(G_i)\right)\times\left(\bigcup_i V(G_i)\right)$ is a set of *ordered* pairs of boundary marked points (v_1,v_2) of the G_i, which we identify. After the identification of the vertices v_1 and v_2, the corresponding point in the graph is called a node. The vertex v_1 is called the legal side of the node and the vertex v_2 is called the illegal side of the node.

(c) Ghost components do not contain the illegal sides of nodes.

(d) For any component G_i, any boundary component of it contains an odd number of points which are either marked points or legal sides of nodes.

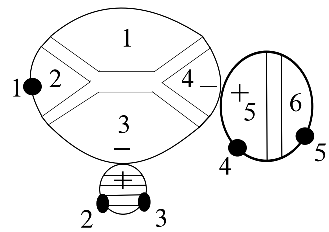


Figure 1: A nodal ribbon graph.

We require that elements of N be disjoint as sets (without ordering).

The set of edges $E(G)$ is composed of the internal edges of the G_i and of the boundary edges. The boundary edges are the boundary segments between successive vertices which are not the illegal sides of nodes. For any boundary edge e, we denote by $m(e)$ the number of the illegal sides of nodes lying on it. The boundary marked points of G are the boundary marked points of the G_i which are not nodes. The set of boundary marked points of G will be denoted by $B(G)$, the set of faces by $I(G)$.

An odd critical nodal ribbon graph is naturally embedded into the nodal surface $\Sigma = \left(\coprod_i \Sigma_i\right)/N$. The genus of the graph is defined as the genus of Σ. A (g,k,l)–odd critical nodal ribbon graph is a connected odd critical nodal ribbon graph, together with a pair of bijections, $\mathtt{m}^B : B(G) \to [k]$ and $\mathtt{m}^I : I(G) \to [l]$, called markings.

Two marked odd critical nodal ribbon graphs $\iota : G \to \Sigma$ and $\iota' : G' \to \Sigma'$ are isomorphic if there is an orientation-preserving homeomorphism $\Phi : (\Sigma, \{z_i\}, \{x_i\}) \to (\Sigma', \{z_i'\}, \{x_i'\})$ of marked surfaces, and an isomorphism of graphs $\phi : G \to G'$, such that

(a) $\iota' \circ \phi = \Phi \circ \iota$, and

(b) the maps preserve the markings.

Figure 1 depicts a nodal graph of genus 0, with 5 boundary marked points, 6 internal marked points, three components, one of which is a ghost, and two nodes, where a plus sign indicates the legal side of a node and a minus sign indicates the illegal side.

Notation 1.3 Denote by $\mathcal{OR}^m_{g,k,l}$ the set of isomorphism classes of odd (g,k,l)–critical nodal ribbon graphs with m legal nodes.

Remark 1.4 In Section 4 we have to consider more general ribbon graphs, and the notions of this subsection are defined in a different but equivalent way.

The goal of this paper is to prove the following theorem.

Theorem 1.5 *Fix $g,k,l \ge 0$ which satisfy conditions (5). Let $\lambda_1,\dots,\lambda_l$ be formal variables. Then we have*

$$2^{\frac{1}{2}(g+k-1)} \sum_{a_1,\dots,a_l\ge 0} \langle \tau_{a_1}\tau_{a_2}\cdots\tau_{a_l}\sigma^k\rangle_g^o \prod_{i=1}^{l} \frac{2^{a_i}(2a_i-1)!!}{\lambda_i^{2a_i+1}} \tag{13}$$
$$= \sum_{m\ge 0} \sum_{G=(\coprod_i G_i)/N\in \mathcal{OR}^m_{g,k,l}} \frac{\prod_i 2^{|V^I(G_i)|+g(G_i)+b(G_i)-1}}{|\mathrm{Aut}(G)|} \prod_{e\in E(G)} \lambda(e),$$

where

$$\lambda(e) := \begin{cases} \dfrac{1}{\lambda_i+\lambda_j} & \text{if } e \text{ is an internal edge between faces } i \text{ and } j, \\ \dfrac{1}{m+1}\dbinom{2m}{m}\lambda_i^{-2m-1} & \text{if } e \text{ is a boundary edge of face } i \text{ and } m(e)=m, \\ 1 & \text{if } e \text{ is a boundary edge of a ghost.} \end{cases}$$

Remark 1.6 The invariants of [31; 35] are defined as integrals of relative Euler classes, relative to canonical boundary conditions, over the moduli of graded surfaces, which are oriented orbifolds with corners. Theorem 1.5 is proven based on these definitions; more precisely, it assumes that the moduli spaces of graded surfaces are oriented orbifolds with corners, that the orientations satisfy some compatibility properties along nodal strata, and that (special) canonical multisections can be found. Since [35], which proves these assumptions in the positive genus case, has not appeared yet, in addition to defining everything we use, we also review the arguments.

First, the fact that the moduli of graded surfaces are smooth orbifolds with corners is a technical result, whose proof imitates of the proof of Theorem 2 of [41], and is provided in Section 2.3.6. Second, the construction of special canonical boundary conditions is similar to the proof of Lemma 3.53(a) in [31], and appears in Section 2.5.

On the other hand, proving that the high genus moduli is orientable, constructing the orientations and showing their properties is more involved, and is based on the discovery of the open Arf invariant in Solomon and Tessler [34]. However, in Sections 5 and 6.2, we provide completely different proofs for the orientability and the orientation properties we need, using the stratification of the moduli and properties of Kasteleyn orientations.

It is also worth mentioning that one of the main results of [31; 35] is the independence of the open intersection numbers on choices. This fact is also a byproduct of the

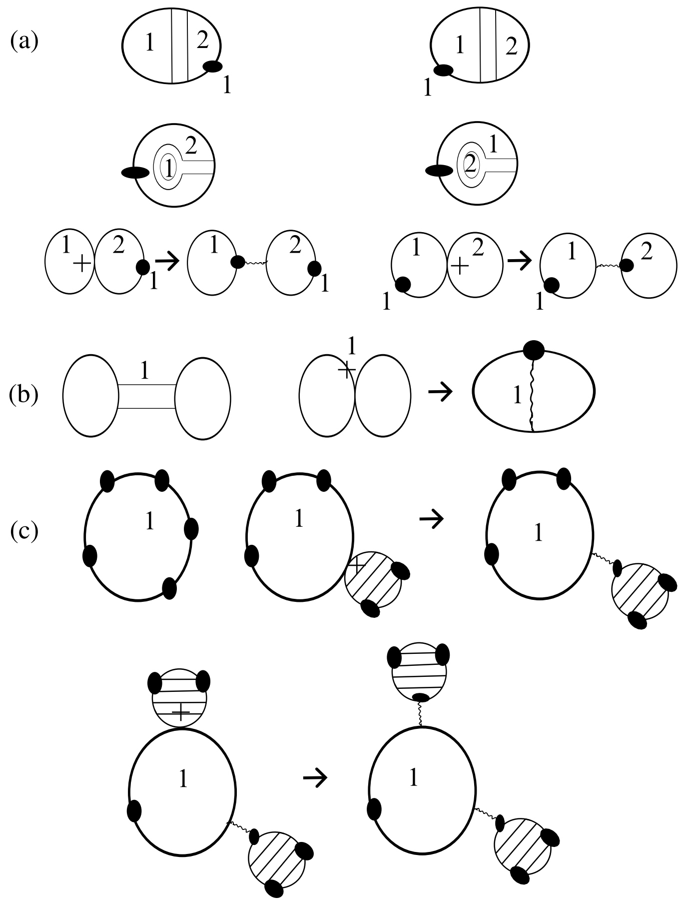


Figure 2: Examples of contributing graphs.

proof of Theorem 1.5, which uses just the defining properties of canonical boundary conditions and not a specific canonical multisection.

1.4.1 Examples

$\langle \tau_1 \tau_0 \sigma \rangle_0 = 1$. Thus, for $g = 0$, $k = 1$ and $l = 2$ the left-hand side of equation (13) with $\lambda_1 = \lambda$ and $\lambda_2 = \mu$ is

$$\frac{2}{\lambda \mu^3} + \frac{2}{\mu \lambda^3}.$$

The right-hand side receives contributions from several graphs; see Figure 2(a). The two nonnodal contributions in the first line are

$$\frac{1}{\lambda(\lambda+\mu)\mu^2} + \frac{1}{\mu(\lambda+\mu)\lambda^2}.$$

The two nonnodal contributions in the second line are

$$\frac{2}{2\lambda^3(\lambda+\mu)}+\frac{2}{2\mu^3(\lambda+\mu)}.$$

The nodal ones sum to

$$\frac{1}{\lambda\mu^3}+\frac{1}{\mu\lambda^3}.$$

And the two sides agree.

The second example is of $\langle\tau_1\rangle_1=\frac{1}{2}$. Consider case (b) in Figure 2. The left-hand side is $1/\lambda^3$. Nonnodal terms do not contribute, as the single relevant graph — the leftmost graph of (b) — is not odd. The nodal contribution is exactly $1/\lambda^3$.

The last example, Figure 2(c), is of $\langle\tau_2\sigma^5\rangle=8$. The left-hand side gives $384/\lambda^5$. Then 24 nonnodal diagrams — one for each cyclic order of the boundary points — contribute $24/\lambda^5$. There are 120 diagrams with a single node, one for each order; each contributes $1/\lambda^5$. There are 120 diagrams with two nodes; each contributes $2/\lambda^5$, where 2 comes from the Catalan term.

1.5 Proof of the conjecture and related works

Some recent developments, related works and open questions are summarized below.

(i) **Proof of the open KdV conjecture** Based on the combinatorial formula presented here, the conjecture of [31] has been proven in [10]: first, the combinatorial formula was transformed to a formula of matrix integrals, and then, by analytical tools and ideas from the theory of integrable hierarchies, the integral was shown to satisfy the open Virasoro constraints, which are equivalent to the open KdV equations by Buryak [5].

(ii) **Boundary descendents** Buryak [6] showed that the string solution of the open KdV equation is closely related to the wave function of the KdV hierarchy. In [5] a more general generating function, which is a tau function of the Burgers–KdV system, was introduced. It was conjectured that this function should correspond to an open intersection theory which includes *descendents of boundary marked points*. Such a theory can be constructed, extending the construction of [35], and, based on the combinatorial construction in this paper and on Buryak and Tessler [10], this theory can be shown to satisfy the Burgers–KdV hierarchy. The definition of the extended theory, its calculation and the proof of its relation with the Burgers–KdV hierarchy will appear soon.

(iii) **Kontsevich–Penner matrix model, Refined open intersection numbers** An alternative description of the solution of the Burgers–KdV equations in terms of matrix integrals was found algebraically by A Alexandrov [1] in terms of the $N = 1$ specification of the Kontsevich–Penner tau function.

Open problem 1 Is there a direct geometric way to derive Alexandrov's solution of the open KdV equations from the geometric construction of [31; 35]?

The combinatorial construction presented here was used in Alexandrov, Buryak and Tessler [2] to write a formula for more refined open intersection numbers. The main conjecture of [2], which is a strengthening of a conjecture of Safnuk [33], is that the generating series of the refined open numbers equals the Kontsevich–Penner tau function.

(iv) **Open r–spin** In recent work of Buryak, Clader and Tessler [8; 7], a far-reaching generalization of [31] to an intersection theory over the moduli of r–spin disks has appeared. The potential of the genus 0 open r–spin integrals was shown to be closely related to the wave function of the rKdV hierarchy, and an all-genus generalization was conjectured. Work in progress with Gross and Kelly generalizes this construction to open FJRW theory, and the genus 0 intersection numbers are explained using mirror symmetry.

Open problem 2 Generalize the formula presented in this work to the case of open r–spin intersection numbers.

(v) **Other interpretations of the theory** There were several related works in the physics literature; we mention two. In [15], Dijkgraaf and Witten provide a physical interpretation to the open intersection theory of [31; 35]. In [3], Bawane, Muraki and Rim describe a solution for the open KdV equations in terms of minimal gravity on the disk.

In [32], Safnuk gives an interpretation of the $N = 1$ specification of the Kontsevich–Penner tau function — which is, as explained above, a solution of the Burgers–KdV hierarchy — in terms of combinatorially defined volumes of moduli spaces.

(vi) **Similar formulas for other OGW invariants** There are two newer works which present formulas for open GW invariants in terms of summation over graphs with boundary nodes and are of the same flavor as the formula given here, and the refined formula of [2]. Zernik [40] presents an equivariant localization calculation of

OGW disk invariants for the pair $(\mathbb{CP}^{2n}, \mathbb{RP}^{2n})$. Buryak, Zernik, Pandharipande and Tessler [11] construct the stationary OGW theory of $(\mathbb{CP}^1, \mathbb{RP}^1)$, derive a localization formula for all intersection numbers, including descendents, and in [9] use it to prove a correspondence with open Hurwitz theory. Both formulas contain corner contributions, in addition to the naive contributions, in resemblance to (13). To the best knowledge of the author, such formulas have not appeared in literature before. Formulas for open GW invariants have appeared in the past, usually in the context of equivariant localization; see the calculations of Katz and Liu [24] as a prototypical example. In the older formulas which involved graph summation, the graphs were dual to topological stable marked surfaces with boundaries (which parametrized fixed-point loci). These surfaces included disk components which were connected by internal nodes to the closed part. There were no boundary nodes. The amplitudes of such graphs were usually similar to the analogous amplitudes in the closed case (and the disk contribution was usually more or less the square root of the sphere contribution). In the formulas of this work and of [2; 11; 40], the boundary nodes contribute an additional factor to the amplitudes. It would be interesting to gain a general understanding of this new type of expression, to understand when are they expected to appear, and to analyze them.

1.6 Plan of the paper

In Section 2 the constructions of [31] and [35] are reviewed. In particular, graded spin surfaces are defined, as well as their moduli space $\overline{\mathcal{M}}_{g,k,l}$, tautological line bundles and special canonical boundary conditions. With these in hand, the open intersection numbers are then defined.

In Section 3 the notions of sphere bundles and angular forms are recalled. We explain how to calculate the integral of the relative Euler class, relative to nowhere-vanishing boundary conditions. The main result of this section is an explicit formula for a representative of the angular form of a sphere bundle. This formula is the starting point of the paper.

Section 4 is devoted to constructing an open analog of Strebel's stratification. Symmetric stable Jenkins–Strebel differentials are defined, and used to stratify the moduli space of open surfaces and then the moduli of graded surfaces. In addition, combinatorial sphere bundles are constructed. It is then shown that special canonical multisections are pulled back from the combinatorial moduli. The main result of this section is that the open descendent integrals can be calculated as integrals over the combinatorial moduli.

Section 5 describes in more detail the cells in the stratification which will eventually contribute to the open descendents. Extended Kasteleyn orientations are defined, and their equivalence classes are shown to be equivalent to the data of a graded spin structure. The Kasteleyn orientations are used to provide a more explicit description of the contributing cells, of the boundary conditions and of the orientation of the moduli. As a byproduct, an alternative proof that the moduli $\overline{\mathcal{M}}_{g,k,l}$ is canonically oriented is given. The analysis of orientations is an important ingredient in the proof.

The last section, Section 6, proves the combinatorial formula, Theorem 1.5. With the aid of the explicit angular form constructed in Section 3, an integral representation of the open gravitational descendent is given. The integral depends explicitly on the boundary conditions. The properties of special canonical multisections are then used to iteratively integrate by parts, until an integrated form of the combinatorial formula, Theorem 6.10, is obtained. Finally, by performing a detailed study of the Kasteleyn orientations and the multiplicative constants they contribute,[1] we are able to apply the Laplace transform to the integrated formula and obtain the main theorem, Theorem 1.5.

Acknowledgements The author would like to thank J Solomon for the encouragement to work on this problem, and for many stimulating discussions. The author would also like to thank A Alexandrov, A Buryak, K F Gutkin, D Kazhdan, A Lavon, A Okounkov, R Pandharipande, E Shustin, A Solomon, A Zernik and D Zvonkine for interesting discussions related to the work presented here. The author wishes to thank the reviewers for careful reading of this work, and for many valuable remarks that vastly improved the article.

The author (incumbent of the Lillian and George Lyttle Career Development Chair) was supported by ISF grant 335/19 and a research grant from the Center for New Scientists of the Weizmann Institute. In addition, he was supported by ISF Grant 1747/13 and ERC Starting Grant 337560 in the group of J P Solomon at the Hebrew University of Jerusalem.

2 The moduli, bundles and intersection numbers

This section briefly summarizes the required definitions and results from [31; 35].

[1]This study also applies to the closed case, and gives a conceptual calculation in terms of discrete spin structures of a constant appearing in Kontsevich's work [25, Appendix C], which was the subject of several other works.

2.1 General conventions and notation

For $l \in \mathbb{N}$ we write $[l] = \{1, 2, \dots, l\}$. The set $[0]$ will denote the empty set.

Throughout this article, a map $\mathtt{m}\colon A \to \mathbb{Z}$ from an arbitrary set A, which is injective away from $\mathtt{m}^{-1}(0)$, will be called a *marking* or a *marking of* A. Given a marking, we shall identify elements of $\mathtt{m}^{-1}(\mathbb{Z} \setminus \{0\})$ with their images.

In what follows, the markings will be used to mark points in surfaces, half-edges in dual graphs and vertices in ribbon graphs. The reason we allow noninjective marking functions is that we will have to perform many graph or surface operations that will create new marked points. There will be no natural way to mark these new points, and therefore we will mark them all by 0.

We will encounter many types of graphs in the next sections. Dual graphs, to be defined in Section 2, will be denoted by capital Greek letters. Ribbon graphs, to be defined in Sections 4 and 5, will be denoted by capital Roman letters.

Many of the objects in this paper, such as surfaces or graphs, will have natural notions of genus, boundary labels and internal labels. A (g, B, I)–object is an object whose genus is g, set of boundary labels is B, and set of internal labels is I. Similarly, in the closed setting, a (g, I)–object is an object whose genus is g and set of internal labels is I.

Given a permutation π on a set S, we write s/π for the π–cycle of $s \in S$. For $a \in S/\pi$, we write $\pi^{-1}(a)$ for the elements which belong to the cycle a.

We shall sometimes use the shorthand notation $\boldsymbol{y}$ to denote a sequence $\{y_i\}_{i \in [r]}$, if the sequence we are referring to is understood from context.

2.2 Open surfaces and their moduli space

2.2.1 Stable open surfaces We recall the notion of a *stable marked open surface*.

Definition 2.1 We define a *smooth pointed surface* to be a triple

$$(\Sigma, \boldsymbol{x}, \boldsymbol{z}) = (\Sigma, \{x_i\}_{i \in \mathcal{B}}, \{z_i\}_{i \in \mathcal{I}}),$$

consisting of

(a) a Riemann surface Σ, possibly with boundary;

(b) an injection $\mathcal{B} \to \partial\Sigma$, with $i \mapsto x_i$, where $\mathcal{B}$ is a finite set;

(c) an injection $\mathcal{I} \to \mathring{\Sigma}$, with $i \mapsto z_i$, where $\mathcal{I}$ is a finite set.

In the case $\partial\Sigma \neq \varnothing$, we say that Σ is an open surface. Otherwise it is closed. We sometimes omit the marked points from our notation. Given a smooth pointed surface Σ, we write $B(\Sigma)$ for the set $\mathcal{B}$, and sometimes also for the set $\{z_i\}_{i\in\mathcal{B}}$. We similarly define $I(\Sigma)$.

A smooth closed pointed surface Σ is called *stable* if

$$2g(\Sigma) + |I(\Sigma)| > 2.$$

A smooth open pointed surface Σ is called *stable* if

$$2g(\Sigma) + |B(\Sigma)| + 2|I(\Sigma)| > 2.$$

Remark 2.2 Σ is canonically oriented, as a Riemann surface. In the case that $\partial\Sigma \neq \varnothing$, it is endowed with a canonical induced orientation.

Definition 2.3 For a pointed Riemann surface $(\Sigma, \{x_i\}_{i\in\mathcal{B}}, \{z_i\}_{i\in\mathcal{I}})$, where in the case that Σ is closed $\mathcal{B} = \varnothing$, we denote by $(\bar{\Sigma}, \{x_i\}_{i\in\mathcal{B}}, \{\bar{z}_i\}_{i\in\mathcal{I}})$ the same surface with opposite complex structure. The *doubling* of an open Σ is

$$\Sigma_{\mathbb{C}} = \Sigma \amalg_{\partial\Sigma} \bar{\Sigma},$$

the surface obtained by the *Schwarz reflection principle along the boundary* $\partial\Sigma$. For an open connected Σ we define the *genus* $g(\Sigma)$ to be the genus of $\Sigma_{\mathbb{C}}$. For Σ closed and connected the genus is just the usual genus. In the case that Σ is disconnected, its genus is defined as the sum of the genera of its connected components.

Remark 2.4 For open surfaces the topological type is determined by two numbers, the doubled genus g and the number of boundary components h, and not only by the genus. The number h is constrained by

$$h = g + 1 \pmod 2, \quad \text{with } 0 \le h \le g + 1,$$

and for any (g, h) satisfying these constraints there is a topological type of open surface.

Definition 2.5 A *prestable* surface is a tuple

$$\Sigma = (\{\Sigma_\alpha\}_{\alpha\in\mathcal{O}\cup\mathcal{C}}, \sim, \mathrm{CB}),$$

where:

(a) $\mathcal{O}$ and $\mathcal{C}$ are finite sets. For $\alpha \in \mathcal{O}$, Σ_α is an open smooth pointed surface; for $\alpha \in \mathcal{S}$, Σ_α is a closed smooth pointed surface.

(b) $\sim = \sim_B \cup \sim_I$, where $\sim_B$ is an equivalence relation on $\bigcup_\alpha B(\Sigma_\alpha)$ with equivalence classes of size at most 2, and $\sim_I$ is an equivalence relation on $\bigcup_\alpha I(\Sigma_\alpha)$ with equivalence classes of size at most 2. We write $B(\Sigma)$ and $I(\Sigma)$ for the equivalence classes of size 1 of $\sim_B$ and $\sim_I$, respectively.

(c) $\mathrm{CB}(\Sigma)$ is a subset of $I(\Sigma)$.

Elements of $B(\Sigma)$ are called *boundary marked points*. Elements of $I(\Sigma)\setminus \mathrm{CB}(\Sigma)$ are called *internal marked points*. The $\sim_B$ (resp. $\sim_I$) equivalence classes of size 2 are called boundary (resp. interior) nodes, and elements which belong to these equivalence classes are called *half-nodes*. Elements of CB are called contracted boundaries. The equivalence classes of $\sim$ (resp. $\sim_B$, $\sim_I$) are collectively called *special* (resp. *special boundary*, *special internal*) *points* of Σ.

We also write $\Sigma = \coprod_{\alpha\in\mathcal{O}\cup\mathcal{C}} \Sigma_\alpha/{\sim}$. If $\mathcal{O}$ is empty and CB is empty, Σ is called a *prestable closed surface*. Otherwise it is called a *prestable open surface*.

A prestable surface is *marked*, if it is also endowed with *markings* $\mathtt{m}^B : B(\Sigma)\to\mathbb{Z}$ and $\mathtt{m}^I : I(\Sigma)\setminus \mathrm{CB}\to\mathbb{Z}$. Write $\mathtt{m}=\mathtt{m}^I\cup\mathtt{m}^B$. Recall that a marking is injective outside of the preimage of 0.

A prestable marked surface is called a *stable marked surface* if each of its constituent smooth surfaces Σ_α is stable.

The *doubled surface* $\Sigma_{\mathbb{C}}$ of a stable open surface is defined as

$$\Sigma_{\mathbb{C}} = \Bigl(\coprod_{\alpha\in\mathcal{O}} (\Sigma_\alpha)_{\mathbb{C}} \coprod_{\alpha\in\mathcal{C}} \Sigma_\alpha \coprod \overline{\Sigma}_\alpha\Bigr)\Big/{\sim_{\mathbb{C}}},$$

where

$$\sim_{\mathbb{C}} = (\sim_B \cup \sim_I \cup \sim_{\bar I} \cup \sim_{\mathrm{CB}})$$

is defined as follows: $\sim_{\bar I}$ identifies internal marked points of $\{\overline{\Sigma}_\alpha\}_{\alpha\in\mathcal{C}}$ if and only if $\sim_I$ identifies the corresponding marked points in $\{\Sigma_\alpha\}_{\alpha\in\mathcal{C}}$, and $\sim_{\mathrm{CB}}$ identifies $z_i\in\Sigma_\alpha$ and $\bar z_i\in\overline{\Sigma}_\alpha$ whenever $i\in\mathrm{CB}(\Sigma)$. $\Sigma_{\mathbb{C}}$ is endowed with an involution ϱ, with $\bar z_i=\varrho(z_i)$, whose fixed-point set is $\partial\Sigma\cup\mathrm{CB}(\Sigma)$, and is such that $\Sigma\simeq\Sigma_{\mathbb{C}}/\varrho$. Write $D(\Sigma)=(\Sigma_{\mathbb{C}},\varrho)$.

Σ is connected if the underlying space $\coprod_{\alpha\in\mathcal{D}\cup\mathcal{S}}\Sigma_\alpha/{\sim}$ is. Σ is smooth if $\mathrm{CB}(\Sigma)=\varnothing$ and $\sim$ has only equivalence classes of size 1.

The *normalization* $\mathrm{Norm}(\Sigma)$ of the stable marked surface Σ is defined to be the surface $(\{\Sigma_\alpha\}_{\alpha\in\mathcal{O}\cup\mathcal{C}}, \sim', \mathrm{CB}', \mathtt{m}')$ where $\sim'$ has only size 1 equivalence classes, CB' is empty, and the marking $\mathtt{m}'$ agrees with $\mathtt{m}$ whenever is defined, and otherwise $\mathtt{m}'^{I} = 0$, $\mathtt{m}'^{B} = 0$. For a marked point marked $i \neq 0$, write Σ_i for the component of $\mathrm{Norm}(\Sigma)$ which contains marked point z_i.

A *topological stable marked surface*, open or closed, is defined in the same way, only with the Σ_α being topological surfaces rather than Riemann surfaces.

In what follows, our default choice of marking function $\mathtt{m}$ is a bijection $\mathtt{m}^I : I(\Sigma) \to [n]$ if Σ is closed, and if Σ is open we usually take bijections $\mathtt{m}^I : I(\Sigma) \setminus \mathrm{CB}(\Sigma) \to [l]$ and $\mathtt{m}^B : B(\Sigma) \to [k]$. Therefore whenever a surface is written as $(\Sigma, z_1, \dots, z_n)$ or $(\Sigma, x_1, \dots, x_k, z_1, \dots, z_l)$, we implicitly mean that it is marked, and that the indices of the marked points represent the markings.

See Figure 3 for examples of prestable surfaces and their normalizations.

We sometimes identify $D(\Sigma)$ and $\Sigma_{\mathbb{C}}$.

Definition 2.6 An *isomorphism* between two prestable marked surfaces

$$\Sigma = (\{\Sigma_\alpha\}_{\alpha \in \mathcal{O} \cup \mathcal{C}}, \sim, \mathrm{CB}, \mathtt{m}) \quad \text{and} \quad \Sigma' = (\{\Sigma'_\alpha\}_{\alpha \in \mathcal{O}' \cup \mathcal{C}'}, \sim', \mathrm{CB}', \mathtt{m}')$$

is a tuple $f = (f^{\mathcal{O}}, f^{\mathcal{C}}, \{f^\alpha\}_{\alpha \in \mathcal{O} \cup \mathcal{C}})$ such that:

(a) The maps $f^{\mathcal{O}} : \mathcal{O} \to \mathcal{O}'$ and $f^{\mathcal{C}} : \mathcal{C} \to \mathcal{C}'$ are bijections between the sets which index the components of the surfaces.

(b) For $\alpha \in \mathcal{O}$, $f^\alpha : \Sigma_\alpha \to \Sigma'_{f^{\mathcal{O}}(\alpha)}$ is a biholomorphism, which induces a bijection on the sets of special points. For $\alpha \in \mathcal{C}$, $f^\alpha : \Sigma_\alpha \to \Sigma'_{f^{\mathcal{C}}(\alpha)}$ is a biholomorphism, which induces a bijection on the sets of special points.

(c) For $x \in \Sigma_\alpha$ and $y \in \Sigma_\beta$, $x \sim y$ if and only if $f^\alpha(x) \sim' f^\beta(y)$.

(d) For any special point $x \in \Sigma_\alpha$, $\mathtt{m}'(f^\alpha(x)) = \mathtt{m}(x)$.

(e) $\bigcup_\alpha f^\alpha(\mathrm{CB}) = \mathrm{CB}'$.

We denote by $\mathrm{Aut}(\Sigma)$ the group of automorphisms of Σ.

An isomorphism between stable topological surfaces is similarly defined, only with the maps f^α required to be homeomorphisms rather than biholomorphisms.

2.2.2 Stable graphs It is useful to encode some of the combinatorial data of stable marked surfaces in graphs.

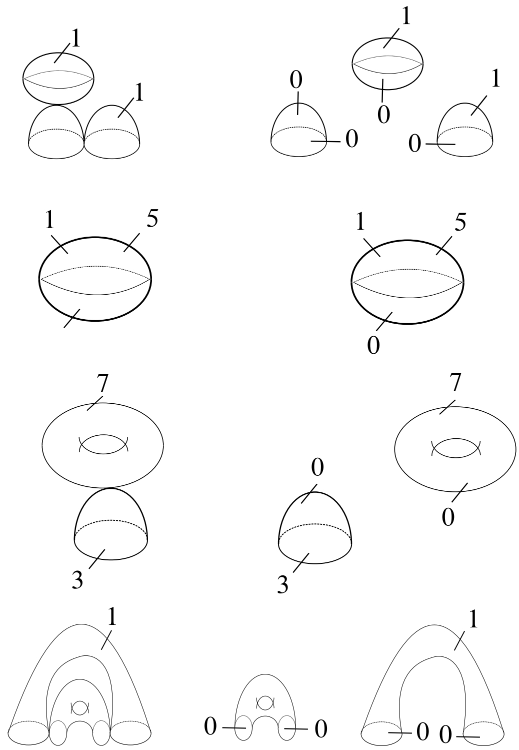


Figure 3: In this diagram in every row the leftmost picture is a prestable surface, and on the right side of the same row is the normalization. In the top row there is a prestable marked surface with boundary, and its normalization into two stable marked disks and a prestable marked sphere. In the second row there is a stable sphere with an (unmarked) contracted boundary. Its normalization is a stable sphere with three markings. In the third row there is a stable surface with boundary which is normalized into a disk and a torus. The last row contains a stable surface whose normalization is the union of a cylinder and a genus 3 surface with boundary.

Definition 2.7 A (not necessarily connected) *prestable dual graph* Γ is a tuple

$$(V = V^O \cup V^C,\ H = H^B \cup H^I,\ \sigma_0,\ \sim\, = \,\sim_B \cup \sim_I,\ g,\ H^{\mathrm{CB}},\ \mathtt{m} = \mathtt{m}^B \cup \mathtt{m}^I),$$

where:

(a) V^O and V^C are finite sets, called the *open and closed vertices*, respectively.

(b) H^B and H^I are finite sets of *boundary and internal half-edges*.

(c) $\sigma_0\colon H \to V$ associates any half-edge to its vertex.

(d) $\sim_B$ is an equivalence relation on H^B with equivalence classes of sizes 1 or 2, and $\sim_I$ is an equivalence relation on H^I with equivalence classes of sizes 1 or 2. Denote by T^B the equivalence classes of size 1 of $\sim_B$, and by T^I the equivalence classes of size 1 of $\sim_I$.

(e) $H^{\mathrm{CB}} \subseteq T^I$.

(f) $g\colon V \to \mathbb{Z}_{\geq 0}$ is a genus assignment.

(g) $\mathtt{m}^B\colon T^B \to \mathbb{Z}$ and $\mathtt{m}^I\colon T^I \setminus H^{\mathrm{CB}} \to \mathbb{Z}$ are markings.

We call T^B *boundary tails*, H^{CB} *contracted boundaries*, and $T^I \setminus H^{\mathrm{CB}}$ *internal tails*. Set $T = T^I \cup T^B$. Now, $\sim_B$ induces a fixed-point-free involution on $H^B \setminus T^B$. Similarly, $\sim_I$ induces a fixed-point-free involution on $H^I \setminus T^I$. We denote this involution on $H \setminus T$ by σ_1. We set $E^B = (H^B \setminus T^B)/{\sim_B}$, the set of boundary *edges*. We define $E^I = (H^I \setminus T^I)/{\sim_I} \cup H^{\mathrm{CB}}$. We put $E = E^I \cup E^B$, the set of edges. We denote by σ_0^B the restriction of σ_0 to H^B; in a similar fashion we define σ_0^I.

We require that for all $h \in H^B$, $\sigma_0(h) \in V^O$.

We say that Γ is connected if its underlying graph (V, E) is connected.

For a vertex v we set $k(v) = |(\sigma_0^B)^{-1}(v)|$. It is defined to be 0 if v is closed. We set $l(v) = |(\sigma_0^I)^{-1}(v)|$. Write $\mathrm{CB}(v)$ for the number of contracted boundaries of v. A dual graph is *closed* if $V^O = H^{\mathrm{CB}} = \varnothing$, and otherwise it is *open*.

The *genus of a stable connected closed dual graph* Γ is defined by

$$g(\Gamma) = \sum_{v \in V^C} g(v) + |E^I| - |V^C| + 1.$$

The *genus of a stable connected open dual graph* Γ is defined by

$$g(\Gamma) = \sum_{v \in V^O} g(v) + 2\sum_{v \in V^C} g(v) + |E^B| + 2|E^I| - |H^{\mathrm{CB}}| - |V^O| - 2|V^C| + 1.$$

A closed vertex $v \in V^C$ is *stable* if

$$2g(v) + l(v) > 2.$$

An open vertex $v \in V^O$ is *stable* if

$$2g(v) + k(v) + 2l(v) > 2.$$

A dual graph Γ is stable if all its vertices are.

The *normalization* $\mathrm{Norm}(\Gamma)$ of the graph Γ is defined to be the unique stable graph $(V', H', \sigma_0', \sim', g', H'^{\mathrm{CB}}, \mathtt{m}')$ with $V' = V$, $H' = H$, $\sigma_0' = \sigma_0$, $g' = g$, $H'^{\mathrm{CB}} = \varnothing$, and $\sim'$ has only classes of size 1. The map $\mathtt{m}'$ agrees with $\mathtt{m}$ whenever $\mathtt{m}$ is defined. Otherwise $\mathtt{m}' = 0$.

For $i \in \mathrm{Image}(\mathtt{m}^I) \setminus \{0\}$, we denote by $v_i(\Gamma)$ the connected component of $\mathrm{Norm}(\Gamma)$ which contains the tail marked i.

It is easy to see that the genus is always nonnegative. Figure 4 illustrates several dual graphs and their normalizations. Note that open vertices without boundary half-edges are allowed.

Definition 2.8 An *isomorphism* between graphs

$$\Gamma = (V, H, \sigma_0, \sim, g, H^{\mathrm{CB}}, \mathtt{m}) \quad \text{and} \quad \Gamma' = (V', H', \sigma_0', \sim', g', H'^{\mathrm{CB}}, \mathtt{m}')$$

is a pair $f = (f^V, f^H)$ such that

(a) $f^V \colon V \to V'$ and $f^H \colon H \to H'$ are bijections,

(b) $g' \circ f = g$,

(c) $h_1 \sim h_2$ if and only if $f(h_1) \sim' f(h_2)$,

(d) $\sigma_0' = f \circ \sigma_0$,

(e) $\mathtt{m}' \circ f = \mathtt{m}$,

(f) $f(H^{\mathrm{CB}}) = H'^{\mathrm{CB}}$.

We denote by $\mathrm{Aut}(\Gamma)$ the group of automorphisms of Γ.

To each stable marked surface Σ we associate an isomorphism class of connected stable graphs as follows. We set

$$V^O = \mathcal{O}, \quad V^C = \mathcal{C}, \quad H^B = \bigcup_\alpha B(\Sigma_\alpha), \quad H^I = \bigcup_\alpha I(\Sigma_\alpha), \quad H^{\mathrm{CB}} = \mathrm{CB}(\Sigma).$$

The definitions of g, $\sim$, σ_0 and $\mathtt{m}$ are straightforward. In particular, a tail marked a is associated to a marked point labeled a. An edge between two vertices corresponds to a node between their corresponding components. See Figure 4 for the dual graphs which correspond to the surfaces of Figure 3. Note that this correspondence is at the level of isomorphism classes of topological stable surfaces, and that a surface is closed precisely if its corresponding graph is closed.

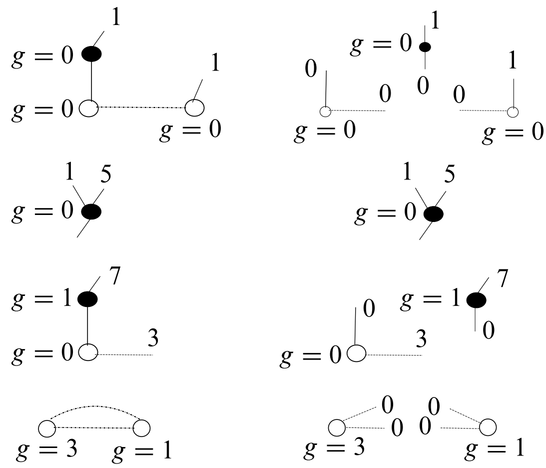


Figure 4: This diagram presents the dual graphs which correspond to the surfaces from Figure 3, under the correspondence of Definition 2.9. Again the right-hand side of each row is the normalization of the left-hand side. Black vertices correspond to closed components, and empty vertices to open. The genus of the vertex is written next to it. Boundary edges or half-edges are drawn as dashed lines, and the other edges or half-edges are internal (the case of contracted boundary is included). The label of a tail is written next to it. The genus of the dual graphs in the left-hand side are, going from top to bottom, $0, 0, 2, 5$.

Definition 2.9 The graph associated to a stable surface Σ is denoted by $\Gamma(\Sigma)$. The *genus* of a stable surface Σ is defined as the genus of $\Gamma(\Sigma)$.

Observe that the genus of a stable closed surface agrees with the standard definition, while the genus of a stable open surface equals the standard genus of its doubled surface. The genus of a stable surface equals the genus of the surface obtained by smoothing its nodes, including the contracted boundaries which are smoothed to boundary components. Observe also that $\mathrm{Norm}(\Gamma(\Sigma)) = \Gamma(\mathrm{Norm}(\Sigma))$, and that for any internal marked point which is marked $i \neq 0$, we have $v_i(\Gamma(\Sigma)) = \Gamma(\Sigma_i)$, where Σ_i is the component of Σ which contains marked point z_i.

Throughout this paper we will sometimes write "graph" instead of "dual graph" when the meaning is clear from the context. Dual graphs will be denoted by capital Greek letters, to help us distinguish them from another kind of graphs we shall meet below, ribbon graphs, which will be denoted by capital Roman letters.

We denote by $\mathcal{G}^{\mathbb{R}}_{g,k,l}$ the set of isomorphism classes of all stable graphs of genus g with k boundary tails, l internal tails, and for which

$$\operatorname{Image}(\mathtt{m}^B) = [k] \quad \text{and} \quad \operatorname{Image}(\mathtt{m}^I) = [l].$$

We write $\mathcal{G}^{\mathbb{R}}$ for the set of isomorphism classes of all stable graphs. Note that the cases $k = 0$ or $l = 0$ are not excluded, as surfaces without boundary or internal marked points will be considered in what follows.

Notation 2.10 Given nonnegative integers k, l with $2g + k + 2l > 2$, denote by $\Gamma^{\mathbb{R}}_{g,k,l}$ the stable graph with $V^O = \{*\}$ and $V^C = \varnothing$, with

$$g(*) = g, \quad T^B = H^B \simeq [k], \quad T^I = H^I \simeq [l],$$

where the equivalences with $[k]$ and $[l]$ are obtained using $\mathtt{m}^B$ and $\mathtt{m}^I$, respectively. We similarly define $\Gamma_{g,n}$ as the closed graph with a single vertex of genus g, and $T^I = H^I \simeq [n]$.

Definition 2.11 A stable dual graph is *effective* if

(a) any internal half-edge is a tail or a contracted boundary,

(b) any vertex without internal tails has exactly three boundary half-edges and genus 0, and

(c) different vertices without internal half-edges are not adjacent.

A surface is called *effective* if it is associated to an effective graph.

The notion of effectiveness will be important later on, when we construct the combinatorial moduli space using Jenkins–Strebel differentials. On moduli strata which correspond to effective dual graphs, the map to the combinatorial moduli is a homeomorphism. This fact will turn out to be useful when we come to translate the geometric intersection numbers to combinatorial expressions.

In the leftmost column of Figure 3, only the sphere from the second row is effective: the surface from the first row has an internal node, and in addition it is not stable; the surface from the third row also has an internal node as well; the surface from the lowest row has a component without internal markings, which is not a disk with three boundary markings. Equivalently, in the leftmost column of Figure 4 only the second graph is effective. Additional examples of effective and noneffective surfaces and graphs are illustrated in Figure 5.

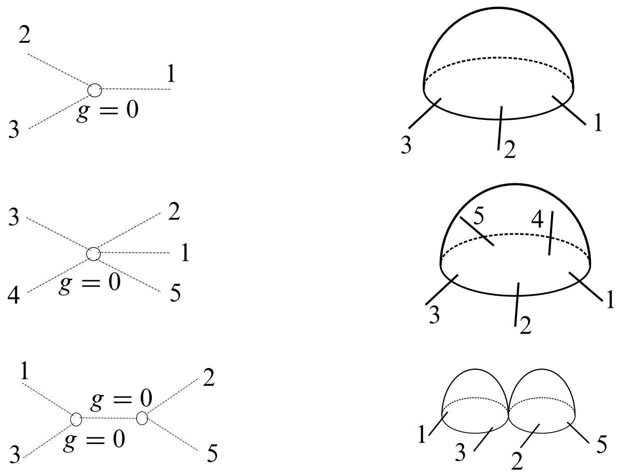


Figure 5: Every row in this diagram illustrates a dual graph and the corresponding surface. Only the first row represent an effective graph/surface. Note that the cyclic order of boundary markings on boundaries cannot be read from the dual graph data.

2.2.3 Some graph operations For the purpose of the next definition, for a vertex v in a dual graph Γ, write $\varepsilon(v)=1$ if v is open, and $\varepsilon(v)=2$ otherwise. For an edge e set $\varepsilon(e)=0$ unless e is an internal edge connecting two open vertices, in which case put $\varepsilon(e)=1$.

Definition 2.12 Consider a stable graph Γ. The *smoothing* of Γ at $f\in E$ is the stable graph

$$d_f\Gamma=\Gamma'=(V',H',\sim',s_0',g',\mathtt{m}'),$$

defined as follows. Suppose $f\notin H^{\mathrm{CB}}(\Gamma)$ is the $\sim$–equivalence class $\{h_1,h_2\}$. Write $\sigma_0(h_1)=v_1$ and $\sigma_0(h_2)=v_2$. The vertex set is given by

$$V'=(V\setminus\{v_1,v_2\})\cup\{v\}.$$

The new vertex v is closed if and only if both v_1 and v_2 are closed. We have that

$$H'=H\setminus\{h_1,h_2\},$$

and $\sim'$ is the restriction of $\sim$ to H'. For $h\in\sigma_0^{-1}(\{v_1,v_2\})$, we define $\sigma_0'(h)=v$; otherwise, $\sigma_0'(h)=\sigma_0(h)$. For any tail t, $\mathtt{m}'(t)=\mathtt{m}(t)$. We set

$$g'(v)=\begin{cases} g(v_1)+1+\varepsilon(f) & \text{if } v_1=v_2,\\ g(v_1)+g(v_2)+\varepsilon(f) & \text{if } v_1\neq v_2 \text{ and } \varepsilon(v_1)=\varepsilon(v_2),\\ \varepsilon(v_1)g(v_1)+\varepsilon(v_2)g(v_2) & \text{otherwise.}\end{cases}$$

When $f \in H^{\mathrm{CB}}$, a contracted boundary of vertex v, then

$$V' = V, \quad H' = H \setminus \{f\}, \quad H'^{\mathrm{CB}} = H^{\mathrm{CB}} \setminus \{f\}.$$

We update $\sim'$, σ_0' and $\mathtt{m}'$ as above. We put $g'(w) = g(w)$ for $w \neq v$, and we put $g'(v) = g(v) + 1$ if v is open, otherwise we set $g'(v) = 2g(v)$ and declare v to be open.

Observe that there is a natural proper injection $H' \hookrightarrow H$, so we may identify H' with a subset of H. This identification induces identifications of tails and of edges. Using the identifications, we extend the definition of smoothing in the following manner. Given a set $S = \{f_1, \dots, f_n\} \subseteq E(\Gamma)$, define the smoothing at S as

$$d_S\Gamma = d_{f_n}(\cdots d_{f_2}(d_{f_1}\Gamma)\cdots).$$

Observe that $d_S\Gamma$ does not depend on the order of smoothings performed.

Definition 2.13 A stable topological surface Σ' is a *smoothing* of a topological stable marked surface Σ at an internal node $z_\nu \sim z_\mu$ if there exists a simple closed path $\gamma \hookrightarrow \Sigma'$, and a map $\varphi\colon \Sigma' \to \Sigma$ which takes γ to the node and restricts to an orientation-preserving homeomorphism $\varphi\colon \Sigma' \setminus \gamma \simeq \Sigma \setminus \{z_\mu, z_\nu\}$. In this case we say that γ is contracted to the node. We say that γ *degenerates* to z_ν when this time γ is an *oriented* simple closed path in Σ', if γ is contracted to the node, and the φ–preimage of a small enough neighborhood of z_ν lies to the left of γ. The definitions of smoothing in a boundary node or degeneration to a boundary half-node are analogous, only with a simple arc that connects two boundary points.

A topological stable surface Σ' is the smoothing of a topological stable surface Σ at a contracted boundary z_ν if there exists a boundary component $\partial\Sigma'_\nu$, and a map $\varphi\colon \Sigma' \to \Sigma$ such that $\varphi(\partial\Sigma'_\nu) = z_\nu$ and $\varphi\colon \Sigma' \setminus \partial\Sigma'_\nu \simeq \Sigma \setminus z_\nu$.

If e is the edge of $\Gamma(\Sigma)$ which corresponds to the node $z_\nu \sim z_\mu$ in Σ, then $\Gamma(\Sigma') = d_e\Gamma(\Sigma)$, where Σ' is the smoothing of Σ in that node; similarly for smoothing in contracted boundaries.

If $\Gamma = d_S\Gamma'$, then H' is canonically a subset of H, and we have a natural identification between $E(\Gamma)$ and $E(\Gamma') \setminus S$.

We can now define boundary maps

$$\partial^! \colon \mathcal{G}^{\mathbb{R}} \to 2^{\mathcal{G}^{\mathbb{R}}} \quad \text{and} \quad \partial \colon \mathcal{G}^{\mathbb{R}} \to 2^{\mathcal{G}^{\mathbb{R}}}$$

by putting

$$\partial^! \Gamma = \{\Gamma' \mid \Gamma = d_S \Gamma' \text{ for some } S \subseteq E(\Gamma')\} \quad \text{and} \quad \partial \Gamma = \partial^! \Gamma \setminus \{\Gamma\}.$$

These maps naturally extend to maps $2^{\mathcal{G}^{\mathbb{R}}} \to 2^{\mathcal{G}^{\mathbb{R}}}$.

2.2.4 Moduli of open surfaces In this paper we consider orbifolds with corners; we follow the definitions of [41, Section 3], which build on the works [22; 21].

Notation 2.14 For $\Gamma \in \mathcal{G}^{\mathbb{R}}$, denote by $\mathcal{M}_{\Gamma}^{\mathbb{R}}$ the set of isomorphism classes of stable marked genus g surfaces with associated graph Γ.

Define

$$\overline{\mathcal{M}}_{\Gamma}^{\mathbb{R}} = \coprod_{\Gamma' \in \partial^! \Gamma} \mathcal{M}_{\Gamma'}^{\mathbb{R}}.$$

We abbreviate

$$\overline{\mathcal{M}}_{g,k,l}^{\mathbb{R}} = \overline{\mathcal{M}}_{\Gamma_{g,k,l}^{\mathbb{R}}}^{\mathbb{R}} \quad \text{and} \quad \mathcal{M}_{g,k,l}^{\mathbb{R}} = \mathcal{M}_{\Gamma_{g,k,l}^{\mathbb{R}}}^{\mathbb{R}}.$$

We similarly define $\overline{\mathcal{M}}_{g,n}$ and $\mathcal{M}_{g,n}$, which are just the usual Deligne–Mumford moduli spaces of stable and smooth curves respectively.

For $i \in \mathrm{Image}(\mathtt{m}^I) \setminus \{0\}$, write $\mathcal{M}_{v_i(\Gamma)}$ for the moduli of the graph $v_i(\Gamma)$, and denote by $v_i \colon \mathcal{M}_{\Gamma} \to \mathcal{M}_{v_i(\Gamma)}$ the natural map which on the level of objects sends $\Sigma \to \Sigma_i$.

The space $\overline{\mathcal{M}}_{g,k,l}^{\mathbb{R}}$ is a compact smooth orbifold with corners of real dimension

$$\dim_{\mathbb{R}} \overline{\mathcal{M}}_{g,k,l}^{\mathbb{R}} = k + 2l + 3g - 3.$$

We attribute this result to Amitai Netser Zernik [41, Section 2]. His setting is slightly different. He considers open stable genus 0 maps to homogeneous varieties, and he proves that the moduli space of these maps is an orbifold with corners. In our case the target space is a point, but the genus is arbitrary. This change does not affect his results or techniques, since they only rely on convexity of the corresponding closed moduli problem, that is, on the fact that the moduli space of (complex) stable maps is a smooth orbifold, which clearly holds for $\overline{\mathcal{M}}_{g,n}$. We review the argument. Consider the sequence

$$\overline{\mathcal{M}}_{g,k,l}^{\mathbb{R}} \xhookrightarrow{(4)} \widetilde{\mathcal{M}}_{g,k,2l}^{\mathbb{R}} \xrightarrow{(3)} \widetilde{\mathbb{R}\mathcal{M}}_{g,k,2l} \xrightarrow{(2)} \overline{\mathbb{R}\mathcal{M}}_{g,k,2l} \xrightarrow{(1)} \overline{\mathcal{M}}_{g,k+2l}. \tag{14}$$

We define the moduli spaces and maps appearing in (14) as follows.

Step 1 First, $\overline{\mathbb{R}\mathcal{M}}_{g,k,2l}$ is the fixed locus of the involution on $\overline{\mathcal{M}}_{g,k+2l}$ defined by

$$(C; z_1, \ldots, z_{k+2l}) \mapsto (\overline{C}; z_1, \ldots, z_k, z_{k+l+1}, \ldots, z_{k+2l}, z_{k+1}, \ldots, z_{k+l}),$$

where $\overline{C}$ is the same smooth curve C, but with the conjugate complex structure. This is a compact smooth real orbifold, as it is the fixed locus of an antiholomorphic involution over a smooth complex orbifold. More details on the fixed-point functor on stacks can be found in [41, Section 2.5]. This orbifold parametrizes isomorphism types of stable marked curves with a conjugation.

Step 2 The next step is to cut $\overline{\mathbb{R}\mathcal{M}}_{g,k,2l}$ along strata which parametrize surfaces with at least one real node. These strata form a real normal crossing divisor, as they are the fixed-point loci of the previous involution, applied to the normal crossing divisor of nodal strata in $\overline{\mathcal{M}}_{g,k+2l}$. The cutting procedure is via the *real hyperplane blowup* of [41, Section 3.3], and it is proven there that the result of this blowup is an orbifold with corners which we denote by $\widetilde{\mathbb{R}\mathcal{M}}_{g,k,2l}$.

Step 3 $\widetilde{\mathbb{R}\mathcal{M}}_{g,k,2l}$ is made of several connected component. Consider those components whose generic point is a real curve C with a conjugation ϱ such that $C \setminus C^{\varrho}$ is disconnected. Then $\widetilde{\mathcal{M}}^{\mathbb{R}}_{g,k,2l}$ is the disconnected two-to-one cover of the union of those connected components, given, at the level of objects (C, ϱ), by the choice of a *distinguished half*, a connected component of $C \setminus C^{\varrho}$. Thanks to the real blowup procedure, this choice extends naturally to the boundary strata. The resulting space is still a compact orbifold with corners, as a degree 2 cover of such a space.

Step 4 $\overline{\mathcal{M}}^{\mathbb{R}}_{g,k,l}$ is the submoduli of $\widetilde{\mathcal{M}}^{\mathbb{R}}_{g,k,2l}$ made of connected components such that the marked points $w_{k+1}, \ldots, w_{k+l}$ lie in the distinguished half. This final space is a compact orbifold with corners, as it is the union of connected components of a compact orbifold with corners.

Set-theoretically $\overline{\mathcal{M}}^{\mathbb{R}}_{g,k,l}$ is naturally identified with the moduli space of stable marked open (g, k, l)–surfaces, and therefore we identify this moduli with $\overline{\mathcal{M}}^{\mathbb{R}}_{g,k,l}$. The construction endows the moduli space $\overline{\mathcal{M}}^{\mathbb{R}}_{g,k,l}$ with topology and an orbifold with corners structure. For the dimension, see, for example, [27, Theorem 1.2].

In general the space $\overline{\mathcal{M}}^{\mathbb{R}}_{g,k,l}$ is nonorientable and disconnected. A stable marked surface with b boundary nodes or contracted boundaries belongs to a corner of the moduli space $\overline{\mathcal{M}}^{\mathbb{R}}_{g,k,l}$ of codimension b. For further reading about the nodal strata of the real and open moduli spaces we refer the reader to [27, Section 3].

Notation 2.15 Denote by $D\colon \overline{\mathcal{M}}^{\mathbb{R}}_{g,k,l} \to \overline{\mathcal{M}}_{g,k+2l}$ the moduli-level doubling map $\Sigma \to \Sigma_{\mathbb{C}}$, which is the composition of the maps of (14).

2.3 Graded surfaces and their moduli space

We present here the extra structure needed for the definition of intersection theory for open Riemann surfaces, following [35; 34].

2.3.1 Smooth graded surfaces Let Σ be a smooth closed genus g surface. A *spin structure* twisted in $\{z_i\}_{i\in\mathcal{I}_1}$, where $\mathcal{I}_1 \subseteq \mathcal{I}$, is a complex line bundle $\mathcal{L} \to \Sigma$ together with an isomorphism

$$b\colon \mathcal{L}^{\otimes 2} \simeq \omega_\Sigma\Bigl(-\sum_{i\in\mathcal{I}_1} z_i\Bigr),$$

where $\omega_\Sigma\bigl(-\sum_{i\in\mathcal{I}_1} z_i\bigr)$ is the canonical bundle twisted in $\{z_i\}_{i\in\mathcal{I}_1}$.

Let Σ be a smooth genus g open surface. A *real spin structure* twisted in $\{x_i\}_{i\in\mathcal{B}_1}$ and $\{z_i\}_{i\in\mathcal{I}_1}$, where $\mathcal{B}_1 \subseteq \mathcal{B}$ and $\mathcal{I}_1 \subseteq \mathcal{I}$, is a triple $(\mathcal{L}, b, \widetilde{\varrho})$, where $(\mathcal{L}, b)$ is a spin structure on the doubled surface $D(\Sigma) = (\Sigma_{\mathbb{C}}, \varrho)$ twisted in $\{x_i\}_{i\in\mathcal{B}_1}$ and $\{z_i, \bar{z}_i\}_{i\in\mathcal{I}_1}$, ie $\mathcal{L} \to \Sigma_{\mathbb{C}}$ is a line bundle and

$$b\colon \mathcal{L}^{\otimes 2} \simeq \omega_{\Sigma_{\mathbb{C}}}\Bigl(-\sum_{i\in\mathcal{B}_1} x_i - \sum_{i\in\mathcal{I}_1}(z_i + \bar{z}_i)\Bigr)$$

is an isomorphism, where $\omega_{\Sigma_{\mathbb{C}}}\bigl(-\sum_{i\in\mathcal{B}_1} x_i - \sum_{i\in\mathcal{I}_1}(z_i + \bar{z}_i)\bigr)$ is the canonical bundle twisted in $\{x_i\}_{i\in\mathcal{B}_1}$ and $\{z_i, \bar{z}_i\}_{i\in\mathcal{I}_1}$. The map $\widetilde{\varrho}\colon \mathcal{L} \to \mathcal{L}$, is an involution which lifts $d\varrho$, the induced involution on $\omega_{\Sigma_{\mathbb{C}}}$.

The maps $\widetilde{\varrho}$ and $d\varrho$ restrict to conjugations on the fibers of

$$\mathcal{L} \to \Sigma^{\varrho}_{\mathbb{C}} \simeq \partial\Sigma, \quad \omega_{\Sigma_{\mathbb{C}}}\Bigl(-\sum_{i\in\mathcal{B}_1} x_i\Bigr) \to \Sigma^{\varrho}_{\mathbb{C}} \simeq \partial\Sigma.$$

These conjugations give rise to a ϱ–invariant real subbundle. The real line bundle

$$\omega_{\Sigma_{\mathbb{C}}}\Bigl(-\sum_{i\in\mathcal{B}_1} x_i\Bigr)^{\varrho} \to \Sigma^{\varrho}_{\mathbb{C}}$$

is oriented: take any nowhere-vanishing section $\xi \in \Gamma(T\Sigma^{\varrho}_{\mathbb{C}} \to \Sigma^{\varrho}_{\mathbb{C}})$ which points in the direction of the orientation on $\Sigma^{\varrho}_{\mathbb{C}}$, induced from its identification with $\partial\Sigma$. The orientation of $\omega^{\varrho}_{\Sigma_{\mathbb{C}}}|_{\Sigma^{\varrho}_{\mathbb{C}}\setminus i\in\mathcal{B}_1}$ is defined by a section $\widehat{\xi}$ which satisfies $\widehat{\xi}(\xi) > 0$.

Such a section is said to be *positive*. Thus, using b, it is seen that for any connected component of $\Sigma_{\mathbb{C}}^{\varrho} \setminus \{x_i\}_{i\in\mathcal{B}_1}$, either $\hat{\xi}$ or $-\hat{\xi}$ has a root in $\mathcal{L}^{\tilde{\varrho}}$. In the case that for each connected component of $\Sigma_{\mathbb{C}}^{\varrho} \setminus A$, where $A \supseteq \{x_i\}_{i\in\mathcal{B}_1}$ is a finite set of points, the positive sections have roots in $\mathcal{L}^{\tilde{\varrho}}$, we say that $(\mathcal{L}, \tilde{\varrho})$ is *compatible* away from A. In the case $A = \{x_i\}_{i\in\mathcal{B}_1}$, we say that the structure is compatible.

Proposition 2.16 *If $\mathcal{B}_1 \neq \varnothing$ then there are no compatible real twisted spin structures.*

Proof Suppose $i \in \mathcal{B}_1$. Let U be a contractible ϱ–invariant neighborhood of x_i which contains no other marked points. One can find a ϱ–invariant section $s \in \Gamma(\mathcal{L} \to U)$ which vanishes nowhere in U, possibly after replacing U by a smaller neighborhood. In ϱ–anti-invariant local coordinates around x_i, the real section $z\,dz$ generates $\omega_{\Sigma_{\mathbb{C}}}(U)$. Write $f(z) = z\,dz/b(s^{\otimes 2})$; this is a nowhere-vanishing holomorphic function in U. Moreover, f is conjugation invariant, and hence real on U^{ϱ}. In particular, it does not change sign there. But this is impossible for a compatible structure since $z\,dz$ is positive on exactly one component of $U^{\varrho} \setminus \{x_i\}$. $\square$

Given a compatible real spin structure, a *lifting* of the spin structure is a choice of a section in

$$\Gamma(S^0(\mathcal{L}^{\tilde{\varrho}}) \to \Sigma_{\mathbb{C}}^{\varrho} \setminus \{x_i\}_{i\in\mathcal{B}}),$$

where S^0 stands for the rank 0 sphere bundle. We say that the lifting *alternates* in x_j, and that x_j is a *legal* point, if this choice cannot be extended to $\Gamma(S^0(\mathcal{L}^{\tilde{\varrho}}) \to \Sigma_{\mathbb{C}}^{\varrho} \setminus \{x_i\}_{i\in\mathcal{B}\setminus\{j\}})$. Otherwise the lifting does not alternate in x_j, and x_j is an *illegal* point.

Definition 2.17 A *twisted closed smooth spin surface* is a closed smooth surface $(\Sigma, \{z_i\}_{i\in\mathcal{I}})$, together with a twisted spin structure twisted in $\{z_i\}_{i\in\mathcal{I}_1}$. In the case $\mathcal{I}_1 = \varnothing$, we call it a *closed smooth spin surface*.

A *twisted open smooth spin surface* is a smooth open surface $(\Sigma, \{x_i\}_{i\in\mathcal{B}}, \{z_i\}_{i\in\mathcal{I}})$, together with a compatible twisted real spin structure twisted in $\{z_i\}_{i\in\mathcal{I}_1}$. In the case that $\mathcal{I}_1 = \varnothing$, we call it an *open smooth spin surface*. A (*twisted*) *smooth spin surface with a lifting* is a (twisted) open spin surface, together with a lifting. A lifting with all boundary points legal is called a *grading*. A surface with a nontwisted spin structure (that is, $\mathcal{I}_1 = \mathcal{B}_1 = \varnothing$) and a grading is called a *graded surface*. An isomorphism of twisted spin surfaces is an isomorphism of the underlying surfaces and of the line bundles which respects the twists, commutes with the maps between the canonical lines in the expected sense and, in the open case, also with the involutions. An isomorphism of twisted spin surfaces with a lifting is an isomorphism of the twisted spin surfaces which takes the lifting to the lifting in the target, and respects the alternations.

We will see below in Proposition 2.32 that the only obstruction to the existence of a graded spin structure is the parity of $g+k$: in a graded spin structure, $g+k$ must be odd.

2.3.2 Stable graded surfaces We follow the terminology of [19]. Let $\Sigma=\{\Sigma_\alpha\}_{\alpha\in\mathcal{C}}$ be a stable closed surface. A *spin structure* twisted in $\{z_i\}_{i\in\mathcal{I}_1}$, where $\mathcal{I}_1\subseteq\mathcal{I}$, is a rank 1 torsion-free sheaf $\mathcal{L}$ over Σ together with a map

$$b\colon \mathcal{L}^{\otimes 2}\to\omega_\Sigma\biggl(-\sum_{i\in\mathcal{I}_1} z_i\biggr),$$

where $\omega_\Sigma\bigl(-\sum_{i\in\mathcal{I}_1} z_i\bigr)$ is the dualizing sheaf, twisted in $\{z_i\}_{i\in\mathcal{I}_1}$.

We require that

(a) $\deg(\mathcal{L})=\frac{1}{2}(\deg(\omega_\Sigma)-|\mathcal{I}_1|)$,

(b) b is an isomorphism on the locus where $\mathcal{L}$ is locally free, and

(c) for any point p where $\mathcal{L}$ is not free, the length of $\operatorname{coker}(b_p)$ is 1.

In particular, b is an isomorphism away from nodes. Nodes where b is not an isomorphism are called *Neveu–Schwarz* (NS); at these nodes the last requirement says exactly that b vanishes in order 2. The other nodes are called *Ramond*.

Let $\Sigma=\{\Sigma_\alpha\}_{\alpha\in\mathcal{C}\cup\mathcal{O}}$ be a stable open (g,k,l)–surface. A *real spin structure* twisted in $\{x_i\}_{i\in\mathcal{B}_1}$ and $\{z_i\}_{i\in\mathcal{I}_1}$, with $\mathcal{I}_1\subseteq\mathcal{I}$ and $\mathcal{B}_1\subseteq\mathcal{B}$, is a triple $(\mathcal{L},b,\widetilde{\varrho})$, where $(\mathcal{L},b)$ is a spin structure over the doubled surface $D(\Sigma)=(\Sigma_{\mathbb{C}},\varrho)$, and $\widetilde{\varrho}\colon\mathcal{L}\to\mathcal{L}$ is an involution which lifts $d\varrho$, the induced involution on $\omega_{\Sigma_{\mathbb{C}}}$. Thus, in particular, b is a map

$$b\colon \mathcal{L}^{\otimes 2}\to\omega_{\Sigma_{\mathbb{C}}}\biggl(-\sum_{i\in\mathcal{B}_1} x_i-\sum_{i\in\mathcal{I}_1}(z_i+\bar z_i)\biggr),$$

where $\omega_{\Sigma_{\mathbb{C}}}\bigl(-\sum_{i\in\mathcal{B}_1} x_i-\sum_{i\in\mathcal{I}_1}(z_i+\bar z_i)\bigr)$ is the dualizing sheaf, twisted in $\{x_i\}_{i\in\mathcal{B}_1}$ and $\{z_i,\bar z_i\}_{i\in\mathcal{I}_1}$, and

$$\deg(\mathcal{L})=\tfrac{1}{2}\bigl(\deg(\omega_{\Sigma_{\mathbb{C}}})-2|\mathcal{I}_1|-|\mathcal{B}_1|\bigr).$$

Remark 2.18 Suppose that Σ is a nodal curve, open or closed, and z is a node with preimages $z_\nu, z_\mu \in \mathrm{Norm}(\Sigma)$. Then there are natural residue maps

$$\mathrm{res}_\eta \colon (\mathrm{Norm}^*\omega_\Sigma)_{z_\eta} \simeq \mathbb{C}.$$

These induce an isomorphism $a\colon (\mathrm{Norm}^*\omega_\Sigma)_{z_\mu} \simeq (\mathrm{Norm}^*\omega_\Sigma)_{z_\nu}$, by

$$\mathrm{res}(v) + \mathrm{res}(a(v)) = 0.$$

In the Ramond case, we also have an isomorphism $\widetilde{a}\colon (\mathrm{Norm}^*\mathcal{L})_{z_\mu} \to (\mathrm{Norm}^*\mathcal{L})_{z_\nu}$, and $\mathrm{res}(b(v^{\otimes 2})) + \mathrm{res}(b(\widetilde{a}(v)^{\otimes 2})) = 0$. For more details see [19].

When $z \in \Sigma \subset \Sigma_\mathbb{C}$ is a contracted boundary which is Ramond, $d\varrho$ and $\widetilde{\varrho}$ lift to complex antilinear isomorphisms between the fibers of $\mathrm{Norm}^*\omega_{\Sigma_\mathbb{C}}$ and $\mathrm{Norm}^*\mathcal{L}$ in $z_\pm$, where z_+ is the preimage of z in $\mathrm{Norm}(\Sigma)$, and z_- is the preimage of z in $\mathrm{Norm}(\overline{\Sigma})$. By composing with a and $\widetilde{a}$ we get antilinear involutions on the fibers at z_+. This defines real lines, which we denote by $(\omega_\Sigma^\varrho)_{z_+}$ and $(\mathcal{L}^{\widetilde{\varrho}})_{z_+}$, together with maps

$$\mathrm{res}\colon (\omega_\Sigma^\varrho)_{z_+} \simeq \sqrt{-1}\mathbb{R},$$

where $\sqrt{-1}$ is the root of -1 in the upper half-plane, and

$$b^2\colon (\mathcal{L}^{\widetilde{\varrho}})_{z_+} \to (\omega_\Sigma^\varrho)_{z_+},$$

defined by $b^2(v) = b(v^{\otimes 2})$.

We say that the real spin structure is *compatible in a contracted boundary* z if z is a Ramond node of $\Sigma_\mathbb{C}$ and the image of b^2 is in the positive imaginary half-line $\mathrm{res}^{-1}(\sqrt{-1}\mathbb{R}_{\geq 0})$.

The real spin structure is *compatible* if it is compatible in contracted boundaries and away from special boundary points. Compatibility away from special points is defined as in the smooth case.

A *lifting* of a compatible real spin structure is a choice of a section

$$s \in \Gamma\biggl(S^0(\mathcal{L}^{\widetilde{\varrho}}) \to \Sigma_\mathbb{C}^\varrho \setminus \biggl(\bigcup_{\alpha \in \mathcal{O}} B(\Sigma_\alpha)\biggr)\biggr),$$

where S^0 stands for the rank 0 sphere bundle. The notions of alternations and of legal marked point or a legal half-node are as in the smooth case.

The definition of the lifting includes, for any contracted boundary node z, a *choice of a lifting for the contracted boundary*, ie with the above notation and identifications, a choice of direction in $(\mathcal{L}^{\widetilde{\varrho}})_{z_+}$ which is mapped by $\mathrm{res} \circ b^2$ to the ray $\sqrt{-1}\mathbb{R}_{\geq 0}$.

Proposition 2.19 (a) *A real spin structure on a stable surface, twisted or not, induces a real spin structure, possibly twisted, on any open component of the normalization, and a possibly twisted spin structure on any closed component of it. For any node of Σ, the induced structure is either twisted in both of its preimages in the normalization, or not twisted in both. The former case is the Ramond case, the latter is Neveu–Schwarz. If there are no Ramond nodes then the spin structures on the closed components of the normalization, together with the real spin structures on its open components, determine the real spin structure on Σ.*

(b) *If the real spin structure is compatible, then so is the induced structure on any open component of the normalization. In this case, in particular, there are no twists in boundary marked points, and no boundary Ramond nodes. In the case that there are no Ramond internal nodes but there may be contracted boundaries, compatible spin structures on the normalization determine the compatible spin structure on Σ.*

(c) *A lifting on Σ induces a lifting on the normalization. A lifting on the normalization, together with a choice of a direction in $(\mathrm{res} \circ b^2)^{-1}(\sqrt{-1}\mathbb{R}_{\geq 0})$ for the preimage z_+ of any contracted boundary, induces a lifting on Σ.*

Proof The fact that the twisted spin structure induces one on the normalization by pullback, and is induced by one, when there are no Ramond nodes is already true in the closed case; see for example [19]. Moreover, it is shown there that given the structures on the normalization and the identifications of the stalks in preimages of nodes — see Remark 2.18 — the twisted spin structure on the surface is determined. The involution extends uniquely by continuity.

The second claim follows from the fact that one can examine compatibility away from special points. Ramond boundary nodes cannot appear by Proposition 2.16. If z is a contracted boundary, there is a single, up to sign, possible identification map $\widetilde{a}$, as in Remark 2.18. Now, if $\widetilde{a}$ makes the contracted boundary compatible, with respect to the involution, $-\widetilde{a}$ will make it not compatible, and vice versa. The last statement is evident. □

Definition 2.20 A closed stable surface $(\Sigma, \{z_i\}_{i \in \mathcal{I}})$, together with a spin structure twisted in $\{z_i\}_{i \in \mathcal{I}_1}$, is called a *twisted closed stable spin surface*. In the case that $\mathcal{I}_1 = \varnothing$, we call it a *stable closed spin surface*. A *twisted open stable spin surface* is a stable open surface $(\Sigma, \{x_i\}_{i\in\mathcal{B}}, \{z_i\}_{i\in\mathcal{I}})$, together with a compatible real spin structure twisted in $\{z_i\}_{i\in\mathcal{I}_1}$. In the case that $\mathcal{I}_1 = \varnothing$, we call it a *stable open spin surface*. A (*twisted*) *stable spin surface with a lifting* is a (twisted) open spin surface, together with a lifting such that for any boundary node, exactly one half-node is legal. If all the boundary marked points are legal, the lifting is called a *grading*. A (twisted) stable spin surface with a grading is *effective* if the underlying surface is, and, for any component of the normalization of genus zero with 3 special boundary points and no special internal points, its special points are legal. A *stable graded surface* is a (nontwisted) stable spin surface with a grading. The isomorphism notions are as in the smooth case.

The legality condition on the nodes may seem peculiar at first glance. However this is the condition which allows smoothing of the stable graded surface at a boundary node. The closed analog of it is that the twists at the two half-nodes of the same node must agree. In a nutshell, as we will see in the next subsection, in a twisted spin surface any closed path which does not pass through special points has a well-defined notion of parity. By pinching the surface in that path, a node is formed, and this node is NS or Ramond according to the parity of the pinched path. Similarly, any *oriented* arc between boundary points which avoids special points also has a well-defined notion of parity. We will see in Proposition 2.31 that this parity changes if the orientation of the arc changes. By pinching the arc one obtains a surface with a new boundary node. The boundary node is NS, but the legality of its half-nodes is determined by the parity of the corresponding oriented arcs. See Lemma 2.39 for an exact statement. Interestingly, when the node is separating the legality can be determined from the parity considerations of Proposition 2.32. Since in $g = 0$ all nodes are separating, the genus 0 theory could have been defined without referring to the graded spin structure. These points will be discussed more in [34].

Notation 2.21 Denote by $\mathrm{Spin}(\Sigma)$ the set of isomorphism classes of graded spin structures on a stable open surface Σ.

The definition of graded surfaces, together with Proposition 2.19, yields a corollary.

Corollary 2.22 *If Σ has no internal nodes, there is a bijection between* $\mathrm{Spin}(\Sigma)$ *and*

(a) *isomorphism types of spin structures with a lifting on* $\mathrm{Norm}(\Sigma)$, *twisted precisely at preimages of contracted boundaries, such that any boundary marked point*

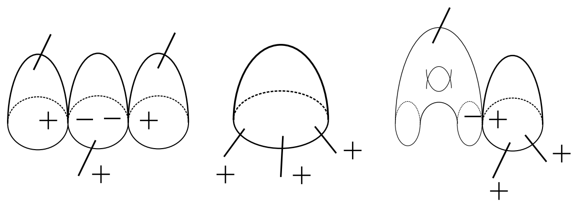


Figure 6: In this figure we consider three stable graded spin surfaces. In these cases the underlying surfaces are effective. The symbol $+$ near a boundary marking or a boundary half-node indicates a legal point, while $-$ indicates an illegal point. We omit the labels. The graded surface on the left is noneffective, since in the normalization the middle component has one legal boundary marking, two illegal boundary markings and no internal markings. On the other hand, the remaining two graded surface are effective.

of Σ *is legal as a point of* $\mathrm{Norm}(\Sigma)$, *and for any node of* Σ *exactly one half-node in* $\mathrm{Norm}(\Sigma)$ *is legal*;

(b) *a choice of a direction in* $(\mathrm{res}\circ b^2)^{-1}(\sqrt{-1}\mathbb{R}_{\geq 0})$ *for the preimage* z_+ *of any contracted boundary.*

2.3.3 An alternative definition for the smooth case In this subsection we provide an alternative definition for smooth spin surfaces with a lifting. This definition will be easier to work with. Let $(\Sigma, \{x_i\}_{i\in\mathcal{B}}, \{z_j\}_{j\in\mathcal{I}})$ be a smooth, open or closed, pointed Riemann surface. Choose any Riemannian metric on it.

Notation 2.23 Denote by $T^1\Sigma$ the S^1–bundle of $T\Sigma$. For a simple smooth arc or a simple smooth closed path $\gamma\subset\Sigma$, we denote the S^0–bundle of $T\gamma$ by $T^1\gamma$.

When the arc or path γ is oriented, $T^1\gamma$ will stand for the unit-length oriented tangent vector field to γ. In particular, we shall use the notation $T^1\partial\Sigma$ for the branch of $T^1\partial\Sigma$ which covers the direction of the induced orientation on the boundary.

We consider $T^1\Sigma$ as the S^1–subbundle of unit-length vectors of $T\Sigma$; similarly for $T^1\gamma$. We also consider $T^1\gamma$ as a S^0–subbundle of $T^1\Sigma|_\gamma$. In what follows we use these identifications without mentioning a choice of metric. Different metrics will give rise to equivalent structures, and in fact, one can make these definitions metric independent by considering the S^0– and S^1–bundles as subquotients of the corresponding vector bundles.

For a point $p\in\Sigma$, a vector $w\in T_p\Sigma$ and an angle $\theta\in\mathbb{R}/2\pi\mathbb{R}$, let $r_\theta w = r_\theta(p)w$ be the operator of rotation by θ in the counterclockwise direction. We shall omit p from the notation when it is clear from context. The operator $r_\theta(p)$ is induced on $T^1_p\Sigma$, and we shall use the same notation.

If u and w are two tangent vectors at p, denote the counterclockwise angle from u to w by $\measuredangle(u,w)$.

For a smooth arc $\gamma\colon[0,1]\to\Sigma$, there exists a canonical trivialization

$$\varsigma\colon[0,1]\times S^1\to T^1\Sigma|_\gamma$$

defined by

$$\varsigma(t,\theta)=(\gamma(t),e^{i\theta}v_t),\quad \text{where } v_t=(T^1)_{\gamma(t)}\gamma.$$

This trivialization defines a continuous family of maps

$$\{p(\gamma)^t_s\colon T^1_{\gamma(s)}\Sigma\to T^1_{\gamma(t)}\Sigma\}_{0\le s,t\le 1},$$

uniquely determined by the condition

$$p_2\bigl(\varsigma^{-1}(\gamma(s),v)\bigr)=p_2\bigl(\varsigma^{-1}(\gamma(t),p(\gamma)^t_s v)\bigr),$$

where p_2 is the projection on the second coordinate. One can extend the trivialization to the piecewise smooth context by approximation. In the case that $s=0$ and $t=1$, we omit them from the notation and write $p(\gamma)$. One can easily verify, in the piecewise smooth case, that if γ is composed of smooth subarcs $\gamma_i\colon[a_i,a_{i+1}]\to\Sigma$, where $a_0=0<a_1<\cdots<a_n=1$, and θ_{i+1} is $\measuredangle(\dot\gamma_i|_{\gamma_{i+1}(a_{i+1})},\dot\gamma_i|_{\gamma_i(a_{i+1})})$, then

$$p(\gamma)=p(\gamma_{n-1})r_{\theta_{n-1}}p(\gamma_{n-2})\cdots r_{\theta_1}p(\gamma_0).$$

We shall denote such γ by $\gamma_1\to\gamma_2\to\cdots\to\gamma_n$. For a *closed* piecewise smooth path γ, we slightly change the definition of p to be

$$p(\gamma)=r_{\theta_0}p(\gamma_{n-1})r_{\theta_{n-1}}p(\gamma_{n-2})\cdots r_{\theta_1}p(\gamma_0),$$

and note that this is in fact the identity map. We shall denote such γ by $\gamma_1\to\gamma_2\to\cdots\to\gamma_n\to\gamma_1$.

Definition 2.24 A *twisted spin structure* $\mathbb{S}\to\Sigma\setminus\{z_j\}_{j\in I}$ on a smooth marked Σ is an S^1–bundle on $\Sigma\setminus\{z_j\}_{j\in I}$ together with a degree 2 cover bundle map

$$\pi=\pi^{\mathbb{S}}\colon\mathbb{S}\to T^1\Sigma|_{\Sigma\setminus\{z_j\}_{j\in I}}.$$

For a point $p \in \Sigma$, a vector $w \in \mathbb{S}_p$ and an angle $\theta \in \mathbb{R}/4\pi\mathbb{Z}$, let $R_\theta w = R_\theta(p)w$ be the operator of rotation by θ in the counterclockwise direction. We shall omit p from the notation when it is clear from context.

The *parallel transport* along $\gamma\colon [0,1] \to \Sigma$ is the unique continuous family of maps

$$\{P(\gamma)_s^t \colon \mathbb{S}_{\gamma(s)} \to \mathbb{S}_{\gamma(t)}\}_{0 \le s,t \le 1}$$

which covers $\{p(\gamma)_s^t\}$. We shall sometimes call $P(\gamma)_0^1 v$ the *parallel transport of* v *along* γ, and write it as $P(\gamma)v$.

Remark 2.25 R covers r in the sense that if $\pi(s) = v$ for $s \in \mathbb{S}_p$ and $v \in T_p^1\Sigma$, then

$$\pi(R_\theta(p)s) = r_\theta(p)v = r_{\theta (\operatorname{mod} 2\pi)}(p)v.$$

Observe that $R_\alpha R_\beta = R_{\alpha+\beta}$. In addition, P and R commute:

$$R_\theta(\gamma(t))P(\gamma)_s^t v = P(\gamma)_s^t R_\theta(\gamma_s)v.$$

Definition 2.26 A (twisted) spin structure $\mathbb{S}$ is associated with a function

$$q = q^{\mathbb{S}} \colon H_1(\Sigma \setminus \{z_j\}_{j\in I}, \mathbb{Z}_2) \to \mathbb{Z}_2$$

defined as follows. For $x \in H_1(\Sigma \setminus \{z_j\}_{j\in I}, \mathbb{Z}_2)$, take a piecewise smooth connected representative γ. Then $p(\gamma)$ is the identity. Hence $P(\gamma)$ is either the identity or minus the identity. We define $q(x) = q(\gamma)$ to be 1 in the former case, and 0 otherwise.

For any internal marked point z_j, take a small disk D_j which surrounds it and contains no other marked points in its closure. We define the *twist* in z_j to be $q(\partial D_j)$.

The following well-known theorem was proven by Johnson [20]. It states that q is a *quadratic enhancement* of the Poincaré pairing $\langle \alpha, \beta\rangle$.

Theorem 2.27 *The function q is well defined on $H_1(\Sigma \setminus \{z_j\}_{j\in I}, \mathbb{Z}_2)$. For $\alpha, \beta \in H_1(\Sigma \setminus \{z_j\}_{j\in I}, \mathbb{Z}_2)$, we have*

$$q(\alpha+\beta) = q(\alpha) + q(\beta) + \langle \alpha, \beta\rangle.$$

Proposition 2.28 *If $\gamma\colon [0,1] \to \Sigma \setminus \{z_j\}_{j\in I}$ is a piecewise smooth closed curve which bounds a contractible domain, then $P(\gamma)_0^1 = R_{2\pi}$. Moreover, suppose Σ is a disk with a piecewise smooth boundary γ. Let $\mathbb{S} \to T^1\Sigma|_\gamma$ be a double cover by an S^1–bundle $\mathbb{S}$. Then $\mathbb{S}$ can be extended to a nontwisted spin structure on Σ if and only if $P(\gamma)_0^1 = R_{2\pi}$. In this case the extension is unique. In particular, the spin structure can be extended to a marked point z_i if and only if its twist is 0, in which case the extension is unique.*

The first part follows from Theorem 2.27 by taking $\alpha = \beta = [\gamma]$. The other parts are also simple and will be omitted.

Definition 2.29 Let $(\Sigma, \mathbb{S})$ be an open marked Riemann surface together with a (twisted) spin structure. Suppose $\partial\Sigma \neq \varnothing$. A *lifting* is a choice of a section

$$s\colon \partial\Sigma \setminus \{x_i\}_{i\in\mathcal{B}} \to \mathbb{S}|_{\partial\Sigma\setminus\{x_i\}_{i\in\mathcal{B}}}$$

which covers the *oriented* $T^1(\partial\Sigma \setminus \{x_i\}_{i\in\mathcal{B}})$.

For $j \in \mathcal{B}$, suppose $i\colon \left(-\frac{1}{2}, \frac{1}{2}\right) \to \partial\Sigma$ is a smooth orientation-preserving embedding with $i(0) = x_j$ and $x_b \notin i\left(-\frac{1}{2}, \frac{1}{2}\right)$ for $b \neq j$. In the case that

$$\lim_{x\to 0^-} s(x) \neq \lim_{x\to 0^+} s(x),$$

we say that the structure *alternates* in x_j, and that x_j is a *legal point*. Otherwise x_j is *illegal* and the structure does not alternate. We extend the definition of s to the boundary marked points by $s(x) = \lim_{x\to 0^+} s(x)$.

A *smooth spin surface with a lifting* $(\Sigma, \{x_i\}_{i\in\mathcal{B}}, \{z_i\}_{i\in\mathcal{I}}, \mathbb{S}, s)$ is a smooth open Riemann surface together with a spin structure and a lifting. A *smooth graded surface* is a smooth spin surface with a lifting, such that all boundary marked points are legal.

The notion of alternation can be generalized in the following manner.

Definition 2.30 A *bridge* is a piecewise smooth simple arc which meets the boundary only at its two distinct endpoints $x, y \in \partial\Sigma \setminus \{x_i\}_{i\in\mathcal{B}}$. Suppose we orient the bridge and parametrize it as

$$\gamma\colon [0,1] \to \Sigma, \quad \text{with } \gamma(0) = x, \gamma(1) = y.$$

Define $Q(\gamma) \in \mathbb{Z}_2$ by the equation

$$R_{2\pi-\alpha_y}(y)P(\gamma)R_{\alpha_x}(x)s(x) = R_{2\pi Q(\gamma)}(y)s(y), \tag{15}$$

where

$$\alpha_x = \measuredangle((T^1)_x\partial\Sigma, (T^1)_x\gamma) \in [0,\pi],$$
$$\alpha_y = \measuredangle((T^1)_y\partial\Sigma, (T^1)_y\gamma) \in [\pi,2\pi].$$

$Q(\gamma)$ depends on the orientation but not on the parametrization. An oriented bridge with $Q = 1$ is called *a legal side of the bridge*, otherwise it is called an *illegal side*.

Proposition 2.31 *Let Σ be a smooth open spin surface with a lifting. Let γ be a bridge and denote by $\overline{\gamma}$ the same bridge with opposite orientation. Then $Q(\gamma)+Q(\overline{\gamma})=1$. Thus, any bridge has exactly one legal side and exactly one illegal side.*

Proof Work with the notation of Definition 2.30. For $w\in\{x,y\}$, α'_w is defined by $\alpha'_w=\measuredangle((T^1)_w\partial\Sigma,(T^1)_w\overline{\gamma})$. Observe that $\alpha'_x=\alpha_x+\pi$ and $\alpha'_y=\alpha_y-\pi$. Apply $R_{2\pi Q(\overline{\gamma})}(y)$ to the left-hand side of (15). By Remark 2.25, the left-hand side becomes

$$R_{2\pi Q(\overline{\gamma})}(y)R_{2\pi-\alpha_y}(y)P(\gamma)R_{\alpha_x}(x)s(x)=R_{2\pi-\alpha_y}(y)P(\gamma)R_{\alpha_x}(x)R_{2\pi Q(\overline{\gamma})}(x)s(x).$$

Using equation (15) for $\overline{\gamma}$, Remark 2.25 again, and the relations between pairs α_x,α'_x and α_y,α'_y, the last expression simplifies to $R_\pi P(\overline{\gamma})R_\pi P(\gamma)s(y)$. By Proposition 2.28 applied to the piecewise smooth closed curve $\gamma\to\overline{\gamma}\to\gamma$, this is just $R_{2\pi}(y)s(y)$.

Applying $R_{2\pi Q(\overline{\gamma})}(y)$ to the right-hand side of (15), we obtain $R_{2\pi(Q(\gamma)+Q(\overline{\gamma}))}(y)s(y)$. Thus,

$$R_{2\pi}(y)s(y)=R_{2\pi(Q(\gamma)+Q(\overline{\gamma}))}(y)s(y),$$

and the claim follows. □

Proposition 2.32 (a) *Suppose $(\Sigma,\{z_i\}_{i\in\mathcal{I}},\mathbb{S})$ is a genus g closed spin surface. Suppose that exactly l_1 marked points have twist 1. Then l_1 is even. For any closed Riemann surface $(\Sigma,\{z_i\}_{i\in\mathcal{I}})$, there exist 2^{2g} distinct nontwisted spin structures on Σ.*

(b) *Suppose $(\Sigma,\{x_i\}_{i\in\mathcal{B}},\{z_i\}_{i\in\mathcal{I}},\mathbb{S},s)$ is a genus g open spin surface with a lifting. Suppose that exactly k_+ of the boundary marked points are legal, and l_1 internal marked points have twist 1. Then*

$$l_1=g+1+k_+\ (\mathrm{mod}\ 2).$$

For any $(\Sigma,\{x_i\}_{i\in\mathcal{B}},\{z_i\}_{i\in\mathcal{I}})\in\mathcal{M}^{\mathbb{R}}_{g,k,l}$ with $2\mid g+k+1$, there exist exactly 2^g graded structures on Σ.

Proof For the first claim, let $\{C_i\}$ be a family of nonintersecting circles around each marked point. Then $\sum C_i$ is homologous to 0. By Theorem 2.27, $q\bigl(\sum C_i\bigr)=\sum q(C_i)=0$. For the number of spin structures, see for example [19].

Regarding the second claim, let C_i be as above, and for any boundary component $\partial\Sigma_b$, let C_b be a curve surrounding this boundary, disjoint from it, but isotopic to it in $\Sigma\setminus\boldsymbol{z}$.

By the definitions of q and Q one easily sees that $q(C_b)$ is 1 plus the number of legal marked points of $\partial\Sigma_b$. Again

$$\sum q(C_i) + \sum q(C_b) = 0 \text{ (mod 2)},$$

but this sum equals $l_1 + k_+ + b$ (mod 2), where b is the number of boundaries. It is easy to see that $b = g + 1$ (mod 2). For the number of graded structures; see [34]. We will also obtain it as a byproduct in Section 5.1; see the end of Example 5.18. □

Lemma 2.33 *The definitions given in this subsection of smooth spin surfaces with a lifting*, *twisted or not*, *and graded or not*, *are equivalent to the analogous ones given in Section 2.3.1.*

Starting with a real spin structure $\mathcal{L}$ in the sense of Section 2.3.1, $\mathbb{S}$ is just the S^1–bundle of $\mathcal{L}^*$, and the lifting is the reduction of the lifting to that bundle. See [34] for more details, and for the rather straightforward proof of equivalence.

2.3.4 A comment about the alternative definition in the stable case In the stable case, by Proposition 2.19, the sheaf $\mathcal{L}$ and the grading data determine the spin structures and liftings on the normalization, hence by Lemma 2.33 determine the data of $\mathbb{S}$ and s for each component. However, it is determined by it, again, using the same lemma and proposition, only when there are no Ramond nodes. Even when there are such nodes, the data of $\mathbb{S}$ and s for each component determine $\mathcal{L}$ and the grading data up to a finite choice of identification maps between stalks of half-nodes and liftings at the preimages of the contracted boundaries, as explained in the proof of Proposition 2.19. Therefore, since working with the S^1–bundle and its lifting is more convenient, throughout this paper we shall usually write $(\Sigma, \mathbb{S}, s)$ to indicate a spin structure with a lifting, and leave $\mathcal{L}$ implicit. We shall sometimes even leave $\mathbb{S}$ and s implicit.

2.3.5 Spin graphs It is useful to encode some of the combinatorial data of spin surfaces with a lifting in graphs.

Definition 2.34 A (*pre*)*stable spin graph* Γ *with a lifting* is a (pre)stable graph

$$\Gamma = (V, H, \sim\, = \,\sim_B \cup \sim_I),$$

together with a twist map $\mathrm{tw}\colon H^I \to \mathbb{Z}_2$ and an alternation map $\mathrm{alt}\colon H^B \to \mathbb{Z}_2$. We require that

(a) $\mathrm{tw}(h) = \mathrm{tw}(\sigma_1(h))$ for any $h \in H^I \setminus T^I$,

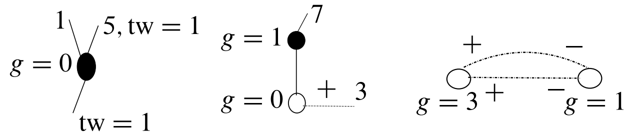


Figure 7: Three examples of graded dual graphs. The numbers stand for the markings, and all twists are 0 unless "tw = 1" is written next to an element of H^I. In order to avoid confusion, legal half-edges, the elements $h \in H^B$ with $\mathrm{alt}(h) = 1$, are decorated by $+$ signs.

(b) $\mathrm{alt}(h) + \mathrm{alt}(\sigma_1(h)) = 1$ for any $h \in H^B \setminus T^B$,

(c) $\mathrm{tw}(h) = 1$ for all $h \in H^{\mathrm{CB}}$,

(d) $\sum_{h \in (\sigma_0^B)^{-1}(v)} \mathrm{alt}(h) + \sum_{h \in (\sigma_0^I)^{-1}(v)} \mathrm{tw}(h) = g(v) + 1 \pmod 2$ for $v \in V^O$,

(e) $\sum_{h \in \sigma_0^{-1}(v)} \mathrm{tw}(h) = 0$ for $v \in V^C$.

A boundary half-edge h, and in particular a tail with $\mathrm{alt}(h) = 0$, is said to be *illegal*, otherwise it is *legal*.

We say that the graph is *stable* if Γ is stable. We call Γ a *graded graph* if $\mathrm{alt}(t) = 1$ for all $t \in T^B$ and $\mathrm{tw}(t) = 0$ for all $t \in T^I \setminus H^{\mathrm{CB}}$.

Γ is *effective* if its underlying graph is effective, $\mathrm{alt}(t) = 1$ for all $t \in T^B$, and for any $v \in V^O$ without internal half-edges, its three boundary half-edges have $\mathrm{alt} = 1$.

The normalization $\mathrm{Norm}(\Gamma)$ is just the normalization of the underlying graph Γ, with the maps tw and alt defined on the tails of $\mathrm{Norm}(\Gamma)$ by their values on the corresponding half-edges of Γ. As in the spinless case, whenever an internal tail of Γ is marked $i \neq 0$, the graph $v_i(\Gamma)$ is the component of $\mathrm{Norm}(\Gamma)$ which contains tails i, but with the additional data of tw and alt.

When it is clear from the context that the dual graph under consideration is a spin graph with a lifting, we sometimes omit the maps tw and alt from the notation.

Definition 2.35 An *isomorphism* between spin graphs with a lifting $(\Gamma, \mathrm{tw}, \mathrm{alt})$ and $(\Gamma', \mathrm{tw}', \mathrm{alt}')$ is a tuple

$$f = (f^V, f^H)$$

such that

(a) $f : \Gamma \to \Gamma'$ is an isomorphism of stable graphs,

(b) $\mathrm{tw}' = \mathrm{tw} \circ f^H$ and $\mathrm{alt}' = \mathrm{alt} \circ f^H|_{H^B}$.

We denote by $\mathrm{Aut}(\Gamma)$ the group of the automorphisms of $\Gamma = (\Gamma, \mathrm{tw}, \mathrm{alt})$.

We denote by $\mathcal{G}$ the set of isomorphism classes of all spin graphs with a lifting. We have a natural forgetful map

$$\widetilde{\mathrm{for}}_{\mathrm{spin}}\colon \mathcal{G} \to \mathcal{G}^{\mathbb{R}}, \quad \text{where } \widetilde{\mathrm{for}}_{\mathrm{spin}}(\Gamma, \mathrm{tw}, \mathrm{alt}) = \Gamma.$$

Write $\mathrm{for}_{\mathrm{spin}}$ for its restriction to graded graphs. We denote by $\mathcal{G}_{g,k,l}$ the set of isomorphism classes of graded graphs with $\mathrm{Image}(\mathtt{m}^B) = [k]$ and $\mathrm{Image}(\mathtt{m}^I) = [l]$. Define $\Gamma_{g,k,l}$ as the unique connected graded dual graph with a single open vertex of genus g, exactly k boundary tails marked by $[k]$, exactly l internal tails marked by $[l]$, $H^{\mathrm{CB}} = \varnothing$, and no further half-edges.

To each graded stable marked surface Σ we associate a graded stable graph $(\Gamma, \mathrm{tw}, \mathrm{alt})$ as follows. First, $\Gamma = \Gamma(\Sigma)$. Let $w \in \Sigma_\alpha$ be any special point of this component. It corresponds to some half-edge h. If $h \in H^I$, then $\mathrm{tw}(h)$ is defined to be the twist in w. If $h \in H^B$, then $\mathrm{alt}(h) = 1$ if and only if h is legal. For brevity we denote the graded stable graph corresponding to Σ by $\Gamma(\Sigma)$, omitting tw and alt from the notation. Note that $\mathrm{Norm}(\Gamma(\Sigma)) = \Gamma(\mathrm{Norm}(\Sigma))$, and whenever an $i \neq 0$ marks an internal marked point, then $v_i(\Gamma(\Sigma)) = \Gamma(\Sigma_i)$.

We can also extend the graph operations to the graded case. The *smoothing* of a stable spin graph with a lifting $(\Gamma, \mathrm{alt}, \mathrm{tw})$, at $f \in E$ is the stable graph

$$d_f\Gamma = (\Gamma', \mathrm{alt}', \mathrm{tw}')$$

such that $d_f(\Gamma) = \Gamma'$. Recall that we may identify H' as a subset of H. We define tw' and alt' as the restrictions of tw and alt with respect to this identification. Given a set $S = \{f_1, \dots, f_n\} \subseteq E(\Gamma)$, define the smoothing at S as

$$d_S\Gamma = d_{f_n}(\cdots d_{f_2}(d_{f_1}\Gamma)\cdots).$$

Note that again in the case $\Gamma = d_S\Gamma'$, H' is canonically identified as a subset of H, and alt and tw respect this identification.

Again we define $\partial^!\colon \mathcal{G} \to 2^{\mathcal{G}}$ and $\partial\colon \mathcal{G} \to 2^{\mathcal{G}}$ by

$$\partial^!\Gamma = \{\Gamma' \mid \Gamma = d_S\Gamma' \text{ for some } S \subseteq E(\Gamma')\} \quad \text{and} \quad \partial\Gamma = \partial^!\Gamma \setminus \{\Gamma\}.$$

And again these maps naturally extend to maps $2^{\mathcal{G}} \to 2^{\mathcal{G}}$.

2.3.6 $\overline{\mathcal{M}}_{g,k,l}$

Notation 2.36 For $\Gamma \in \mathcal{G}$, denote by $\mathcal{M}_\Gamma$ the set of isomorphism classes of marked stable spin surfaces with a lifting, associated to graph Γ.

Define

$$\overline{\mathcal{M}}_\Gamma = \coprod_{\Gamma' \in \partial^! \Gamma} \mathcal{M}_{\Gamma'}.$$

Define $\overline{\mathcal{M}}_{g,k,l} = \overline{\mathcal{M}}_{\Gamma_{g,k,l}}$. Similarly define $\mathcal{M}_{g,k,l}$ as the subspace parametrizing smooth surfaces.

For a marking i, denote by $v_i : \mathcal{M}_\Gamma \to \mathcal{M}_{v_i(\Gamma)}$ the canonical map $[\Sigma] \to [\Sigma_i]$. Observe that in order to define this map we have used Proposition 2.19. If Σ has a contracted boundary, then Σ_i has a marked Ramond point which corresponds to it. The passage from Σ to Σ_i *forgets the lifting at contracted boundaries.*

Theorem 2.37 [35] *The space $\overline{\mathcal{M}}_{g,k,l}$ is a compact smooth orbifold with corners of real dimension $3g - 3 + k + 2l$. It is endowed with a canonical orientation.*

We note that $\overline{\mathcal{M}}_{g,k,l}$ is in general disconnected. Different connected components correspond to different topologies with the same doubled genus, to different partitions of the boundary points between boundary components, and sometimes also to different connected components of graded spin structures.

The main difficulty in this theorem is the proof of orientability. The properties of the canonical orientation will be detailed in Theorem 2.53 below. In Theorem 5.32, Proposition 5.48 and Corollary 5.49 below we will provide a different proof for the orientability and for the properties of the canonical orientations. We now briefly review the proof that $\overline{\mathcal{M}}_{g,k,l}$ is a compact smooth orbifold with corners. As in the spinless case, we rely on [41]. We also refer the reader to [8, Lemma 3.5], where a similar procedure, also based on [41], is applied to the moduli of r–spin disks.

Our starting point is the fact that in the closed setting the moduli space $\overline{\mathcal{M}}^{1/2}_{g,n}$ of twisted spin curves is a smooth orbifold; see, for example [18]. Consider the sequence

$$(16)\quad \overline{\mathcal{M}}_{g,k,l} \xrightarrow{(5)} \widehat{\mathcal{M}}_{g,k,l} \xrightarrow{(4)} \widetilde{\mathcal{M}}_{g,k,l} \xrightarrow{(3)} \widetilde{\mathbb{R}\mathcal{M}}_{g,k,l} \xrightarrow{(2)} \overline{\mathbb{R}\mathcal{M}}^{1/2}_{g,k+2l} \xrightarrow{(1)} \overline{\mathcal{M}}'^{1/2}_{g,k+2l}.$$

As in the spinless case, we explain the notation throughout the steps below.

Step 1 First, $\overline{\mathcal{M}}'^{1/2}_{g,k+2l}$ is the suborbifold of $\overline{\mathcal{M}}^{1/2}_{g,k+2l}$, the moduli of stable marked 2–spin curves, given by the condition that all the markings have twist 0. Inside this space, $\overline{\mathbb{R}\mathcal{M}}^{1/2}_{g,k+2l}$ is the fixed locus of the involution defined by

$$(C; w_1, \dots, w_{k+2l}, \mathbb{S}) \mapsto (\overline{C}; w_1, \dots, w_k, w_{k+l+1}, \dots, w_{k+2l}, w_{k+1}, \dots, w_{k+l}, \overline{\mathbb{S}}),$$

where $\overline{C}$ and $\overline{\mathbb{S}}$ are the same as C and $\mathbb{S}$ but with the conjugate complex structure. Here k is required to satisfy $2 \nmid g + k$. As the fixed locus of an antiholomorphic involution,

$\overline{\mathbb{R}\mathcal{M}}^{1/2}_{g,k+2l}$ is a smooth compact real orbifold. It parametrizes isomorphism types of marked spin curves with an involution $\widetilde{\varrho}$ covering the conjugation ϱ on C, and 0 twists.

Step 2 The next step is to cut $\overline{\mathbb{R}\mathcal{M}}^{1/2}_{g,k+2l}$ along the real simple normal crossings divisor consisting of curves with at least one real node, via the real hyperplane blowup of [41]. As in the spinless case, this yields an orbifold with corners $\widetilde{\mathbb{R}\mathcal{M}}_{g,k,l}$.

Step 3 Consider the subset of $\widetilde{\mathbb{R}\mathcal{M}}^{1/2}_{g,k+2l}$ whose generic point is a smooth marked real spin curve with nonempty real locus. Then $\widetilde{\mathcal{M}}_{g,k,l}$ is the disconnected 2-to-1 cover of this subset given, as in the spinless case, by the choice of a distinguished half Σ, a connected component of $C \setminus C^{\varrho}$. Note that $C = D(\Sigma)$.

Step 4 Inside $\widetilde{\mathcal{M}}_{g,k,l}$, we denote by $\widehat{\mathcal{M}}_{g,k,l}$ the union of connected components such that the marked points $w_{k+1}, \dots, w_{k+l}$ lie in the distinguished half, and the spin structure is compatible. The generic point in this orbifold has isotropy $\mathbb{Z}_2$, coming, at the level of objects, from scaling the fibers of $\mathbb{S}$ by -1.

Step 5 Finally, $\overline{\mathcal{M}}_{g,k,l}$ is the degree 2 cover of $\widehat{\mathcal{M}}_{g,k,l}$ given by a choice of grading. The choice of the grading cancels the global $\mathbb{Z}_2$ isotropy, since the -1 map is no longer an automorphism, as it does not preserve the grading. As a cover, $\overline{\mathcal{M}}_{g,k,l}$ is also endowed by an orbifold with corners structure.

For any $\Gamma \in \mathcal{G}_{g,k,l}$, $\overline{\mathcal{M}}_{\Gamma}$ is a suborbifold with corners which is the closure of $\mathcal{M}_{\Gamma}$. The map $\mathrm{For}_{\mathrm{spin}}$ is an orbifold branched cover. A graded surface with b boundary nodes and contracted boundaries belongs to a corner of the moduli space $\overline{\mathcal{M}}_{g,k,l}$ of codimension b. Thus $\partial\overline{\mathcal{M}}_{g,k,l}$ consists of graded stable surfaces with at least one boundary node or contracted boundary. For details see [35]. We should note that the same argument applies for the more general setting of the moduli space of twisted spin surfaces with a lifting. These more general moduli spaces are also smooth orbifolds with corners, but in general they are not orientable.

Remark 2.38 By Proposition 2.32, the degree of the map $\mathrm{For}_{\mathrm{spin}}$ is 2^g. The automorphism group of the underlying surface acts on the set of spin structures. When the surface is smooth this group is generically trivial, but when it is not, it may happen that the fiber of $\mathrm{For}_{\mathrm{spin}}$ is of cardinality smaller than 2^g. Still, even in this case its weighted cardinality, which takes into account the isotropies, is 2^g, so that the orbifold degree in the smooth case is constant. When the topology becomes nodal the number of graded spin structures on a given underlying surface may change. But still, for any graded dual graph Γ the degree of $\mathrm{For}_{\mathrm{spin}}$ restricted to $\mathcal{M}_{\Gamma}$ is generically constant, and when isotropy groups are taken into account, it is always constant. This constant is a power of 2 which can be calculated from the graph structure of Γ using, for example, Proposition 2.19 and the first paragraph in its proof, which relate spin structures on a stable surface and twisted spin structures on its normalization.

The *universal curve* $\overline{\mathcal{C}}_{g,k,l} \to \overline{\mathcal{M}}_{g,k,l}$ is the space whose fiber over $[\Sigma] \in \overline{\mathcal{M}}_{g,k,l}$ is Σ. Its topology can be defined as in the closed case.

The following simple lemma is useful for understanding the geometry of $\overline{\mathcal{M}}_{g,k,l}$; see [34; 35] for details.

Lemma 2.39 (a) *The maps q and Q are isotopy invariants, in the sense that if $(\Sigma_s)_{0\le s\le 1}$ is a path in $\overline{\mathcal{M}}_{g,k,l}$, and $(\gamma_{t,s})_{0\le s,t\le 1}$ is a continuous family of simple paths $\gamma_{\cdot,s} \subseteq \Sigma_s \hookrightarrow \overline{\mathcal{C}}_{g,k,l}$ which miss the special points and which are either all bridges or all closed. If they are all bridges then $Q(\gamma_{\cdot,s})$ is fixed for any continuous choice of orientations on $\gamma_{\cdot,s}$, and if they are all closed, then $q(\gamma_{\cdot,s})$ is fixed.*

(b) *Suppose now that $(\Sigma_s)_{0\le s\le 1}$ is a path in $\overline{\mathcal{M}}_{g,k,l}$ and $(\gamma_{t,s})_{0\le s,t\le 1}$ is a continuous family of paths $\gamma_{\cdot,s} \subseteq \Sigma_s \hookrightarrow \overline{\mathcal{C}}_{g,k,l}$ which for $s < 1$ are simple and miss the special points and are either all bridges or all closed. Assume $\gamma_{\cdot,1}$ is a constant path mapped to a node or a contracted boundary. If $\gamma_{\cdot,s}$ are all closed, then the node is internal or a contracted boundary and for any $s < 1$, its twist is $q(\gamma_{\cdot,s})$. If $\gamma_{\cdot,s}$ are all open, then the node is a boundary node. In this case, the illegal sides of the bridges degenerate to the illegal half-node, in the sense of Definition 2.13.*

In particular, by Proposition 2.31, exactly one of the half-nodes of each boundary node is legal.

(c) *Two graded spin structures on Σ without a Ramond node which give rise to the same pair (q, Q) are isomorphic.*

Remark 2.40 A classification of all pairs (q, Q) is given in [34].

Notation 2.41 We denote by $\widetilde{\mathrm{For}}_{\mathrm{spin}}$ the canonical map

$$\widetilde{\mathrm{For}}_{\mathrm{spin}} \colon \overline{\mathcal{M}}_\Gamma \to \overline{\mathcal{M}}^{\mathbb{R}}_{\mathrm{for}_{\mathrm{spin}}(\Gamma)}$$

defined by forgetting the twisted spin structure and the lifting. Write $\mathrm{For}_{\mathrm{spin}}$ for the restriction to graded moduli. The definitions of $\widetilde{\mathrm{For}}_{\mathrm{spin}}$, $\mathrm{For}_{\mathrm{spin}}$ make sense also when Γ is closed (and then the lifting is trivial).

We end this subsection with a brief illustration of the phenomenon underlying the branched cover property of the map $\mathrm{For}_{\mathrm{spin}}$. The branching phenomenon occurs along strata which parametrize surfaces with *internal* nodes, and therefore happens, from the same geometric reasoning, also in the setting of the closed 2–spin intersection theory. We shall explain it in this setting, for simplicity of notation.

Let Σ_0 be a curve with a single nonseparating node, and let Σ_1 be its smoothing, so that Σ_0 is obtained from Σ_1 by pinching at some simple smooth closed path γ. Let $(\Sigma_t)_{t\in[0,1]}$ be a path in the moduli of curves, interpolating between Σ_1 and Σ_0. This path induces an identification of $H_1(\Sigma_t,\mathbb{Z}_2)$ for $t>0$, which in the limit $t\to 0$ corresponds to the surjection obtained by taking the quotient $[\gamma_t]=0$, where $[\gamma_t]$ is the generator of $H_1(\Sigma_t,\mathbb{Z}_2)$ which corresponds to $[\gamma]\in H_1(\Sigma_1,\mathbb{Z}_2)$ under this isomorphism. Let α be any element of $H^1(\Sigma_1,\mathbb{Z}_2)$ satisfying $\langle\alpha,\gamma\rangle=1$. Denote by α' the element in $H^1(\Sigma_0,\mathbb{Z}_2)$ which corresponds to α after the pinching, via the aforementioned surjection. Let B_1 be an ordered basis of $H_1(\Sigma_1,\mathbb{Z}_2)$ whose first two elements are $[\gamma]$ and $[\alpha]$, and whose remaining basis elements do not intersect $[\gamma]$. For $t>0$, define B_t as the image of B_1 under the isomorphism, and extend to $t=0$ via the mentioned surjection. Now choose any spin structure of Σ_0 which gives all markings twist 0 and makes the node NS. Recall that spin structures on smooth curves are determined by the map q of Definition 2.26, using the rule of Theorem 2.27, and any map which satisfies this rule gives rise to such a spin structure. Recall also that spin structures on Σ_0 which give all markings twist 0 and make the node NS are in bijection with spin structures on the normalization of Σ_0 giving all of its special points twist 0. Assign a number $q(\beta)$ to any element $\beta\in B_1\setminus\{\gamma\}$, and put $q(\gamma)=0$. Recall Lemma 2.39. The identifications between the different B_t with $t>0$ define a spin structure $\mathbb{S}_t$ on Σ_t for any $t>0$. It extends to a spin structure on Σ_0 with an NS node. We can also define spin structures $\mathbb{S}'_t$ for $t>0$, whose restrictions to B_t are the same except for the elements which correspond to α, on which they are opposite. Both $(\Sigma_t,\mathbb{S}_t)_{t\in[0,1]}$ and $(\Sigma_t,\mathbb{S}'_t)_{t\in[0,1]}$ are paths in the moduli of 2–spin curves which have the same limit point $(\Sigma_0,\mathbb{S}_0)=(\Sigma_0,\mathbb{S}'_0)$, and which cover the same path $(\Sigma_t)_{t\in[0,1]}$ in the moduli of curves. The existence of the paths is due to the fact that $q(\alpha')$ is undefined, and this data loss is the reason for the appearance of the branched cover phenomenon.

In the case of a separating NS node, this argument no longer works, however in this case the automorphism group of the spin structure becomes larger: scaling the fibers of the spin bundle by -1 on each one of the two components is an automorphism. This growth of the automorphism group implies that the orbifold degree of the restriction of $\mathrm{For}_{\mathrm{spin}}$ to such strata decreases.

2.4 The line bundles $\mathbb{L}_i$

Definition 2.42 Let Γ be a stable graph with an internal tail marked $i \neq 0$. The line bundle $\mathbb{L}_i \to \overline{\mathcal{M}}_\Gamma^{\mathbb{R}}$ is the line bundle whose fiber at $(\Sigma, \{x_j\}_{j\in\mathcal{B}}, \{z_j\}_{j\in\mathcal{I}}) \in \overline{\mathcal{M}}_\Gamma^{\mathbb{R}}$ is $T^*_{z_i}\Sigma$. This bundle can also be defined by pulling back the corresponding relative cotangent line over the closed moduli space, via the doubling map.

Let Γ be a spin graph with a lifting and an internal tail marked $i \neq 0$. The line bundle $\mathbb{L}_i \to \overline{\mathcal{M}}_\Gamma$ is the line bundle whose fiber at $(\Sigma, \{x_j\}_{j\in\mathcal{B}}, \{z_j\}_{j\in\mathcal{I}}) \in \overline{\mathcal{M}}_\Gamma$ is $T^*_{z_i}\Sigma$. Equivalently, this bundle can be defined as the pullback of $\mathbb{L}_i \to \overline{\mathcal{M}}^{\mathbb{R}}_{\widetilde{\mathrm{for}}_{\mathrm{spin}}(\Gamma)}$ by the map $\widetilde{\mathrm{For}}_{\mathrm{spin}}$.

2.5 Boundary conditions and intersection numbers

We begin with a simple observation.

Observation 2.43 *Let $(\Sigma, \mathbb{S}, s)$ be a smooth marked surface with a spin structure and a lifting, Σ' the marked surface obtained by forgetting points $\{x_b\}_{b\in\mathcal{B}'}$, where $\mathcal{B}'$ is a subset of illegal boundary marked points. Then $\mathbb{S}$ is canonically a (twisted) spin structure for Σ', and s canonically extends to a lifting on Σ'. In particular, a marked point is legal for $(\Sigma', \mathbb{S}, s)$ if and only if it is legal for $(\Sigma, \mathbb{S}, s)$.*

Definition 2.44 Consider $\Gamma \in \mathcal{G}_{g,k,l}$ and $i \in [l]$, and let $v = i/\sigma_0$ be the vertex of Γ which contains the tail marked i. Define a graph $v_i^*(\Gamma)$ as follows; it will be called the *abstract vertex of i in* Γ, or just the *abstract vertex* for short.

(a) $V(v_i^*(\Gamma)) = \{*\}$, a singleton. It is open if and only if v is.

(b) $T^I(v_i^*(\Gamma)) = (\sigma_0^I)^{-1}(v)$. Any internal tail of $v_i^*(\Gamma)$ which corresponds to a tail marked by $j \in [l]$ is marked j, otherwise it is marked 0. The twist of any tail of $v_i^*(\Gamma)$ is the same as the twist of the corresponding half-edge of v. Also, $H^{\mathrm{CB}} = \varnothing$.

(c) $T^B(v_i^*(\Gamma)) = \{h \in (\sigma_0^B)^{-1}(v) \mid \mathrm{alt}(h) = 1\}$, and all of these boundary tails are marked 0.

(d) $g(v_i^*(\Gamma)) = g(v)$ and $E(v_i^*(\Gamma)) = \varnothing$.

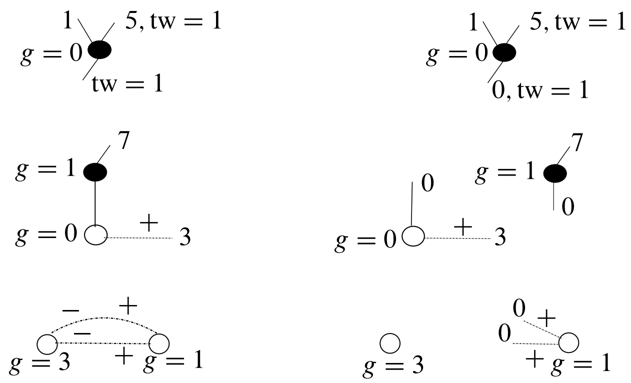


Figure 8: In this figure, to the right side of each dual graded spin graph its corresponding abstract vertices are shown. Again, half-edges h with $\mathrm{alt}(h) = 1$ are decorated with the symbol $+$.

Let $\mathrm{for}_{\mathrm{illegal}} \colon \mathcal{G} \to \mathcal{G}$ be the map which forgets all tails $t \in T^{\boldsymbol{B}}$ with $\mathrm{alt}(t) = 0$. As a consequence of Observation 2.43, it induces a map at the level of moduli spaces, which will be denoted by $\mathrm{For}_{\mathrm{illegal}}$.

Write $\Phi_{\Gamma,i} = \mathrm{For}_{\mathrm{illegal}} \circ v_i \colon \mathcal{M}_\Gamma \to \mathcal{M}_{v_i^*(\Gamma)}$. This map extends to a map $\overline{\mathcal{M}}_\Gamma \to \overline{\mathcal{M}}_{v_i^*(\Gamma)}$, and we also denote the extension by $\Phi_{\Gamma,i}$.

At the level of surfaces, $\Phi_{\Gamma,i}(\Sigma)$ for $\Sigma \in \overline{\mathcal{M}}_\Gamma$ is the graded smooth surface obtained from Σ by normalizing the nodes which correspond to the edges of Γ, taking the component of z_i, forgetting all illegal half-nodes which were formed, renaming all remaining special points by 0, and forgetting the lifting at preimages of contracted boundaries; see Figure 9.

Observation 2.45 *For Γ as above, the two orbifold line bundles $\mathbb{L}_i \to \mathcal{M}_\Gamma$ and $\Phi^*_{\Gamma,i}(\mathbb{L}_i \to \mathcal{M}_{v_i^*(\Gamma)})$ are canonically isomorphic.*

For a proof, see [31]; it is proven there for the $g = 0$ case, but the same argument works in general.

In order to define the open intersection numbers we need to define special canonical multisections, following [31; 35]. We first recall what multisections are, and refer the reader to [7, Appendix A] for more details and references.

Definition 2.46 Let $E \to M$ be an orbibundle over an orbifold with corners, and identify E with its total space. A *multisection* is a function $\kappa \colon E \to \mathbb{Q}_{\geq 0}$ which satisfies the following properties. For any $p \in M$, let $(F \to U)/G$ be a local model for $E \to M$

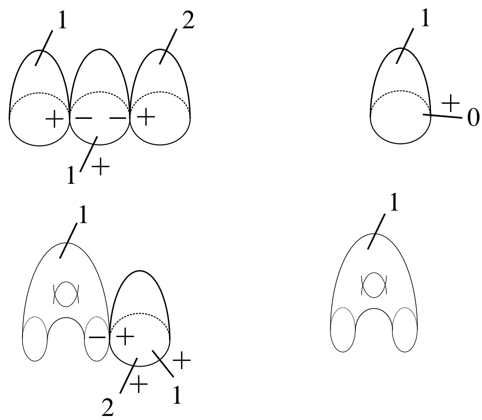


Figure 9: In this figure, the graded surface on the right of each row is the image (as a moduli point) of $\Phi_{\Gamma,1}(\Sigma)$, where Σ is the corresponding surface to the left of the same row, and Γ is the dual graph which corresponds to Σ.

in a neighborhood of p, where $U \simeq \mathbb{R}^m \times \mathbb{R}^{n-m}_{\geq 0}$, p is identified with 0, $F \simeq U \times \mathbb{R}^h$, the map $\pi: F \to U$ is the projection, and G is a finite group acting linearly on the pair, commuting with π. Denote by $\widehat{\kappa}$ the pullback of κ to a G–invariant function on F. Then:

(a) For all $y \in U$,
$$\sum_{v \in \pi^{-1}(y)} \widehat{\kappa}(v) = 1.$$

(b) We can find sections $s_1, \dots, s_N: U \to F$, perhaps after replacing U with a smaller neighborhood of 0, and nonnegative rational numbers $\mu_1, \dots, \mu_N$, such that for all $y \in U$ and $v \in \pi^{-1}(y)$,
$$\widehat{\kappa}(v) = \sum_{i \mid s_i(y) = v} \mu_i.$$

The sections $s_1, \dots, s_N$ are called *local branches* and the numbers $\mu_1, \dots, \mu_N$ are their *weights*. The locus where $\kappa \neq 0$, which is locally the union of its local branches, is called the *support* of the multisection. The elements in the support of κ which lie in the fiber E_p of E over p form the set of *values* of the multisection at p.

Although the support does not, in general, capture all the information of the multisection, we usually refer to the multisection κ by its support s, and write $s(x)$ for the values of the multisection at x. If $N = 1$ for all $p \in M$, then the multisection is just a usual section. The multisection is smooth (piecewise smooth) if all its local branches are smooth (piecewise smooth). Many of the natural operations and properties of sections of vector bundles generalize to multisections of orbibundles in a natural way. These include addition of multisections, multiplication by functions $f : M \to \mathbb{R}$, and most transversality statements. We say that the multisection is nowhere vanishing if none of its branches vanishes, or equivalently $\kappa(x, 0) = 0$ for all $x \in M$. The multisection is transverse to zero if all its branches are transverse to the zero section, and it has isolated zeroes, if all its local branches have isolated zeroes. A point x is a zero of the multisection if $\kappa(x, 0) \neq 0$, that is, at least one of the local branches at x vanishes at x. The zero locus of a multisection is the set of its zeroes.

Definition 2.47 Suppose $A \subseteq \mathcal{G}_{g,k,l}$ is a collection of graphs with at least one boundary edge. A piecewise smooth multisection s of $\mathbb{L}_i \to \bigcup_{\Gamma\in A} \overline{\mathcal{M}}_\Gamma$ is called *special canonical* on $\bigcup_{\Gamma\in A} \overline{\mathcal{M}}_\Gamma$ if, for all $\Lambda \in \partial^!\Gamma$,

$$s|_{\mathcal{M}_\Lambda} = \Phi^*_{\Lambda,i} s^{v_i^*(\Lambda)}$$

for some piecewise smooth multisection $s^{v_i^*(\Lambda)}$ of $\mathbb{L}_i \to \mathcal{M}_{v_i^*(\Lambda)}$.

In the case that $A \subseteq \mathcal{G}_{g,k,l}$ is the collection of all graphs with at least one boundary edge, we say that s as above is *special canonical*.

A multisection $s = \bigoplus_{i\in[l], j\in[a_i]} s_{ij}$ of $\bigoplus_i \mathbb{L}_i^{\oplus a_j}$ is special canonical if each component s_{ij} is special canonical.

Intuitively, being special canonical means that the multisection depends only on the irreducible component of z_i in the normalization, after forgetting the locations of the illegal boundary half-nodes and the liftings at contracted boundaries.

Still following [7, Appendix A], let $p \in M$ be an internal point, and let s be a multisection with isolated zeroes. We assume that E and M are oriented and $\mathrm{rk}(E) = \dim(M)$. Take a local model $(F \to U)/G$ for the neighborhood of p as in Definition 2.46. Choose a metric on U, a metric on the fibers $\mathbb{R}^h$, and let $\pi' : F \to \mathbb{R}^h$ be the projection on the $\mathbb{R}^h$ component. Let B be a small ball around 0 (which is identified with p) which contains no zero of s except possibly 0. Denote by S the unit sphere in $\mathbb{R}^h$. We use the orientations of M and E to endow S and ∂B with the induced orientations as the boundaries of oriented balls. We define $\deg_p(s_i)$, the local degree of s_i at p, as the degree of the map $t : \partial B \to S$, where

$$t(x) = \frac{\pi'(s_i(x))}{|\pi'(s_i(x))|}.$$

This definition is independent of choices. The *weight* of p in the zero locus of s is defined as

$$\epsilon_p = \frac{1}{|G|} \sum_{i=1}^{N} \mu_i \deg_p(s_i). \tag{17}$$

If s has a finite zero locus $\{p_1, \dots, p_t\}$, then the *weighted signed zero count* of s is $\sum_{i=1}^{t} \epsilon_{p_i}(s) \in \mathbb{Q}$.

Let s be a piecewise smooth multisection of $E \to \partial M$, where $E \to M$ is an oriented orbibundle over a compact oriented orbifold with corners. Suppose s vanishes nowhere. For *any* piecewise smooth multisection $\tilde{s}$ extending s to the interior of M with isolated zeroes, the weighted signed zero count of $\tilde{s}$ is the same. This follows from standard cobordism arguments — see for example [16, Section 3] for the case $\partial M = \varnothing$; the addition of boundary does not complicate the argument[2] — and it is also a consequence of Proposition 3.3, whose proof is sketched below. We denote this number by $\int_M e(E, s)$ and call it the *integral of the relative Euler class* of E relative to s.

Remark 2.48 The *relative Euler class* $e(E, s) \in H^n(M, \partial M, \mathbb{Q})$, where $E \to M$ is an oriented orbibundle over a compact oriented orbifold with corners with $\mathrm{rk}(E) = \dim(M) = n$, is defined whenever s is a nowhere-vanishing boundary condition for $E \to M$. Integrating, or capping with the fundamental class, gives by Poincaré–Lefschetz duality an element of $H_0(M, \mathbb{Q}) \simeq \mathbb{Q}$. This element is precisely what we defined as the integral of the relative Euler class. For our needs the definition of the relative Euler class itself is not required. See the appendix in [7] for more details and references.

The integral relative Euler class can be defined for orbifold sphere bundles rather than orbifold vector bundles, for example by using an embedding of the sphere bundle into the vector bundle using a choice of a metric for the vector bundle, and inducing the boundary conditions by this embedding. The resulting integrals are the same when working with a vector bundle E or with its associated sphere bundle. We shall use these two notions interchangeably throughout the paper.

Observation 2.49 *Suppose that $E \to M$ is an oriented orbibundle over a compact oriented piecewise smooth orbifold with corners with $\mathrm{rk}(E) = \dim(M) = n$, and that*

[2] In [16] the definition of multisections is slightly different, as a section to the symmetric product of the orbifold vector bundle. However, a multisection in our terminology induces in a natural way a multisection in the terminology of that paper, and the definitions of the zero counts agree.

s is a nowhere-vanishing multisection of $E \to \partial M$. Let $f : N \to M$ be a surjection between compact oriented piecewise smooth orbifolds with corners of dimension n, which maps ∂N onto ∂M. Suppose that f is generically of degree one, meaning that outside of a subspace $K \subset M$ which is a union of finitely many compact suborbifolds of M of real codimension one, f is injective. Then

$$\int_N e(f^*E, f^*s) = \int_N e(E, s).$$

Indeed, standard transversality arguments show that a generic piecewise extension of s to M will have no zeroes in K. Using the pullback to N of such a generic extension proves the claim.

The following theorem has appeared in [31] in the genus 0 case, and will appear in [35] for all genera.

Theorem 2.50 *Suppose $a_1, \dots, a_l \geq 0$ are integers which sum to $\frac{1}{2}(k + 2l + 3g - 3)$. Then one can choose multisections $\{s_{ij}\}_{i \in [l], j \in [a_i]}$ such that*

(a) *For all i and j, s_{ij} is a special canonical multisection of $\mathbb{L}_i \to \partial\overline{\mathcal{M}}_{g,k,l}$.*
(b) *The multisection $s = \bigoplus_{i,j} s_{ij}$ vanishes nowhere.*

Moreover, for any two choices $\{s_{ij}\}$ and $\{s'_{ij}\}$ which satisfy the above requirements, we have

$$\int_{\overline{\mathcal{M}}_{g,k,l}} e\Big(\bigoplus_i \mathbb{L}_i^{\oplus a_j}, s\Big) = \int_{\overline{\mathcal{M}}_{g,k,l}} e\Big(\bigoplus_i \mathbb{L}_i^{\oplus a_j}, s'\Big),$$

where $s' = \bigoplus_{i,j} s'_{ij}$.

For completeness, and since [35] is yet to appear, we will shortly review the proof of first claim in the theorem. We will not review the "Moreover" part, since it will be a consequence of our main theorem, Theorem 1.5, which calculates the integral of the relative Euler class, and obtains an answer which does not involve the special canonical multisection, without relying on the assumption that the integral is independent of the multisection.

The proof that nowhere-vanishing special canonical boundary conditions exist has two steps. The first step shows that for any boundary point $p \in \partial\overline{\mathcal{M}}_{g,k,l}$ there exists a special canonical multisection none of whose branches vanishes at p. This step is the heart of the argument; it is similar but not identical to [31, Proposition 3.49(a)] and we will review it in the next paragraph. The second step uses the multisections constructed in the first step to construct nowhere-vanishing boundary conditions: Using the first step and compactness, one can find finitely many canonical multisections $s_1, \dots, s_N$ of $E = \bigoplus_{i \in [l], j \in [a_i]} \mathbb{L}_i^{\oplus a_i}$ such that for any boundary point $p \in \partial \overline{\mathcal{M}}_{g,k,l}$ and any choice of local branches s_i' of s_i at p, the vectors $(s_i')_p$ for $i \in [N]$ span the fiber E_N. Then, by a standard transversality argument, a generic linear combination of $s_1, \dots, s_N$ will be a nowhere-vanishing canonical multisection. By generic we mean that the subset of linear combinations of $s_1, \dots, s_N$ with this property is residual in the set of all possible linear combinations. The proof of this step is identical to [31, Lemma 3.53(a)], and we refer the interested reader there.

We turn to explain the first step. Fix $p \in \partial \overline{\mathcal{M}}_{g,k,l}$ and $i \in [l]$. Suppose p belongs to the stratum $\mathcal{M}_\Gamma$ for some graded spin dual graph Γ corresponding to the graded surface Σ. Let $u \in (\mathbb{L}_i)_p$ be an arbitrary nonzero vector. Finally, let $[\Sigma']$ be the image of $p = [\Sigma]$ under the map $\Phi_{\Gamma,i}$, and write $G = \mathrm{Aut}(\Sigma)$. The action of G lifts to an action on the cotangent of the i^{th} marking, that is, on $(\mathbb{L}_i)_{[\Sigma']}$, the fiber of $\mathbb{L}_i$ at $[\Sigma']$. By Observation 2.45, the fibers of $\mathbb{L}_i$ at $[\Sigma']$ and $[\Sigma]$ are isomorphic, canonically up to the action of G on $(\mathbb{L}_i)_{[\Sigma']}$. Thus, the G–action lifts also to $(\mathbb{L}_i)_{[\Sigma]}$. Write

$$\boldsymbol{u} = \{u_1, \dots, u_m\} = \{g \cdot u \mid g \in G\}.$$

We will construct a special canonical multisection of $\mathbb{L}_i$ whose branches at p have values $\boldsymbol{u}$, with equal weights. Set

$$V_{g,k,l} = \{v_i^*(\Lambda) \mid \Lambda \in \partial^! \Gamma_{g,k,l}\},$$

ie $V_{g,k,l}$ is the collection of abstract vertices $v_i^*(\Lambda)$ for any graded spin graph Λ that corresponds to a stratum of $\overline{\mathcal{M}}_{g,k,l}$. We will construct for any $v \in V_{g,k,l}$ a special canonical multisection s^v for $\mathbb{L}_i \to \overline{\mathcal{M}}_v$. These multisections are required to be compatible in the following sense. Let $v \in V_{g,k,l}$, and let $\Lambda \in \partial v$ be a graph which corresponds to a boundary stratum of $\overline{\mathcal{M}}_v$. Let $v' = v_i^*(\Lambda)$. It is easy to see that $v' \in V_{g,k,l}$. We require, for all such v and Λ, that

$$s^v|_{\mathcal{M}_\Lambda} = \Phi_{\Lambda,i}^* s^{v'}. \tag{18}$$

These constraints for different Λ are compatible. See the explanation at the beginning of the proof of [31, Proposition 3.49], which extends to our setting. This construction will provide, in particular, a construction of a special canonical multisection for $v_i^*(\Gamma_{g,k,l})$, which is the same graded dual graph as $\Gamma_{g,k,l}$ except that the boundary tails are marked 0.

The pullback of this section by the canonical map

$$\overline{\mathcal{M}}_{\Gamma_{g,k,l}} \to \overline{\mathcal{M}}_{v_i^*(\Gamma_{g,k,l})}$$

which changes the boundary markings to 0 will be the required multisection.

Write $v^* = v_i^*(\Gamma)$ and $a = \dim(\mathcal{M}_{v^*})$, where Γ is the dual graph which corresponds to Σ. The construction of multisections s^v for $v \in V_{g,k,l}$ will be by induction on $d = \dim \mathcal{M}_v$. The basis is $d = -1$, which holds trivially since there are no such vertices. Suppose we have constructed multisections with the required properties for all v' with $\dim \mathcal{M}_{v'} < d$. Consider $v \in V_{g,k,l}$ with $d = \dim \mathcal{M}_v$. Note that v need not be an open vertex, and may even have internal tails with $\mathrm{tw} = 1$. Write $\Upsilon = \coprod_{\Lambda \in \partial v} \mathcal{M}_\Lambda$. Define first $s^v|_\Upsilon$ according to (18), where the right-hand side of the compatibility equations is already defined by induction. We now extend s^v to the whole moduli space $\overline{\mathcal{M}}_v$. Here we separate into cases. If $v \neq v^*$, we extend arbitrarily. If $v = v^*$ we extend arbitrarily, but under the requirement that $s^{v^*}_{[\Sigma']} = \boldsymbol{u}$, meaning that each u_i appears in some branches of s^{v^*}, and with the same total weight. This can be done for example in the following way. Let $\rho\colon \overline{\mathcal{M}}_{v^*} \to [0,1]$ be a smooth function which is 1 near $[\Sigma']$ and 0 near Υ. Let s' be an arbitrary extension of the already defined $s^{v^*}|_\Upsilon$ to $\overline{\mathcal{M}}_{v^*}$, and s'' an arbitrary multisection of $\mathbb{L}_i \to \overline{\mathcal{M}}_{v^*}$ which has the required values $[\Sigma']$. Then one can take

$$s^{v^*} = \rho s'' + (1-\rho)s'.$$

The induction follows,[3] and thus also the proof.

For the benefit of the reader we now explain the difference between this proof and the proof of [31, Proposition 3.49(a)], and the intuitive reason for why canonical boundary conditions should give rise to well-defined intersection numbers. In [31] there were no contracted boundaries and all boundary nodes were separating. In this case the definition of canonical boundary conditions can be given without spin structure, by using only parity considerations: for each node, precisely one half-node is forgotten, and the forgotten half-nodes are chosen in the unique way which leaves on each connected component of genus s of the normalized surface a total number of unforgotten special boundary points whose parity is $s+1 \pmod 2$. This numerical reasoning cannot work

[3]In the proof of the corresponding claim in [31], the multisections were also required to satisfy some invariance under symmetry groups. In our case, since we work with orbifolds and orbibundles, this invariance is part of the definition of being a multisection; see the appendix of [7]. In [31], the orbifoldness was implicit, and was a result of forgetting the boundary markings. In higher genus, even the moduli with injective markings is an orbifold.

when there are nonseparating nodes. However, as it turns out, this parity notion neatly generalizes to the notion of a graded spin structure, and the forgotten half-nodes are precisely the illegal ones. The importance of this scheme is that it forces the boundary conditions to be pulled back from a real codimension-two space rather than from a codimension-one space (the codimension is with respect to the dimension of the whole moduli space).

This idea cannot work in the case of moduli strata which parametrize surfaces with a contracted boundary component. However, for such surfaces, for any contracted boundary component there are two possible choices of liftings. Moreover, by the "Moreover" part of Theorem 2.53 below, the boundary strata of the moduli which correspond to the different choices of liftings come with opposite orientations. Since, in the definition of the base, the lifting in such points is forgotten, the boundary conditions should be the same for these two boundary strata.[4]

These two properties are strong enough to guarantee that the integrals are well defined: The dimension reduction, together with a standard transversality argument, enables one to construct a homotopy between any two choices of canonical boundary conditions s and s' which does not vanish on boundary strata which correspond to surfaces with a boundary node. It may vanish on boundary strata which correspond to surfaces with contracted boundaries, but these vanishings cancel in pairs, which differ in the liftings of these contracted boundaries. This homotopy argument thus shows that s and s' determine the same integral. In the course of the proof of Theorem 1.5 this independence will become manifest.

Based on Theorem 2.50 we can now define open intersection numbers.

Definition 2.51 With the notation of Theorem 2.50, define the open intersection number

$$\langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle_g := 2^{-\frac{1}{2}(g+k-1)} \int_{\overline{\mathcal{M}}_{g,k,l}} e\Big(\bigoplus_i \mathbb{L}_i^{\oplus a_j}, s \Big),$$

where s is a nowhere-vanishing special canonical multisection.

[4] Essentially this discussion says that such codimension-one boundaries can be glued, and that the integrals can be calculated with respect to the glued moduli space. In an earlier version of this manuscript we chose this path, but we believe that this gluing is less elegant than the equivalent choice of unglued boundaries we make here. The cost of this choice is that there are now additional boundary conditions to impose and to analyze.

The power of 2 is a normalization factor chosen in [31], which makes some initial conditions nicer but has no geometric or algebraic importance.

Since we define the intersection numbers to be 0 unless the numerical condition of Theorem 2.50 holds, the genus is determined from knowing k, l and $a_1, \dots, a_l$, and for this reason we will usually omit it from the notation and simply write $\langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle$ for $\langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle_g$.

2.6 The orientation of $\overline{\mathcal{M}}_{g,k,l}$

As mentioned above, the spaces $\overline{\mathcal{M}}_{g,k,l}$ were proved to be orientable, and moreover were given canonical orientations. In order to state properties of these orientations that will be required for later, we need the following definition.

Definition 2.52 Let M be an oriented orbifold with corners. Then ∂M is also orientable. The *induced orientation* on ∂M is defined by the exact sequence

$$0 \to N \to TM|_{\partial M} \to T\partial M \to 0,$$

where N, the dimension-one normal bundle of ∂M in M, is oriented by taking the outward normal as a positive direction and the orientation on TM as the given one.

For the benefit of the reader, we recall the construction of the induced orientation also in terms of local coordinates. Let p be a boundary point which is not a corner. A local neighborhood of p is diffeomorphic to $(\mathbb{R}_{\geq 0} \times \mathbb{R}^{n-1})/G$ for some finite group G which acts on $\mathbb{R}^n$, and under the diffeomorphism, p is mapped to the origin. By the orientability assumption G acts in an orientation-preserving manner, and we may assume that the orientation induced on $\mathbb{R}^n$ by the diffeomorphism is the standard one. Since p is a boundary point, $\{0\} \times \mathbb{R}^{n-1}$ is preserved by G, and since G acts on $\mathbb{R}_{\geq 0} \times \mathbb{R}^{n-1}$, by definition $\mathbb{R}_{\geq 0} \times \mathbb{R}^{n-1}$ is preserved. Take an oriented frame $(v_1, v_2, \dots, v_n)$ for $\mathbb{R}^n$ which is in the class of the standard orientation, so that v_1 has negative first coordinate and the remaining vectors of the frame have first coordinate 0. Then $(v_2, \dots, v_n)$ is a frame for $\{0\} \times \mathbb{R}^{n-1}$. For $g \in G$, $(gv_1, \dots, gv_n)$ is in the same orientation class as the original frame. Since $\mathbb{R}_{\geq 0} \times \mathbb{R}^{n-1}$ is preserved under G, gv_1 has a negative first coordinate. Since $\{0\} \times \mathbb{R}^{n-1}$ is preserved under G, the first coordinate of each gv_i for $i \geq 2$ is 0, and we obtain that

$$(gv_2, \dots, gv_n) \quad \text{and} \quad (v_2, \dots, v_n)$$

are in the same orientation class. This class is defined as the orientation frame, which defines the local orientation at p. We extend the orientation to the whole boundary by continuity.

The next theorem, proven in [35], describes some useful properties of the canonical orientations of $\overline{\mathcal{M}}_{g,k,l}$, properties that characterize these orientations uniquely.

Theorem 2.53 *There is a unique choice of orientations $\mathfrak{o}_\Gamma$ for any graded graph Γ all of whose connected components contain a single vertex, satisfying the following requirements:*

(a) *The zero-dimensional spaces $\overline{\mathcal{M}}_\Gamma$ for $\Gamma \in \{\Gamma_{0,1,1}, \Gamma_{0,3,0}\}$ are oriented positively.*

(b) *If $\Gamma = \{\Gamma_1, \dots, \Gamma_r\}$, where the Γ_i are the connected components, then*

$$\mathfrak{o}_\Gamma = \boxtimes_{i=1}^{r} \mathfrak{o}_{\Gamma_i}.$$

(c) *Let Γ be a graded stable graph with a single boundary edge e, and put $\Lambda = d_e\Gamma$. Denote by Γ' the graph obtained by detaching that edge into two tails t and t', with $\mathrm{alt}(t) = 1$ and $\mathrm{alt}(t') = 1$, and forgetting the tail t. Note that we have a fibration $\mathcal{M}_\Gamma \to \mathcal{M}_{\Gamma'}$ whose fiber over the graded surface $\Sigma \in \mathcal{M}_{\Gamma'}$ is naturally identified with $\partial\Sigma \setminus \{x_i\}_{i\in B(\Gamma')}$. Then the induced orientation on $\mathcal{M}_\Gamma$ as a codimension-one boundary of $\overline{\mathcal{M}}_\Lambda$ agrees with the orientation on $\mathcal{M}_\Gamma$ induced by the fibration $\mathcal{M}_\Gamma \to \mathcal{M}_{\Gamma'}$, where the base is given the orientation $\mathfrak{o}_{\Gamma'}$ and the fiber over Σ gets the orientation of $\partial\Sigma$.*

Moreover, these orientations have the following additional property. For Γ as above, let C be a connected component of $\overline{\mathcal{M}}_\Gamma$ which parametrizes surfaces with at least one boundary component containing no boundary markings. Let C' be another connected component which parametrizes surfaces that differ from those of C only at the grading in that boundary component, which is opposite. There is a natural map $\Psi\colon C \to C'$ which maps a stable graded marked surface to the same surface but with the opposite grading at this boundary component. Let $C_\times$ and $C'_\times$ be the boundary strata of C and C', respectively, which parametrize surfaces in which this boundary component is contracted, and let $\mathfrak{o}_{C_\times}$ and $\mathfrak{o}_{C'_\times}$ be the respective orientations induced on these subspaces. Then Ψ maps $C_\times$ bijectively onto $C'_\times$, and

$$\mathfrak{o}_{C'_\times} = -\Psi_*\mathfrak{o}_{C_\times}.$$

The difficulty in this theorem lies in the existence and the "Moreover" parts, which will be proven by other means below. Given the existence, the uniqueness follows easily using induction on dimension. In [35] the behavior of the orientations with respect to strata with internal nodes is also explained, but it is not needed here.

3 Sphere bundles and relative Euler class

Given a rank n complex vector bundle $\pi: E \to M$ and a metric on it, one can define the sphere bundle $\pi: S = S(E) = S^{2n-1}(E) \to M$ whose fiber S_p at $p \in M$ is the set of unit-length vectors in E_p, the fiber of E at p, with the induced orientation. Given a sphere bundle $S \to M$, its *linearization* is the space

$$S \times \mathbb{R}_{\geq 0}/\sim,$$

where $(v, r) \sim (v', r')$ if either $r = r' = 0$, or $v = v'$ and $r = r'$. This space can be endowed with a natural linear structure, a metric and a projection to M. When $S = S(E)$, the linearization of S recovers E. The sphere bundle of E can be defined also without referring to a metric, by removing the zero section and taking the quotient by the $\mathbb{R}_+$ action. Different metrics give rise to isomorphic sphere bundles.

Definition 3.1 An *angular form* for E (or for S) is a $(2n-1)$–form Φ on S which satisfies the following two requirements:

(a) $\int_{S_p} \Phi = 1$ for all $p \in M$.

(b) $d\Phi = -\pi^*\Omega$, where Ω is some $2n$–form on M.

The form Ω is a local representative of the top Chern form of $E \to M$, and will be called the *Euler form* which corresponds to Φ. Denote by Φ also the form on $E \setminus M$, where we identify E and its total space, defined by $P^*\Phi$, where $P: E \setminus M \to S(E)$ is the map

$$(p, v) \to \left(p, \frac{v}{|v|}\right) \quad \text{for } p \in M \text{ and } v \in E \setminus M.$$

It is straightforward that:

Observation 3.2 *The form $|v|\Phi$ extends to a form on all the total space of E.*

We will use the following claim.

Proposition 3.3 *Let $E \to M$ be a real oriented rank $2n$ vector bundle on a smooth oriented manifold with boundary M of real dimension $2n$. Let Φ be an angular form, and let Ω be its corresponding Euler form. Given a nowhere-vanishing section $s \in \Gamma(E \to \partial M)$, one can define the integral of the relative Euler class, and it holds that*

$$\int_M e(E,s) = \int_M \Omega + \int_{\partial M} s^*\Phi.$$

Moreover, the statement also holds if $E \to M$ is an orbifold vector bundle over an orbifold with corners and s is a nowhere-vanishing multisection over the boundary.

This claim is well known, in the case of manifolds, and the extension to orbifolds is straightforward. We briefly recall the proof of the claim for manifolds, referring the reader to [4, Chapter 11] for further details, then we explain the changes required for handling the orbifold case. As usual we are interested in the integral of the relative Euler class, rather than the class itself.

We wish to calculate $\int_M e(E,s)$, the weighted number of zeroes of an extension of s to M to a section with isolated zeroes. Let $\bar{s}$ be such an extension, and let $p_1,\dots,p_m$ be its zeroes. By choosing diffeomorphisms from neighborhood of $p_1,\dots,p_m$ to open sets in $\mathbb{R}^n$, for small enough r we can define $M_r = M \setminus \bigcup_{i=1}^m B_r(p_i)$, where $B_r(p)$ is the ball around p, and sections s_r which are the restrictions of $\bar{s}$ to ∂M_r. By taking r to be even smaller we may assume that the balls are disjoint. By Stokes' theorem, $\bar{s}$ being a global section over M_r, and the definition of the angular form, we get

$$\begin{aligned}\int_M \Omega &= \lim_{r\to 0}\int_{M_r}\Omega = \lim_{r\to 0}\int_{M_r}\bar{s}^*\pi^*\Omega \\ &= -\lim_{r\to 0}\int_{M_r}\bar{s}^* d\Phi = -\lim_{r\to 0}\int_{\partial M_r} s_r^*\Phi \\ &= -\int_{\partial M} s^*\Phi + \sum_{i=1}^m \lim_{r\to 0}\int_{\partial B_r(p_i)} s_r^*\Phi.\end{aligned}$$

For each $i = 1,\dots,m$ and small enough r, $\int_{\partial B_r(p_i)} s_r^*\Phi$ is the order of vanishing of $\bar{s}$ at p_i; see [4, Theorem 11.16]. Thus, the right-hand side of the previous equation equals $\int_M e(E,s) - \int_{\partial M} s^*\Phi$, as needed.

The argument works also in the orbifold case. One first shows that Stokes' theorem generalizes to the case of orbifolds with corners and multisections of the vector bundle $\Lambda^\bullet(T^*M)$ instead of sections of this bundle (differential forms). For differential forms over orbifolds with corners this is shown, for example, in [36]. The extension to multisections is proven similarly. Then the integral around p_i becomes, in the local model and notation of Definition 2.46,

$$\sum_{i=1}^{N} \mu_i \int_{\partial B(0)} \bar{s}_i^* \Phi,$$

with $B \subset U$ a small ball around 0, and $\bar{s}_i$ and μ_i the local branches and weights. But this is precisely the weight (17) in the definition of $\int_M e(E,s)$, so again the result follows.

Suppose now that $E = \bigoplus_{i=1}^n L_i$ is the sum of n complex line bundles L_i. Choose a metric for E for which the line bundles L_i are pairwise orthogonal. Write α_i for an angular form for $S_i = S(L_i)$, and ω_i for the corresponding Euler form, ie the curvature of L_i. Define the functions

$$r_i : E \to \mathbb{R}$$

to be the length of the projection of $(p,v) \in E$ to L_i. The sphere bundle can be described as the set of vectors which satisfy $\sum r_i^2 = 1$. For convenience, denote by ω_i and $r_i\alpha_i$ the pullbacks of ω_i and $r_i\alpha_i$ to the total space of E and of $S(E)$, where for the latter form we use Observation 3.2.

As far as we know, the following theorem has not appeared in the literature before.

Theorem 3.4 *The form*

$$\Phi = \sum_{k=0}^{n-1} 2^k k! \sum_{i\in[n]} r_i^2 \alpha_i \wedge \sum_{\substack{I\subseteq[n]\setminus\{i\} \\ |I|=k}} \bigwedge_{j\in I} (r_j\, dr_j \wedge \alpha_j) \wedge \bigwedge_{h\notin I\cup\{i\}} \omega_h \tag{19}$$

is an angular form for E, whose corresponding Euler form is $\bigwedge_{i=1}^n \omega_i$.

Proof We first need to show that the integration on a fiber gives 1. Since the ω_i are pulled back from the base for all i, the only term in Φ that may have a nonzero integral over a fiber is the term

$$\Phi^{\mathrm{top}} = 2^{n-1}(n-1)! \sum_{i\in[n]} r_i^2\alpha_i \wedge \bigwedge_{j\neq i} (r_j\, dr_j \wedge \alpha_j).$$

We wish to show that for an arbitrary $p \in M$, we have $\int_{S(E_p)} \Phi^{\mathrm{top}} = 1$. We first integrate all the α_i terms. By using that α_i is an angular form for L_i the integral of α_i is 1, and we are left with calculating

$$\int_{\sum r_i^2=1} 2^{n-1}(n-1)! \sum_{i\in[n]} r_i^2 \wedge \bigwedge_{j\neq i} r_j\, dr_j.$$

By changing the variables to $t_i = r_i^2$ with $dt_i = 2r_i\, dr_i$, the integral becomes

$$\begin{aligned}(n-1)! \int_{\substack{\sum t_i = 1\\ t_1,\dots,t_n \geq 0}} \sum_{i=1}^{n} t_i \wedge \bigwedge_{j \neq i} dt_j &= n! \int_{\substack{\sum_{i=1}^{n-1} t_i \leq 1\\ t_1,\dots,t_{n-1} \geq 0}} \biggl(1 - \sum_{i=1}^{n-1} t_i\biggr) \wedge \bigwedge_{1 \leq j \leq n-1} dt_j\\ &= n! \int_{\substack{\sum_{i=1}^{n} t_i \leq 1\\ t_1,\dots,t_n \geq 0}} \bigwedge_{1 \leq j \leq n} dt_j,\end{aligned}$$

where in the first equality we have used the symmetric role of the variables t_i and then eliminated t_n, and in the second equality we have used that

$$1 - \sum_{i \leq n-1} t_i = \int_{0 \leq s \leq 1 - \sum_{i \leq n-1} t_i} ds.$$

The left-hand side is just $n!$ times the Euclidean volume of the n–simplex

$$\{t_1 + \cdots + t_n \leq 1 \mid t_1, \dots, t_n \geq 0\}.$$

It is well known that this volume is $1/n!$, and the first property of the angular form follows.

For the second property, we will now show that when calculating $d\Phi$, one gets a telescopic sum which turns out to be equal to $\bigwedge \omega_i$. Write

$$S_{I,i} := 2^k k!\, r_i^2 \alpha_i \wedge \bigwedge_{j \in I} (r_j\, dr_j \wedge \alpha_j) \wedge \bigwedge_{h \notin I \cup \{i\}} \omega_h$$

for the contribution for given I with $i \notin I$, where $k = |I|$. Taking the derivative, as ω_i and $r_i\, dr_i$ are closed, only r_i^2 or α_j may contribute. We obtain

$$dS_{I,i} = d_1 S_{I,i} + d_2 S_{I,i} + \sum_{l \in I} d_{3,l} S_{I,i},$$

where

$$\begin{aligned} d_1 S_{I,i} &:= 2^{k+1} k!\, r_i\, dr_i \wedge \alpha_i \wedge \bigwedge_{j \in I} (r_j\, dr_j \wedge \alpha_j) \wedge \bigwedge_{h \notin I \cup \{i\}} \omega_h,\\ d_2 S_{I,i} &:= -2^k k!\, r_i^2 \omega_i \wedge \bigwedge_{j \in I} (r_j\, dr_j \wedge \alpha_j) \wedge \bigwedge_{h \notin I \cup \{i\}} \omega_h,\\ d_{3,l} S_{I,i} &:= -2^k k!\, r_i^2 \alpha_i \wedge r_l\, dr_l \wedge \omega_l \wedge \bigwedge_{j \in I \setminus \{l\}} (r_j\, dr_j \wedge \alpha_j) \bigwedge_{h \notin I \cup \{i\}} \omega_h \quad \text{for } l \in I. \end{aligned}$$

The third contribution appears only when $I \neq \varnothing$.

Now, fixing I, one has

$$\sum_{i \in I} d_1 S_{I \setminus \{i\}, i} = k 2^k (k-1)! \bigwedge_{j \in I} (r_j\, dr_j \wedge \alpha_j) \wedge \bigwedge_{h \notin I} \omega_h, \tag{20}$$

$$\begin{aligned}
(21)\qquad \sum_{i\notin I} d_2 S_{I,i} &= -\sum_{i\notin I} 2^k k!\, r_i^2 \bigwedge_{j\in I}(r_j\,dr_j\wedge\alpha_j)\wedge\bigwedge_{h\notin I}\omega_h\\
&= -\Big(1-\sum_{i\in I} r_i^2\Big)2^k k! \bigwedge_{j\in I}(r_j\,dr_j\wedge\alpha_j)\wedge\bigwedge_{h\notin I}\omega_h\\
&= -2^k k!\Bigg(\bigwedge_{j\in I}(r_j\,dr_j\wedge\alpha_j)\wedge\bigwedge_{h\notin I}\omega_h\\
&\qquad -\sum_{i\in I} r_i^3\,dr_i\wedge\alpha_i\wedge\bigwedge_{j\in I\setminus\{i\}}(r_j\,dr_j\wedge\alpha_j)\wedge\bigwedge_{h\notin I}\omega_h\Bigg),
\end{aligned}$$

where we have used $\sum r_i^2 = 1$ in the second equality. And, fixing I and $i\notin I$,

$$\begin{aligned}
(22)\quad &\sum_{l\notin I\cup\{i\}} d_{3,l} S_{I\cup\{l\},i}\\
&= -\sum_{l\notin I\cup\{i\}} 2^{k+1}(k+1)!\, r_i^2\alpha_i\wedge r_l\,dr_l\wedge\omega_l\wedge\bigwedge_{j\in I}(r_j\,dr_j\wedge\alpha_j)\wedge\bigwedge_{h\notin I\cup\{i,l\}}\omega_h\\
&= -2^{k+1}(k+1)!\sum_{l\in I\cup\{i\}} r_l\,dr_l\wedge r_i^2\alpha_i\wedge\bigwedge_{j\in I}(r_j\,dr_j\wedge\alpha_j)\wedge\bigwedge_{h\notin I\cup\{i\}}\omega_h\\
&= -2^{k+1}(k+1)!\, r_i^3\,dr_i\wedge\alpha_i\wedge\bigwedge_{j\in I}(r_j\,dr_j\wedge\alpha_j)\wedge\bigwedge_{h\notin I\cup\{i\}}\omega_h,
\end{aligned}$$

where the identity $\sum r_i\,dr_i = 0$ was used for the second equality. The last passage follows from noting that except for the $l=i$ term, for all other $l\in I$ we will get a monomial with two dr_l terms.

Summing equations (20), (21) and (22) over all possibilities for I, and in (22) also for $i\notin I$, we see that:

- (20) vanishes if $I=\varnothing$. For $I\neq\varnothing$, the contribution of (20) cancels with the first term on the right-hand side of (21) for the same I.
- For a given $J\neq\varnothing$, the sum of (22) over all pairs (I,i) with $i\in J$ and $I = I\setminus\{i\}$ cancels with the second term of (21) with $I=J$.
- For $I=\varnothing$, the second term of (21) vanishes.

Thus, the only term which is left uncanceled is $\bigwedge\omega_i$, coming from the first term of (21) with $I=\varnothing$. Hence, as needed,

$$d\Phi = \sum_{I,i} dS_{I,i} = -\bigwedge\omega_i. \qquad\square$$

Remark 3.5 In what follows we will sometimes use forms on $S(E)$ which are defined similarly to Φ, but depend a subset of its arguments. For this reason it will be useful to extend Φ and similar expressions to multilinear functions in the variables r_i, dr_i, α_i and ω_i for $i = 1, \dots, n$, without imposing $\sum r_i^2 = 1$ and $\sum r_i \, dr_i = 0$.

Without these constraints the right-hand side of (21) gets a correction of

$$2^k k! \biggl(1 - \sum_{h \in [n]} r_h^2\biggr) \Bigl(\bigwedge_{j \in I} r_j \, dr_j \wedge \alpha_j \Bigr) \wedge \bigwedge_{h \notin I} \omega_j,$$

while the right-hand side of (22) gets a correction of

$$2^{k+1}(k+1)! \biggl(\sum_{l \in [n]} r_l \, dr_l \biggr) \wedge r_i^2 \alpha_i \wedge \Bigl(\bigwedge_{j \in I} r_j \, dr_j \wedge \alpha_j \Bigr) \wedge \bigwedge_{h \in [n] \setminus (I \cup \{i\})} \omega_h.$$

Summing the first correction over all I, and adding the sum of the second correction over all I with $i \notin I$, we obtain

$$\begin{aligned} Z = \biggl(1 - \sum_{h \in [n]} r_h^2\biggr) \wedge \sum_{m \geq 0} 2^m m! \sum_{\substack{|I| = m \\ I \subseteq [n]}} \Bigl(\bigwedge_{j \in I} r_j \, dr_j \wedge \alpha_j \Bigr) \wedge \bigwedge_{j \in [n] \setminus I} \omega_j \\ + \biggl(\sum_{h \in [n]} r_h \, dr_h \biggr) \wedge \sum_{i \in [n] \setminus \{h\}} r_i^2 \alpha_i \wedge \sum_{m \geq 0} 2^{(m+1)}(m+1)! \\ \wedge \sum_{\substack{|I| = m \\ I \subseteq [n] \setminus \{i, h\}}} \Bigl(\bigwedge_{j \in I} r_j \, dr_j \wedge \alpha_j \Bigr) \wedge \bigwedge_{j \in [n] \setminus (I \cup \{i\})} \omega_j. \end{aligned}$$

Therefore, without imposing $\sum r_i^2 = 1$ and $\sum r_i \, dr_i = 0$ we have

$$d\Phi = Z - \bigwedge_{i=1}^{n} \omega_i.$$

Clearly Z vanishes if we do make these assumptions.

Construction–Notation 1 Suppose that $S_1, \dots, S_l \to M$ are piecewise smooth S^1–bundles over a piecewise smooth orbifold with corners. Denote by $S(S_1, \dots, S_l) \to M$ the $(2l-1)$–sphere bundle on M whose fibers are

$$S(S_1, \dots, S_l)_x = \Bigl\{ (r_1, P_1, r_2, P_2, \dots, r_l, P_l) \Bigm| P_i \in (S_i)_x, \, r_i \geq 0, \, \sum r_i^2 = 1 \Bigr\} \Big/ {\sim},$$

where $\sim$ is the equivalence relation generated by

$$(r_1, P_1, \dots, 0, P_i, \dots, r_l, P_l) \sim (r_1, P_1, \dots, 0, P_i', \dots, r_l, P_l),$$

equipped with the natural topology.

4 Symmetric Jenkins–Strebel stratification

In the remainder of the article all open spin surfaces we will encounter, twisted or not, will have a lifting. Similarly, we will encounter several types of graphs: the dual graphs we have defined above, ribbon graphs and nodal graphs. These graphs will also be classified as open or closed and will sometimes carry spin structures, twisted or not. All the open spin graphs we shall meet will have a lifting. For this reason we will sometimes slightly abuse notation and omit the suffix "with a lifting" from the terminology. We will also usually omit the addition "twisted". It will be clear from the context if we mean a closed or open object, twisted or not, etc.

4.1 JS stratification for the closed moduli

4.1.1 JS differential and the induced graph In this subsection we briefly describe the stratification of moduli of closed stable curves, following [25; 42; 28].

Let Σ be a nodal Riemann surface with $2g-2+n \geq 0$. A meromorphic section γ of the tensor square of the cotangent bundle defined on each component of the normalization of Σ can be written in a local coordinate z as $f(z)\,dz^2$. If γ has a double pole at $w \in \Sigma$, the *residue* of γ at w is the coefficient of $dz^2/(z-w)^2$ in the expansion of γ around w. The residue is independent of the choice of the local coordinate. A *quadratic differential* γ is such a section which has at most double poles, all the poles are located either at the marked points or at the nodes, and for any node, the residues of γ at its two branches are the same.

Let γ be a quadratic differential, and $w \in \Sigma$ a point which is neither a zero nor a pole. In a neighborhood U we can take its unique (up to sign) square root $\sqrt{\gamma}$. This is a 1–form, hence can be integrated along a path. This defines a map

$$g\colon U \to \mathbb{C}, \quad g(z) = \int_w^z \sqrt{\gamma},$$

where the integral is taken along any path in U.

A *horizontal trajectory* is the preimage of $\mathbb{R} \subset \mathbb{C}$, and it is a smooth path containing w in its interior. It turns out that the notion of horizontal trajectories can be defined also in the case where w is a zero of order $d \geq -1$, where as usual a zero of order $-m$ is a pole of order m. In this case there are exactly $d+2$ horizontal rays leaving w. When w is a pole of order 2, if its residue is $-(p/2\pi)^2$ for some $p \in \mathbb{R}_+$, there is a family of nonintersecting horizontal trajectories surrounding it whose union is a topological open disk, punctured at w. Moreover, with respect to the metric defined by $|\sqrt{\gamma}|$, the perimeter of each of these trajectories is p.

Example 4.1 Let Σ be the Riemann sphere. For all $p > 0$,

$$\gamma_p = -\left(\frac{p}{2\pi}\right)^2 \left(\frac{dz}{z}\right)^2$$

is a quadratic differential, whose only poles are at 0 and ∞ and whose horizontal lines are the sets $|z| = r$ for $r > 0$, whose lengths are indeed p. Their union is an open punctured disk. It should be noted that actually this is the only quadratic differential on the sphere, up to scaling, which is invariant under the reflection in the equator whose only poles are at 0 and ∞.

Definition 4.2 Let $(\Sigma, z_1, \dots, z_n, z_{n+1}, \dots, z_{n+n_0})$ be a marked genus g nodal Riemann surface with $2g-2+n \geq 0$, where the subscript of z_i indicates its marking. Let $p_1, \dots, p_n$ be positive reals, and $p_i = 0$ for $i > n$. A *marked component* is a smooth component of the curve with at least one marked point z_i, with $i \in [n]$. The other components are called *unmarked*. A *Jenkins–Strebel differential*, or a *JS*–differential for short, is a quadratic differential γ such that:

(a) γ is holomorphic outside of special points. At nodes it has at most simple poles, and at the i^{th} marked point it has a double pole with residue $-(p_i/2\pi)^2$. In particular, if $p_i = 0$ there is at most a simple pole at that point.

(b) γ vanishes identically on unmarked components.

(c) Let Σ' be any marked component of Σ. When $p_i \neq 0$, if D_i is the punctured disk which is the union of horizontal trajectories surrounding $z_i \in \Sigma'$, then

$$\bigcup_{i \in [n]} \overline{D}_i = \Sigma'.$$

The following theorem was proved in [37] for the smooth case; the nodal case was treated in [28; 42].

Theorem 4.3 *Given a stable marked surface $(\Sigma, z_1, \dots, z_{n+n_0})$ with $n > 0$ and $\boldsymbol{p} = (p_1, \dots, p_{n+n_0}) \in \mathbb{R}_+^n \times (0, \dots, 0)$ as above, the JS differential exists and is unique.*

Given $(\Sigma, \boldsymbol{z})$ and $\boldsymbol{p}$ as above, define the decorated surface $\widetilde{\widetilde{\Sigma}}$ and the map $K_{n_0}\colon \Sigma \to \widetilde{\widetilde{\Sigma}}$ as follows. $\widetilde{\widetilde{\Sigma}}$ is obtained from Σ by contracting any unmarked component to a point, and decorating any such point by its *genus defect* and *marking defect*. The genus defect is the genus of the preimage of the point in Σ, and if that preimage is a single point, it is defined to be 0. The marking defect is the set of marked points in this preimage, which is labeled by a subset of $[n+n_0] \setminus [n]$. We should stress that γ need not vanish on a preimage of a node in the normalization, but it can have at most a simple pole there. Thus, from the discussion about horizontal trajectories, each node or unmarked component and in particular any point z_i for $i > n$ must be mapped to a point which touches at least one horizontal trajectory. Note also that an unmarked component always touches a node (unless $n = 0$ and then the whole surface is unmarked).

The JS differential γ induces a metric graph on $\widetilde{\widetilde{\Sigma}}$ whose vertices are zeroes of order $d \geq -1$[5] of γ, including the images of unmarked components, and whose edges are the horizontal trajectories, with their intrinsic length. These embedded graphs can be fully described.

Definition 4.4 A graph $G = (V, H, s_0, s_1, g, f)$, where

(a) V is the set of vertices, H is the set of half-edges,

(b) s_0 is a permutation of the half-edges emanating from each vertex,

(c) s_1 is a fixed-point-free involution of H,

(d) g is a map $g\colon V \to \mathbb{Z}_{\geq 0}$, called the *genus defect*, and

(e) f is a map $f\colon [n+n_0] \setminus [n] \to V$,

is called a $(g,(n,n_0))$*–stable closed ribbon graph*. The *faces* of the graph are s_2–equivalence class of half-edges, where $s_2 = s_0^{-1} s_1$. We write $F = H/s_2$. The edges are $E = H/s_1$. The *genus* of G can be defined as follows. Glue disks along the faces to obtain a surface $\widetilde{\widetilde{\Sigma}}$. The genus of G is the (arithmetic) genus of $\widetilde{\widetilde{\Sigma}}$ plus the sum of genus defects in vertices. The *marking defect* of a vertex v is defined as $f^{-1}(v)$. We require that:

(a) For a vertex v of degree 1 or of degree 2, but such that the assigned permutation is a transposition,
$$g(v) + |f^{-1}(v)| \geq 1.$$

[5] We consider a simple pole as a zero of order -1, and a point which is neither a zero nor a pole to be a zero of order 0.

(b) The genus of the graph is g.

(c) The number of faces is n.

A *stable metric ribbon graph* is a stable ribbon graph together with a metric

$$\ell\colon E \to \mathbb{R}_+.$$

We usually write ℓ_e instead of $\ell(e)$.

A graph is *smooth* if all the vertices' permutations s_0 are cyclic, all genus defects are 0, and all marking defects are of size at most 1. The ribbon graph is connected if the underlying graph is. We define isomorphisms and automorphisms in the expected way. Write $\mathrm{Aut}(G)$ for the automorphism group of G.

Note that case (a) above occurs when v is either the image of a contracted unmarked component, or the image of one of the points p_i for some $i > n$.

Remark 4.5 To a stable metric ribbon graph one can associate in a natural way a decorated metric space made of a disjoint union of closed intervals, one for each $e \in E$, modulo the identification of endpoints dictated by the graph structure. The vertices, which are the equivalence classes of endpoints of intervals, are endowed with genus and marking defects, and the closed interval which corresponds to the edge e is associated to a metric structure which makes it isometric to the interval $[0, \ell_e] \subset \mathbb{R}$. The associated decorated metric space is unique up to the expected notion of isomorphism. Stable metric ribbon graphs which arise from a JS differential (we will see in Theorem 4.8 below that all stable metric ribbon graphs arise this way) are endowed with this additional structure of isometries between the embedded edges and intervals of $\mathbb{R}$. This will be used below, when we give coordinates to the combinatorial S^1–bundles. For more details we refer the reader to [42].

Notation 4.6 Throughout this article, given a ribbon graph, possibly with extra structure such as a graded ribbon graph, or a nodal graph, which will be defined later, we shall write $[h]$ for the class of the half-edge or the edge h under the action of the automorphism group. We similarly write $[A]$ for a subset of edges or half-edges.

Remark 4.7 If $\mathrm{Norm}\colon \mathrm{Norm}(\Sigma) \to \Sigma$ is the normalization of Σ, and γ is the JS differential on Σ with prescribed perimeters, then $\mathrm{Norm}^*\gamma$ is a JS differential, hence the unique JS differential, on $\mathrm{Norm}(\Sigma)$, with the same perimeters, and such that marked points which are preimages of nodes have 0–perimeter.

4.1.2 Combinatorial moduli For a closed stable ribbon graph G, write $\mathcal{M}_G$ for the set of all metrics on G, write $\mathcal{M}_G(\boldsymbol{p})$ for the set of all such metrics where the i^{th} face has perimeter p_i. We have that $\mathcal{M}_G \simeq \mathbb{R}_+^{E(G)}/\mathrm{Aut}(G)$ canonically, and this identification endows it with a smooth structure.

For $e \in E(G)$, the edge between vertices v_1 and v_2, define the graph $\partial_e G$, the edge contraction, as follows. Write h_1, h_2 for the two half-edges of e. Set $V(\partial_e G) = V(G) \setminus \{v_1, v_2\} \cup \{v_1 v_2\}$ and $H(\partial_e G) = H(G) \setminus \{h_1, h_2\}$. The maps s_1', g' and f' are just s_1, g and f when restricted to vertices and half-edges of G. For the new vertex $v = v_1 v_2$, set $f'(v) = f(v_1) \cup f(v_2)$, and set $g'(v) = g(v_1) + g(v_2)$ whenever $v_1 \neq v_2$, otherwise it is $g(v_1) + \delta$, where $\delta = 1$ if h_1 and h_2 belong to different s_0–cycles, and otherwise $\delta = 0$. For any half-edge h, with $h/s_1 \neq e$, define $s_2'(h)$ to be the first half-edge among $s_2(h), s_2^2(h), \dots$ which is not a half-edge of e. We then put $s_0' = s_1'(s_2')^{-1}$.

Edge contractions commute with each other, and allow us to define a cell complex $\overline{\mathcal{M}}_G = \coprod_{G'} \mathcal{M}_{G'}$, where the union is over all graphs obtained from G by edge contractions, and we glue the cell $\mathcal{M}_{G'}$ of $G' = \partial_{e_1,\dots,e_r} G$ to the cell $\mathcal{M}_G$ along $\ell_{e_1} = \dots = \ell_{e_r} = 0$. We similarly define $\overline{\mathcal{M}}_G(\boldsymbol{p})$.

Write $\mathcal{M}^{\mathrm{comb}}_{g,(n,n_0)} = \coprod \mathcal{M}_G$, where the union is taken over smooth closed $(g,(n,n_0))$ ribbon graphs. Write $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,(n,n_0)} = \coprod \overline{\mathcal{M}}_G/{\sim} = \coprod \mathcal{M}_G$, where the union is taken over all closed stable $(g,(n,n_0))$ ribbon graphs, and $\sim$ is induced by edge contractions. Define $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,(n,n_0)}(\boldsymbol{p})$ and $\mathcal{M}^{\mathrm{comb}}_{g,(n,n_0)}(\boldsymbol{p})$ by constraining the perimeters to be p_i. In all cases we define the cell attachment using edge contractions, and the resulting spaces are piecewise smooth Hausdorff orbifolds; see [28; 42] for details.

Set $\mathrm{comb} = \mathrm{comb}_{n_0}$ as the canonical maps

$$\mathrm{comb}\colon \overline{\mathcal{M}}_{g,n+n_0} \times \mathbb{R}_+^n \to \overline{\mathcal{M}}^{\mathrm{comb}}_{g,(n,n_0)} \quad \text{and} \quad \mathrm{comb}_{\boldsymbol{p}}\colon \overline{\mathcal{M}}_{g,n+n_0} \to \overline{\mathcal{M}}^{\mathrm{comb}}_{g,(n,n_0)}(\boldsymbol{p}),$$

which send a stable curve and a set of perimeters to the corresponding graph.

We have, from [25; 28; 42]:

Theorem 4.8 *Suppose $n > 0$. The maps* comb *and* $\mathrm{comb}_{\boldsymbol{p}}$ *are continuous surjections of topological orbifolds. The map* $\mathrm{comb}_{\boldsymbol{p}}$ *takes the fundamental class to a fundamental class. Moreover, the cell complex topology described above is the finest topology with respect to which* comb *is continuous. The maps are isomorphisms onto their images when restricted to $\mathcal{M}_{g,n+n_0} \times \mathbb{R}_+^n$ and $\mathcal{M}_{g,n+n_0}$.*

More generally, suppose Γ is a closed dual graph with the property that any vertex without a tail marked by $[n]$ is of genus zero, and has exactly three half-edges, and any two such vertices are not adjacent. Then, with the same proofs, comb *and* $\mathrm{comb}_{\boldsymbol{p}}$ *restricted to $\mathcal{M}_\Gamma \times \mathbb{R}_+^n$ and $\mathcal{M}_\Gamma$ are isomorphisms onto their image.*

4.1.3 Tautological line bundles and associated forms

Definition 4.9 Suppose $p_i > 0$. Define the space

$$\mathcal{F}_i(\boldsymbol{p}) \to \overline{\mathcal{M}}_{g,n}^{\mathrm{comb}}(\boldsymbol{p})$$

as the collection of pairs (G,ℓ,q), where $(G,\ell) \in \overline{\mathcal{M}}_{g,n}^{\mathrm{comb}}(\boldsymbol{p})$ and q is a boundary point of the i^{th} face. These spaces glue together to the bundle $\mathcal{F}_i \to \overline{\mathcal{M}}_{g,n}^{\mathrm{comb}}$. Define ϕ_j to be the distance from q to the j^{th} vertex, taken along the arc from q in the counterclockwise direction, so that $0 < \phi_1 < \phi_2 < \cdots < \phi_N < p_i$, where N is the number of edges in the i^{th} face, counted with multiplicities, and the distances are measured using the identifications of the edges with subintervals of $\mathbb{R}$; see Remark 4.5. Write $\ell_j = \phi_{j+1} - \phi_j$. Orient the fibers with the clockwise orientation.

Define the following one-form and two-form on each cell of $\overline{\mathcal{M}}_{g,(n,n_0)}^{\mathrm{comb}}(\boldsymbol{p})$:

$$\alpha_i = \sum_{j=1}^{N} \frac{\ell_j}{p_i} d\left(\frac{\phi_j}{p_i}\right) \quad \text{and} \quad \omega_i = -d\alpha_i = \sum_{1 \le a < b \le N} d\left(\frac{\ell_a}{p_i}\right) \wedge d\left(\frac{\ell_b}{p_i}\right). \tag{23}$$

Later we will integrate forms which are made out of α_i and ω_i, and we will perform Laplace transform over $\boldsymbol{p}$. For this reason it will be convenient to define the scaled versions of α_i and ω_i, which do not contain p_i in their denominators. We thus put

$$\bar{\alpha}_i = p_i^2 \alpha_i, \quad \bar{\omega}_i = p_i^2 \omega_i, \quad \bar{\omega} = \sum_i \bar{\omega}_i.$$

The bundles $\mathcal{F}_i$ carry natural piecewise smooth structures. Moreover, [25] says the following; see also [42, Theorem 5].

Theorem 4.10 (a) *For $i \in [n]$, we have* $\mathrm{comb}^* \mathcal{F}_i \simeq S^1(\mathbb{L}_i)$ *canonically.*

(b) *The forms α_i and ω_i are a piecewise smooth angular one-form and Euler two-form for $\mathcal{F}_i$.*

Remark 4.11 In [25], $\mathcal{F}_i$ was given the opposite orientation and the equivalence was hence with the bundle $S^1(\mathbb{L}_i^*)$, which is canonically $S^1(\mathbb{L}_i)$ with the opposite orientation.

Thus, combined with Theorem 4.8, we see that all descendents may be calculated combinatorially on $\overline{\mathcal{M}}_{g,n}^{\mathrm{comb}}$. In fact, all descendents can be calculated as integrals over the highest-dimensional cells of $\overline{\mathcal{M}}_{g,n}^{\mathrm{comb}}$. These are parametrized by trivalent ribbon graphs.

4.2 JS stratification for the open moduli

4.2.1 Symmetric JS differentials The next definition is motivated by Definition 4.2 and Example 4.1.

Definition 4.12 Let $(\Sigma, \{z_i\}_{i\in\mathcal{I}\cup\mathcal{P}_0}, \{x_i\}_{i\in\mathcal{B}})$ be a stable open marked Riemann surface, and let $\boldsymbol{p} = (p_i)_{i\in\mathcal{I}\cup\mathcal{P}_0} \in \mathbb{R}_+^{\mathcal{I}} \times (0,\dots,0)$. A *symmetric JS differential* on Σ is the restriction to Σ of the unique JS differential of $D(\Sigma)$ whose residues at z_i and $\bar z_i$ are $-(p_i/2\pi)^2$, which are 0 for $i \in \mathcal{P}_0$. We extend the definition to the case $g = 0$, $\mathcal{I} = [1]$ and $\mathcal{P}_0 = \mathcal{B} = \varnothing$, where the differential is defined to be the restriction of the section γ_{p_1} of Example 4.1.

Existence and uniqueness follow from Theorem 4.3 and the discussion in Example 4.1.

As before, the symmetric JS differential defines a cell decomposition of $D(\Sigma)$ in the smooth case, and in general a metric graph embedded in $\widetilde{D(\Sigma)}$, the surface obtained from $D(\Sigma)$ by contracting components with no z_i or $\bar z_i$ with $i \in \mathcal{I}$, whose complement is a disjoint union of disks. Note that $\widetilde{D(\Sigma)}$ inherits the conjugation from $D(\Sigma)$, which we also denote by ϱ. The uniqueness forces the decomposition to be ϱ–invariant.

Lemma 4.13 *The ϱ–fixed locus of $\widetilde{D(\Sigma)}$ is a union of (possibly closed) horizontal trajectories and isolated vertices. Any ϱ–fixed point is a zero the differential of an even order, possibly* 0.

Proof The case $g=0, \mathcal{I}=[1], \mathcal{P}_0=\mathcal{B}=\varnothing$ follows from the discussion in Example 4.1. In other cases, take an arbitrary point in $\widetilde{D(\Sigma)}^{\varrho}$. It cannot belong to the disk cell of any z_j, since otherwise it would have belonged to the cell of $\bar z_j$ as well. Thus, $\widetilde{D(\Sigma)}^{\varrho}$ is contained in the one-skeleton of the decomposition. Consider $p \in \widetilde{D(\Sigma)}^{\varrho}$. If p is an isolated vertex in the ϱ–fixed locus, then by connectivity it must be incident to some non-ϱ–fixed horizontal trajectory which, without loss of generality, lies in the image of Σ^o in $\widetilde{D(\Sigma)}$. Suppose it touches r such trajectories. Then it also touches their ϱ–conjugate trajectories, which lie in the image of $\overline{\Sigma}^o$ in $\widetilde{D(\Sigma)}$. Thus, $2r$ horizontal trajectories emanate from p, for $r \geq 1$, hence p is a zero of order $2r-2 \geq 0$. The second case is that p is not isolated, so it lies in the image of $\partial\Sigma$ in $\widetilde{D(\Sigma)}$, which, as explained, is contained in the 1–skeleton. In this case, at least two horizontal trajectories which are contained in the image of $\partial\Sigma$ emanate out of p, one to its left and one to its right. In addition, there are also $r \geq 0$ such trajectories in the image of Σ^o, and because of symmetry there are also r such trajectories in the image of $\overline{\Sigma}^o$. In total, there are $2r+2$ horizontal trajectories emanating from p, which means that it is a zero of order $2r \geq 0$. □

Lemma 4.13 has the following corollary.

Corollary 4.14 *Suppose Σ and $\mathbf{p}$ are as above, and γ is the associated symmetric JS differential. Assume that for some $i \in \mathcal{B}$, forgetting x_i makes no component of Σ unstable. Denote by Σ' the resulting surface, and let $\iota\colon \Sigma' \to \Sigma$ be the natural map between the surfaces. Then if γ and γ' are the unique JS differentials for Σ and Σ', respectively, with the prescribed perimeters, then*

$$\gamma' = \iota^*\gamma.$$

Indeed, both γ and γ' are JS differentials on Σ', since there is no pole in x_i. Hence they must be equal.

Remark 4.7 has the following consequence.

Corollary 4.15 *If* $\mathrm{Norm}\colon \mathrm{Norm}(\Sigma) \to \Sigma$ *is the normalization of* Σ, *and* γ *is the JS differential on* Σ *with prescribed perimeters, then* $\mathrm{Norm}^*\gamma$ *is the unique JS differential on* $\mathrm{Norm}(\Sigma)$ *with the same perimeters and such that marked points which are preimages of nodes have perimeter zero.*

Remark 4.16 Although throughout the article we will be mainly interested in internal markings with positive perimeters, markings of perimeter zero occur naturally when one considers normalizations; see Proposition 4.34. In the open intersection theory the normalizations are crucial for the definition of intersection numbers, Definition 2.47, and therefore considering markings with zero perimeters is unavoidable. In addition, since boundary markings carry no descendents, we to not lose from fixing their perimeters to be zero, and it simplifies calculations. For these reasons, throughout this section we shall allow marked points to have perimeter zero, at the cost of making the notation somehow more cumbersome.

4.2.2 Open ribbon graphs

Notation 4.17 Let I and B be finite sets. Let $\mathcal{IT}(g, I, B)$ denote the set of isotopy types of open connected genus g smooth oriented marked surfaces, with I being the set of internal marked points and B being the set of boundary marked points. Write $\mathcal{IT}(g, I)$ for the set of isotopy types of closed connected genus g smooth oriented marked surfaces, which is just a singleton.

Definition 4.18 An *open ribbon graph* is a tuple

$$G = (V = V^I \cup V^B, H = H^I \cup H^B, s_0, s_1, f = f^I \cup f^B \cup f^{P_0}, g, d),$$

where:

(a) V^I is the set of internal vertices, V^B the set of boundary vertices.

(b) H^B is the set of boundary half-edges, H^I is the set of internal half-edges; s_1 is a fixed-point-free involution on H whose equivalence classes are the edges, E. E^B is the set of edges which contain a boundary half-edge.

(c) s_0 is a permutation assigned to each vertex, and should be thought of as a cyclic order of the half-edges issuing from each vertex. We write s_0 also for the product of all these permutations.

We denote by $\widetilde{V}$ the set of cycles of s_0. Write $\widetilde{V}^I$ for cycles which do not contain boundary half-edges. Set $\widetilde{V}^B = \widetilde{V} \setminus \widetilde{V}^I$. Denote by $N : \widetilde{V} \to V$ the map which takes a cycle to the vertex which contains its half-edges, and let N^{P_0} and N^B be the restrictions to $\widetilde{V}^I$ and $\widetilde{V}^B$, respectively.

(d) $f^B : \mathcal{B} \to V^B$ is a map from a finite set $\mathcal{B}$.

(e) $f^{P_0} : \mathcal{P}_0 \to V$ is a map from a finite set $\mathcal{P}_0$.

(f) $f^I : \mathcal{I} \hookrightarrow H/s_2$ is an injection, where $s_2 := s_0^{-1} s_1$.

(g) $g : V \to \mathbb{Z}_{\geq 0}$ is a map called the *genus defect*.

(h) For any $v \in V^B$, we have an element

$$d(v) \in \mathcal{IT}\big(g(v), (f^{P_0})^{-1}(v) \cup (N^{P_0})^{-1}(v), (f^B)^{-1}(v) \cup (N^B)^{-1}(v)\big).$$

For any $v \in V^I$, the element $d(v)$ is the unique element in

$$\mathcal{IT}\big(g(v), (f^{P_0})^{-1}(v) \cup (N^{P_0})^{-1}(v)\big).$$

This d is called the *topological defect* of v.

Write $\deg(v)$ for the degree of the vertex v. A *closed contracted component* is a vertex $v \in V^I$ with

$$2g(v) + |(f^{P_0})^{-1}(v)| + |N^{-1}(v)| > 2.$$

Denote their collection by $\mathrm{Cont}^C(G)$. An *open contracted component* is a vertex $v \in V^B$ with

$$2(g(v) + |(f^{P_0})^{-1}(v)| + |(N^{P_0})^{-1}(v)|) + |(f^B)^{-1}(v)| + |(N^B)^{-1}(v)| > 2.$$

Denote their collection by $\mathrm{Cont}^O(G)$.

We have the following requirements.

(a) Any half-edge appears in the permutation s_0 of exactly one vertex. We define a graph whose vertices are the elements of V and whose half-edges are the elements of H. A half-edge is connected to a vertex if and only if it appears in the vertex's permutation s_0.

(b) $N(\widetilde{V}^B) \subseteq V^B$.

(c) If $h \in H^B$, then $s_1 h \notin H^B$.

(d) s_2 preserves the partition $H = H^I \cup H^B$. The image of f^I is exactly H^I/s_2.

(e) For $v \in V^I$, if $\deg(v) = 1$, or $\deg(v) = 2$ but $|N^{-1}(v)| = 1$, then $|(f^{P_0})^{-1}(v)| + g(v) \geq 1$.

(f) For $v \in V^B$, if v has at least one boundary edge and $\deg(v) = 2$ then

$$|(f^{P_0})^{-1}(v)| + |(f^B)^{-1}(v)| + g(v) \geq 1.$$

(g) Any vertex of degree 0 is a *contracted component*.

We call the elements of H^B/s_2 *boundary components*, and the elements of $F = H^I/s_2$ *faces*. The number of boundary components is $b(G) = |H^B/s_2|$. The *marking defect* of $v \in V$ is defined as $(f^{P_0})^{-1}(v) \cup (f^{\mathcal{B}})^{-1}(v)$. The sets $\mathcal{I}$, $\mathcal{P}_0$ and $\mathcal{B}$ are called the sets of internal markings, internal markings of perimeter zero, and boundary markings, respectively. The set $\mathcal{B}$ is also denoted by $B(G)$; define $I(G)$ and $P_0(G)$ similarly. An *internal node* is either a contracted component with at least one edge and no boundary edges, or an internal vertex whose assigned permutation is not transitive. A boundary vertex v without boundary half-edges, with an empty marking defect and such that $g(v) = 0$ and $|N^{-1}(v)| = 1$ is called a *contracted boundary*. We denote the collection of those boundary vertices by $\mathrm{CB}(G)$. A boundary vertex v which is either a contracted component with at least one boundary edge, or whose assigned permutation is not transitive, is called a *boundary node*. A *boundary marked point* is an image of f^B which is not a node. An *internal marked point of perimeter zero* is an image of f^{P_0} which is not a node. A *boundary half-node* is an $(N^B)^{-1}$–preimage of a node. Denote their collection by $\mathrm{HN}(G)$. A vertex which is either a node or a contracted component, or the f–image of a unique element in $\mathcal{P}_0 \cup \mathcal{B}$, is called a *special point*.

We write $i(h) = h/s_2$ and $H_i = \{h \in H \mid i(h) = i\}$.

An *open metric ribbon graph* is an open ribbon graph together with a positive metric $\ell \colon E \to \mathbb{R}_+$. We sometimes write ℓ_h, with $h \in H$, instead of ℓ_{h/s_1}.

Markings of an open ribbon graph are markings

$$\mathtt{m}^I : \mathcal{I} \cup \mathcal{P}_0 \to \mathbb{Z} \quad \text{and} \quad \mathtt{m}^B : \mathcal{B} \to \mathbb{Z}$$

such that $\mathtt{m}^I(\mathcal{P}_0) = 0$ and $\mathtt{m}^I(\mathcal{I}) \subset \mathbb{Z}_{\neq 0}$. A graph together with a marking is called a *marked graph*.

An isomorphism of marked graphs and an automorphism of a marked graph are the expected notions. $\mathrm{Aut}(G)$ denotes the group of automorphisms of G. A metric is *generic* if (G, ℓ) has no automorphisms.

A ribbon graph is said to be closed if $V^B = 0$, and it is said to be connected if the underlying graph is connected.

The maps f^B and f^{P_0} should be thought of as the respective associations of the boundary marked points and the internal marked points of perimeter zero to the vertices of the graph formed by the symmetric JS differential. Requirements (e) and (f) in this definition are the open counterparts of requirement (a) of Definition 4.4. Note that a half-edge h is canonically oriented away from its basepoint h/s_0. Throughout the paper we identify boundary marked points, which are vertices, with their (unique) preimages in $B(G) = \mathcal{B}$.

Remark 4.19 Here, unlike in the closed case, the genus defect is not enough to classify surfaces with contracted components. In particular, there are several topologies for a given genus, as mentioned in Remark 2.4, and the set of topologies grows as we add boundary marked points, which may be divided between different boundary components.

Figure 10 shows some examples of ribbon graphs.

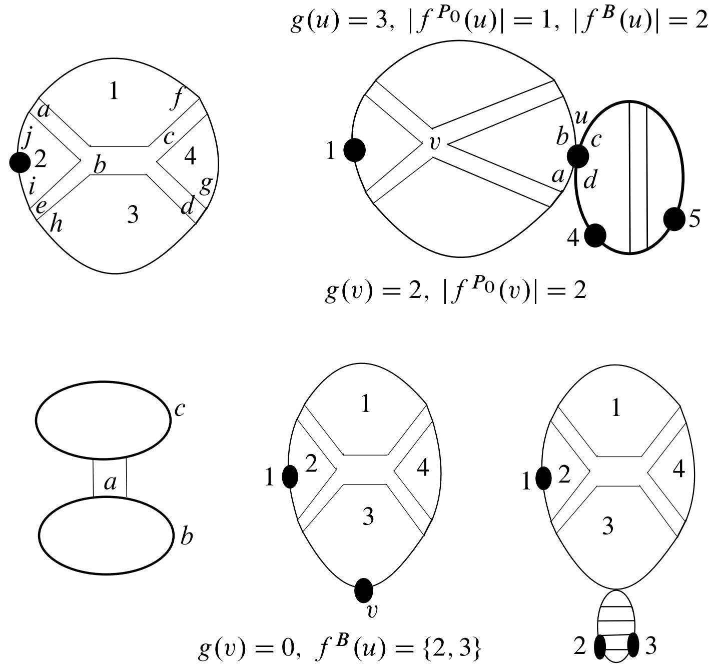


Figure 10: Examples of ribbon graphs. Internal edges are drawn as strips. Top left is a ribbon graph with one boundary marking and four internal markings (the name of a half-edge appears next to the vertex from which it emanates). Its underlying surface is a disk, and the boundary edges are $s_1 f$, $s_1 g$, $s_1 h$, $s_1 i$, $s_1 j$. The cyclic orders in the internal vertices are $s_1 a, s_1 b, e$ and $s_1 b, s_1 d, c$. Face 1, for example, is the s_2–cyclic order a, b, c, f. Bottom right is a ribbon graph on a cylinder. It has one face, the s_2–cycle $a, c, s_1 a, b$ and two boundary components, each made of a single boundary edge, $s_1 b$ and $s_1 c$. The ribbon graph at top right has one boundary node u, which is also an open contracted component, and an internal node v, which is also a contracted component. The permutation of half-edges at u is $(ab)(cd)$. The contracted component is open, of genus defect 3, has an internal marking of perimeter zero, and four special boundary points: the markings $1, 2$ and the half-nodes (ab), (cd). The topological defect can be any topology which corresponds to doubled genus 3, one internal marking and four boundary markings. The node v has genus defect 2 and two perimeter-zero internal markings. The center bottom picture has an open contracted component at v, it is a contracted disk with two boundary markings $2, 3$ and a boundary half-node and no special internal points. Contracted components which are disks with three boundary markings and no internal markings will play an important role in what follows. We shall therefore draw such components as disks cut by parallel lines, as in the bottom right picture.

Notation 4.20 By gluing disks along the faces, any open ribbon graph gives rise to a topological open oriented surface Σ_G. This surface is a union of smooth surfaces, identified in a finite number of points. One can easily define its double, $D(\Sigma_G) = (\Sigma_G)_{\mathbb{C}}$, as in the nontopological case.

Definition 4.21 The genus of the open graph G is defined by

$$g(G) := g((\Sigma_G)_{\mathbb{C}}) + \sum_{v \in V^B} g(v) + 2 \sum_{v \in V^I} g(v).$$

The graph is *stable* if $2g - 2 + |\mathcal{B}| + 2(|\mathcal{I}| + |\mathcal{P}_0|) > 0$.

For a stable open surface $(\Sigma, \{z_i\}_{i \in \mathcal{I} \cup \mathcal{P}_0}, \{x_i\}_{i \in \mathcal{B}})$, define the marked components to be components with at least one z_i, for some $i \in \mathcal{I}$. The other components are unmarked. Define the decorated surface $\widetilde{\Sigma} = K_{\mathcal{B},\mathcal{P}_0}(\Sigma)$, and the map $K_{\mathcal{B},\mathcal{P}_0} \colon \Sigma \to \widetilde{\Sigma}$ to be the surface obtained by contracting unmarked components to points, and $K_{\mathcal{B},\mathcal{P}_0}$ is the quotient map. We decorate any point p in $\widetilde{\Sigma}$ by its genus defect, marking defect and the topological defect, which can be defined by the genus, boundary markings and topological type of the surface obtained by smoothing the nodes in $K^{-1}_{\mathcal{B},\mathcal{P}_0}(p)$.

Remark 4.22 This definition agrees with the one given for closed surfaces, in the sense that one can also define the doubling D of $\widetilde{\Sigma}$ in a natural way, and then $D(\widetilde{\Sigma}) \simeq \widetilde{D(\Sigma)}$.

Definition 4.23 A *ghost* is a ribbon graph without half-edges. A *smooth open ribbon graph* is a stable open ribbon graph such that none of its connected components contains a node or a contracted boundary.

A stable ribbon graph, open or closed, is *effective* if

(a) any genus defect is 0,

(b) there are no internal nodes, and

(c) contracted components or ghost components v must have

$$(N^{P_0})^{-1}(v) = \varnothing \quad \text{and} \quad |(N^B)^{-1}(v)| + |(f^B)^{-1}(v)| = 3.$$

The graph is *trivalent* if

(a) it is effective,

(b) $\mathcal{P}_0 = \varnothing$,

(c) it has no contracted boundaries,

(d) all vertices which are not special boundary points are trivalent, and

(e) for every special boundary point, all the s_0–cycles are of length 2.

A boundary marked point or a boundary half-node in a trivalent graph G which is not a ghost is said to *belong to* a face i if its unique internal half-edge belongs to that face.

In Figure 10 the diagrams on the left represent smooth graphs, and all but the top right are effective.

Remark 4.24 The only nonzero open intersection number which does not involve internal markings is the genus 0 intersection number with three boundary markings, $\langle\sigma^3\rangle_0^o$. The graph which corresponds to this picture is precisely the trivalent ghost.

The following proposition is a consequence of Lemma 4.13, and the closed theory; the proof is in the appendix.

Proposition 4.25 *Let Σ be a stable open marked Riemann surface. The unique symmetric JS differential of Σ defines a unique metric graph (G,ℓ) embedded in $K_{\mathcal{B},\mathcal{P}_0}(\Sigma)$. This graph is an open ribbon graph, whose vertices are $K_{\mathcal{B},\mathcal{P}_0}$–images of zeroes of the differential, and whose edges are $K_{\mathcal{B},\mathcal{P}_0}$–images of horizontal trajectories. The boundary edges, if there are any, are embedded in the boundary and cover it, and the defects of vertices agree with the defects of their image in $K_{\mathcal{B},\mathcal{P}_0}(\Sigma)$; in particular, boundary nodes go to boundary nodes. Under this embedding the orientation of any half-edge $h \in s_1 H^B$ agrees with the orientation induced on $\partial K_{\mathcal{B},\mathcal{P}_0}(\Sigma)$. Topologically, $K_{\mathcal{B},\mathcal{P}_0}(\Sigma) \simeq \Sigma_G$.*

Moreover, any stable $(g,\mathcal{B},\mathcal{I}\cup\mathcal{P}_0)$–metric graph is the graph associated to some stable open $(g,\mathcal{B},\mathcal{I}\cup\mathcal{P}_0)$–surface and a set of perimeters $\boldsymbol{p}$. This surface is unique if the graph is smooth or effective.

We sometimes identify the graph with its image under the embedding. In particular, throughout this article we shall consider an edge as a trajectory in the surface, and a half-edge h as a trajectory oriented outward from h/s_0.

Notation 4.26 With the notation of the above observation, denote by $\mathrm{comb}^{\mathbb{R}}_{\boldsymbol{p}}$ the map between surfaces and open metric ribbon graphs defined by $(G,\ell) = \mathrm{comb}^{\mathbb{R}}_{\boldsymbol{p}}(\Sigma)$. Write also $(G,\ell) = \mathrm{comb}^{\mathbb{R}}(\Sigma,\boldsymbol{p})$.

Definition 4.27 The *normalization* $\mathrm{Norm}(G)$ of a stable connected open ribbon graph G is the unique smooth, not necessarily connected, open ribbon graph, defined in the following way. If G is smooth, $\mathrm{Norm}(G) = G$. Otherwise the vertex set is $\widetilde{V}^I \cup \widetilde{V}^B \cup \mathrm{Cont}^C(G) \cup \mathrm{Cont}^O(G)$, contracted components are isolated vertices in the graph, and the half-edges are $H^I \cup H^B$. The genus and topological defects of vertices in $\widetilde{V}^I \cup \widetilde{V}^B$ are 0.

For a contracted component v, the genus and topological defects are given by

$$g^{\mathrm{Norm}(G)}(v) = g(v) \quad \text{and} \quad d^{\mathrm{Norm}(G)}(v) = d(v).$$

The marking defect and the maps $f^{P_0,v}$ and $f^{B,v}$ are derived from $d^{\mathrm{Norm}(G)}(v)$. In particular, $B(v) = (N^B)^{-1}(v) \cup (f^B)^{-1}(v)$. The permutations s_0^v and s_1^v are the trivial permutations, and $I(v) = \varnothing$.

For any connected component C of $\mathrm{Norm}(G)$ not in $\mathrm{Cont}^C(G) \cup \mathrm{Cont}^O(G)$, define $s_0 = s_0^C$, $s_1 = s_1^C$ and $f^I = f^{I,C}$ as those induced from G. Let $P_0(C)$ be the union of the set of elements of $\mathcal{P}_0$ which map to vertices whose unique N–preimage is in C, and the set of preimages of internal nodes of C, ie internal vertices v of C such that $|N^{-1}(N(v))| > 1$. In other words, we can write $P_0(C) = (P_0(C) \cap \mathcal{P}_0) \cup (P_0(C) \setminus \mathcal{P}_0)$. We define $f^{P_0} = f^{P_0,C} \colon P_0(C) \to V^I(C)$ as follows. On $P_0(C) \cap \mathcal{P}_0$ we put $f^{P_0,C}(p_i) = N^{-1}(f^{P_0}(p_i))$, where f^{P_0} of the right-hand side is the function from the definition of G, while on $P_0(C) \setminus \mathcal{P}_0$, the preimages of nodes, we set $f^{P_0,C}(v) = v$. Define $B(C)$ and $f^B = f^{B,C} \colon B(C) \to V^B(C)$ similarly.

The normalization $\mathrm{Norm}(G)$ of a marked graph is the marked graph whose underlying graph is the normalization of the underlying graph of G, and new marked points are marked 0.

Write $\mathrm{Norm} \colon \mathrm{Norm}(G) \to G$ for the evident normalization map.

Observe that the normalization of a trivalent graph is trivalent, and that if v is a contracted component which touches at least one edge in G, then $|\mathrm{Norm}^{-1}(v)| = |N^{-1}(v)| + 1$.

Figure 11 shows the normalizations of the graphs in the right column of Figure 10.

Notation 4.28 There is a canonical injection $B(G) \hookrightarrow B(\mathrm{Norm}(G))$. There is a fixed-point-free involution on $B(\mathrm{Norm}(G)) \setminus B(G)$, which we also denote by s_1, which on preimages of a node that is not a contracted component just interchanges its two preimages. If v is a contracted component, its new boundary markings correspond to elements $u \in (N^B)^{-1}(v)$. Any such u corresponds also to a unique marking w in another *noncontracted* component. Write $s_1 u = w$ and $s_1 w = u$.

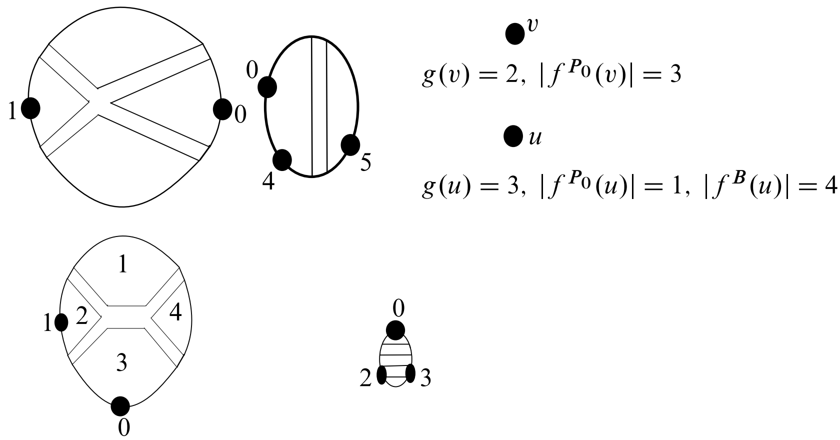


Figure 11: The normalizations of the graphs in the right column of Figure 10. The upper normalization has four components; two are contracted components. The one which corresponds to v has three internal points of perimeter 0: the original two and the node. The one which corresponds to u has four boundary markings: the original two and two that corresponds to half-nodes. The lower normalization is made of two components. New special points in both normalizations are labeled 0.

4.2.3 Moduli of open metric graphs For a stable open ribbon graph G, denote by $\mathcal{M}_G^{\mathbb{R}}$ the set of all metrics on G, and write $\mathcal{M}_G^{\mathbb{R}}(\boldsymbol{p})$ for the set of all such metrics where the i^{th} face has perimeter p_i. Note that $\mathcal{M}_G^{\mathbb{R}} = \mathbb{R}_+^{E(G)}/\mathrm{Aut}(G)$ canonically.

Construction–Notation 2 For $e \in E(G)$ the edge between vertices v_1 and v_2, one can define the graph $\partial_e G$ as the graph obtained by contracting e to a point, identifying its vertices to give a new vertex $v_1 v_2$ and updating the permutations and marking defects as in the closed case. When v_1 and v_2 are internal, then so is $v_1 v_2$. The genus defect is updated as in the closed case, and this determines the whole defect. Suppose v_1 is a boundary vertex. Then so is $v_1 v_2$. If $v_2 \neq v_1$, then $g(v_1 v_2) = g(v_1) + g(v_2)$ if $v_2 \in V^B$, and otherwise $g(v_1 v_2) = g(v_1) + 2g(v_2)$. When $v_1 = v_2$, let h_1 and h_2 be the half-edges of e. Let $\tilde{h}_i \in N^{-1}(v_1)$ be the s_0–cycle of h_i. Then $g(v_1 v_2) = g(v_1) + \delta$, where

$$\delta = \begin{cases} 0 & \text{if } \tilde{h}_1 = \tilde{h}_2, \\ 1 & \text{if } \tilde{h}_1 \neq \tilde{h}_2, \quad \text{where } \tilde{h}_1, \tilde{h}_2 \in \tilde{V}^B, \\ 2 & \text{otherwise.} \end{cases}$$

We have

$$d(v_1 v_2) \in \mathcal{IT} = \mathcal{IT}(g(v_1 v_2), I_{v_1 v_2}, B_{v_1 v_2})$$

or

$$d(v_1v_2) \in \mathcal{IT} = \mathcal{IT}(g(v_1v_2), I_{v_1v_2}),$$

where

$$B_{v_1v_2} = (f^B)^{-1}(v_1v_2) \cup (N^B)^{-1}(v_1v_2),$$
$$I_{v_1v_2} = (f^{P_0})^{-1}(v_1v_2) \cup (N^{P_0})^{-1}(v_1v_2).$$

These two sets are already known from what we have constructed so far. In particular, whenever $\mathcal{IT}$ is trivial — which is always the case for internal vertices, and for boundary vertices it happens when $2g(v_1v_2) + 2|I_{v_1v_2}| + |B_{v_1v_2}| \le 2$ — we know $d(v_1v_2)$. For brevity we will not describe the general update of the topological defect. We do describe a special case of particular importance. Suppose that $e \in E^B$ and that $v_1 \neq v_2$ are boundary vertices with $d(v_i) \in \mathcal{IT}(0, \varnothing, B_i)$, where $|B_i| = 2$. This is the case when each v_i is a marked point or a boundary node which is not a contracted component. Write $B_i = \{\tilde{h}_i, a_i\}$, where $\tilde{h}_i$ is as above. Suppose $h_2 \in H^B$, that is, its orientation disagrees with the orientation of the boundary. Then $d(v_1v_2) \in \mathcal{IT}(0, \varnothing, \{a, a_1, a_2\})$, where a is the new cycle of s_0h_2 obtained from concatenating $\tilde{h}_1$ and $\tilde{h}_2$ after erasing h_1 and h_2, and $d(v_1v_2)$ is the element which corresponds to cyclic order $a \to a_1 \to a_2$.

Suppose $E' = \{e_1, \dots, e_r\} \subseteq E$. Then there is an identification between $E(G) \setminus E'$ and $E(\partial_{e_1,\dots,e_r} G)$. Throughout this paper we shall use this identification without further comment.

Figure 12 illustrates several examples of edge contractions.

For a stable open ribbon graph G, we define the orbifold cell complex $\overline{\mathcal{M}}^{\mathbb{R}}_G$ as the cell complex whose cells are $\mathcal{M}^{\mathbb{R}}_{G'}$ for all graphs G' obtained from G by edge contractions. The cell $\mathcal{M}^{\mathbb{R}}_G$ which corresponds to contracting the empty subset of $E(G)$ is included. If G' and G'' are two such cells, and G'' is obtained from G' by contracting the edges $\{e_1, \dots, e_r\}$, then the corresponding cell $\mathcal{M}^{\mathbb{R}}_{G''}$ is the boundary of the cell $\mathcal{M}^{\mathbb{R}}_{G'}$ glued to it along $\ell_{e_1} = \cdots = \ell_{e_r} = 0$. In this case we say that $\mathcal{M}^{\mathbb{R}}_{G''}$ is a face of $\mathcal{M}^{\mathbb{R}}_{G'}$. Write $\overline{\mathcal{M}}^{\mathbb{R}\,\mathrm{comb}}_{g,k,l} = \coprod \overline{\mathcal{M}}^{\mathbb{R}}_G/\sim \,= \coprod \mathcal{M}^{\mathbb{R}}_G$, where the union is over all open (g,k,l)–ribbon graphs, and $\sim$ is induced by the canonical injections $\overline{\mathcal{M}}^{\mathbb{R}}_{G'} \hookrightarrow \overline{\mathcal{M}}^{\mathbb{R}}_G$ over pairs (G, G') where G' is obtained from G by edge contractions. Write $\mathcal{M}^{\mathbb{R}\,\mathrm{comb}}_{g,k,l}$ for the locus which is the union over smooth graphs. Define $\overline{\mathcal{M}}^{\mathbb{R}}_G(\boldsymbol{p})$, $\overline{\mathcal{M}}^{\mathbb{R}\,\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$ and $\mathcal{M}^{\mathbb{R}\,\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$ by restricting perimeters to be p_i. In these cases we also define the cell attachments using edge contractions.

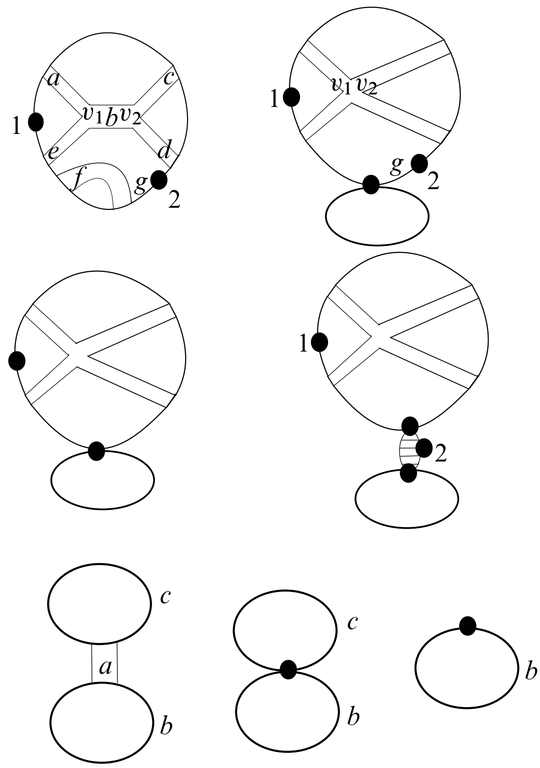


Figure 12: Examples of edge contractions. Contracting the internal edges b and f of the smooth graph on the top left gives rise to the nodal graph on the top right. The vertex $v_1 v_2$ corresponds to the permutation $(ae)(cd)$. By further contracting the boundary edge g between the boundary node and the marked point 2, we obtain the graph on the left in the middle row. The boundary node there corresponds to a contracted component which contains two nodes and the marking 2. The graph on the right-hand side of the same row is equivalent to the left one, only that the ghost is illustrated and there the cyclic order of half-nodes is seen. At bottom left a genus 1 ribbon graph is drawn. After contracting the edge a we obtain a nodal graph. Further contracting c, we obtain the graph on its right, which contains an open contracted component. The genus defect of the contracted component is 1 and its topological defect is that of a cylinder with one special boundary point: the node.

The pointwise maps $\mathrm{comb}^{\mathbb{R}}$ induce moduli maps

$$\mathrm{comb}^{\mathbb{R}} \colon \overline{\mathcal{M}}^{\mathbb{R}}{}_{g,k,l} \times \mathbb{R}_+^l \to \overline{\mathcal{M}}^{\mathbb{R}}{}^{\mathrm{comb}}_{g,k,l} \quad \text{and} \quad \mathrm{comb}^{\mathbb{R}}_{\boldsymbol{p}} \colon \overline{\mathcal{M}}^{\mathbb{R}}{}_{g,k,l} \to \overline{\mathcal{M}}^{\mathbb{R}}{}^{\mathrm{comb}}_{g,k,l}(\boldsymbol{p}),$$

which send a stable open surface and a set of perimeters to the corresponding graph.

Lemma 4.29 *$\overline{\mathcal{M}}^{\mathbb{R}\,\mathrm{comb}}_{g,k,l}$ with the cell structure defined above is a piecewise smooth Hausdorff orbifold with corners. This is the finest topology on the moduli of (g,k,l)–graphs such that the map $\mathrm{comb}^{\mathbb{R}}$ is continuous. $\overline{\mathcal{M}}^{\mathbb{R}\,\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$ is compact for any $\boldsymbol{p}$. We have $\mathrm{comb}^{\mathbb{R}}\colon \mathcal{M}^{\mathbb{R}}_{g,k,l}\times\mathbb{R}^{l}_{+}\simeq\mathcal{M}^{\mathbb{R}\,\mathrm{comb}}$. Moreover, the analogous claims remain true if we declare some, but not all, of the internal marked points to have zero perimeter. In fact, for any effective dual graph Γ, the map $\mathrm{comb}^{\mathbb{R}}$ restricted to $\mathcal{M}^{\mathbb{R}}_{\Gamma}\times\mathbb{R}^{l}_{+}$ is an isomorphism onto its image.*

The proof is similar to the closed case; see [42; 28] for a proof of the analogous theorem.

4.3 JS stratification for the graded moduli

4.3.1 Graded ribbon graphs For a metric, open or closed ribbon graph, (G,ℓ), write

$$\widetilde{Z}_{G,\ell}=\pi_0(\widetilde{\mathrm{For}}^{-1}_{\mathrm{spin}}((\mathrm{comb}^{\mathbb{R}})^{-1}(G,\ell))\quad\text{and}\quad Z_{G,\ell}=\pi_0(\mathrm{For}^{-1}_{\mathrm{spin}}((\mathrm{comb}^{\mathbb{R}})^{-1}(G,\ell)),$$

where the maps $\widetilde{\mathrm{For}}_{\mathrm{spin}}$ and $\mathrm{For}_{\mathrm{spin}}$ are defined in Notation 2.41. For any two generic metrics ℓ and ℓ', the sets $Z_{G,\ell}$ and $Z_{G,\ell'}$ are isomorphic; see Remark 2.38. When G has nontrivial automorphisms the sets are noncanonically isomorphic. For any G, let Z_G be the set $Z_{G,\ell}$ for a fixed generic ℓ. Define $\widetilde{Z}_G$ similarly.

Definition 4.30 A *metric spin ribbon graph with a lifting* (G,z,ℓ) is a metric ribbon graph together with an element $z\in\widetilde{Z}_{G,\ell}$. The graph is called *graded* when $z\in Z_{G,\ell}$. A graded graph is a pair (G,z), $z\in Z_G$. Similarly, in the closed setting, a *metric spin ribbon graph* (G,z,ℓ) is a metric ribbon graph together with $z\in\widetilde{Z}_{G,\ell}$.

The *normalization* $\mathrm{Norm}(G,z,\ell)$ of (G,z,ℓ) is the smooth, not necessarily connected graph $\coprod(G_i,\ell_i,z_i)$, where the (G_i,ℓ_i) are the components of $\mathrm{Norm}(G,\ell)$, and the $z_i\in\widetilde{Z}_{G_i,\ell_i}$ are induced from z by Proposition 2.19. A half-node is *legal* if it is legal as a marked point in the graded structure of $\mathrm{Norm}(G,z)$.

By Proposition 4.25, a graded surface, together with perimeters $\{p_i\}_{i\in\mathcal{I}}$, defines a unique graded metric graph (G,z,ℓ), where (G,ℓ) is embedded in $K_{\mathcal{B},\mathcal{P}_0}(\mathrm{For}_{\mathrm{spin}}(\Sigma))$, as in Proposition 4.25, and z is the class of graded spin structures which contains the graded structure of Σ. When (G,ℓ) is generic and effective, all possible automorphisms of (G,ℓ) leave all half-edges in place, and may only act nontrivially on isolated contracted components, which are of genus 0. Thus, the action of this automorphism group on $Z_G = Z_{G,\ell}$ is trivial, and hence, in this case, Z_G is isomorphic to $\mathrm{Spin}(\Sigma)$, and any element z of it corresponds to a unique graded structure.

Moreover, by Corollary 2.22, if in addition G has no contracted boundaries, then Z_G is in one-to-one correspondence with isomorphism classes of tuples $(\mathbb{S}_1, \dots, \mathbb{S}_r)$, where each $\mathbb{S}_i$ is a spin structure with a lifting on the i^{th} component of $\mathrm{Norm}(\Sigma)$ such that all original boundary marked points are legal and for any boundary node of Σ exactly one half is legal.

Definition 4.31 A spin ribbon graph with a lifting (G, z), with or without a metric ℓ, is called *effective* if G is effective, and z is a spin structure with a lifting in which for every contracted component $v \in V(G)$, all boundary marked points of the isolated component in $\mathrm{Norm}^{-1}(v)$ are legal. In the case that v is not isolated, it is equivalent to all half-nodes in $(N^B)^{-1}(v)$ being illegal. An effective graded graph (G, z) is *trivalent* if G is trivalent. The graph is *smooth* if its underlying graph is. These definitions extend to the closed case, without the assumptions on boundary nodes.

Denote by $\mathcal{SR}^0$ the set of isomorphism classes of graded smooth trivalent ribbon graphs, and write $\mathcal{R}^0$ for the set of their underlying open ribbon graphs. Denote by $\mathcal{SR}^0_{g,k,l} \subseteq \mathcal{SR}^0$ the subset whose faces are marked $[l]$ and whose boundary points are marked by $[k]$. Define $\mathcal{R}^0_{g,k,l}$ similarly.

Let $\mathcal{OSR}^0_{g,k,l}$ be the collection of all graphs in $\mathcal{SR}^0_{g,k,l}$ with an odd number of boundary marked points on each boundary component. Define $\mathcal{OR}^0_{g,k,l}$ similarly.

Note that in a trivalent graph, by definition if v is a contracted component, the unique ghost component in $\mathrm{Norm}^{-1}(v)$ has all marked points *legal*.

Recall that smooth graded surfaces have no internal markings of twist 1 or illegal boundary markings. Therefore an immediate corollary of Proposition 2.32, which can be taken as an alternative definition of $\mathcal{R}^0_{g,k,l}$, is:

Corollary 4.32 *$\mathcal{R}^0_{g,k,l} \neq \varnothing$ precisely when $2 \mid g+k-1$. Every trivalent smooth graph satisfying this constraint belongs to $\mathcal{R}^0_{g,k,l}$.*

Notation 4.33 We define the map comb between graded surfaces and graded metric ribbon graphs by

$$\mathrm{comb}(\Sigma, \mathbb{S}, s, \boldsymbol{p}) = (G, z, \ell),$$

where $(G, \ell) = \mathrm{comb}^{\mathbb{R}}(\Sigma, \boldsymbol{p})$ and $z \in Z_{G,\ell}$ is the corresponding class. Write $\mathrm{comb}_{\boldsymbol{p}} = \mathrm{comb}(-,-,-,\boldsymbol{p})$. Write $\mathrm{For}^{\mathrm{comb}}_{\mathrm{spin}}(G, z, \ell) = (G, \ell)$.

Proposition 4.34 *Suppose* $\mathrm{comb}(\Sigma, \boldsymbol{p}) = (G, z, \ell)$.

(a) *Then* $\mathrm{comb}(\mathrm{Norm}(\Sigma), \boldsymbol{p}) = \mathrm{Norm}(G, z, \ell)$, *where preimages of nodes in* Σ *will be internal markings of perimeter zero.*

(b) *Suppose* Σ' *is obtained from* Σ *by forgetting an illegal marked point* x_ν *whose removal makes no component unstable. Suppose that* x_ν *is mapped to vertex* v *of* G. *Write* $(G', z', \ell') = \mathrm{comb}(\Sigma', \boldsymbol{p})$. *Then* (G', ℓ') *is obtained from* (G, z, ℓ) *by the following procedure. If* $\deg(v) = 2$, *and* v *has a zero genus defect and marking defect* $\{\nu\}$, *remove* v *from the graph, unite its two edges* e_1 *and* e_2 *to one edge* e, *define* $\ell'(e) = \ell(e_1) + \ell(e_2)$ *and for the other edges put* $\ell' = \ell$. *Otherwise the graph and metric do not change, but the marking* ν *is removed from the marking defect of* v. *The point* z' *is the image of* z *under the natural map* $Z_{G,\ell} \to Z_{G',\ell'}$ *obtained from Observation 2.43 with* $\mathcal{B}' = \{\nu\}$.

Proof The first item is a consequence of Corollary 4.15. The second follows from Corollary 4.14 and Observation 2.43. □

4.3.2 Combinatorial moduli for graded surfaces, bundles and forms Denote by $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ the set of metric graded (g,k,l)–ribbon graphs. Write $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$ for the subspace of graphs with fixed perimeters $\boldsymbol{p}$. Define $\mathcal{M}^{\mathrm{comb}}_{g,k,l}$ as the subspace of smooth graphs. Define similarly $\mathcal{M}^{\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$. The pointwise maps comb induce moduli maps

$$\mathrm{comb}\colon \overline{\mathcal{M}}_{g,k,l} \times \mathbb{R}^l_+ \to \overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l} \quad \text{and} \quad \mathrm{comb} = \mathrm{comb}_{\boldsymbol{p}}\colon \overline{\mathcal{M}}_{g,k,l} \to \overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}(\boldsymbol{p}),$$

which send a stable graded surface and a set of perimeters to the corresponding graph. Endow these spaces with the finest topology such that comb is continuous.

We now study the cell structure of $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$. Recall that a metric ℓ is generic if the metric graph has no automorphisms. In particular, in the open and connected setting, metrics which give all edges different lengths are generic. For a generic $\ell \in \mathcal{M}^{\mathbb{R}}_G$, choose $z \in Z_G = Z_{G,\ell}$, and define $\mathcal{M}_{(G,z)}$ to be the connected component of $(\mathrm{For}^{\mathrm{comb}}_{\mathrm{spin}})^{-1}(\mathcal{M}^{\mathbb{R}}_G)$ which contains (G, z, ℓ).

The map $\mathrm{For}^{\mathrm{comb}}_{\mathrm{spin}}$ is continuous. Moreover, by the same reasoning as in the noncombinatorial case (see the discussion in the end of Section 2.3.6), it is an orbifold branched cover, and over any $\mathcal{M}^{\mathbb{R}}_G$ it is an orbifold cover.

Thus, $(\mathrm{For}^{\mathrm{comb}}_{\mathrm{spin}})^{-1}(\mathcal{M}^{\mathbb{R}}_G)$ is an orbibundle over $\mathcal{M}^{\mathbb{R}}_G$, with a generic fiber Z_G. Since

$$\mathcal{M}^{\mathbb{R}}_G = \mathbb{R}^{E(G)}_+/\mathrm{Aut}(G),$$

such a bundle must be of the form

$$(\mathrm{For}_{\mathrm{spin}}^{\mathrm{comb}})^{-1}(\mathcal{M}_G^{\mathbb{R}}) \simeq (\mathbb{R}_+^{E(G)} \times Z_G)/\mathrm{Aut}(G)$$

for some action of $\mathrm{Aut}(G)$ which we now explain.

Let $C \subseteq \mathcal{M}_G^{\mathbb{R}}$ be the locus of generic metrics, and $\overline{C} \subseteq \mathbb{R}^{E(G)}$ its preimage under the quotient by $\mathrm{Aut}(G)$. Except from some borderline cases, which can be treated separately, its complement is of real codimension at least 3. Over C the fiber of the bundle is always of size $|Z_G|$. Denote this fiber bundle by E, and let $\overline{E} \to \overline{C}$ be its pullback to $\overline{C}$. Now $\pi_1(\overline{C})$ is trivial, as $\mathbb{R}^{E(G)} \setminus \overline{C}$ is of codimension at least 3. Thus $\overline{E}$ must be trivial, and is hence isomorphic to $\overline{C} \times Z_G$.

Let $\overline{\ell} \in \overline{C}$ be any point, and let ℓ be its image in C. Recall that, as an orbispace, $\mathrm{Aut}(G) \simeq \pi_1(\overline{C}/\mathrm{Aut}(G), \ell)$, and this isomorphism can be made explicit as follows: for $g \in \mathrm{Aut}(G)$, choose any path $\overline{\gamma}^g : [0,1] \to \overline{C}$ with $\overline{\gamma}_0^g = \overline{\ell} \in \mathbb{R}_+^{E(G)}$ and $\overline{\gamma}_1^g = g \cdot \overline{\ell}$, and set γ_g to be its $\overline{\gamma}^g$ to C.

Parallel transport $z = z_0$ along γ^g to get z_1. This can be done as the fiber is zero-dimensional. Define $g \cdot (\overline{\ell}, z) = (g \cdot \overline{\ell}, z_1)$. This action is independent of choices, and can be defined continuously over all $\overline{E}$. This gives us the orbibundle structure over C. Again by continuity, it can be uniquely extended to an action on $\mathbb{R}_+^{E(G)} \times Z_G$.

In particular, we have defined an action of $\mathrm{Aut}(G)$ on Z_G. Define the group $\mathrm{Aut}(G, z)$ as the subgroup of $\mathrm{Aut}(G)$ which leaves z invariant. Then $\mathcal{M}_{(G,z)} \simeq \mathbb{R}_+^{E(G)}/\mathrm{Aut}(G, z)$. Define $\mathcal{M}_{(G,z)}(\boldsymbol{p})$ as the subspace where the perimeters are $\boldsymbol{p}$.

For $e \in E(G)$, define the *edge contraction* to be $\partial_e(G, z) = (\partial_e G, \partial_e z)$, where $\partial_e z \in Z_{\partial_e G}$ using the cell structure of $\overline{\mathcal{M}}_G^{\mathbb{R}}$ and the topology of $\overline{\mathcal{M}}_{g,k,l}^{\mathrm{comb}}$. Explicitly, fix $\boldsymbol{p}$ and take an arbitrary continuous path $([\Sigma_t])_{t \in [0,1]} \subset \overline{\mathcal{M}}_{g,k,l}$ so that $\mathrm{comb}([\Sigma_t]) \in \mathcal{M}_{(G,z)}$ for $t > 0$ and $\mathrm{For}_{\mathrm{spin}}(\mathrm{comb}([\Sigma_0])) \in \mathcal{M}_{\partial_e G}^{\mathbb{R}}$. Suppose that $\mathrm{comb}([\Sigma_0]) \in \mathcal{M}_{(\partial_e G, z')}$. Then $z' = \partial_e z$, and this definition is easily seen to be independent of choices.

An explicit combinatorial description for the special case of trivalent graphs appears in Section 5.1.2.

As in the spinless case $\overline{\mathcal{M}}_{(G,z)}$, the closure of $\mathcal{M}_{(G,z)}$ in $\overline{\mathcal{M}}_{g,k,l}^{\mathrm{comb}}$, is the union of cells $\mathcal{M}_{(G',z')}$ where (G', z') is obtained from (G, z) by edge contractions, and the attachment of the cells is also defined via the edge contractions, ie $\mathcal{M}_{(G',z')}$ is glued to $\mathcal{M}_{(G,z)}$ along $\ell_{e_1} = \cdots = \ell_{e_r} = 0$, where $e_1, \dots, e_r$ are the edges of G which are contracted to obtain G'. In this case we say that $\mathcal{M}_{(G',z')}$ is a face of $\mathcal{M}_{(G,z)}$.

We similarly define $\overline{\mathcal{M}}_{(G,z)}(\boldsymbol{p})$. Again as in the spinless case we can now define the orbifold cell complex structure on $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$, as

$$\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l} = \coprod \overline{\mathcal{M}}_{(G,z)}/\sim\, = \coprod \mathcal{M}_{(G,z)},$$

where the union is over all connected components which correspond to graded (g,k,l)–ribbon graphs, and $\sim$ is induced by edge contractions. We similarly define the orbifold cell complex structure on $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$. In both cases the cell structure agrees with the topology. Denote the quotient-by-$\sim$ map by Ξ.

A graph (G,z) corresponds to a boundary stratum of $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$, that is $\mathcal{M}_{(G,z)} \subseteq \mathrm{comb}(\partial\overline{\mathcal{M}}_{g,k,l} \times \mathbb{R}^l_+)$ if and only if it has at least one boundary node or contracted boundary. In this case we call it a *boundary graph*. All of the above constructions extend to the setting of spin ribbon graphs with a lifting, and to (closed) spin ribbon graphs.

Lemma 4.35 *Suppose* $2 \mid g+k-1$. *Then* $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ *and* $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$ *are piecewise smooth Hausdorff orbifolds with corners, and the latter is compact.*

The maps comb *and* $\mathrm{comb}_{\boldsymbol{p}}$ *are isomorphisms onto their images when restricted to the open dense subsets* $\mathcal{M}_{g,k,l} \times \mathbb{R}^l_+$ *and* $\mathcal{M}_{g,k,l}$.

The map $\mathrm{comb}_{\boldsymbol{p}}$ *induces an orientation on* $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$, *and* $\deg(\mathrm{comb}_{\boldsymbol{p}}) = 1$ *with this orientation.*

Analogous claims are true if we declare some, but not all, of the internal marked points to have perimeter zero. Analogous claims are also true if we allow some internal markings to be Ramond or if we consider the case of closed (twisted) spin surfaces. In addition, for an effective dual spin graph with a lifting Γ, *the maps* comb *and* $\mathrm{comb}_{\boldsymbol{p}}$ *restricted to* $\mathcal{M}_\Gamma \times \mathbb{R}^l_+$ *and* $\mathcal{M}_\Gamma$ *are isomorphisms onto their images.*

The proof is similar to the closed case and will be omitted. The orientation on $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ will be constructed explicitly later.

The combinatorial S^1–bundles $\mathcal{F}_i$ for $i \in [l]$ are defined as in Definition 4.9. Again these carry a natural piecewise smooth structure, compatible with the natural piecewise smooth structures on $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$. The forms α_i, ω_i, $\overline{\alpha}_i$, $\overline{\omega}_i$ and $\overline{\omega}$ are defined as in Definition 4.9 and equation (23).

Definition 4.36 Let $S \subseteq \mathbb{N}$ be a finite set. An (S,l)–set L is a function $L\colon S \to [l]$. We write $S = \mathrm{Dom}(L)$. In the case that $S = [d]$, we simply write it as (d,l)–set. We say that L is an l–set if the set S is understood from the context.

Given two l–sets L and L', we write

$$L' \subseteq L,$$

and say that L' is a subset of L, writing $L' \subseteq L$, if

$$\mathrm{Dom}(L') \subseteq \mathrm{Dom}(L) \quad \text{and} \quad L|_{\mathrm{Dom}(S')} = L'.$$

In this case we define the l–set $L \setminus L'$ by

$$L \setminus L'\colon \mathrm{Dom}(L) \setminus \mathrm{Dom}(L') \to [l], \quad (L \setminus L')(s) = L(s).$$

In the case that $j \in \mathrm{Dom}(L)$, we write $j \in L$. For $i \in [l]$ we put

$$L_i = L^{-1}(i).$$

The (S,l)–sets will be used to encode direct sums of tautological lines as follows.

Notation 4.37 Recall Construction–Notation 1. To any (S,l)–set L we associate a vector bundle E_L and a sphere bundle S_L given by

$$E_L = \sum_{i \in S} \mathbb{L}_{L(i)} \to \overline{\mathcal{M}}_{g,k,l} \quad \text{and} \quad S_L = S((\mathcal{F}_{L(i)})_{i \in S}).$$

We will also consider the sphere bundle $S(E_L)$ associated to E_L.

Define an angular form Φ_L for S_L by formula (19), and using Kontsevich's forms for the copy $\mathcal{F}_{L(i)}$ of the $L(i)^{\text{th}}$ S^1–bundle. Explicitly,

$$\Phi_L(\{r_i\}_{i \in S}, \{\widehat{\alpha}_i\}_{i \in S}, \{\widehat{\omega}_i\}_{i \in S}) \\ = \sum_{k=0}^{|S|-1} 2^k k! \sum_{i \in S} r_i^2 \widehat{\alpha}_i \sum_{\substack{I \subseteq S \setminus \{i\} \\ |I| = k}} \bigwedge_{j \in I} (r_j\, dr_j \wedge \widehat{\alpha}_j) \wedge \bigwedge_{h \notin I \cup \{i\}} \widehat{\omega}_h,$$

where $\widehat{\omega}_i$ is Kontsevich's two-form $\omega_{L(i)}$ and $\widehat{\alpha}_i$ is a *copy* of Kontsevich's one-form $\alpha_{L(i)}$. We refer to it as a copy since, for $i_1, i_2 \in L_j$, both $\widehat{\alpha}_{i_1}$ and $\widehat{\alpha}_{i_2}$ are given by the same formula of the angular 1–form of $\mathcal{F}_j$, but with different ϕ variables. Write

$$\omega_L = -d\Phi_L = \bigwedge_{i \in S} \omega_{L(i)}, \quad p^{2L} = \prod_{i \in S} p^2_{L(i)}, \quad \overline{\omega}_L = \boldsymbol{p}^{2L} \omega_L, \quad \overline{\Phi}_L = \boldsymbol{p}^{2L} \Phi_L.$$

When $S \neq [d]$ we will sometimes omit the assumption that $\sum_{i\in S} r_i^2 = 1$, and then $-d\Phi_L$ gets a correction; see Remark 3.5.

When it is not clear from context, we write α_j^G to indicate the specific graph G. The same remark goes for the other forms.

Exactly as in the closed case, we have:

Lemma 4.38 (a) *For $i \in [l]$, there is a canonical isomorphism* $\mathrm{comb}^*\mathcal{F}_i \simeq S^1(\mathbb{L}_i)$. *As a result,* $\mathrm{comb}^* S_L \simeq S(E_L)$ *canonically.*

(b) *The forms α_i and ω_i are a piecewise smooth angular one-form and Euler two-form for $S^1(\mathbb{L}_i)$. Φ_L is an angular form of S_L, and ω_L is its Euler form.*

(c) *For $(G,z) \in \mathcal{SR}^0_{g,k,l}$, there is a canonical identification*

$$(\mathcal{F}_i \to \overline{\mathcal{M}}_{(G,z)}) \simeq \Xi^*(\mathcal{F}_i \to \overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}).$$

Similarly for the bundles S_L.

Notation 4.39 Recall Proposition 4.34. Let (G,z,ℓ) be a metric spin ribbon graph with a lifting. Define the graph $\widetilde{\mathcal{B}}(G,z,\ell) = (\widetilde{\mathcal{B}}G, \widetilde{\mathcal{B}}z, \widetilde{\mathcal{B}}\ell)$ by first taking the normalization of (G,z,ℓ), and then forgetting isolated components, the lifting data in contracted boundaries, and the new illegal marked points. Let $\widetilde{\mathcal{B}}\colon \mathcal{M}_{(G,z)} \to \mathcal{M}_{(\widetilde{\mathcal{B}}G,\widetilde{\mathcal{B}}z)}$ be the induced map on the moduli.

Observation 4.40 *For any spin ribbon graph with a lifting (G,z), and face marked i, we have $\mathcal{F}_i \to \mathcal{M}_{(G,z)} \simeq \widetilde{\mathcal{B}}^*(\mathcal{F}_i \to \mathcal{M}_{\widetilde{\mathcal{B}}(G,z)})$ canonically, and a similar claim holds for S_L.*

The observation follows from the natural identification of the boundary of the i^{th} faces in G and $\widetilde{\mathcal{B}}G$.

Proposition 4.41 *A special canonical multisection s of $S(E_L)$ is a pullback of a multisection s' of S_L.*

Proof Take $\mathcal{M}_\Gamma \subseteq \partial\overline{\mathcal{M}}_{g,k,l}$ and let $i_1,\ldots,i_r$ be labels of internal tails, one for each vertex of Γ. Now

$$\mathrm{comb}(\mathcal{M}_\Gamma \times \mathbb{R}_+^l) = \coprod_{(G,z)} \mathcal{M}_{(G,z)},$$

where the union is taken over some graded graphs (G, z). Consider one of them; denote it by (G, z). Write

$$\Phi_\Gamma = \prod_{j=1}^{r} \Phi_{\Gamma,i}.$$

The diagram

$$\begin{tikzcd} \mathrm{comb}^{-1}\mathcal{M}_{(G,z)} \arrow[r, "\Phi_\Gamma"] \arrow[d, "\mathrm{comb}"] & \mathrm{comb}^{-1}\mathcal{M}_{(\widetilde{\mathcal{B}}G,\widetilde{\mathcal{B}}z)} \arrow[d, "\mathrm{comb}"] \\ \mathcal{M}_{(G,z)} \arrow[r, "\widetilde{\mathcal{B}}"] & \mathcal{M}_{(\widetilde{\mathcal{B}}G,\widetilde{\mathcal{B}}z)} \end{tikzcd} \tag{24}$$

commutes, by Proposition 4.34. Now $(\widetilde{\mathcal{B}}G, \widetilde{\mathcal{B}}z)$ is smooth, hence the right vertical arrow is an isomorphism, by Lemma 4.35. A special canonical multisection over $\mathcal{M}_\Gamma \times \mathbb{R}_+^l$ is pulled back via Φ_Γ, from $S(E_L) \to \prod_{j=1}^r \mathcal{M}_{v_i^*(\Gamma)} \times \mathbb{R}_+^l$. Let s be special canonical; we now construct s' with $s = \mathrm{comb}^* s'$. Write $s|_{\mathrm{comb}^{-1}\mathcal{M}_{(G,z)}} = \Phi_\Gamma^*(\mathrm{comb}^*(s''))$, where s'' is a multisection of $S_L \to \mathcal{M}_{(\widetilde{\mathcal{B}}G,\widetilde{\mathcal{B}}z)}$. Define $s'|_{\mathcal{M}_{(G,z)}} = \widetilde{\mathcal{B}}^* s''$. These multisections for different strata evidently glue. $\square$

Definition 4.42 A *special canonical multisection* of $S_L \to \overline{\mathcal{M}}_{g,k,l}^{\mathrm{comb}}$ is a multisection s with $\mathrm{comb}^* s$ special canonical. A *special canonical multisection* of $S_L \to \overline{\mathcal{M}}_{(G,z)}$ is a Ξ–pullback of a special canonical multisection on $\overline{\mathcal{M}}_{g,k,l}^{\mathrm{comb}}$. Write $s^{(G,z)}$ for the restriction of s to $\overline{\mathcal{M}}_{(G,z)}$.

The proof of the proposition yields the following immediate corollary.

Corollary 4.43 *Suppose (G, z) is a boundary (g, k, l)–graded ribbon graph, and s is a special canonical multisection of S_L, where L is a (d, l)–set, restricted to the boundary cell $\mathcal{M}_{(G,z)}$. Then $s = \widetilde{\mathcal{B}}^* s'$, where s' is a multisection of $S_L \to \mathcal{M}_{\widetilde{\mathcal{B}}(G,z)}$.*

The main result of this section is that the descendents can be calculated over the combinatorial moduli.

Lemma 4.44 *Let s be a special canonical multisection for $S(E_L)$. Denote by s' the multisection on S_L with $s = \mathrm{comb}^* s'$. Then*

$$\int_{\overline{\mathcal{M}}_{g,k,l}} e(S(E_L), s) = \int_{\overline{\mathcal{M}}_{g,k,l}^{\mathrm{comb}}(\boldsymbol{p})} e(S_L, s'). \tag{25}$$

The orientations are those induced on the combinatorial moduli by comb_*.

The proof is an immediate consequence of Lemmas 4.35 and 4.38 and Observation 2.49.

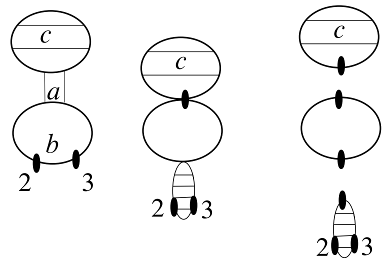


Figure 13: Bridges and their contractions. On the left, three compatible bridges are drawn, a, b and c. In the center, b and c are contracted, and on the right, the normalization is presented. If h_b is the boundary half-edge which corresponds to b, then $\partial_b h$ corresponds to the half-node in the ghost component of the normalization. If h_1 and h_2 are the half-edges of c, then $\partial_c h_1$ and $\partial_c h_2$ are the two half-nodes in the normalization of the node which corresponds to c.

4.3.3 Intersection numbers as integrals over the combinatorial moduli We can now use the natural piecewise linear structure on $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ and the associated bundles to write an explicit integral formula for them.

Definition 4.45 A *boundary loop* in a graded graph (G,z) is a boundary edge which is a loop. We denote the collection of these elements by $\mathrm{Loop}(G)$. A *bridge* in a graded graph (G,z) is either a boundary edge between two distinct special legal boundary points or an internal edge between two boundary vertices; see Figure 13 and the left-hand sides of Figure 14 rows four and five. Denote by $\mathrm{Br}(G,z)$ the set of bridges of (G,z). Usually we shall omit z from the notation and write $\mathrm{Br}(G)$ instead. A *compatible sequence of bridges* $\{e_1,\dots,e_r\}$ is a sequence of bridges such that e_{i+1} is a bridge in $\partial_{e_1,\dots,e_i} G$ for all i.

Suppose e is a bridge and $h \in H^I$ satisfies $h/s_1 = e$. Set $h' = s_2 h$. We define $\partial_e h \in \mathrm{HN}(\partial_e G)$ (recall HN was defined in Definition 4.18) to be the unique vertex $v \in V(\mathrm{Norm}(\partial_e G))$ with $h'/s_0 = v$, where we consider h' as an edge of $\mathrm{Norm}(\partial_e G)$, using the canonical identification; see Figure 13 and the right-hand sides of the fourth and fifth rows of Figure 14. When there is $h \in H^B$ with $h/s_1 = e$, contracting e creates a contracted component v, which is identified with a ghost component of $\mathrm{Norm}(G)$; see Figure 13 and row four of Figure 14 again. We denote by $\partial_e h \in B(v)$ the marking which is the s_0–cycle of $s_2(s_1 h)$ in $(N^B)^{-1}(v)$. This is equivalent to writing $\partial_e h = s_1 \partial_e (s_1 h)$, recalling Notation 4.28.

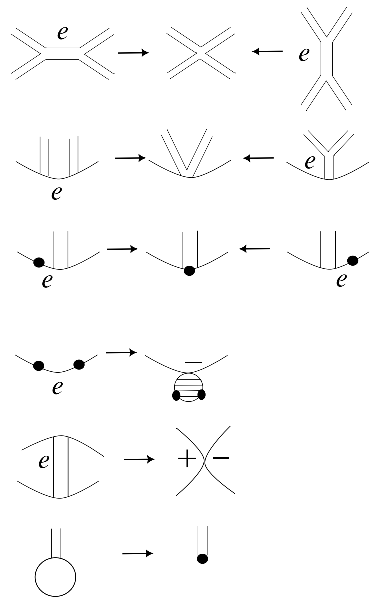


Figure 14: Edge contractions and Feynman moves. In rows four and five, bridge contractions are presented. In the bottom row, a boundary contraction is shown. In the first three rows it is shown how the other types of contracted edges can be obtained as the result of two different contractions.

The following observation is immediate.

Observation 4.46 (a) *We have that* $\dim \mathcal{M}_{(G,z)}(\boldsymbol{p}) = \dim \overline{\mathcal{M}}_{g,k,l}$ *if and only if* $(G,z) \in \mathcal{SR}^{\mathbf{0}}_{g,k,l}$.

(b) *In addition*, (G,z) *is a boundary graph if and only if it can be represented as* $\partial_{e_1,\dots,e_r}(G',z')$, *where* $(G',z') \in \mathcal{SR}^{\mathbf{0}}_{g,k,l}$ *and at least one* e_i *is a bridge or a loop. The only boundary graphs* (G,z) *whose moduli is of full dimension* $\dim \overline{\mathcal{M}}_{g,k,l} - 1$ *are those which can be written as* $\partial_e(G',z')$ *for* $(G',z') \in \mathcal{SR}^{\mathbf{0}}_{g,k,l}$ *and* $e \in \mathrm{Br}(G') \cup \mathrm{Loop}(G')$.

(c) *If* $\{e_1,\dots,e_r\}$ *is a compatible sequence of bridges in a trivalent graph* (G,z), *then* $\partial_{e_1,\dots,e_r}(G,z)$ *is trivalent. Any trivalent graph can be written in the form* $\partial_{e_1,\dots,e_r}(G,z)$, *where* (G,z) *is smooth trivalent and* $\{e_1,\dots,e_r\}$ *is a compatible sequence of bridges. This representation is unique up to reordering the bridges in the sequence.*

See rows four and five of Figure 14 for examples.

Recall Definition 2.51. Using Observation 4.46, Lemma 4.44 and Proposition 3.3, we immediately get:

Lemma 4.47 *Let L be a (d,l)–set, where $d = \frac{1}{2}(3g-3+k+2l)$, and let s be a special canonical multisection for S_L. Then*

$$2^{\frac{1}{2}(g+k-1)}\langle\tau_{a_1}\cdots\tau_{a_l}\sigma^k\rangle = \sum_{(G,z)\in\mathcal{SR}^0_{g,k,l}}\int_{\mathcal{M}_{(G,z)}(\boldsymbol{p})}\omega_L + \sum_{\substack{(G,z)\in\mathcal{SR}^0_{g,k,l}\\ [e]\in[\mathrm{Br}(G)\cup\mathrm{Loop}(G)]}}\int_{\overline{\mathcal{M}}_{\partial_e(G,z)}(\boldsymbol{p})}s^*\Phi_L.$$

Equivalently,

$$\boldsymbol{p}^{2L}2^{\frac{1}{2}(g+k-1)}\langle\tau_{a_1}\cdots\tau_{a_l}\sigma^k\rangle = \sum_{(G,z)\in\mathcal{SR}^0_{g,k,l}}\int_{\mathcal{M}_{(G,z)}(\boldsymbol{p})}\bar{\omega}_L + \sum_{\substack{(G,z)\in\mathcal{SR}^0_{g,k,l}\\ [e]\in[\mathrm{Br}(G)\cup\mathrm{Loop}(G)]}}\int_{\overline{\mathcal{M}}_{\partial_e(G,z)}(\boldsymbol{p})}s^*\bar{\Phi}_L.$$

The orientations are those induced on the combinatorial moduli by comb_*.

Remark 4.48 The formalism of piecewise linear forms and their integration is treated, for instance, in [42].

Construction–Notation 3 For later purposes we now define *Feynman moves* in edges. Suppose that G is a trivalent graph, and let $e \in E \setminus \mathrm{Br}(G)$. If e is a boundary edge, we require that least one of its vertices is not a special point. If e is a boundary loop, define the graph $G_e := G$. Otherwise, define G_e as the graph obtained from G by first contracting e and then reopening it in the unique different possible way; see the first three rows of Figure 14.

Let (G,z) be a graded trivalent graph. For a boundary loop e define the graded structure $z_e \in Z_G$ as the graded structure which is identical to z except that the lifting on the boundary component e is opposite. For an edge $e \notin \mathrm{Br}(G)\cup\mathrm{Loop}(G)$, write $z_e \in Z_{G_e}$ for the graded structure on G_e, defined by the following proposition.

Proposition 4.49 *For (G,z) and e as above, there is a unique graded structure z_e such that if G is smooth, $\mathcal{M}_{(G_e,z_e)}$ is the unique codimension-zero cell of $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ adjacent to $\mathcal{M}_{(G,z)}$ along $\mathcal{M}_{\partial_e(G,z)}$. For nonsmooth G, write $(G,z) = \partial_{e_1,\dots,e_r}(H,w)$, where $e_1,\dots,e_r \in E(H)$, with (H,w) trivalent and smooth. Then*

$$(G_e,z_e) = \partial_{e_1,\dots,e_r}(H_e,w_e).$$

Proof For a smooth trivalent G and an edge e, $\partial_e\mathcal{M}_{(G,z)}$ is a codimension-one face; hence, since $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ is an orbifold with corners, this face must be adjacent to at most one additional codimension-zero cell. Since e is neither a boundary loop nor a bridge, this face is not contained in the boundary of the moduli; hence it is adjacent to two codimension-zero cells. Since $\mathrm{For}^{\mathrm{comb}}_{\mathrm{spin}}$ is continuous, this cell must be of the form $\mathcal{M}_{(G,z_e)}$ for some graded structure $z_e \in Z_G \setminus z$, or of the form $\mathcal{M}_{(G_e,z_e)}$ for $z_e \in Z(G_e)$. The map $\mathcal{M}^{\mathrm{comb}}_{g,k,l} \simeq \mathcal{M}_{g,k,l} \to \mathcal{M}^{\mathbb{R}}_{g,k,l} \simeq \mathcal{M}^{\mathbb{R},\mathrm{comb}}_{g,k,l}$, when restricted to the open dense set of generic metrics, is a (nonbranched) covering map, as there are no automorphisms to the objects, and since the neighboring cell in $\mathcal{M}^{\mathbb{R},\mathrm{comb}}_{g,k,l}$ to $\mathcal{M}^{\mathbb{R}}_{G}$ along $\partial_e\mathcal{M}^{\mathbb{R}}_{G}$ is $\mathcal{M}^{\mathbb{R}}_{G_e}$, the neighboring cell of $\mathcal{M}_{(G,z)}$ along the boundary $\partial_e\mathcal{M}_{(G,z)}$ must be $\mathcal{M}_{(G_e,z_e)}$. The rest of the claim follows from the cell structure and Observation 4.46(c). $\square$

The operations $G \to G_e$ and $(G,z) \to (G_e,z_e)$ are called *Feynman moves*.

5 Trivalent and critical nodal graphs

It follows from Lemma 4.47 that all intersection numbers can be calculated as integrals over the highest-dimensional cells of $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$, and of $\partial\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$. In the next section we will describe an iterative integration formula for the integrals. We will see that the cells that contribute to this iterative process are those which correspond to trivalent graded ribbon graphs. Analyzing their contribution is done by using a new type of graph, which we define below and name *critical nodal graphs*. It turns out that both for trivalent graded graphs, and for critical nodal graphs, the extra data of the graded spin structure can be described in an explicit combinatorial manner. In this section we shall provide this combinatorial interpretation, use it to describe the boundary conditions and to write an explicit expression for the canonical orientations.

5.1 Kasteleyn orientations

From here until the end of this subsection fix a graph $G \in \mathcal{R}^0_{g,k,l}$, where $\mathcal{R}^0_{g,k,l}$ was defined in Definition 4.31.

Definition 5.1 Consider the set $\mathcal{A}$ of all assignments $H^I \to \mathbb{Z}_2$. A *vertex flip* is the involution $f_v\colon \mathcal{A} \to \mathcal{A}$ defined as follows. For $A \in \mathcal{A}$, f_vA is the assignment which satisfies the following condition: $f_vA(h) \neq A(h)$ if and only if exactly one of the vertices of h, h/s_0 and $s_1(h)/s_0$ is v.

A *Kasteleyn orientation on G* is an assignment $K \in \mathcal{A}$ which satisfies the following conditions:

(a) If h belongs to a boundary edge, that is, $s_1 h \in H^B$, then
$$K(h) = 1.$$

(b) For other half-edges h,
$$K(h) + K(s_1(h)) = 1.$$

(c) For every face i,
$$\sum_{h \in H_i} K(h) = 1.$$

For convenience extend K to H^B by 0, so that property (b) holds for any half-edge. $K(G)$ will stand for the set of all Kasteleyn orientations of G. Vertex flips act on the set $K(G)$. Two Kasteleyn orientations are *equivalent* if they differ by vertex flips. Write $[K(G)]$ for the set of equivalence classes of Kasteleyn orientations, and $[K]$ for the equivalence class of K.

Observation 5.2 *Equivalent assignments give the same value to any half-edge of a bridge.*

Definition 5.3 The *legal side* of a bridge e is the half-edge $h \in s_1^{-1}(e)$ with $K(h) = 0$. The other side is *illegal*.

The main goal of this subsection is to show that there is a natural bijection between $\mathcal{SR}^0_{g,k,l}$ and $\{(G, [K]) \mid G \in \mathcal{R}^0_{g,k,l}, [K] \in [K(G)]/\mathrm{Aut}(G)\}$.

We first show how a graded structure induces an element in $[K(G)]$. Take a graded surface $(\Sigma, \mathbb{S}, s)$ whose corresponding embedded ribbon graph, defined by the JS differential, is G.

Definition 5.4 Let $v \in V^I$, and let $\{h_i\}_{i=1,2,3}$ be its three half-edges, ordered so that $s_0 h_i = h_{i+1}$. A *choice of lifting for* v is a choice of lifts $l_{h_i} \in \mathbb{S}_v$ for the *oriented* $T_v^1 h_i$ (see Notation 2.23) such that

$$l_{h_{i+1}} = R_{\theta_i + 2\pi} l_{h_i} \quad \text{for } i = 1, 2, \tag{26}$$

where $\theta_i = \measuredangle(T_v h_i, T_v h_{i+1})$.

Let $\partial\Sigma_b$ be a boundary component. Write $H_b = \{h_i\}_{i=1}^m$, where the $h_i \in H^I$ are the half-edges which are embedded in $\partial\Sigma_b$, ordered so that $h_{i+1} = s_1(s_2^{-1}(s_1(h_i)))$. Put $v_i = h_i/s_0$. A *lifting for* $\partial\Sigma_b$ is the unique choice of lifts $l_h \in \mathbb{S}_{v_i}$ of $T^1_{v_i}h$, for any i and any $h \in H_{v_i}$, satisfying the following requirements:

(a) For $h = h_i \in s_1 H_b$, we have $l_h = s(v_i)$.

(b) If v_i is not a marked point, let $f = s_0 h_i$ and put $\theta = \measuredangle(h_i, f)$. Then
$$l_f = R_{\theta+2\pi} l_{h_i} \quad \text{and} \quad l_{s_0^{-1}h_i} = R_\pi l_{h_i}.$$

(c) If v_i is a marked point, $l_{s_0^{-1}h_i} = R_{3\pi} l_{h_i}$.

A *choice of lifting* is a choice of lifting for any vertex, and a lifting for any boundary component of the graph.

Note that given a choice of lifting in a vertex v, (26) holds also for $i = 3$, since composing (26) for $i = 1, 2, 3$ yields
$$R_{6\pi+\sum_{i=1}^3 \theta_i} l_{h_1} = R_{8\pi} l_{h_1} = l_{h_1},$$
where the first equality follows from $\sum \theta_i = 2\pi$, and the last equality uses that $R_{4\pi}$ is the identity map. This also shows that a choice of a lifting for an internal vertex does not depend on the choice of which half-edge is taken to be h_1. In addition, note that a lifting of a boundary does not depend on choices.

Figure 15 illustrates the three types of liftings described above.

A consequence of the definition of the graded boundary conditions is the following.

Observation 5.5 *Consider a lifting for the boundary $\partial\Sigma_b$. With the above notation, if v_i is a marked point, then $l_{h_i} = R_{2\pi} P(h_{i-1}) l_{h_{i-1}}$. If v_i is a boundary vertex which is not a marked point, then $l_{h_i} = P(h_{i-1}) l_{h_{i-1}}$. In both cases, $R_\pi P(h_{i-1}) l_{h_{i-1}} = l_{s_1(h_{i-1})} = l_{s_0^{-1}h_i}$.*

Remark 5.6 Iterating Observation 5.5 over all boundary vertices, we are led to the single constraint $l_{h_i} = R_{2k_b\pi} l_{h_i}$, where k_b is the number of boundary marked points of the boundary component $\partial\Sigma_b$. By unwinding the alternations in boundary marked points, we see that $q(\gamma) = k_b + 1$ for γ a simple closed path isotopic to $\partial\Sigma_b$.

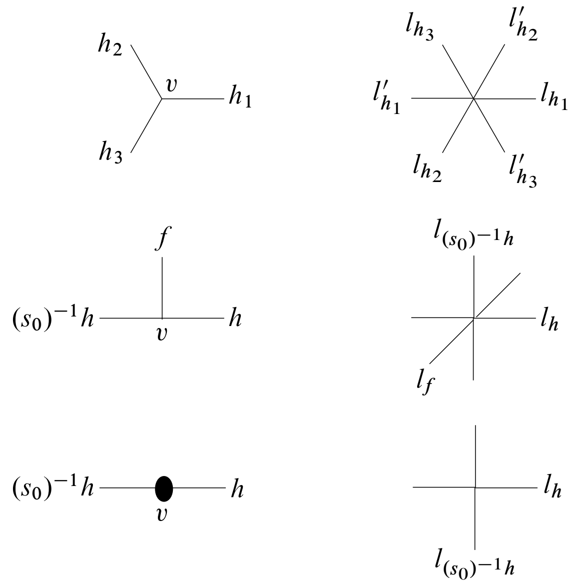


Figure 15: In this figure the three types of liftings from Definition 5.4 are illustrated. The left column represents the local picture at the surface, while the right column represents the corresponding picture at the level of the spin fiber. Each vector on the left-hand side has two preimages on the right-hand side (where the angles between consecutive vectors on the right are half of those from the left). In the top row an internal trivalent vertex v is drawn. For v there are two possible lifts: $\{l_{h_1}, l_{h_2}, l_{h_3}\}$ and $\{l'_{h_1}, l'_{h_2}, l'_{h_3}\}$. In the middle row, v is a trivalent boundary vertex and in the bottom row v is a boundary marked point. In both of these cases the horizontal line in the left column represents the boundary, and in both cases l_h is determined from the data of the grading, so there is no choice in the liftings, and they are as in the figure.

A choice of a lifting induces an assignment $K \in \mathcal{A}$ as follows. $K(h) = 1$ if $s_1 h \in H^B$. For an internal half-edge h, considered as an arc from u to v, we have lifts l_h and $l_{s_1(h)}$ of $T_u^1 h$ and $T_v^1 s_1 h$, respectively. Now, $R_\pi P(h) l_h$ also covers $T_v^1 s_1 h$, hence it equals either $l_{s_1(h)}$ or $R_{2\pi} l_{s_1(h)}$. In the first case we define $K(h) = 1$, otherwise $K(h) = 0$. Write $K(\Sigma, \mathbb{S}, s)$ for the set of all assignments of G induced by choices of liftings.

Definition 5.7 A *vertex lift flip* in a vertex $v \in V^I$ is the involution of the set of choices of lifts which takes one choice to the choice that differs exactly in the lift at v.

Lemma 5.8 *If C and C' are two choices of lift which differ by a vertex lift flip in v, the corresponding assignments K and K' differ by a vertex flip f_v. The vertex flips act* *commutatively freely transitively on $K(\Sigma, \mathbb{S}, s)$. The correspondence between choices of lift and $K(\Sigma, \mathbb{S}, s)$ is a bijection. As a conclusion, $|K(\Sigma, \mathbb{S}, s)| = 2^{|V^I(G)|}$.*

Proof The first assertion as well as the commutativity and transitivity of the action are straightforward. The rest will follow from proving that the action is free. In order to show this, note that we can think of $K(\Sigma, \mathbb{S}, s)$ as subset of $\mathbb{Z}_2^{H^I}$. This is a vector space, and a vertex flip f_v is just an addition of an element $\widetilde{f_v} \in \mathbb{Z}_2^{H^I}$ which is s_1–invariant and zero everywhere except for edges with exactly one of their ends being v. Thus, we can also think of $\widetilde{f_v}$ as a function from E to $\mathbb{Z}_2$ which vanishes identically on boundary edges. In other words, $\widetilde{f_v}$ is canonically a 1–cochain with coefficients in $\mathbb{Z}_2$ relative to boundary. In fact, if δ is the coboundary operator on the relative cochain complex defined on Σ by the 1–skeleton G, then $\widetilde{f_v} = \delta e_v$, where e_v is the 0–cochain which is 1 only at v. If the action of vertex flips were not free, there would be a subset $A \subseteq V^I$ such that

$$\sum_{v \in A} \widetilde{f_v} = 0,$$

or equivalently

$$\delta \sum_{v \in A} e_v = 0,$$

so $\sum_{v \in A} e_v$ would be δ–closed in $H^0(\Sigma, \partial\Sigma) \simeq H_2(\Sigma)^*$, by Poincaré–Lefschetz duality. But $H_2(\Sigma) = 0$, which means $A = \varnothing$. □

We now study $K(\Sigma, \mathbb{S}, s)$ more carefully.

Proposition 5.9 *Fix $K \in K(\Sigma, \mathbb{S}, s)$.*

For $h \in H^I$, put $v = h/s_0$, $u = (s_1 h)/s_0$, $f = s_0^{-1} s_1 h$ and, if u is not a marked point, $f' = s_0 s_1 h$. Write $\theta = \measuredangle(P(h) T_v^1 h, T_u^1 f) \in (-\pi, \pi)$ and $\alpha = \measuredangle(f', f) \in (0, 2\pi)$ if u is not a marked point. Let l_h and l_f denote the lifts of $T_v^1 h$ and $T_u^1 f$, respectively, induced by K, and when u is not a marked point, let $l_{f'}$ be the lift of $T_u^1 f'$. Finally, let $\varepsilon = K(h)$. Then we have the following equalities:

(a) $l_f = R_{2\pi\varepsilon+\theta} P(h) l_h$.

(b) *If u is not a marked point, $l_{f'} = R_{2\pi(1+\varepsilon)+\theta-\alpha} P(h) l_h$ and $\theta - \alpha \in (-\pi, \pi)$.*

For $h \in H^B$ from v to u, write $f = s_2 h$. If u is a marked point, then $R_{2\pi} P(h) l_h = l_f$. If u is not a marked point, write $f' = s_0 s_1 h$ and $\theta = \measuredangle(P(h) T_v^1 h, T_u^1 f') \in (-\pi, 0)$. Then $P(h) l_h = l_f$ and $R_{\theta+2\pi} l_h = l_{f'}$.

Proof We prove it for $h \in H^I$; the proof for boundary half-edges is similar and follows from Observation 5.5. We have

$$\begin{aligned} K(h) = \varepsilon \iff & R_\pi P(h) l_h = R_{(1+\varepsilon)2\pi} l_{s_1(h)} \\ \iff & R_\pi P(h) l_h = R_{(1+\varepsilon)2\pi}(R_{2\pi+\pi-\theta} l_f) \\ \iff & R_\theta P(h) l_h = R_{\varepsilon 2\pi} l_f, \end{aligned}$$

where the equivalence in the second line follows from the definition of a choice of lift in a vertex, while the equivalence in the last line is a consequence of Remark 2.25. The second claim follows from $l_{f'} = R_{-2\pi-\alpha} l_f$ and the cyclic order of the half-edges. □

We now prove:

Lemma 5.10 *If $K \in K(\Sigma, \mathbb{S}, s)$, then K is a Kasteleyn orientation.*

Proof Property (a) of Kasteleyn orientations is just Observation 5.5. Property (b) is reduced, thanks to Remark 2.25 and the construction of K, to

$$R_\pi P(s_1(h)) R_\pi P(h) = R_{2\pi},$$

but this follows from Proposition 2.28 applied to the piecewise smooth closed curve $h \to \bar{h} \to h$, where $\bar{h}$ is h with the opposite orientation.

To show property (c), let $h_1, \dots h_m$ be an ordering of H_i such that $s_2(h_j) = h_{j+1}$. Set $v_j = h_j / s_0$. Let l_{h_j} be the lift of $T^1_{v_j} h_j$ determined by K, using Lemma 5.8. Proposition 2.28 applied to the piecewise smooth curve $\gamma_i = h_1 \to h_2 \to \cdots h_m \to h_1$ is equivalent to $P(\gamma_i) l_{h_1} = R_{2\pi} l_{h_1}$. Put $\theta_{j+1} = \measuredangle\big(P(h_j) T^1_{v_j} h_j, T^1_{v_{j+1}} h_{j+1}\big) \in (-\pi, \pi)$. Now, by Proposition 5.9,

$$R_{\theta_{j+1}} P(h_j) l_{h_j} = R_{\varepsilon_j 2\pi} l_{h_{j+1}}, \quad \text{with } \varepsilon_j \in \mathbb{Z}_2,$$

where $\varepsilon_j = K(h_j)$. Iterating this equation for $j = 1, \dots, m$, we get

$$\begin{aligned} l_{h_1} &= R_{2\pi\varepsilon_m + \theta_1} P(h_m) R_{2\pi\varepsilon_{m-1} + \theta_m} P(h_{m-1}) \cdots R_{2\pi\varepsilon_1 + \theta_2} P(h_1) l_{h_1} \\ &= R_{2\pi \sum_{i=1}^m \varepsilon_i} R_{\theta_1} P(h_m) R_{\theta_m} P(h_{m-1}) \cdots R_{\theta_2} P(h_1) l_{h_1}. \end{aligned}$$

On the other hand, $R_{\theta_1} P(h_m) R_{\theta_m} P(h_{m-1}) \cdots R_{\theta_2} P(h_1) = R_{2\pi(1+q(\gamma_i))}$ by the definition of q. But $q(\gamma_i) = 0$, since γ_i is trivial in the homology of Σ. So the sum $\sum_{i=1}^m \varepsilon_i = \sum_{i=1}^m K(h_i)$ must be odd. □

Theorem 5.11 *Let G and Σ be as above. There is a bijection between* $\mathrm{Spin}(\Sigma)$, *the set of isomorphism classes of graded spin structures on Σ, and $[K(G)]$.*

Proof Given a graded spin structure $(\mathbb{S}, s)$ on Σ, we have constructed an equivalence class of Kasteleyn orientations, and this equivalence class depends only on the isomorphism type of $(\mathbb{S}, s)$, so that we get a map

$$[K]\colon \mathrm{Spin}(\Sigma) \to [K(G)].$$

We shall construct a map Spin in the other direction.

Fix $K \in K(G)$. We first construct the restriction of the spin bundle to G, the 1–skeleton of Σ. For any vertex v, write

$$N_v = \bigcup_i \{h_i'\},$$

where h_i' are the open half-edges emanating from v, after removing their second endpoint. We define $\mathrm{Spin}(K)|_{N_v}$ as the trivial spin cover of $T^1\Sigma|_{N_v}$. On any fiber of $\mathrm{Spin}(K)$ there is an action of $\mathbb{R}/4\pi\mathbb{Z}$; denote it by R_θ.

For a vertex v, choose sections $l_{h_i} : h_i' \to \mathrm{Spin}(K)|_{h_i'}$ which cover $T_v^1 h_i$ so that for any $h_i \notin H^B$,

$$R_{2\pi+\theta_i}(v) l_{h_i}(v) = l_{s_0(h_i)}(v),$$

where $\theta_i = \measuredangle(T_v^1 h_i, T_v^1 s_0(h_i))$.

The transition map $g_{e', s_1(e)'}\colon \mathrm{Spin}(K)|_{e'} \to \mathrm{Spin}(K)|_{s_1(e)'}$ is given by identifying $R_{2K(e)\pi-\pi} l_h$ and $l_{s_1 h}$, and extending using the $\mathbb{R}/4\pi\mathbb{Z}$–action.

It follows from construction and from property (c) of Kasteleyn orientations that for each $i \in [l]$, the spin structure on the boundary of face i of G, which is a topological disk, satisfies Proposition 2.28, and hence can be extended uniquely to the face. Thus, we have constructed a spin structure on Σ. The section $\{l_h\}_{h \in s_1 H^B}$ is evidently a grading. Call this graded spin structure $\mathrm{Spin}(K)$. It can be verified easily that equivalent Kasteleyn orientations give rise to isomorphic graded spin structure, and that the maps $[K]$ and Spin are inverse to each other. □

Knowing now that the data of an equivalence class of Kasteleyn orientations is equivalent to the data of a graded spin structure, we may try to calculate q and Q using K.

Definition 5.12 Let $\gamma = (h_1 \to \cdots \to h_m (\to h_1))$ be an open (closed) directed path in $G \in \mathcal{R}^0_{g,k,l}$ without backtracking; that is, the directed edge $s_1 h$ cannot follow h in the path. Put $v_i = h_i / s_0$. We say that γ makes a *bad turn* at v_i if either

(a) $h_{i-1} \in H^I$ and $h_i \neq s_2 h_{i-1}$, or

(b) $h_{i-1} \in H^B$ and $h_i = s_0 s_1 h_{i-1}$,

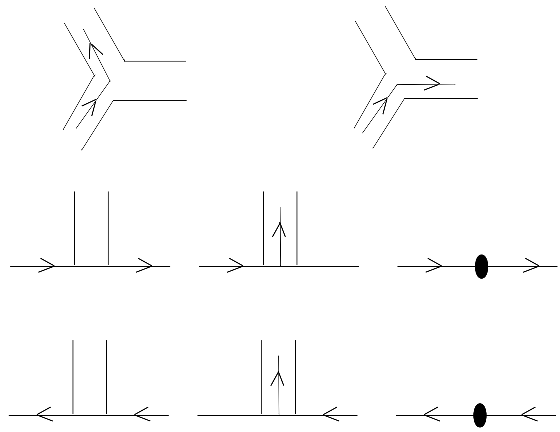


Figure 16: Good and bad turns. In this figure a line with an arrow represents a half-edge in a directed path, and the orientation is always counterclockwise. In the top row an internal vertex is drawn; the left shows a good turn, the right a bad turn. In the middle row the horizontal line is the boundary, and the surface lies above it. The oriented half-edges in the boundary belong to $s_1 H^B$. Only the leftmost image represents a bad turn. In the bottom row the oriented half-edges in the boundary component are boundary half-edges. The image on the left is a good turn, while the other two are bad.

where $i-1$ is taken modulo m in the closed case. Otherwise it makes a *good turn*. $\mathrm{BT}(\gamma)$ is the number of bad turns.

See Figure 16 for illustrations of good and bad turns.

Proposition 5.13 *Fix* $[K]$. *With the conventions of the previous definition*:

(a) *For* γ *closed*, $q(\gamma) = q_K(\gamma) := 1 + \sum_i K(h_i) + \mathrm{BT}(\gamma)$ *for any* $K \in [K]$.

(b) *For* γ *open, with* $h_1, h_m \in s_1 H^B$, *let* $\widetilde{\gamma}$ *be the subarc obtained from* γ *after removing small neighborhoods of its endpoints. Then* $Q(\widetilde{\gamma}) = Q_K(\gamma) := 1 + \sum_i K(h_i) + \mathrm{BT}(\gamma)$ *for any* $K \in [K]$.

We defined $\widetilde{\gamma}$ in order to avoid marked points as endpoints.

Proof Fix $K \in [K]$. Recall the correspondence between Kasteleyn orientations and lifts (Lemma 5.8), and take the corresponding lift l. Put $\theta_{j+1} = \measuredangle(P(h_j)T^1 h_j, T^1 h_{j+1}) \in (-\pi, \pi)$, write $\varepsilon_j = K(h_j)$, and define $\mathrm{bt}_{j+1} \in \mathbb{Z}_2$ to be 1 if and only if γ makes a bad turn in v_{j+1}, and otherwise 0. Proposition 5.9 is equivalent, in this notation, to

$$R_{\theta_{j+1}} P(h_j) l_{h_j} = R_{(\varepsilon_j + \mathrm{bt}_{j+1})2\pi} l_{h_{j+1}}. \tag{27}$$

When γ is closed, iterating (27) for $j = 1, \dots, m$ we get that

$$\begin{aligned} l_{h_1} &= R_{2\pi(\varepsilon_m+\mathrm{bt}_1)+\theta_1} P(h_m) R_{2\pi(\varepsilon_{m-1}+\mathrm{bt}_m)+\theta_m} P(h_{m-1}) \cdots R_{2\pi(\varepsilon_1+\mathrm{bt}_2)+\theta_2} P(h_1) l_{h_1} \\ &= R_{2\pi\sum_{i=1}^m \varepsilon_i+\mathrm{bt}_i} R_{\theta_1} P(h_m) R_{\theta_m} P(h_{m-1}) \cdots R_{\theta_2} P(h_1) l_{h_1} \\ &= R_{2\pi(\mathrm{BT}(\gamma)+\sum_{i=1}^m \varepsilon_i)} R_{(1+q(\gamma))2\pi} l_{h_1} \\ &= R_{2\pi(q(\gamma)+1+\mathrm{BT}(\gamma)+\sum_{i=1}^m \varepsilon_i)} l_{h_1}, \end{aligned}$$

where the final equality uses the definition of q, Definition 2.26.

Similarly, when γ is open, iterating (27) over $j = 1, \dots, m-1$ and applying the same reasoning, this time using Definition 2.30, we obtain, as needed,

$$l_{h_m} = R_{2\pi(\mathrm{BT}(\gamma)+\sum_{i=1}^{m-1} \varepsilon_i+Q(\gamma))} l_{h_1} = R_{2\pi(1+\mathrm{BT}(\gamma)+\sum_{i=1}^m \varepsilon_i+Q(\gamma))} l_{h_1},$$

where we used $\varepsilon_m = K(h_m) = 1$. □

Remark 5.14 The first case of the proposition appeared before in [12]. Although the formula depends on the orientation of γ, the result is orientation-independent in the closed case. Indeed, flipping the orientation changes each $K(h)$ to $K(s_1 h) = K(h) + 1$ and interchanges the sets of good turns and of bad turns. Thus, the total change is the number of edges plus the number of vertices of γ, that is, a change by $2m = 0$. A similar argument shows that in the open case the result changes by 1 when the orientation is flipped.

Definition 5.15 An automorphism $\phi\colon G \to G$ defines an action ϕ_* on $K(G)$ and $[K(G)]$ by

$$(\phi_* K)(h) = K(\phi^{-1}(h)).$$

An automorphism ϕ of $(G, [K])$ is an automorphism ϕ of G for which $\phi_*[K] = [K]$. We write $\mathrm{Aut}(G, [K])$ for the group of these automorphisms.

Proposition 5.16 *For any $G \in \mathcal{SR}^0_{g,k,l}$, the map*

$$\coprod_{z \in Z_G/\mathrm{Aut}(G)} \mathcal{M}_{(G,z)} \to \coprod_{[K]\in[K(G)]/\mathrm{Aut}(G)} \mathbb{R}_+^{E(G)}/\mathrm{Aut}(G,[K])$$

which takes a metric graded graph (G, z, ℓ) to $([K], \ell)$, where $[K]$ is the Kasteleyn orientation associated to the graded spin structure of $\mathrm{comb}^{-1}(G, z, \ell)$, is a homeomorphism.

Proof It is enough to show that along a path $(\Sigma_t)_{0\le t\le 1}$ in $\mathrm{comb}^{-1}(\mathcal{M}_{(G,z)})$, the equivalence classes $[K_t]=[K_t(\Sigma_t,\mathbb{S}_t,s_t)]\in[K(G)]$ are the same. Take $K_0\in[K(\Sigma_0,\mathbb{S}_0,s_0)]$. This determines the maps Q_0 and q_0, by Proposition 5.13 and the fact that any piecewise smooth path may be isotoped to a nonbacktracking one on the 1–skeleton $G\hookrightarrow\Sigma_0$. Now, varying $(\Sigma_t,\mathbb{S}_t,s_t)$ is equivalent to varying the metric ℓ_t on G in the component $\mathcal{M}_{(G,z)}$ continuously. But then it is evident that the maps Q_t and q_t determined by K_0 on the paths in the resulting embedded graph do not change. By Lemma 2.39 we see that $[K_t]=[K_0]$. □

In light of Proposition 5.16, we can redefine $\mathcal{SR}^0$ and the related combinatorial moduli spaces.

Notation 5.17 From now on we write

$$\mathcal{SR}^0_{g,k,l}=\{(G,[K])\mid G\in\mathcal{R}^0_{g,k,l},[K]\in[K(G)]/\mathrm{Aut}(G)\}.$$

Define $\mathcal{M}_{(G,[K])}=\mathbb{R}_+^{E(G)}/\mathrm{Aut}(G,[K])$, the moduli of metrics on G together with a fixed equivalence class of Kasteleyn orientations. We have that $\mathcal{M}_{(G,[K])}\hookrightarrow\overline{\mathcal{M}}_{(G,z)}$ for a unique $z\in Z_G$, as in Proposition 5.16. We therefore set $\overline{\mathcal{M}}_{(G,[K])}=\overline{\mathcal{M}}_{(G,z)}$. Define analogously $\mathcal{M}_{(G,[K])}(\boldsymbol{p})$ and $\overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p})$.

Example 5.18 Fix a connected component C of $\overline{\mathcal{M}}^{\mathbb{R}}_{g,k,l}$. Suppose that smooth surfaces in C have b boundary components and write $g_s=\frac{1}{2}(g-b+1)$. Let k_j for $j=1,\dots,b$ be the number of boundary marked points on boundary component j, for some locally defined numbering of the boundary components. One ribbon graph which corresponds to surfaces in C is the graph $G\in\mathcal{R}^0_{g,k,l}$ with

$$V=\{v^-_{j,j+1}\}_{j=2,\dots,b}\cup\{v^+_{j,j+1}\}_{j\in[b-1]}\cup\{p_{j,i}\}_{j\in[b],i\in[k_j]}\cup\{v^\pm_i\}_{i=2,\dots,l}\cup\{u^\pm_i,w^\pm_i\}_{i\in[g_s]}.$$

See also Figure 17. Only the v^-_i are internal vertices, while the vertices $p_{j,i}$, $v^+_{j,j+1}$ and $v^-_{j-1,j}$ belong to the j^{th} boundary component. The other boundary vertices belong to the first boundary. So

$$H^I=\bigcup_{i\in[b]}H_{\mathrm{bdry},i}\cup H_{\mathrm{bridges}}\cup H_{\mathrm{genus}}\cup H_{\text{internal marked}},$$

where:

(a) $H_{\mathrm{bdry},j}=\{e_{j,i}\}_{0\le i\le k_j+(1-\delta_{jb})}$ for $j\neq 1$ are the boundary edges of boundary component j and of face 1, and $e_{j,i}/s_0=p_{j,i}$ for $1\le i\le k_j$. In addition,

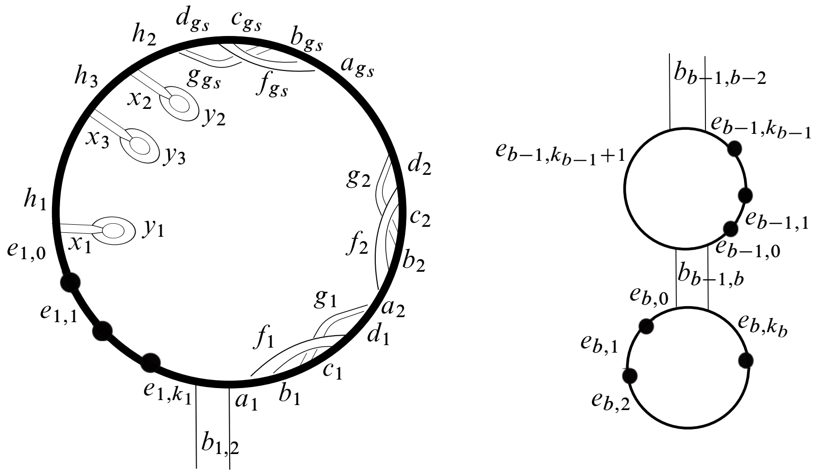


Figure 17

$e_{j,0}/s_0 = v^+_{j,j+1}$ and $(s_1 e_{j,0})/s_0 = p_{j,1}$. For $j \neq b, 1$, the edge e_{j,k_j} connects p_{j,k_j} to $v^-_{j-1,j}$, and we have $e_{j,k_j+1}/s_0 = v^-_{j-1,j}$ and $s_1(e_{j,k_j+1})/s_0 = v^+_{j,j+1}$. For $j = b$, we have $e_{j,k_j}/s_0 = v^-_{b-1,b}$. They are ordered so that $e_{j,i+1} = s'_2 e_{j,i}$, where $s'_2(e) := s_1(s_2^{-1}(s_1(e)))$ for $e \in s_1 H^B$.

(b) $H_{\mathrm{bdry},1} = a_1, b_1, c_1, d_1, a_2, \dots, d_{g_s}, h_2, \dots, h_l, e_{10}, e_{1,1}, \dots, e_{1,k_1}$ is the set of boundary edges of the first boundary, which all belong to face 1, ordered by s'_2 order. The boundary vertices, in counterclockwise order starting from $v^+_{1,2}$, the vertex of the bridge, are

$$v^+_{1,2}, u^+_1, w^+_1, u^-_1, w^-_1, u^+_2, \dots, w^-_{g_s}, v^+_2, \dots, v^+_l, p_{1,1}, \dots, p_{1,k_1}.$$

The adjacency relation is thus $a_1/s_0 = v^+_{1,2}$. For $i > 1$, we have

$$a_i/s_0 = w^-_{i-1}, \quad b_i/s_0 = u^+_i, \quad c_1/s_0 = w^+_i, \quad d_1/s_0 = u^-_i.$$

Next, $h_2/s_0 = w^-_{g_s}$, and $h_i/s_0 = v^+_{i-1}$ for $i > 1$. Finally, $e_{1,0}/s_0 = v^+_l$, and $e_{1,i}/s_0 = p_{1,i}$ for $i > 0$.

(c) $H_{\mathrm{bridges}} = \{b_{j,j+1}, \bar{b}_{j,j+1}\}_{j \in [b-1]}$ is the set of bridges between consecutive boundaries. We have

$$b_{j,j+1}/s_0 = v^+_{j,j+1}, \quad \bar{b}_{j,j+1} = s_1 b_{j,j+1}, \quad \bar{b}_{j,j+1}/s_0 = v^-_{j,j+1}.$$

(d) $H_{\mathrm{genus}} = \{f_i, \bar{f}_i, g_i, \bar{g}_i\}_{i \in g_s}$ is a set of internal half-edges of face 1 such that f_i goes from u^+_i to u^-_i and satisfies $\bar{f}_i = s_1 f_i$, and g_i goes from w^+_i to w^-_i and satisfies $\bar{g}_i = s_1 g_i$.

(e) $H_{\text{internal marked}} = \{x_i, \overline{x}_i, y_i, \overline{y}_i\}_{i=2,\dots,l}$ is the following set: y_i is the unique edge of face i satisfying $y_i/s_0 = v_i^-$, and $\overline{y}_i = s_1 y_i$. The third half-edge of v_i^- is x_i, and $\overline{x}_i = s_1 x_i$, while $\overline{x}_i/s_0 = v_i^+$.

We now describe $K(G)$. First of all, $K(h) = 1$ if $s_1 h \in H^B$ or $h = y_i$. There is no constraint on $K(x_i)$, but different values are equivalent by flips in v_i^-. Since there are no more internal vertices, for all other edges there are no constraints and no relations. Thus there is a total of $2^{2g_s+b-1} = 2^g$ different graded spin structures in this case. Since this is a topological invariant, for any generic open genus g surface in C there are 2^g graded structures. Thus, for any generic open genus g surface which satisfies condition (5) there are 2^g graded structures.

Remark 5.19 In [34] a notion of parity is defined for smooth graded surfaces with an odd number of boundary points for each component. It is defined as follows. Given such a graded surface $(\Sigma, \mathbb{S}, s)$, choose a symplectic basis $\{\alpha_i, \beta_i\}_{i\in[g_s]}$ to $H_1(\Sigma, \mathbb{Z}_2)/H_0(\partial\Sigma, \mathbb{Z}_2)$. The quadratic form q factors through this quotient. Define $\operatorname{Arf}(\Sigma) = \sum q(\alpha_i)q(\beta_i) \pmod 2$. This is an isotopy invariant. A spin structure is said to be even if the Arf is 0, otherwise it is odd. This notion is generalized, also in [34], to give the *open Arf invariant*, which is defined for any graded surface, and specializes to the parity if there is an odd number of markings on each boundary.

For example, with the notation of Example 5.18, suppose that each k_j is odd. A possible choice for the symplectic basis is

$$\alpha_i = b_i \to c_i \to \overline{f}_i \to b_i, \ \beta_i = c_i \to d_i \to \overline{g}_i \to c_i.$$

Now, by Proposition 5.13,

$$q(\alpha_i) = 1 + K(b_i) + K(c_i) + K(\overline{f}_i) + \mathrm{BT}(\alpha_i) = K(\overline{f}_i),$$

since there is one bad turn. Similarly, $q(\beta_i) = K(\overline{g}_i)$. Therefore,

$$\operatorname{Arf}(\Sigma) = \sum_{i\in[g_s]} K(\overline{f}_i)K(\overline{g}_i).$$

A simple calculation now shows that the difference between even and odd spin gradings in this case is $2^{g_s+b-1} = 2^{\frac{1}{2}(g+b-1)}$.

Remark 5.20 Kasteleyn orientations are named after W Kasteleyn, who used them to analyze dimer statistics; see for example [23]. The connection between Kasteleyn orientations and spin structures on closed surfaces is established in [26; 12].

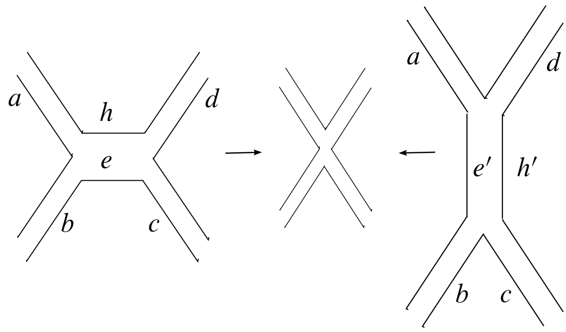


Figure 18: G, $\partial_e G$ and G_e. The middle graph is $\partial_e G$. We draw an half-edge inside the face which contains it.

5.1.1 Adjacent Kasteleyn orientations Recall Construction–Notation 3. In the cell structure of $\overline{\mathcal{M}}^{\text{comb}}_{g,k,l}$, the cell $(G,[K])$ is adjacent to cells of the form $(G_e,[K_e])$ for edges $e \notin \mathrm{Br}(G) \cup \mathrm{Loop}(G)$ with $[K_e] \in [K(G_e)]$, by Proposition 4.49. We now describe $[K_e]$ explicitly in terms of $[K]$.

Fix a Kasteleyn orientation $K \in [K]$. Write h for the unique half-edge such that $K(h) = 1$ and $h/s_1 = e$. Write

$$a = s_0(h), \quad b = s_0^2(h), \quad c = s_1(s_0(s_1(h))), \quad d = s_1(s_0^2(s_1(h)));$$

see Figure 18. For brevity write $\bar{x}$ for $s_1(x)$. Apart from some borderline cases, which may be treated separately, we may assume all these vertices and half-edges are distinct, and then, using vertex flips if needed, we may also restrict ourselves to the case where $K(\bar{d}) = 1$. Note that $E(G) \setminus e = E(G_e) \setminus e'$ canonically for some $e' \in E(G_e)$. We therefore identify these sets, and also identify $H(G) \setminus \{h, s_1 h\}$ and $H(G_e) \setminus s_1^{-1} e'$. In G_e, let v_1' be the vertex from which a and $\bar{d}$ issue, and let v_2' be the vertex from which b and $\bar{c}$ issue. We may take the half-edge h' to be the third half-edge from v_1'. Define the assignment $K' \colon H^I(G) \to \mathbb{Z}_2$ by

$$K'(h') = 1, \quad K'(\bar{h}') = 0, \quad K'(d) = K(d) + 1 = 1, \quad K'(\bar{d}) = K(\bar{d}) + 1 = 0,$$

and $K'(f) = K(f)$ for any other half-edge f.

For later purposes, define, for a boundary loop e and a Kasteleyn orientation $K \in [K]$, an assignment K' by $K'(h) = K(h)$ for any h with $h/s_1 \neq f$, where f is the unique edge which shares a vertex with e, and otherwise $K'(h) = K(h) + 1$.

Lemma 5.21 *In both cases, $K' \in [K(G_e)]$, and moreover, $K' \in [K_e]$.*

Proof The claim is straightforward when e is a boundary loop. Suppose that $e \notin \mathrm{Br}(G) \cup \mathrm{Loop}(G)$. The first assertion is simple; we focus the second one. Write $C(G)$ and $C(G')$ for the set of closed paths without backtracking in G and G', respectively. Write $O(G)$ and $O(G')$ for the set of open directed paths without backtracking in G and G', respectively, which connect boundary vertices which are not marked points. We have bijections $f_C : C(G) \to C(G')$ and $f_O : O(G) \to O(G')$, defined as follows. For a path $(e_1 \to e_2 \to \cdots \to e_m) \in C(G)$, the path $f_C(e_1 \to e_2 \to \cdots \to e_m) \in C(G')$ is defined by erasing any appearance of e in the sequence and adding e' any time we have a move $f \to f'$ where the third edge of the vertex between f and f' is e. The inverse map is defined similarly, but changing the roles of e and e'. The map f_O is defined in the same way.

Using Proposition 5.13 it is straightforward to verify that $q_K(\gamma) = q_{K'}(f_C(\gamma))$ for any $\gamma \in C(G)$, and $Q_K(\gamma) = Q_{K'}(f_C(\gamma))$ for any $\gamma \in O(G)$.

Now, let $(\Sigma_t, \mathbb{S}_t, s_t)_{t\in[0,1]}$ be a continuous path in $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$, with

$$(\Sigma_t, \mathbb{S}_t, s_t) \in \mathrm{comb}^{-1}(\mathcal{M}_{(G_t,z_t)}), \quad \text{where } G_t = \begin{cases} G & \text{if } t < \frac{1}{2}, \\ \partial_e G & \text{if } t = \frac{1}{2}, \\ G' & \text{if } t > \frac{1}{2}, \end{cases}$$

and where the graded structure $z_0 \in Z_G$ corresponds to the Kasteleyn orientation $[K]$. In light of Lemma 2.39, Proposition 5.16 and isotopy arguments, the Kasteleyn orientation on G' defined by $(\Sigma_t, \mathbb{S}_t, s_t)_{t\in(\frac{1}{2},1)}$ is the unique class of Kasteleyn orientations for which $q(\gamma_t)$ or $Q(\gamma_t)$ is constant for any continuous family $(\gamma_t \subseteq \Sigma_t)$ of closed paths or bridges. By performing an isotopy, we may assume that γ_t is in fact a path in the graph G_t. It is easy to see that for ε small enough, $f_C(\gamma_{\frac{1}{2}-\varepsilon}) = \gamma_{\frac{1}{2}+\varepsilon}$ if the γ_t are closed, or $f_O(\gamma_{\frac{1}{2}-\varepsilon}) = \gamma_{\frac{1}{2}+\varepsilon}$ if they are open. In the first case, $q_{[K]}(\gamma_{\frac{1}{2}-\varepsilon}) = q_{[K']}(\gamma_{\frac{1}{2}+\varepsilon})$, while in the second the same equation holds for Q. By Lemma 2.39(c) and Theorem 5.11, the graded structure z_t for $t > \frac{1}{2}$ must correspond to $[K']$. □

5.1.2 Trivalent graphs

Definition 5.22 Recall Definition 4.23. Let G be a trivalent graph. Recall that a half-node is an $(N^B)^{-1}$–preimage of a node, and that their collection is denoted by $\mathrm{HN}(G)$. An *extended Kasteleyn orientation* on G is a map $K : H(G) \cup \mathrm{HN}(G) \to \mathbb{Z}_2$ such that:

(a) For any $h \in H^B$, $K(h) = 0$.

(b) For any $h \in H$, $K(h) + K(s_1 h) = 1$.

(c) For any node v, if $|N^{-1}(v)| = 3$, then $K|_{N^{-1}(v)} = 1$. Otherwise $K(v_{i,1}) + K(v_{i,2}) = 1$, where $N^{-1}(v) = \{v_{i,1}, v_{i,2}\}$.

(d) For any face f, $\sum K(x) = 1$, where the variable x is taken from the set of half-edges with $x/s_2 = f$, together with the set of half-nodes which belong to f.

Two extended Kasteleyn orientations are equivalent if they differ by the action of internal vertex flips. Write $[K]$ for the equivalence class of K. Define $K(G)$ and $[K(G)]$ as the sets of extended Kasteleyn orientations and the set of equivalence classes of extended Kasteleyn orientations. Write $\mathrm{Aut}(G,[K])$ for the automorphism subgroup of G which preserves $[K]$.

Item (c) above deals with the case that v is a contracted component whose normalization contains at least three half-nodes. In the trivalent case, this can only happen if the unique contracted component in $\mathrm{Norm}^{-1}(v)$ is a ghost, and its three marked points are legal. Therefore there are exactly three corresponding half-nodes in the noncontracted parts, and they are illegal.

With the exact same techniques as for Section 5.1, together with Corollary 2.22, we obtain:

Lemma 5.23 *For a trivalent G and a metric ℓ, there is a natural bijection between $[K(G)]$ and $\mathrm{Spin}((\mathrm{comb}^{\mathbb{R}})^{-1}(G,\ell))$. The induced map*

$$\coprod_{z \in Z_G/\mathrm{Aut}(G)} \mathcal{M}_{(G,z)} \to \coprod_{[K]\in[K(G)]/\mathrm{Aut}(G)} \mathbb{R}_+^{E(G)}/\mathrm{Aut}(G,[K])$$

*is a homeomorphism. In particular, $Z_G \simeq [K(G)]$ canonically. A half-node v in (G,z) is **illegal** if and only if $K(v) = 1$ for any $K \in [K]$ which corresponds to z.*

From now on we denote trivalent graphs (G,z) by $(G,[K])$, for the corresponding $[K] \in [K(G)]$.

Definition 5.24 Define $\mathcal{M}_{(G,[K])} := \mathbb{R}_+^{E(G)}/\mathrm{Aut}(G,[K])$, the moduli of metrics on $\mathcal{M}_G$, together with a fixed equivalence class of Kasteleyn orientations. Define $\overline{\mathcal{M}}_{(G,[K])} := \overline{\mathcal{M}}_{(G,z)}$, for the unique z which corresponds to $[K]$ by the above lemma. For $f_1,\dots,f_s \in E(G)$, set $\partial_{f_1,\dots,f_s}\overline{\mathcal{M}}_{(G,[K])}$ to be the face of $\overline{\mathcal{M}}_{(G,[K])}$ defined by setting the coordinates $\ell_{f_1},\dots,\ell_{f_s}$ to 0. For $p_1,\dots,p_l > 0$, define $\mathcal{M}_{(G,[K])}(\boldsymbol{p})$ and $\overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p})$ by setting the perimeters to these values.

Suppose G is a trivalent graph $K \in K(G)$, and let $e \in \mathrm{Br}(G)$. In the case that e is a boundary edge, let h_1 be its internal half-edge, $h_1/s_1 = e$, with $h_1 \in H^I$. In the case that e is an internal edge, write $s_1^{-1}(e) = \{h_1, h_2\}$, where $K(h_i) = i \pmod 2$. Define $\partial_e K$ to be the unique map $\partial_e K\colon H(\partial_e G) \cup \mathrm{HN}(\partial_e G) \to \mathbb{Z}_2$ which agrees with K on any half-edge $h' \notin s_1^{-1}e$, and such that $\partial_e K(\partial_e h_i) = i \pmod 2$. In a similar way, one can define $\partial_{e_1,\dots,e_r} K$ for a compatible sequence of bridges.

Observation 5.25 *For any trivalent $(G,[K])$, and bridge e, the graph $(\partial_e G, [\partial_e K])$ is a well-defined trivalent graph, in particular $\partial_e K \in [K(\partial_e G)]$. Moreover, the map $\partial_e\colon [K(G)] \to [K(\partial_e G)]$ is a bijection.*

In addition, for any trivalent connected graph $(G,[K])$, there is a unique smooth trivalent graph $(G',[K'])$ and a unique (up to order) compatible sequence of bridges $e_1,\dots,e_r$ with $(G,[K]) = \partial_{e_1,\dots,e_r}(G',[K'])$.

With the same techniques as in the proof of Lemma 5.21, one obtains:

Lemma 5.26 *Let G be a trivalent graph, and let $e_1,\dots,e_r$ be a compatible sequence of bridges. Under the identification of Lemma 5.23 between Z_H and $[K(H)]$, for $H = G, \partial_{e_r} G, \dots, \partial_{e_1,\dots,e_r} G$, we have that*

$$\overline{\mathcal{M}}_{\partial_{e_1,\dots,e_r}(G,[K])} \simeq \partial_{e_1,\dots,e_s} \overline{\mathcal{M}}_{\partial_{e_{s+1},\dots,e_r}(G,[K])}$$

canonically.

In what follows we shall identify $\overline{\mathcal{M}}_{(G,z)}$ and the corresponding $\overline{\mathcal{M}}_{(G,[K])}$ without further notice.

5.2 Orientation

In this subsection we construct an orientation to $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$. We do it by writing an explicit formula for the orientation of each highest-dimensional cell of $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$ — that is, for cells $\mathcal{M}_{(G,[K])}(\boldsymbol{p})$ where $G \in \mathcal{R}^0$, $[K] \in [K(G)]$ — and then showing that on codimension-one faces between two such cells, the induced orientations disagree. We also discuss the induced orientation on the boundary, and prove that these orientations are the ones induced from $\overline{\mathcal{M}}_{g,k,l}$ by comb_*.

For $G \in \mathcal{R}^0_{g,k,l}$, we have a map

$$A_G\colon \mathbb{R}_+^{E(G)} \to \mathbb{R}^{F(G)} = \mathbb{R}^{[l]}, \tag{28}$$

which takes as input a collection of edge lengths and outputs the face perimeters, and

$$\mathcal{M}_{(G,[K])}(\boldsymbol{p}) = A_G^{-1}(\boldsymbol{p})/\mathrm{Aut}(G,[K]).$$

In particular, orienting $\mathcal{M}_{(G,[K])}$ is equivalent to orienting $\ker(A_G)/\mathrm{Aut}(G,[K])$. Using the exact sequence

$$0 \to \ker(A_G) \to \mathbb{R}^{E(G)} \to \mathbb{R}^{F(G)} = \mathbb{R}^{[l]} \to 0, \tag{29}$$

we see that orienting $\mathbb{R}^{E(G)}$ and $\mathbb{R}^{[l]}$, or equivalently, ordering $E(G)$ and $[l]$, up to even permutations, gives an orientation to $\mathcal{M}_{(G,[K])}(\boldsymbol{p})$, as long as the action of $\mathrm{Aut}(G,[K])$ preserves the orientation.

Fix any order for $[l]$, for example $1, 2, \ldots, l$. Choose any Kasteleyn orientation $K \in [K]$. Define $\mathfrak{o}_i = \mathfrak{o}_{(G,K,i)}$ by

$$\bigwedge_{\substack{K(h)=1 \\ h/s_2=i}} d\ell_h,$$

that is, we take the wedge of $d\ell_h$ over half-edges h of face i with $K(h) = 1$. The wedge is taken counterclockwise. Because there is an odd number of half-edges of the i^{th} face with $K = 1$, the element $\mathfrak{o}_i$ is well defined, and independent of which half-edge appears first. In addition, $\mathfrak{o}_i$ is an odd-degree form.

Definition 5.27 Choose any Kasteleyn orientation K. Put

$$\mathfrak{o}_{(G,K)} = \bigwedge_{i=1}^{l} \mathfrak{o}_i.$$

Define $\overline{\mathfrak{o}}_{(G,K)}$ as the orientation on $\ker(A_G)$ induced from the exact sequence (29) when $\mathbb{R}^{E(G)}$ is oriented by $\mathfrak{o}_{(G,K)}$ and $\mathbb{R}^{[l]}$ by $\bigwedge_{i=1}^{l} dp_i$.

Remark 5.28 Since both dp_i and $\mathfrak{o}_i$ are odd variables, choosing another order on $[l]$ does not change $\overline{\mathfrak{o}}_G$.

Lemma 5.29 *The orientation $\overline{\mathfrak{o}}_{(G,K)}$ depends only on $[K]$.*

Before we get to the proof, we add a few auxiliary definitions.

Definition 5.30 Let G be any open ribbon graph. A *good ordering* is a bijection $n\colon H^I \to |H^I|$ which satisfies the following properties. First, if $i(h) < i(h')$, that is, h belongs to face marked i and h' to face marked $i' > i$, then $n(h) < n(h')$. Thus, half-edges of the same face are clustered together. Second, the ordering n, when restricted to half-edges of a single face, agrees with the counterclockwise ordering.

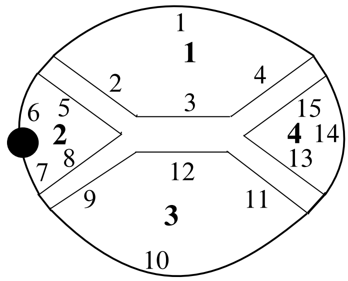


Figure 19: A good ordering of internal half-edges: the bold numbers in the middles of the faces are the labels of the faces, the smaller ones next to the half-edges are the half-edges numbers in the ordering. The numbers of half-edges in face i are smaller than those of face j if $i < j$. In each face the numbers of half-edges agree with the cyclic order induced by the face's orientation.

Let n be a good ordering, as in Definition 5.30, and $K \in K(G)$ a Kasteleyn orientation. Define $H_K = \{h \in H^I \mid K(h) = 1\}$. We also define $n_K \colon |H^I| \to \mathbb{Z}$ by

$$n_K(i) = |\{h \in H_K \mid n(h) < i\}|.$$

Figure 19 illustrates a good ordering. Note that the restriction of a good ordering to a subset of H^I induces an order on its elements.

Proof of Lemma 5.29 Take any $K \in [K]$. We recall from Lemma 5.8 that any other element of $[K]$ can be obtained from K by successive flips in vertices. It will thus suffice to prove that the orientations induced by K and K' are the same when K and K' differ by a single flip in vertex v. It will be enough to prove that $\mathfrak{o}_{(G,K)} = \mathfrak{o}_{(G,K')}$.

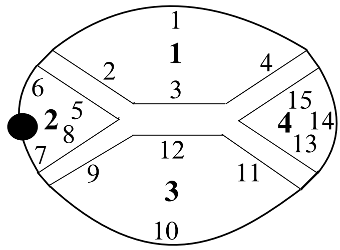
Fix a good ordering n. By definition,

$$\mathfrak{o}_{(G,K)} = \bigwedge_{e \in H_K} d\ell_e,$$

where the order of the wedging is the order n restricted to H_K. The sign difference between $\mathfrak{o}_{(G,K)}$ and $\mathfrak{o}_{(G,K')}$ can be found geometrically by the following procedure, also illustrated in Figure 20. Define

$$L_K = \{(n(h), 0) \mid h \in H_K\} \quad \text{and} \quad L_{K'} = \{(n(h), 1) \mid h \in H_{K'}\} \subseteq \mathbb{R}^2.$$

For any $e \in E$ draw the chord $c(e)$ between $(n(h_0), 0) \in L_K$ and $(n(h_1), 1) \in L_{K'}$, where $h_0/s_1 = h_1/s_1$. By definition the change of signs between $\mathfrak{o}_{G,K}$ and $\mathfrak{o}_{G,K'}$ is just the parity of the number of intersections of these chords (slightly perturbed, if



$$H_{K'} = 1, 5, 6, 7, 9, 10, 12, 13, 14, 15$$
$$H_K = 1, 2, 3, 6, 7, 8, 10, 13, 14, 15$$

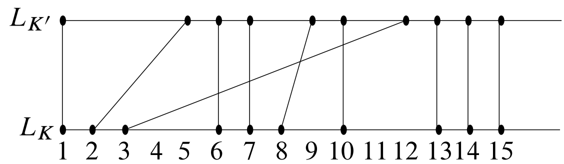


Figure 20: In this diagram H_K and $H_{K'}$ are listed for two Kasteleyn orientations K and K' for the graph G in the picture, whose half-edges are identified with their n–value for some good order n. The Kasteleyn orientations K and K' can be read from H_K and $H_{K'}$, and they differ by a flip in the left internal vertex. Below the chord diagram of L_K and $L_{K'}$ is drawn, and the number of intersections is indeed even.

necessary). We shall prove that this number is always even. Note that for all edges except for those issuing from v, the chords are parallel and vertical.

Let h_1 be a half-edge of v. Put $h_2 = s_0(h_1)$, $h_3 = s_0^2(h_1)$ and $\bar{h}_j = s_1(h_j)$. Apart from some borderline cases which can be treated separately, we may assume that we are in the scenario

$$\begin{aligned} n(\bar{h}_2) &= i_1, & n(h_1) &= i_1 + 1, & n(\bar{h}_3) &= i_2, \\ n(h_2) &= i_2 + 1, & n(\bar{h}_1) &= i_3, & n(h_3) &= i_3 + 1. \end{aligned}$$

Thus, the chord c_{h_j} is either the chord between $(i_j + 1, 0)$ and $(i_{j-1}, 1)$, or the chord between $(i_j + 1, 1)$ and $(i_{j-1}, 0)$. It is easy to see that the number of vertical chords it intersects is the size of

$$I_j = \{h \in H_K \setminus \{h_i, \bar{h}_i\}_{i=1,2,3} \mid n(h) \in (a_j, b_j)\},$$

where $a_j = \min(n_K(i_j + 1), n_K(i_{j-1}))$ and $b_j = \max(n_K(i_j + 1), n_K(i_{j-1}))$. For exactly one $j \in \{1, 2, 3\}$ we have $I_j = I_{j+1} \cup I_{j+2}$, where addition is modulo 3, and the union is disjoint. Thus, any vertical chord either misses the chords c_{h_j} or meets exactly two of them. In addition, it can be checked directly that the chords c_{h_j} intersect each other an even number of times. The lemma follows. $\square$

Corollary 5.31 *For any $G \in \mathcal{R}^0_{g,k,l}$ and $[K] \in [K(G)]$, the group $\mathrm{Aut}(G,[K])$ acts in an orientation-preserving manner. In particular, the orientation $\overline{\mathfrak{o}}_{(G,K)}$ induces, for any $\boldsymbol{p}$, an orientation on $\mathcal{M}_{(G,[K])}$.*

Denote this orientation by $\overline{\mathfrak{o}}_{(G,[K])}$. The main theorem of this subsection is:

Theorem 5.32 *The orientations $\overline{\mathfrak{o}}_{(G,[K])}$ induce a canonical orientation on the space $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$.*

Proof We shall show that the orientations $\mathfrak{o}_G$ for $G \in \mathcal{SR}^0_{g,k,l}$ are compatible on codimension-one faces. This will show that a suborbifold of $\mathcal{M}^{\mathrm{comb}}_{g,k,l}$ which differs from $\mathcal{M}^{\mathrm{comb}}_{g,k,l}$ in codimension-two cells is oriented, hence also $\mathcal{M}^{\mathrm{comb}}_{g,k,l}$ is. Since $\mathcal{M}^{\mathrm{comb}}_{g,k,l}$ itself differs from $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ by codimension-two strata in the interior, and in codimension-one boundary, this argument will show that $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ is also endowed with a canonical orientation.

We therefore have to show that for any $(G,[K]) \in \mathcal{SR}^0_{g,k,l}$ and $e \notin \mathrm{Br}(G) \cup \mathrm{Loop}(G)$ with $(G',[K']) = (G_e,[K_e])$, the orientations induced on $\partial_e \mathcal{M}_{(G,[K])}$ by $\mathcal{M}_{(G,[K])}$ and by $\mathcal{M}_{(G',[K'])}$ *disagree*.

Put $H^I = H^I(G)$ and $H'^I = H^I(G')$. Note that we have a natural identification of $E(G) \setminus e$ and $E(G') \setminus e'$, for some edge e', so from now on we treat them as the same set. Choose a good ordering n for H^I. There exists a good ordering n' of H'^I which, when restricted to $H'^I \setminus s_1^{-1}(e')$, defines the same order as the restriction of n to $H'^I \setminus s_1^{-1}(e') \simeq H^I \setminus s_1^{-1}(e)$. Fix a Kasteleyn orientation $K \in K(G)$ and set $h \in s_1^{-1}(e)$ with $K(h) = 1$. Write

$$a = s_0(h), \quad b = s_0^2(h), \quad c = s_1(s_0(s_1(h))), \quad d = s_1(s_0^2(s_1(h)));$$

see Figure 21. For brevity write $\bar{x}$ for $s_1(x)$. Apart from some borderline cases which may be treated separately, we may assume all these vertices and half-edges are distinct, and then, using vertex flips if needed, we may also restrict ourselves to the case where $K(\bar{d}) = 1$. In this case we can assume n was chosen in such a way that

$$n(\bar{a}) = i, \quad n(h) = i+1, \quad n(\bar{d}) = i+2, \quad n(d) = m, \quad n(\bar{c}) = m+1,$$
$$n(c) = p, \quad n(\bar{h}) = p+1, \quad n(b) = p+2, \quad n(\bar{b}) = j, \quad n(a) = j+1,$$

for some i, m, p and j, as in Figure 21.

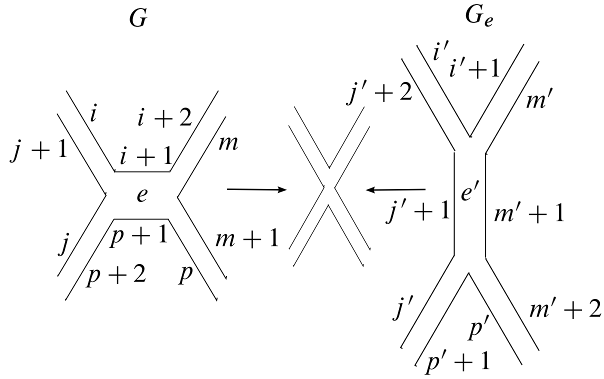


Figure 21: The restrictions of the good orderings n and n' to the half-edges of G and G_e.

A canonical outward normal for $\mathcal{M}_{\partial_e G} \hookrightarrow \overline{\mathcal{M}}_G$ is just $-d\ell_e$. We see that the induced orientation on $\mathcal{M}_{\partial_e G}$ is just

$$(-1)^{n_K(n(h))+1} \bigwedge_{f \in H_K \setminus \{h\}} d\ell_f = (-1)^{n_K(i+1)+1} \bigwedge_{f \in H_K \setminus \{h\}} d\ell_f, \tag{30}$$

where as usual the wedge is taken in the order n_K induced by n.

In G', let v_1' be the vertex from which a and $\overline{d}$ issue, and let v_2' be the vertex from which b and $\overline{c}$ issue. We may take the half-edge h' to be the third half-edge from v_1'. Then, for some i', m', p' and j', we have

$$n'(\overline{a}) = i', \quad n'(\overline{d}) = i'+1, \quad n'(d) = m', \quad n'(h') = m'+1, \quad n'(\overline{c}) = m'+2,$$
$$n'(c) = p', \quad n'(b) = p'+1, \quad n'(\bar{b}) = j', \quad n'(\overline{h}') = j'+1, \quad n'(a) = j'+2.$$

By Lemma 5.21 we have a representative K' of $[K_e]$, described by

$$K'(h') = 1, \quad K'(\overline{h}') = 0, \quad K'(d) = K(d)+1 = 1, \quad K'(\overline{d}) = K(\overline{d})+1 = 0,$$

and $K'(f) = K(f)$ for any other half-edge f. As above, a canonical outward normal for $\mathcal{M}_{\partial_{e'} G'} \hookrightarrow \overline{\mathcal{M}}_{G'}$ is just $-d\ell_{e'}$. We see that the induced orientation on $\mathcal{M}_{\partial_{e'} G'}$ is

$$(-1)^{n_{K'}(n'(h'))+1} \bigwedge_{f \in H_{K'} \setminus \{h'\}} d\ell_f = (-1)^{n'_{K'}(m'+1)+1} \bigwedge_{f \in H_{K'} \setminus \{h'\}} d\ell_f. \tag{31}$$

The choice of n, n' and K' makes the terms $\bigwedge_{f \in H_K \setminus \{h\}} d\ell_f$ and $\bigwedge_{f \in H_{K'} \setminus \{h'\}} d\ell_f$ differ only in the relative location of $d\ell_d$. By our assumptions on $K(\overline{d})$ and $K'(\overline{d})$, the difference is just the difference between $n_K(\overline{d}) - 1 = n_K(i+2) - 1$ and $n'_{K'}(d) = n'_{K'}(m')$. We subtracted 1 from $n_K(\overline{d})$ because we did not want to count h which occurs before $\bar{d}$ in the order n. Now, $n_K(i+2)-1 = n_K(i+1)$, as $n(h)=i$, $K(h)=1$. Similarly, $n'_{K'}(m') = n'_{K'}(m'+1)-1$, since $n'(d)=m'$ and $K'(d)=1$.

The total difference between the two orientations is thus

$$(-1)^{n'_{K'}(m'+1)+1+n'_{K'}(m'+1)-1+n_K(i+1)+1+n_K(i+1)} = -1,$$

as claimed. □

Remark 5.33 The spaces $\mathcal{M}_{g,k,l}$ and $\mathcal{M}^{\mathrm{comb}}_{g,k,l}(\boldsymbol{p})$ are homeomorphic, therefore the last theorem gives, in fact, another proof that $\overline{\mathcal{M}}_{g,k,l}$ is oriented. Later we shall see that the orientation constructed here agrees with the orientation of [35].

Corollary 5.34 *For $G \in \mathcal{SR}^0_{g,k,l}$ and e an internal edge which is not a bridge, the two orientations on $\partial_e\overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p}) \simeq \partial_e\overline{\mathcal{M}}_{(G_e,[K_e])}(\boldsymbol{p})$, induced as boundaries of $\mathcal{M}_{(G,[K])}(\boldsymbol{p})$ and $\mathcal{M}_{(G_e,[K_e])}(\boldsymbol{p})$, are opposite.*

5.3 Critical nodal graphs and their moduli

5.3.1 Critical nodal ribbon graphs In this subsection we describe effective and critical nodal graphs. They will parametrize strata which will participate in the analysis of the intersection numbers and will contribute to the combinatorial formula. For completeness we first describe slightly more general graphs.

Definition 5.35 A *nodal spin ribbon graph with a lifting* (*graded nodal ribbon graph*), or a *nodal graph* for short, is a spin ribbon graph with a lifting (graded ribbon graph) (G,z), together with a subset $\mathcal{V}$ of legal points in $B(\mathrm{Norm}(G)) \setminus B(G)$. We call $\mathcal{V}$ the set of *legal nodes* of the nodal graph and $s_1\mathcal{V}$ the illegal nodes, where s_1 was defined in Notation 4.28. The vertices and edges of the nodal graph are the vertices and edges of $\mathrm{Norm}(G,z)$ after forgetting the illegal nodes $s_1\mathcal{V}$. A metric is a metric on these edges. If e is an edge in the nodal graph $(G,z,\mathcal{V})$, contracting the edge e yields the nodal graph $\partial_e(G,z,\mathcal{V})$ whose underlying graph is $\partial_e(G,z)$, and whose legal nodes are those legal nodes in $\partial_e(G,z)$ which remain special points in $\mathrm{Norm}(\partial_e(G,z))$ after the contraction, where we use the natural correspondence between special points in $\mathrm{Norm}(G,z)$ and in $\mathrm{Norm}(\partial_e(G,z))$.

The *components* of the nodal graph are the connected components created after removing $s_1\mathcal{V}$. More precisely, define an equivalence relation $\sim_N$ on the components of $\mathrm{Norm}(G,z)$ as follows. Components $C_1, C_2 \in \pi_0(\mathrm{Norm}(G,z))$ are neighbors if one of them contains a legal point $u \notin \mathcal{V}$ such that $s_1 u$ belong to the other component. For $C_1, C_2 \in \pi_0(\mathrm{Norm}(G,z))$, we write $C_1 \sim_N C_2$ if they can be connected in a path of neighboring components. The components of the nodal graph are defined to be the Norm–image of $\sim_N$–equivalence classes.

In the case that the underlying graph is effective we have a more convenient definition.

Definition 5.36 An *effective nodal spin ribbon graph with a lifting* (*effective graded nodal ribbon graph*), or an *effective nodal graph*, is a tuple $(G_i, z_i, m, \mathcal{V} = \{\mathcal{V}_e\})$, or (G,z) for short, where

- (a) (G_i, z_i) is an effective spin ribbon graph with a lifting (effective graded ribbon graph),
- (b) $m\colon \bigcup_i s_1 H^B(G_i) \to \mathbb{Z}_{\geq 0}$, and
- (c) the maps $\mathcal{V}_e\colon [m(e)] \to \bigcup_i B(G_i)$ for $e \in \bigcup_i s_1 H^B(G_i)$ are injections.

We require the sets $\mathcal{V}_e = \mathcal{V}_e([m(e)])$ to be disjoint. Denote by $C(G_i, z_i, m, \{\mathcal{V}_e\})$ the different graded components of the graph, that is, the collection of (G_i, z_i).

Let G be the graph obtained by choosing $m(e)$ points $p_{e,1}, \dots, p_{e,m(e)}$ on e, ordered according to the orientation of the boundary and identifying $p_{e,i}$ with $\mathcal{V}_e(i)$. The effective nodal graph is said to be connected if G is connected.

Write $E(G) = \bigcup_i E(G_i)$; similarly define $H^I(G)$, $H^B(G)$, $V(G)$ and $F(G)$. For a boundary edge $e = h/s_1$ where $h \in s_1 H^B$, we sometimes write $m(e) = m(h)$. Vertices in the image of $\mathcal{V}_e$ are called legal nodes and their set is denoted by $\mathcal{V}(G)$. The boundary marked points of G are boundary marked points of the G_i which are not legal nodes. Denote them by $B(G)$. Define $I(G) = \bigcup_i I(G_i)$.

An effective nodal ribbon graph is naturally embedded into the (topological) nodal surface $\Sigma = \bigl(\coprod_i \Sigma_i\bigr)/\sim$, defined as follows. Σ_i is the topological open marked surface into which G_i embeds, and in the case that G_i is a ghost it is a point. We identify G_i with its image in Σ_i. We add $m(e)$ points $p_{e,1}, \dots, p_{e,m(e)}$ along the edge e, and quotient by $p_{e,i} \sim \mathcal{V}_e(i)$. The genus of the graph is defined to be the (doubled) genus of Σ.

A *marked effective nodal graph* is an effective nodal graph together with markings $\mathtt{m}^B\colon B(G) \to \mathbb{Z}$ and $\mathtt{m}^I\colon I(G) \to \mathbb{Z}$.

A *graded critical nodal ribbon graph* is an effective nodal graph such that each $(G_i, z_i) \in \mathcal{SR}^0$. In this case we use the Kasteleyn notation for components, $(G_i, [K_i])$ rather than (G_i, z_i), and we denote the whole graph by $(G, [K])$ for short.

A graded critical nodal graph G is *odd* if each $G_i \in \mathcal{OSR}^0$.

The notion of an isomorphism is the expected one. Write $\mathcal{SR}^m_{g,k,l}$ for the collection of isomorphism classes of marked critical nodal graded ribbon graphs G with m nodes and genus g such that $\mathtt{m}^B : B(G) \simeq [k]$ and $\mathtt{m}^I : I(G) \simeq [l]$. Let $\mathcal{OSR}^m_{g,k,l}$ be the subset of such graphs which are odd. Write $\mathrm{Aut}(G,[K])$ for the group of automorphisms of $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$.

Define *nongraded critical nodal ribbon graphs* $G = (G_i, m, \mathcal{V})$ in the same way, only without the data of Kasteleyn orientations, so that each G_i belongs to $\mathcal{R}^0$ rather than to $\mathcal{SR}^0$. Denote by $\mathcal{R}^m_{g,k,l}$ the collection of isomorphism classes of nongraded critical nodal ribbon graphs G with m nodes and genus g such that $\mathtt{m}^B : B(G) \simeq [k]$ and $\mathtt{m}^I : I(G) \simeq [l]$. Let $\mathcal{OR}^m_{g,k,l}$ be the subset of such graphs which are odd. Write $\mathrm{Aut}(G)$ for the group of automorphisms of $G \in \mathcal{R}^m_{g,k,l}$.

A metric on a nodal ribbon graph is an assignment of positive lengths to its edges.

A bridge $e \in E(G)$ is an edge which is a bridge in one component G_i of G. An *effective bridge* is a bridge with $m(e) = 0$, if m is defined. Let $\mathrm{Br}(G,[K])$ be the collection of bridges, and $\mathrm{Br}^{\mathrm{eff}}(G,[K])$ the collection of effective bridges. As in the nonnodal case, for brevity we shall usually omit $[K]$ from the notation for Br and $\mathrm{Br}^{\mathrm{eff}}$. We similarly define boundary loops as boundary loops in one component G_i of G, and effective loops are boundary loops e with $m(e) = 0$. Write $\mathrm{Loop}(G)$ and $\mathrm{Loop}^{\mathrm{eff}}(G)$ for the collection of boundary loops and effective loops, respectively.

When it is understood from context whether or not the critical nodal graph is graded or nongraded, we omit the words graded/nongraded, and just say critical nodal.

Remark 5.37 It is simple to verify that when $(G, z, m, \mathcal{V})$ is effective, Definitions 5.35 and 5.36 are equivalent. We shall therefore use Definition 5.36, which is more explicit, whenever possible. It is also straightforward to verify that the definition of $\mathcal{OR}^m_{g,k,l}$ agrees with the one given in Notation 1.3.

In a metric effective nodal ribbon graph, the data of distances between illegal nodes to other vertices is absent. On the other hand, the discrete data of which illegal node lies on which edge, and the relative order of illegal nodes on a given edge, are included. See the example at the bottom of Figure 22.

Observation 5.38 *Under the forgetful map* $\mathrm{for}_{\mathrm{spin}} : \mathcal{SR}^m_{g,k,l} \to \mathcal{R}^m_{g,k,l}$, *which forgets the Kasteleyn orientation, odd graphs go to odd graphs and the preimage of* G *is canonically* $[K(G)]/\mathrm{Aut}(G)$.

5.3.2 Trivalent graphs versus graded critical nodal graphs In the analysis required for proving Theorem 1.5, we will mainly need to analyze critical graded nodal graphs and effective graphs which are obtained from them by contracting a single edge and possibly forgetting some data. We will now describe operations between nodal and nonnodal ribbon graphs. Although these operations can be defined in full generality, we are interested only in cases where their output is trivalent or effective. We will therefore restrict our definitions to this setting, leaving the relatively straightforward details of the more general setting to the interested reader.

Given a connected effective spin ribbon graph with a lifting (G, z), we define an effective nodal graph $\mathcal{X}(G, z)$ as follows. Its components are the components of $\mathrm{Norm}(G, z)$, after erasing every illegal boundary point and concatenating its two edges to one edge. Note that under this map a contracted boundary becomes a Ramond marking of perimeter zero. Suppose e is an edge obtained by concatenating $e_1, \dots, e_{m+1}$ in the described process, and in this order. Define $m(e) = m$. Suppose v_i is the vertex between e_i and e_{i+1}. Then $\mathcal{V}_e(i) = s_1 v_i$, where we use Notation 4.28. When $(G, z) = (G, [K])$ is critical trivalent, we denote $\mathcal{X}(G, z)$ by $\mathcal{X}(G, [K])$. It is easy to verify that:

Observation 5.39 *The map $\mathcal{X}$ is a surjection from the collection of connected effective spin ribbon graphs to the collection of nodal connected effective spin ribbon graphs all of whose components are smooth. It restricts to a bijection between connected trivalent graphs and connected graded critical nodal ribbon graphs. For any connected effective spin ribbon graph (G, z), there is a bijection between bridges (boundary loops) in (G, z) and effective bridges (effective loops) in $\mathcal{X}(G, z)$.*

We now extend the definition of $\mathcal{X}$ to metric effective spin ribbon graphs. For such a graph (G, z, ℓ), define the effective nodal metric graph $\mathcal{X}(G, z, \ell) = (\mathcal{X}(G, z), \mathcal{X}\ell)$ by $\mathcal{X}\ell_e = \ell_e$ if the edge e is an edge of $\mathrm{Norm}(G, z)$; otherwise, if e is the union of $e_1, \dots, e_{m+1}$, define $\mathcal{X}\ell_e = \sum_{i=1}^{m+1} \ell_{e_i}$. Note that the perimeters are left unchanged.

We also define a map from effective nodal graphs to effective spin ribbon graphs: given an effective nodal graph $(G, z, m, \mathcal{V})$, define the spin ribbon graph $\widetilde{\mathcal{B}}(G, z)$ as the graph obtained by forgetting the data of m and $\mathcal{V}$, and applying $\widetilde{\mathcal{B}}$ to each component (G_i, z_i). The analogous definition holds for metric effective nodal graphs.

If $(G, z, m, \mathcal{V})$ is an effective nodal graph and e is either an internal edge or a boundary edge with $m(e) = 0$, then $\partial_e(G, z, m, \mathcal{V})$ is the nodal graph whose underlying ribbon graph is the graph obtained by contracting e, and the data of m and $\mathcal{V}$ is induced from G

by the usual identification of edges of $\partial_e G$ as a subset of edges of G. Similarly, when $(G,[K],m,\mathcal{V})$ is critical trivalent and e is either an internal edge or an effective loop, we define $(G_e,[K_e],m',\mathcal{V}')$ as the critical trivalent graph whose underlying graph is $(G_e,[K_e])$, putting $m'=m$ and $\mathcal{V}'=\mathcal{V}$, where we again use the identification between edges of G and G_e.

Notation 5.40 Suppose that $(G,[K])\in\mathcal{SR}^m_{g,k,l}(\boldsymbol{p})$ and that $e=\{h_1,h_2=s_1h_1\}\in\mathrm{Br}^{\mathrm{eff}}(G)\cup\mathrm{Loop}^{\mathrm{eff}}(G)$, with $K(h_1)=0$. Define the nodal ribbon graph $\mathcal{B}_e(G,[K])$ as follows. Suppose G is made of the components $G_1,\dots,G_n$. Without loss of generality assume e is an edge of component G_n. Write $v_i=\partial_e(h_i)$ for the vertex obtained by contracting h_i in $\partial_e G_n$. Write $x=s_2h_1$ and $y=s_1(s_2^{-1}h_1)\in H^I(\partial_e G_n)$.

The first $n-1$ components of the graph $\mathcal{B}_e(G,[K])$ are $G_i'=G_i$ for $i\le n-1$, and for these components we have $K_i'=K_i$, $m'=m$ and $\{\mathcal{V}_f'\}=\{\mathcal{V}_f\}$.

When e is a boundary loop, $(G_n',z_n)=\widetilde{\mathcal{B}}\partial_e(G_n,[K_n])$, and also in this component $m'=m$ and $\{\mathcal{V}_f'\}=\{\mathcal{V}_f\}$, where we use the natural identifications between edges of G_n other than e and edges of G_n'.

If e is an effective bridge, then in the case that the normalization $\mathrm{Norm}(\partial_e G_n)$ is disconnected, let G_n' be the component which does not contain v_2, and let K', m' and $\mathcal{V}'$ be the induced maps. Note that G_n' may be a ghost. Define the component G_{n+1}' as the graph obtained by the component of v_2 in $\mathrm{Norm}(\partial_e G_n)$ after gluing the half-edges x/s_1 and y/s_1 to a new edge xy, and removing the vertex v_2. The updated Kasteleyn orientation is the unique Kasteleyn orientation which gives any internal half-edge its value under K_n. For any half-edge $e'\ne xy$, we have $m'(e')=m(e')$ and $m(xy)=m(x)+m(y)+1$. Similarly, $\mathcal{V}'(e')=\mathcal{V}(e')$ for $e'\ne xy$, while

$$\mathcal{V}'_{xy}(a)=\begin{cases}\mathcal{V}_y(a) & \text{if } a\le m(y),\\ v_1 & \text{if } a=m(y)+1,\\ \mathcal{V}_x(a-m(y)-1) & \text{if } a>m(y)+1.\end{cases}\tag{32}$$

If $\partial_e G_n\setminus\{v_e\}$ is connected, set G_n' to be the component of v_1 in the normalization, where again edges x and y are glued and v_2 is removed, and K', m' and $\mathcal{V}'$ are defined in the same way as above.

There is a canonical surjection, which we shall also denote by $\mathcal{B}_e$,

$$E(G)\cup\mathcal{V}(G)\to E(\mathcal{B}_eG)\cup\mathcal{V}(\mathcal{B}_eG).$$

It takes e to v_1, and all other edges to the corresponding edges, so that it is one-to-one except on the edges x and y, which go to xy.

Given a metric ℓ on the graph, with $\ell_e = 0$, the graph $\mathcal{B}_e(G, [K], \ell)$ is the graded nodal ribbon graph with underlying graph $\mathcal{B}_e(G, [K])$, and the metric is induced from ℓ if e is a boundary loop, while if e is a bridge, then with the same notation as above, $(\mathcal{B}_e\ell)_{e'} = \ell_{e'}$ for $e' \neq x, y$, and $\mathcal{B}_e\ell_{xy} = \ell_x + \ell_y$. For convenience we usually denote $\mathcal{B}_e\ell$ by ℓ as well.

A *compatible sequence of effective bridges* $e_1, \dots, e_r$ is a sequence of bridges such that e_{i+1} is an effective bridge in $\mathcal{B}_{e_i} \cdots \mathcal{B}_{e_1} G$ for all i. For such a sequence define $\mathcal{B}_{e_1,\dots,e_r}(G, [K], \ell) = \mathcal{B}_{e_r} \cdots \mathcal{B}_{e_1}(G, [K], \ell)$, and the map $\mathcal{B}_{e_1,\dots,e_r} = \mathcal{B}_{e_r} \circ \cdots \circ \mathcal{B}_{e_1}$.

The next observation follows easily from Observations 5.39 and 5.25.

Observation 5.41 *If* $(G, [K]) \in \mathcal{SR}^m_{g,k,l}$ *and* $e \in \mathrm{Loop}^{\mathrm{eff}}(G)$, *then* $\mathcal{B}_e G$ *is an effective nodal ribbon graph.*

If $(G, [K]) \in \mathcal{SR}^m_{g,k,l}$ *and* $e \in \mathrm{Br}^{\mathrm{eff}}(G)$, *then* $\mathcal{B}_e G \in \mathcal{SR}^{m+1}_{g,k,l}$.

Moreover, for any $(G, [K]) \in \mathcal{SR}^{m+1}_{g,k,l}$, *and any legal node* v, *there exists a unique graph* $(H, [K']) \in \mathcal{SR}^m_{g,k,l}$ *and an edge* $e \in \mathrm{Br}^{\mathrm{eff}}(H)$ *with* $\mathcal{B}_e(H, [K']) = (G, [K])$ *and* $\mathcal{B}_e e = v$. *In addition, if* $(G, [K])$ *is connected trivalent and* $e \in \mathrm{Br}(G, [K])$, *then*

$$\mathcal{X}(\partial_e(G, [K])) = \mathcal{B}_e(\mathcal{X}(G, [K])),$$

where we use the identification of bridges of Observation 5.39.

Notation 5.42 Recall Notation 4.6. For $(G, [K]) \in \mathcal{SR}^{m+1}_{g,k,l}$, denote by $\mathcal{B}^{-1}_{h,a}(G, [K]) = \mathcal{B}^{-1}_{[h],a}(G, [K])$ the isomorphism class of triples $(H, [K'], e)$ where $H \in \mathcal{SR}^m_{g,k,l}$, $\mathcal{B}_e(H, [K']) = (G, [K])$, and $\mathcal{B}_e e = \mathcal{V}_h(a)$ for $h \in s_1(H^B(G))$ and $a \in [m(h)]$. Let

$$\mathcal{B}^{-1} G = \{\mathcal{B}^{-1}_{[h],a}(G, [K]) \mid [h] \in [s_1(H^B(G))], a \in [m(h)]\}.$$

In other words, $(H, [K'], e) = \mathcal{B}^{-1}_{h,a}(G, [K])$ should be thought as the graph $(H, [K'])$ obtained by canceling the $\mathcal{B}$ operation, ie by returning the a^{th} forgotten illegal node of h, gluing it with its legal side, and then uncontracting the resulting node to obtain the bridge e.

5.3.3 The moduli space of critical nodal graphs, the line bundles and the boundary conditions

Definition 5.43 For an effective nodal ribbon graph $(G, z, m, \mathcal{V})$ define $\mathcal{M}_{(G,z,m,\mathcal{V})} \simeq \mathbb{R}_+^{E(G)}/\mathrm{Aut}(G, z, m, \mathcal{V})$ to be the moduli of positive metrics on G, and $\mathcal{M}_{(G,z,m,\mathcal{V})}$ as the subspace in which the i^{th} perimeter equals $p_i > 0$, $i \in [l]$. In particular, given

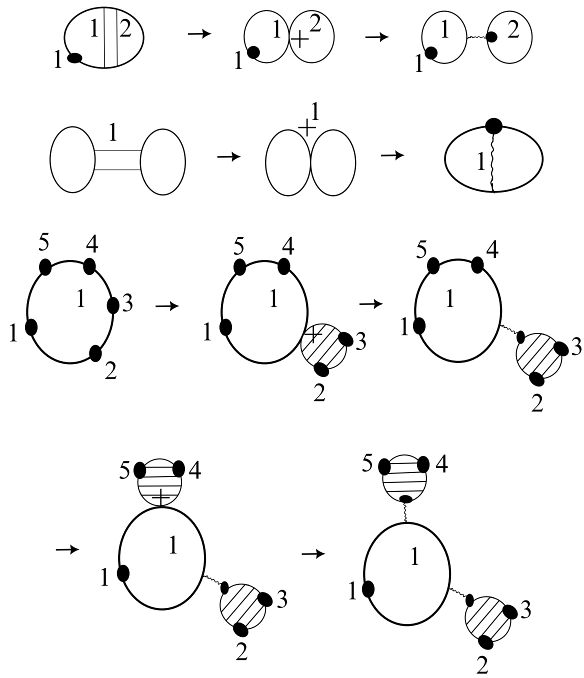


Figure 22: This diagram presents trivalent graphs, their effective bridge contractions and the operation $\mathcal{B}$. The $+$ sign represents a legal side of node and, after performing $\mathcal{B}$, the wiggly lines contain the data of $\mathcal{V}$, namely, which edges contain which legal nodes, and in what order. At top left an effective trivalent smooth graph $(G,[K])$ on a disk is shown, at top center its bridge e is contracted, then at top right $\mathcal{B}_e(G,[K])$ is drawn. The second row describes a similar scenario, but for a graph on a cylinder. The third row presents a graph on a disk. First the bridge between boundary markings 2 and 3 is contracted, and then the bridge between 4 and 5 is contracted. These bridges are compatible. The bridges between 2 and 3 and 3 and 4, on the other hand, are not compatible with each other.

$(G,[K]) \in \mathcal{SR}^m_{g,k,l}$, we have $\mathcal{M}_{(G,[K])} \simeq \mathbb{R}_+^{E(G)}/\mathrm{Aut}(G,[K])$. Define $\overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$ and $\overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}(\boldsymbol{p})$ as the cell complexes whose cells correspond to nodal ribbon graphs obtained from $(G,z,m,\mathcal{V})$ by edge contractions, and the gluing maps are induced by these edge contractions.

For $e \in E(G)$, write $\partial_e\overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$ for the face of $\overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$ where e is contracted, ie the length of the edge e is set to be 0. The boundary of $\overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$ can be written as

$$\partial\overline{\mathcal{M}}_{(G,z,m,\mathcal{V})} = \bigcup_{[e]\in[E(G)]} \partial_e\overline{\mathcal{M}}_{(G,z,m,\mathcal{V})},$$

where $[E(G)] = E(G)/\mathrm{Aut}(G,z,m,\mathcal{V})$, as in Notation 4.6. We similarly define $\partial_{e_1,\dots,e_r}\overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$.

The maps $\widetilde{\mathcal{B}}$, $\mathcal{X}$ and $\mathcal{B}_{e_1,\dots,e_r}$ on metric graphs induce moduli level maps. We denote these maps by the same letters. When $e_1,\dots,e_r$ are understood from the context, we denote the former map by $\mathcal{B}$.

Note that $\overline{\mathcal{M}}_{\partial_e(G,z,m,\mathcal{V})} \simeq \partial_e\overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$, and that $\mathcal{B}_{e_1,\dots,e_r}$ factors $\widetilde{\mathcal{B}}$. The maps $\mathcal{B}$, $\widetilde{\mathcal{B}}$ and $\mathcal{X}$ are easily seen to be piecewise linear submersions.

Definition 5.44 For an effective nodal $(G,z,m,\mathcal{V})$ and $i \in [l]$, the S^1–orbibundle $\mathcal{F}_i \to \overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$ is defined to be the set of pairs (ℓ, x) where $\ell \in \overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$ and x is a point on the i^{th} face, with the natural topology. For a (d,l)–set L, write $S_L \to \overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$ for the sphere bundle associated to $\{S_{L(i)} \mid i \in [d]\}$, as in Construction–Notation 1. We define the forms α_i, ω_i, $\overline{\alpha}_i$ and $\overline{\omega}_i$ as the pullbacks of the corresponding forms defined on the component which contains face i.

If $(G',z',m',\mathcal{V}')$ is obtained from $(G,z,m,\mathcal{V})$ by edge contractions, we have the usual natural identification between $\mathcal{F}_i \to \overline{\mathcal{M}}_{(G',z',m',\mathcal{V}')}$ and the restriction of $\mathcal{F}_i \to \overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$ to the corresponding cell.

By the constructions we immediately get:

Observation 5.45 *For any effective spin ribbon graph (G',z') and $i \in [l]$, we have a natural identification*

$$(\mathcal{F}_i \to \overline{\mathcal{M}}_{(G',z')}) \simeq \mathcal{X}^*(\mathcal{F}_i \to \overline{\mathcal{M}}_{\mathcal{X}(G',z')}),$$

while for an effective nodal spin ribbon graph (G,z) and $i \in [l]$, we have a natural identification

$$(\mathcal{F}_i \to \overline{\mathcal{M}}_{(G,z)}) \simeq \widetilde{\mathcal{B}}^*(\mathcal{F}_i \to \overline{\mathcal{M}}_{\widetilde{\mathcal{B}}(G',z')}).$$

As a consequence:

(a) *For $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ and $e \notin \mathrm{Br}(G) \cup \mathrm{Loop}(G)$, there is a canonical identification*

$$(\mathcal{F}_i \to \overline{\mathcal{M}}_{\partial_e(G,[K])}) \simeq (\mathcal{F}_i \to \partial_e\overline{\mathcal{M}}_{(G,[K])}) \simeq (\mathcal{F}_i \to \partial_e\overline{\mathcal{M}}_{(G_e,[K_e])}),$$

and similarly for the bundles S_L.

(b) *For $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ and $e \in \mathrm{Br}^{\mathrm{eff}}(G)$, there is a canonical identification*

$$(\mathcal{F}_i \to \overline{\mathcal{M}}_{\partial_e(G,[K])}) \simeq (\mathcal{F}_i \to \partial_e\overline{\mathcal{M}}_{(G,[K])}) \simeq \mathcal{B}_e^*(\mathcal{F}_i \to \overline{\mathcal{M}}_{\mathcal{B}_e(G,[K])}),$$

and similarly for the bundles S_L.

(c) *For* $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ *and* $e \in \mathrm{Loop}(G)$, *there is a canonical identification*

$$(\mathcal{F}_i \to \overline{\mathcal{M}}_{\partial_e(G,[K])}) \simeq (\mathcal{F}_i \to \partial_e \overline{\mathcal{M}}_{(G,[K])}) \simeq (\Psi^{\mathrm{comb}})^*(\mathcal{F}_i \to \partial_e \overline{\mathcal{M}}_{(G_e,[K_e])}),$$

and similarly for the bundles S_L.

Proposition 5.46 *Let* s *be a special canonical multisection of* $S_L \to \overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$. *Let* A *be the collection of effective graded* (g,k,l)*–boundary ribbon graphs, so that* s *restricts, in particular, to multisections* $s^{(G,z)}$ *for all* $(G,z) \in A$. *Then* s *induces multisections* $s^{(G,z,m,\mathcal{V})}$ *of* $S_L \to \overline{\mathcal{M}}_{(G,z,m,\mathcal{V})}$ *for all effective nodal ribbon graphs* $(G,z,m,\mathcal{V}) \in \mathcal{X}(A)$, *which satisfy the following relations*:

- *For any effective graded* (G',z'),
$$s^{(G',z')} = \mathcal{X}^* s^{\mathcal{X}(G',z')}.$$
- *For any effective nodal* $(G,z,m,\mathcal{V})$,
$$s^{(G,z,m,\mathcal{V})} = \widetilde{\mathcal{B}}^* s',$$
where s' *is a multisection of* $S_L \to \overline{\mathcal{M}}_{\widetilde{\mathcal{B}}(G,z)}$.

In particular:

(a) *For any* $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ *and* $e \notin \mathrm{Br}(G) \cup \mathrm{Loop}(G)$,
$$s^{(G,[K])}|_{\partial_e \overline{\mathcal{M}}_{(G,[K])}} = s^{(G,[K])}|_{\partial_e \overline{\mathcal{M}}_{(G_e,[K_e])}}.$$

(b) *For any* $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ *and* $e \in \mathrm{Br}^{\mathrm{eff}}(G)$,
$$s^{(G,[K])}|_{\partial_e \overline{\mathcal{M}}_{(G,[K])}} = \mathcal{B}_e^* s^{\mathcal{B}_e(G,[K])}.$$

(c) *For any* $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ *and* $e \in \mathrm{Loop}^{\mathrm{eff}}(G)$,
$$s^{(G,[K])}|_{\partial_e \overline{\mathcal{M}}_{(G,[K])}} = (\Psi^{\mathrm{comb}})^* s^{(G_e,[K_e])}.$$

Here we compare multisections using the identifications of Observation 5.45.

Proof Let s be a special canonical multisection as above. Consider an effective nodal $(G,z,m,\mathcal{V}) \in \mathcal{X}(A)$. Then $(G,z,m,\mathcal{V})$ can be written as $\mathcal{X}(G',z')$ for some effective boundary graph. Now $s^{\mathcal{X}(G',z')} = \widetilde{\mathcal{B}}^* s^{\widetilde{\mathcal{B}}\mathcal{X}(G',z')}$. We have a factorization

$$\begin{tikzcd} \mathcal{M}_{(G',z')} \arrow[r, "\mathcal{X}"] \arrow[dr, "\widetilde{\mathcal{B}}"'] & \mathcal{M}_{(G,z,m,\mathcal{V})} \arrow[d, "\widetilde{\mathcal{B}}"] \\ & \mathcal{M}_{\widetilde{\mathcal{B}}(G',z')} \end{tikzcd}$$

The identifications of bundles S_L, see Observations 4.40 and 5.45, are also compatible with this diagram. Since s is canonical, by Corollary 4.43,

$$s^{(G',z')} = \widetilde{\mathcal{B}}^* s^{\widetilde{\mathcal{B}}(G',z')} = \mathcal{X}^* \widetilde{\mathcal{B}}^* s^{\widetilde{\mathcal{B}}(G',z')}.$$

Define $s^{(G,z,m,\mathcal{V})}$ as the pullback of $s^{\widetilde{\mathcal{B}}(G',z')}$ along the vertical map $\widetilde{\mathcal{B}}$. Clearly $s^{\mathcal{X}(G',z')} = \mathcal{X}^* s^{(G,z)}$.

By Observation 5.39, $\mathcal{SR}^m_{g,k,l} \subseteq \mathcal{X}(A)$. The "In particular" cases are now immediate from the definition and Observation 5.45. In the first and third item we use that $\widetilde{\mathcal{B}}(G,[K]) = \widetilde{\mathcal{B}}(G_e,[K_e])$, while in the second that $\mathcal{B}_e = \widetilde{\mathcal{B}}$ in that case. □

The cells $\overline{\mathcal{M}}_{(G,[K])}$ for graded nodal graphs also carry canonical orientations.

Definition 5.47 We define orientations for $\overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p})$, $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ by

$$\bar{\mathfrak{o}}_{(G,[K])} = \prod_{C \in C(G,[K])} \bar{\mathfrak{o}}_C,$$

$$\mathfrak{o}_{(G,[K])} = \bigwedge_{i\in[l]} dp_i \wedge \bar{\mathfrak{o}}_{(G,[K])} = \bigwedge_{i\in[l]} \bigwedge_{\substack{K(h)=1\\ h/s_2=i}} d\ell_h,$$

with the wedge product over half-edges of face i taken counterclockwise.

Proposition 5.48 *Let $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ and $e \in \mathrm{Br}^{\mathrm{eff}}(G)$. Suppose that $(G',[K']) = \mathcal{B}_e(G,[K]) \in \mathcal{SR}^{m+1}_{g,k,l}$, and let e' be the unique edge in G' with two $\mathcal{B}_e$–preimages. There are canonical identifications*

$$\partial_e \overline{\mathcal{M}}_{(G,[K])} \simeq \overline{\mathcal{M}}_{\partial_e(G,[K])} \simeq \mathcal{F}_{e'},$$
$$\partial_e \overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p}) \simeq \overline{\mathcal{M}}_{\partial_e(G,[K])}(\boldsymbol{p}) \simeq \mathcal{F}_{e'}(\boldsymbol{p}),$$

where the space $\mathcal{F}_{e'} \to \overline{\mathcal{M}}_{(G',[K'])}$ is the set of pairs (ℓ,x) with $\ell \in \overline{\mathcal{M}}_{(G',[K'])}$ and x a point on e', with the natural topology. Moreover, the orientation on $\partial_e \overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p})$ induced from $\overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p})$, as in Definition 2.52, coincides with the orientation $dx \wedge \mathfrak{o}_{(G',[K'])}$ on $\mathcal{F}_{e'}$, where dx is the orientation on the segment e' considered as a segment in the boundary.

Proof The only part which requires an explanation is the statement regarding orientations. Recall that K' satisfies $K(h) = K'(\mathcal{B}h)$ for any $h/s_1 \neq e$. It is enough to compare orientations of $\partial_e \overline{\mathcal{M}}_{(G,[K])} \simeq \mathcal{F}_{e'} G'$. Suppose h is the legal side of e, that is, the half-edge which satisfies $h/s_1 = e$ and $K(h) = 1$. Write $e_{-1} = (s_2^{-1}h)/s_1$ and $e_1 = (s_2 h)/s_1$.

Then, by recalling the definition of the canonical orientation (Section 5.2), we see that the orientation for $\overline{\mathcal{M}}_{(G,[K])}$ can be written as $d\ell_{e_{-1}} \wedge d\ell_e \wedge d\ell_{e_1} \wedge O$, and the orientation on $\overline{\mathcal{M}}_{G'}$ is $d\ell_{e'} \wedge O$, where O is the wedge of other edge lengths, in some order. Note that $d\ell_{e'} = d\ell_{e_{-1}} + d\ell_{e_1}$. Now, the induced orientation on the boundary $\partial_e \overline{\mathcal{M}}_{(G,[K])}$ is given by $d\ell_{e_{-1}} \wedge d\ell_{e_1} \wedge O$. By considering $\mathcal{F}_{e'}G'$ as the moduli of metrics on the graph obtained from G' by adding a new marked point on e', and with the definition of its orientation, we see that this orientation can be written as $d\ell_{e_{-1}} \wedge d\ell_{e'} \wedge O$, where $d\ell_{e_{-1}}$ comes from the location of the new point on f. And indeed,

$$d\ell_{e_{-1}} \wedge d\ell_{e_1} \wedge O = d\ell_{e_{-1}} \wedge d\ell_{e'} \wedge O. \qquad \square$$

Corollary 5.49 *The map* $\mathrm{comb}\colon \overline{\mathcal{M}}_{g,k,l} \to \overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ *preserves orientation.*

Proof Indeed, by Proposition 5.48, we see that the orientations on $\overline{\mathcal{M}}^{\mathrm{comb}}_{g,k,l}$ satisfy the same requirements of Theorem 2.53. The dimension-zero case can be checked by hand. $\square$

We also have the following corollary of Corollary 5.34.

Corollary 5.50 *For* $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ *and an internal edge* e *which is not a bridge, the orientations on* $\partial_e \overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p}) \simeq \partial_e \overline{\mathcal{M}}_{(G_e,[K_e])}(\boldsymbol{p})$*, induced as boundaries of* $\mathcal{M}_{(G,[K])}(\boldsymbol{p})$ *and* $\mathcal{M}_{(G_e,[K_e])}(\boldsymbol{p})$*, are opposite.*

Corollary 5.50 has an analog for the case that e is a boundary loop. For $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ and $e \in \mathrm{Loop}(G)$, write $\Psi^{\mathrm{comb}}_{(G,[K]),e}$ for the map $\partial_e \overline{\mathcal{M}}_{(G,[K])} \to \partial_e \overline{\mathcal{M}}_{(G_e,[K_e])}$ defined at the level of objects by leaving all the metric graph structure — in particular the edge lengths — invariant, and flipping the lifting in the contracted boundary which corresponds to e. When we write Ψ^{comb} we mean the union of the maps $\Psi^{\mathrm{comb}}_{(G,[K]),e}$ over all possible pairs $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$, for $m \ge 0$ and $e \in \mathrm{Loop}(G)$. The following is an immediate corollary of the "Moreover" part of Theorem 2.53, and Corollary 5.49. We will also provide a direct self-contained proof of this corollary in Section 6.2 below.

Corollary 5.51 *For* $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ *and* $e \in \mathrm{Loop}(G)$*, the induced orientation on* $\partial_e \overline{\mathcal{M}}_{(G_e,[K_e])}(\boldsymbol{p})$ *as a boundary of* $\mathcal{M}_{(G_e,[K_e])}(\boldsymbol{p})$ *is opposite to the orientation on it obtained by taking the* Ψ^{comb}*–pushforward of the orientation on* $\partial_e \overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p})$*, induced as a boundary of* $\mathcal{M}_{(G,[K])}(\boldsymbol{p})$*.*

6 The combinatorial formula

Throughout this section we fix g, k, l and set

$$d = \tfrac{1}{2}\dim_{\mathbb{R}}(\overline{\mathcal{M}}_{g,k,l}) = \tfrac{1}{2}(3g-3+k+2l).$$

We also write, for $G \in \mathcal{SR}^m_{g,k,l}$,

$$\dim(G) = \tfrac{1}{2}\dim_{\mathbb{R}}(\mathcal{M}_G) = \tfrac{1}{2}(3g-3+k+2l-2m).$$

In what follows we shall work with the orientations constructed in Section 5.2. These are the same orientations as the ones constructed in [35], by Corollary 5.49.

Definition 6.1 For $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ define

$$W_G, \widetilde{W}_G \colon \mathcal{M}_{(G,[K])} \to \mathbb{R}$$

by

$$W_G(\ell) = \prod_{e \in s_1 H^B(G)} \frac{\ell_e^{2m(e)}}{(m(e)+1)!} \quad \text{and} \quad \widetilde{W}_G(\ell) = \prod_{e \in s_1 H^B(G)} \frac{\ell_e^{2m(e)}}{m(e)!\,(m(e)+1)!}.$$

6.1 Iterative integration and the integral form of the combinatorial formula

Our approach for producing the explicit formula for intersection numbers will be by an iterative process of integration by parts. Recall Definition 4.36 and Notation 4.37. Given an (S,l)–set $L\colon S \to [l]$ for $S \subseteq [d]$, the t^{th} component of E_L is $\mathbb{L}_{L(t)}$. Each step of the iterative integration process below will involve integrating out (the form corresponding to) one component $\mathbb{L}_{L(t)}$ for some $t \in S$, using integration by parts. The integration by parts will produce new boundary terms for the moduli on which we integrate. Only boundary terms that correspond to contracting an effective bridge e may have a nonzero contribution which does not cancel. Moreover, in order for such an edge to contribute a nonzero contribution, when we integrate out the t^{th} component the illegal side of the half-node obtained by contracting e will have to lie in the face $L(t)$. This is the content of first key lemma, Lemma 6.6. In order to be able to state it, we need to add notation: specifically, notation that will allow us to keep track of which illegal half-node corresponds to the t^{th} component of the vector bundle which we integrate out. For this we present the auxiliary notion of decorations. After performing an iteration of integration by parts, the second key lemma, Lemma 6.7, transforms integrals over the boundaries of the moduli to integrals over the moduli spaces obtained by further forgetting the illegal half-node. Theorem 6.10 essentially iterates these lemmas, and uses some other cancellations to obtain a formula for the open intersection numbers as sums of integrals. It is remarkable that this iterative integration process is performed without appealing to a specific canonical multisection, and in some sense this is the key point of the proof. In addition, it gives an alternative proof of the claim that canonical boundary conditions give rise to well-defined intersection numbers, proven in [31] for genus 0 and in [35] for $g > 0$.

Definition 6.2 A *decoration* D of a graph $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$ is a choice of sets $D_h \subseteq [d]$, for any $h \in s_1 H^B$, which are pairwise disjoint and such that

$$|D_h| = m(h).$$

When $e = h/s_1$ we also write $D_e = D_h$. For an (S,l)–set L, an L–decoration is a decoration for which

$$D_h \subseteq L_{i(h)}.$$

In the next series of claims we shall omit $[K]$ from the notation of graded graphs, to lighten notation.

Denote the collection of all decorations of G by $\mathrm{Dec}(G)$, and the collection of all L–decorations of G by $\mathrm{Dec}(G,L)$.

Let $L(D)$ be the l–subset of L given by $L|_{\bigcup_{h\in s_1 H^B} D_h}$, so that $L(D)_i = \bigcup_{i(h)=i} D_h$.

For $(G,[K]) \in \mathcal{SR}^{m>0}_{g,k,l}$ and a (G,L)–decoration D, define the set

$$\mathcal{B}^{-1}(G,D) \subseteq \{(G',e',D') \mid (G',e') \in \mathcal{B}^{-1}G,\ D' \in \mathrm{Dec}(G',L)\}$$

by setting $(G',e',D') \in \mathcal{B}^{-1}(G,D)$ exactly when $(G',e') \in \mathcal{B}^{-1}G$, $D' \in \mathrm{Dec}(G',L)$ and $D'_e \subseteq D_{\mathcal{B}e}$ for any $e \in E(G') \setminus \{e'\}$. Note that in this case $L(D') \subseteq L(D)$, and the difference is exactly one element.

In the language of the paragraph preceding this definition, $L(D) \setminus L(D')$ is precisely the element $t \in [d]$ which corresponds to the effective bridge e' in the iterative process.

In order to be able to calculate intersection numbers, we must understand the restriction of the forms α_i and ω_i to the boundary.

Suppose that $(G,[K]) \in \mathcal{SR}^m_{g,k,l}$, $e \in \mathrm{Br}^{\mathrm{eff}}(G)$ with h its illegal side, $K(h) = 1$ and $i \in [l]$. On $\mathcal{M}_{\partial_e G}(\boldsymbol{p})$ we have two natural representatives for the angular 1–form,

$\alpha_i^{\partial_e G} = \alpha_i^G|_{\partial_e \mathcal{M}_G}$ and $\mathcal{B}^*\alpha_i^{\mathcal{B}_e G}$. Similarly, we have two natural choices for the induced two-forms, $\omega_i^{\partial_e G} = \omega_i^G|_{\partial_e \mathcal{M}_G}$ and $\mathcal{B}^*\omega_i^{\mathcal{B}_e G}$.

Notation 6.3 Write $\beta_i = \beta_i^{\partial_e G} = \alpha_i^{\partial_e G} - \mathcal{B}^*\alpha_i^{\mathcal{B}_e G}$ and $B_i = B_i^{\partial_e G} = \omega_i^{\partial_e G} - \mathcal{B}^*\omega_i^{\mathcal{B}_e G}$.

Observation 6.4 *With the above notation, if $i \neq i(e)$, then $B_i = \beta_i = 0$. Otherwise we have*

$$p_i^2 \beta_i = \ell_{s_2 h} d\ell_{s_2^{-1} h} \quad \text{and} \quad p_i^2 B_i = d\ell_{s_2^{-1} h} \wedge d\ell_{s_2 h}.$$

Unlike the forms α_i, the form β_i is pulled back from the combinatorial moduli, since it has no angular variables.

Proof For $i \neq i(h)$, the forms restricted from $\mathcal{M}_G$ and those pulled back from $\mathcal{M}_{\mathcal{B}_e G}$ are canonically identified. Suppose $i = i(h)$; we handle B_i. The proof for β_i is similar. We have $\ell_e = 0$, hence also $d\ell_e = 0$ on $\partial_e \mathcal{M}_G$. Thus the only difference between $\omega^{\partial_e G}$ and $\mathcal{B}^*\omega_i^{\mathcal{B}_e G}$ is that the former may contain terms with $d\ell_{s_2 h}$ or $d\ell_{s_2^{-1} h}$, while the latter depends only on their sum, by the definition of $\mathcal{B}_e$. Choose a good ordering n in the sense of Definition 5.30, so that half-edges of the i^{th} face appear first, and some half-edge $h' \neq h, s_2 h$ is the first edge in the ordering. One can always find such a half-edge. Otherwise, the i^{th} face is bounded by exactly two edges, h and $s_2 h$, which therefore must be a boundary half-edge, and in particular $K(s_2 h) = 1$. But then the sum of K on the i^{th} face is even, which is impossible for a Kasteleyn orientation.

In $\mathcal{B}_e G$ we choose a good ordering n' for which h', identified as an edge of $\mathcal{B}_e G$, is the first half-edge. Suppose $s_2^{-1} h$ is the j^{th} half-edge in n, so that h and $s_2 h$ are the $j+1^{\text{st}}$ and $j+2^{\text{nd}}$ edges. Write ℓ_a for $\ell_{n^{-1}(a)}$. Then

$$\begin{aligned} p_i^2 \omega_i^G|_{\partial_e \mathcal{M}_G} &= \sum_{a<b} d\ell_a \wedge d\ell_b \\ &= \sum_{\substack{a<b \\ a,b \neq j, j+1, j+2}} d\ell_a \wedge d\ell_b + \sum_{a<j} d\ell_a \wedge (d\ell_j + d\ell_{j+2}) \\ &\qquad + \sum_{j+2<a} (d\ell_j + d\ell_{j+2}) \wedge d\ell_a + d\ell_j \wedge d\ell_{j+2} \\ &= p_i^2 \mathcal{B}^* \omega_i^{\mathcal{B}_e G} + d\ell_j \wedge d\ell_{j+2}. \end{aligned}$$

In the last equality we used the fact that $\ell^{\mathcal{B}_e G}_{n'^{-1}(j)} = \ell_{n^{-1}(j)} + \ell_{n^{-1}(j+2)}$, and for $a \neq j$,

$$\ell^{\mathcal{B}_e G}_{n'^{-1}(a)} = \ell_{e_{a+w(a)}}, \quad \text{where } w(a) = \begin{cases} 0 & \text{if } a < j, \\ 2 & \text{otherwise.} \end{cases} \qquad \square$$

Notation 6.5 Recall Notation 4.37 and Remark 3.5. For G and e as above, given a (S,l)–set L and $i \in S$, we define the form Φ^i_L on the sphere bundle $S_L \to \partial_e \mathcal{M}_G$ by

$$\Phi^i_L = \Phi(\{r_j\}_{j\in S}, \{\alpha'_j\}_{j\in S}, \{\omega'_j\}_{j\in S}) = \Phi^{\partial_e G}(\{r_j\}_{j\in S}, \{\alpha'_j\}_{j\in S}, \{\omega'_j\}_{j\in S}),$$

where α'_j is a *copy* of $\mathcal{B}^*\alpha^{\mathcal{B}_e G}_{L(j)}$ for $j \neq i$, and $\alpha'_i = \beta_{L(i)}$. Similarly, $\omega'_j = \mathcal{B}^*\omega^{\mathcal{B}_e G}_{L(j)}$, unless $j = i$, and then $\omega'_i = B_{L(i)}$. As usual, $\overline{\Phi}^i_L = p^{2L}\Phi^i_L$. As in Remark 3.5, when $S \subset [d]$ we will also extend the domain of Φ^i_L by allowing $\sum_{i\in S} r_i^2$ to vary.

From now until the end of this subsection, we fix a (d,l)–set L, and let E_L be the corresponding bundle.

Lemma 6.6 *Let s be a special canonical multisection of E_L. Take $G \in \mathcal{SR}^m_{g,k,l}$ arbitrary and e an effective bridge of G, with h its illegal side. Letting D' be an L–decoration of G, write $L' = L(D')$. Then*

$$\int_{\partial_e \mathcal{M}_G(\boldsymbol{p})} s^*(W_G \overline{\Phi}_{L\setminus L'}) = \sum_{j\in(L\setminus L')_{i(h)}} \int_{\partial_e \mathcal{M}_G(\boldsymbol{p})} W_G s^*(\overline{\Phi}^j_{L\setminus L'}).$$

It should be noted that different decorations D' and D'' which determine the same set $L(D') = L(D'')$ will give rise to the same integral. The decorations, as mentioned above, are introduced only in order to keep track of the combinatorics of integrals that will appear in the iterative integration process below.

Proof Write $S = \bigcup_{h\in s_1 H^B} D'_h$, so that $L'\colon S \to [l]$ is a restriction of $L\colon [d] \to [l]$. We first use (19) and Notation 4.37 to write $\Phi_{L\setminus L'}$ explicitly:

$$(33)\quad \Phi_{L\setminus L'}(\{r_i\}_{i\in S^c}, \{\widehat{\alpha}_i\}_{i\in S^c}, \{\widehat{\omega}_i\}_{i\in S^c}) \\ = \sum_{k=0}^{|S^c|-1} 2^k k! \sum_{i\in S^c} r_i^2 \widehat{\alpha}_i \wedge \sum_{\substack{I\subseteq S^c\setminus\{i\} \\ |I|=k}} \bigwedge_{j\in I} (r_j\, dr_j \wedge \widehat{\alpha}_j) \wedge \bigwedge_{h\notin I\cup\{i\}} \widehat{\omega}_h,$$

where $\widehat{\omega}_j$ is Kontsevich's two-form $\omega_{L(j)}$, and $\widehat{\alpha}_j$ is a copy of Kontsevich's one-form $\alpha_{L(j)}$. This is a form of degree $\dim_{\mathbb{R}} \mathcal{M}_{\partial_e G} = \dim_{\mathbb{R}} \mathcal{M}_{\mathcal{B}_e G} + 1$. We obtain $\Phi^i_{L\setminus L'}$ by the same formula, after replacing $\widehat{\alpha}_i$ and $\widehat{\omega}_i$ by $\beta_{L(i)}$ and $B_{L(i)}$, respectively.

Now, the function W_G does not depend on variables of the fiber of the sphere bundle, hence it can be taken out of the pullback. By the definitions of the forms we can write

$$\widehat{\alpha}_j = \mathcal{B}^*\widehat{\alpha}^{\mathcal{B}_e G}_j + \beta_{L(j)} \quad \text{and} \quad \widehat{\omega}_j = \mathcal{B}^*\omega^{\mathcal{B}_e G}_{L(j)} + B_{L(j)},$$

where $\hat{\alpha}_j^{\mathcal{B}_e G}$ is a copy of $\alpha_{L(j)}^{\mathcal{B}_e G}$. We now substitute this in $\bar{\Phi}_{L\setminus L'}$, and expand (33) multilinearly.

Write $i = i(h) \in [l]$. Any term containing β_a or B_a for $a \neq i$ will vanish, by Observation 6.4.

Similarly, any term in the expansion that contains either β_a twice, or B_a twice, or β_i and B_i once, will vanish, as a consequence of a multiple appearance of $d\ell_{s_2^{-1}h}$.

By Proposition 5.46, $s|_{\partial_e \mathcal{M}_G}$ is pulled back from $\mathcal{M}_{\mathcal{B}_e G}$. Now, a term in $s^*\Phi_{L\setminus L'}$ with no B_i or β_i is pulled back from $\mathcal{M}_{\mathcal{B}_e G}$. But its degree is $\dim_{\mathbb{R}} \mathcal{M}_{\mathcal{B}_e G} + 1$. Thus, it vanishes for dimensional reasons.

We are left with terms containing a single β_i or B_i. These β_i or B_i are in fact $\beta_{L(j)}$ or $B_{L(j)}$ for some $j \in S^c$ which is mapped by L to i, meaning $j \in (L \setminus L')_i$. The lemma follows. □

The second main lemma we need is the following.

Lemma 6.7 *Fix $m > 0$, $G \in \mathcal{SR}^m_{g,k,l}$ and $D \in \mathrm{Dec}(G, L)$, and write $L' = L(D)$. Then*

$$\sum_{(G',e',D') \in \mathcal{B}^{-1}(G,D)} \int_{\mathcal{M}_{\partial_{e'} G'}(\boldsymbol{p})} W_{G'} s^*(\bar{\Phi}^{\partial_{e'} G'})^{L' \setminus L(D')}_{L \setminus L(D')} = \int_{\mathcal{M}_G(\boldsymbol{p})} W_G \bar{\omega}_{L\setminus L'} + \int_{\partial \mathcal{M}_G(\boldsymbol{p})} W_G s^*(\bar{\Phi}^G)_{L\setminus L'}. \tag{34}$$

Importantly, $\int_{\mathcal{M}_G(\boldsymbol{p})} W_G \bar{\omega}_{L\setminus L'}$ does not depend on the multisection s, so this lemma pushes the dependence on s to lower-dimensional moduli. After iterating, it will allow us to completely remove the dependence of the integrals on s. This phenomenon is expected, from the geometric point of view, since it was proven in [35; 31] that the intersection numbers should be independent of the specific canonical multisection. And indeed, the lemma is enabled by the properties of canonical multisections, and will not be true for arbitrary, noncanonical, boundary conditions.

Proof For convenience we treat the case $|\mathrm{Aut}(G)| = 1$; the general case is handled similarly, but notation becomes more complicated. Put

$$E' = \{e \in E(G) \mid m(e) > 0\}.$$

Recall Notation 5.42. Suppose $(G', e') \in \mathcal{B}^{-1} G$ is $\mathcal{B}^{-1}_{e,a+1} G$ for some $e \in E'$ and $a + 1 \in [m(e)]$. Fix $h \in D_e$, and let

$$D(G', h) := \{D' \mid (G', D') \in \mathcal{B}^{-1}(G, D),\ h \notin L(D')\}.$$

In words, $D(G',h)$ is the set of decorations of G' in $\mathcal{B}^{-1}(G,D)$ such that the only element of L' that they miss is h. Such decorations are determined by how we split the elements in $D_e \setminus \{h\}$ into sets of sizes a and $m(e)-1-a$ that will decorate the two edges in $\mathcal{B}_{e'}^{-1}e$ — the edges which, after contracting e' and forgetting its illegal side, form e. Thus,

$$|D(G',h)| = \binom{m(e)-1}{a}.$$

Let $e_1 = s_2^{-1}e'$ and $e_2 = s_2 e'$ be the two half-edges of G' mapped under $\mathcal{B}_{e'}$ to e. As explained, $m(e_1) = a$ and $m(e_2) = m(e)-a-1$. Put $\ell'_e = \ell_{e_1}$. For fixed G' and h we have the equality

$$\int_{\mathcal{M}_{\partial_{e'}G'}(\boldsymbol{p})} W_{G'} s^* \overline{\Phi}^{L'\setminus L(D')}_{L\setminus L(D')} = \int_{\mathcal{M}_{\partial_{e'}G'}(\boldsymbol{p})} W_{G'} s^* \overline{\Phi}^{h}_{L\setminus L(D')},$$

hence the left-hand side of this equation is independent of D'. We will now show that

$$\begin{aligned}(35)\qquad &\sum_{D'\in D(G',h)} \int_{\mathcal{M}_{\partial_{e'}G'}(\boldsymbol{p})} W_{G'} s^* \overline{\Phi}^{h}_{L\setminus L(D')} \\ &= \int_{\mathcal{M}_G(\boldsymbol{p})} \binom{m(e)-1}{a} \Bigg(\prod_{f\in E'\setminus\{e\}} \frac{\ell_f^{2m(f)}}{(m(f)+1)!} \Bigg) \\ &\qquad\qquad \cdot \int_0^{\ell_e} \frac{(\ell'_e)^{2a}(\ell_e-\ell'_e)^{2(m(e)-a-1)}}{(a+1)!\,(m(e)-a)!} (A_{e,h}+B_{e,h}+C_e),\end{aligned}$$

where

$$A_{e,h} = r_h^2(\ell_e - \ell'_e)\,d\ell'_e \wedge \sum_{n\ge 0} 2^n n! \sum_{\substack{|I|=n\\ I\subseteq L\setminus L'}} \Big(\bigwedge_{j\in I} r_j\,dr_j \wedge \hat{\alpha}_j\Big) \wedge \bigwedge_{j\in L\setminus(I\cup L')} \overline{\omega}_{L(j)},$$

$$\begin{aligned}B_{e,h} = r_h\,dr_h \wedge (\ell_e-\ell'_e)\,d\ell'_e \wedge &\sum_{i\in L\setminus L'} r_i^2\hat{\alpha}_i \\ &\wedge \sum_{n\ge 0} 2^{(n+1)}(n+1)! \sum_{\substack{|I|=n\\ I\subseteq L\setminus(L'\cup\{i\})}} \Big(\bigwedge_{j\in I} r_j\,dr_j\wedge\hat{\alpha}_j\Big) \wedge \bigwedge_{j\in L\setminus(L'\cup I\cup\{i\})} \overline{\omega}_{L(j)},\end{aligned}$$

$$\begin{aligned}C_e = d\ell'_e \wedge d\ell_e \wedge &\sum_{i\in L\setminus L'} r_i^2\hat{\alpha}_i \\ &\wedge \sum_{n\ge 0} 2^n n! \sum_{\substack{|I|=n\\ I\subseteq L\setminus(L'\cup\{i\})}} \Big(\bigwedge_{j\in I} r_j\,dr_j\wedge\hat{\alpha}_j\Big) \wedge \bigwedge_{j\in L\setminus(L'\cup I\cup\{i\})} \overline{\omega}_{L(j)},\end{aligned}$$

where $\widehat{\alpha}_i$ is a copy of $\overline{\alpha}_{L(i)}$. Before proving this equation, observe that $A_{e,h}$, $B_{e,h}$ and C_e *depend* on the multisection s through the sphere bundle fiber variables $r_i = r_i(s)$ and $\overline{\alpha}_i = \overline{\alpha}_i(s)$, but we omit s from the notation. However, because s is special canonical, it follows from the second item of Proposition 5.46 that $s(x, \ell'_e)$ for $x \in \mathcal{M}_G$ and $\ell'_e \in [0, \ell_e]$ depends only on x and not on ℓ'_e, where we have used the identification of Proposition 5.48. Thus, the same is true for the variable r_i and the form $\widehat{\alpha_i}$. Therefore, importantly, $A_{e,h}$, $B_{e,h}$ and C_e are *independent* of a, and their only dependence on ℓ'_e and $d\ell'_e$ is through the terms which explicitly involve them.

The last equation follows from the following facts. First, the multiplicity

$$\binom{m(e)-1}{a}$$

comes from summing over the different decorations D', which all give the same contribution. Second, the term in $W_{G'}$ for the edge $f \in E' \setminus \{e\}$ is

$$\frac{\ell_f^{2m(f)}}{(m(f)+1)!}.$$

The corresponding terms for e_1 and e_2 are, respectively,

$$\frac{(\ell'_e)^{2a}}{(a+1)!} \quad \text{and} \quad \frac{(\ell_e - \ell'_e)^{2(m(e)-a-1)}}{(m(e)-a)!},$$

Third, Proposition 5.48 reduces the integration over $\mathcal{M}_{\partial_{e'}G'}(\boldsymbol{p})$ to the repeated integral obtained by first integrating over $\mathcal{M}_G(\boldsymbol{p})$ and then over the location of the node on the edge e, which is encoded by ℓ'_e. This inner integration is precisely the integration $\int_0^{\ell_e}$ (with respect to $d\ell'_e$). Next, recall that, with $S = \bigcup_{h \in s_1 H^B} D_h$,

$$\begin{aligned}\overline{\Phi}^h_{L\setminus L'}&(\{r_i\}_{i\in S^c}, \{\widehat{\alpha}_i\}_{i\in S^c}, \{\widehat{\omega}_i\}_{i\in S^c}) \\ &= \sum_{k=0}^{|S^c|-1} 2^k k! \sum_{i \in S^c} r_i^2 \widehat{\alpha}_i \wedge \sum_{\substack{I \subseteq S^c \setminus \{i\} \\ |I|=k}} \bigwedge_{j\in I} (r_j\, dr_j \wedge \widehat{\alpha}_j) \wedge \bigwedge_{f \notin I \cup \{i\}} \widehat{\omega}_f,\end{aligned}$$

where for $j \neq h$, $\widehat{\omega}_j = \overline{\omega}_{L(j)}$ and $\widehat{\alpha}_j$ is a copy of $\overline{\alpha}_{L(j)}$, while $\widehat{\omega}_h = p_h^2 B_{L(h)}$ and $\widehat{\alpha}_h = p_h^2 \beta_h$. Using Observation 6.4, the sum of terms which have $i = h$ in the second summation is precisely $A_{e,h}$. The sum of terms with $i \neq h$ in which I contains h is $B_{e,h}$, while the remaining terms sum to C_e.

We shall use the following proposition.

Proposition 6.8 *We have*

(a) $$\sum_{a=0}^{m-1}\binom{m-1}{a}\int_0^y \frac{x^{2a}(y-x)^{2(m-a)-1}}{(a+1)!\,(m-a)!}\,dx=\frac{y^{2m}}{(m+1)!},$$

(b) $$\sum_{a=0}^{m-1}\binom{m-1}{a}\int_0^y \frac{x^{2a}(y-x)^{2(m-a-1)}}{(a+1)!\,(m-a)!}\,dx=\frac{2y^{2m-1}}{(m+1)!}.$$

Still fixing e and $h\in D_e$, we now apply Proposition 6.8, the fact that $A_{e,h}$, $B_{e,h}$ and C_e are independent of a, and that r_i and $\widehat{\alpha}_i$ are independent of ℓ'_e, to sum equation (35) over $(G'_a,e'_a):=\mathcal{B}_{e,a+1}^{-1}G$, where $a=0,\dots,m(e)-1$.

We obtain

$$\sum_{a=0}^{m(e)-1}\sum_{D'\in D(G'_a,h)}\int_{\mathcal{M}_{\partial_{e'}G'(\boldsymbol{p})}} W_{G'}s^*\Phi^h_{L\setminus L(D')} \tag{36}$$
$$=\int_{\mathcal{M}_G(\boldsymbol{p})}\prod_{f\in E'\setminus\{e\}}\frac{\ell_f^{2m(f)}}{(m(f)+1)!}\biggl\{\frac{\ell_e^{2m(e)}}{(m(e)+1)!}(\widetilde{A}_{e,h}+\widetilde{B}_{e,h})+\frac{2\ell_e^{2m(e)-1}d\ell_e}{(m(e)+1)!}\wedge Y\biggr\},$$

where

$$\widetilde{A}_{e,h}=r_h^2\sum_{m\ge 0}2^m m!\sum_{\substack{|I|=m\\ I\subseteq L\setminus L'}}\Bigl(\bigwedge_{j\in I}r_j\,dr_j\wedge\widehat{\alpha}_j\Bigr)\wedge\bigwedge_{j\in L\setminus(I\cup L')}\bar{\omega}_{L(j)},$$

$$\widetilde{B}_{e,h}=-r_h\,dr_h\wedge\sum_{i\in L\setminus L'}r_i^2\widehat{\alpha}_i$$
$$\wedge\sum_{m\ge 0}2^{(m+1)}(m+1)!\sum_{\substack{|I|=m\\ I\subseteq L\setminus(L'\cup\{i\})}}\Bigl(\bigwedge_{j\in I}r_j\,dr_j\wedge\widehat{\alpha}_j\Bigr)\wedge\bigwedge_{j\in L\setminus(L'\cup I\cup\{i\})}\bar{\omega}_{L(j)}$$

and

$$Y=\sum_{i\in L\setminus L'}r_i^2\widehat{\alpha}_i\wedge\sum_{m\ge 0}2^m m!\sum_{\substack{|I|=m\\ I\subseteq L\setminus(L'\cup\{i\})}}\Bigl(\bigwedge_{j\in I}r_j\,dr_j\wedge\widehat{\alpha}_j\Bigr)\wedge\bigwedge_{j\in L\setminus(L'\cup I\cup\{i\})}\bar{\omega}_{L(j)}.$$

The next step is to eliminate r_h terms, for $h \in L'$. For this, put

$$X = \Big(1 - \sum_{h\in L\setminus L'} r_h^2\Big) \sum_{m\geq 0} 2^m m! \sum_{\substack{|I|=m\\ I\subseteq L\setminus L'}} \Big(\bigwedge_{j\in I} r_j\, dr_j \wedge \hat{\alpha}_j\Big) \wedge \bigwedge_{j\in L\setminus(I\cup L')} \bar{\omega}_{L(j)}$$
$$+ \Big(\sum_{h\in L\setminus L'} r_h\, dr_h\Big) \wedge \sum_{i\in L\setminus(L'\cup\{h\})} r_i^2 \hat{\alpha}_i$$
$$\wedge \sum_{m\geq 0} 2^{(m+1)}(m+1)! \sum_{\substack{|I|=m\\ I\subseteq L\setminus(L'\cup\{i,h\})}} \Big(\bigwedge_{j\in I} r_j\, dr_j \wedge \hat{\alpha}_j\Big) \wedge \bigwedge_{j\in L\setminus(L'\cup I\cup\{i\})} \bar{\omega}_{L(j)}.$$

Then since

$$\sum_{h\in L'} r_h^2 = 1 - \sum_{h\in L\setminus L'} r_h^2 \quad \text{and} \quad \sum_{h\in L'} r_h\, dr_h = -\sum_{h\in L\setminus L'} r_h\, dr_h,$$

we obtain

$$\sum_{\substack{e\in E'\\ h\in D_e}} (\widetilde{A}_{e,h} + \widetilde{B}_{e,h}) = X.$$

Therefore, summing equation (36) over $e \in E'$ and $h \in D_e$ gives

$$\sum_{(G',e',D')\in\mathcal{B}^{-1}(G,D)} \int_{\mathcal{M}_{\partial_{e'}G'}(\boldsymbol{p})} W_{G'} s^* \bar{\Phi}_{L\setminus L(D')}^{L(D)\setminus L(D')} \tag{37}$$
$$= \int_{\mathcal{M}_G(\boldsymbol{p})} \Big(\prod_{f\in E'} \frac{\ell_f^{2m(f)}}{(m(f)+1)!}\Big) X$$
$$+ \int_{\mathcal{M}_G(\boldsymbol{p})} \Big(\sum_{e\in E'} \Big(\prod_{f\in E'\setminus\{e\}} \frac{\ell_f^{2m(f)}}{(m(f)+1)!}\Big) \frac{2m(e)\ell_e^{2m(e)-1}\, d\ell_e}{(m(e)+1)!}\Big) \wedge Y,$$

where the factor $m(e)$ in the last term comes from the cardinality of D_e and the summation over h. Observe that $Y = \Phi_{L\setminus L'}$, where we stress that we do not require $\sum_{h\in L\setminus L'} r_h^2 = 1$, as in Remark 3.5. X here is the same as Z there, after substituting $L \setminus L'$ for $[n]$, $\hat{\alpha}_i$ for α_i and $\bar{\omega}_{L(i)}$ for ω_i. Thus, Remark 3.5 immediately gives that the right-hand side of (37) is

$$\int_{\mathcal{M}_G(\boldsymbol{p})} \Big\{ \prod_{e\in E'} \frac{\ell_e^{2m(e)}}{(m(e)+1)!} \bigwedge_{i\in L\setminus L'} \bar{\omega}_{L(i)} + d\Big(\prod_{e\in E'} \frac{\ell_e^{2m(e)}}{(m(e)+1)!} \bar{\Phi}_{L\setminus L'}\Big)\Big\}.$$

The claim now follows from Stokes' theorem. □

Proof of Proposition 6.8 We first prove part (b). Write

$$f(x)=\sum_{m=0}^{\infty}\frac{x^{2m}}{m!\,(m+1)!}.$$

The identity we need to prove is equivalent to

$$(f*f)(x)=f'(x),$$

where $*$ is the convolution

$$(f*g)(x)=\int_0^x f(y)g(x-y)\,dy.$$

Using the Laplace transform, the last equation is equivalent to

$$F^2(\lambda)=\lambda F(\lambda)-1,$$

where

$$F(\lambda)=\int_0^\infty e^{-\lambda x}f(x)\,dx$$

is the Laplace transform of f. Expanding F we obtain

$$\begin{aligned}(38)\qquad F&=\sum_{m=0}^{\infty}\frac{1}{m!\,(m+1)!}\int_0^\infty e^{-\lambda x}x^{2m}dx=\sum_{m=0}^{\infty}\frac{(2m)!}{m!\,(m+1)!}\lambda^{-2m-1}\\ &=\lambda^{-1}\frac{1-\sqrt{1-4\lambda^{-2}}}{2\lambda^{-2}}=\lambda\frac{1-\sqrt{1-4\lambda^{-2}}}{2}.\end{aligned}$$

The third equality is a consequence of the general binomial formula. Thus, we are left with verifying that

$$F^2(\lambda)=\tfrac{1}{2}\lambda^2(1-\sqrt{1-4\lambda^{-2}})-1=\lambda F(\lambda)-1,$$

which is straightforward.

The first identity is a consequence of the second. Indeed, Write

$$I_m=\sum_{a=0}^{m-1}\binom{m-1}{a}\int_0^y\frac{x^{2a}(y-x)^{2(m-a)-1}}{(a+1)!\,(m-a)!}\,dx,$$

$$J_m=\sum_{a=0}^{m-1}\binom{m-1}{a}\int_0^y\frac{x^{2a}(y-x)^{2(m-a-1)}}{(a+1)!\,(m-a)!}\,dx.$$

It suffices to show that

$$I_m=\tfrac{1}{2}yJ_m.$$

Indeed,

$$(39)\quad \begin{aligned} I_m &= \sum_{a=0}^{m-1} \binom{m-1}{a} \int_0^y \frac{x^{2a}(y-x)^{2(m-a)-1}}{(a+1)!\,(m-a)!}\,dx \\ &= y \sum_{a=0}^{m-1} \binom{m-1}{a} \int_0^y \frac{x^{2a}(y-x)^{2(m-a-1)}}{(a+1)!\,(m-a)!}\,dx \\ &\qquad\qquad - \sum_{a=0}^{m-1} \binom{m-1}{a} \int_0^y \frac{x^{2a+1}(y-x)^{2(m-a-1)}}{(a+1)!\,(m-a)!}\,dx \\ &= yJ_m - \sum_{a=0}^{m-1} \binom{m-1}{a} \int_0^y \frac{(y-t)^{2a+1}t^{2(m-a-1)}}{(a+1)!\,(m-a)!}\,dx \\ &= yJ_m - I_m, \end{aligned}$$

where the second equality follows from opening one $(y-x)$ term, and the third follows from the substitution $t = y - x$. □

In order to be able to write an expression for the open intersection numbers we need the following observation.

Observation 6.9 *Suppose $G \in \mathcal{SR}^m_{g,k,l}$, and let e be an edge with $m(e) > 0$. Then for any decoration D,*

$$\int_{\partial_e \mathcal{M}_G(\boldsymbol{p})} W_G s^* \overline{\Phi}_{L\setminus L(D)} = 0.$$

Proof It follows from the definition of W_G that $W_G|_{\mathcal{M}_{\partial_e G}(\boldsymbol{p})} = 0$ identically. □

We can now state and prove the integral form of the combinatorial formula. We recall that $d = \frac{1}{2}(3g - 3 + k + 2l)$.

Theorem 6.10 *Let $L\colon [d] \to [l]$ be a (d,l)–set, with $a_i = |L_i|$ for $i \in [l]$. Then*

$$(40)\quad \boldsymbol{p}^{2L} 2^{\frac{1}{2}(g+k-1)} \langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle = \sum_{G \in \mathcal{OSR}^*_{g,k,l}} \sum_{D \in \mathrm{Dec}(G,L)} \int_{\mathcal{M}_G(\boldsymbol{p})} W_G \overline{\omega}_{L\setminus L(D)},$$

where the collection $\mathcal{OSR}^m_{g,k,l}$ for $m \geq 0$ is defined in Definition 5.36.

Proof Define

$$A_m = \sum_{(G,[K]) \in \mathcal{SR}^m_{g,k,l}} \sum_{D \in \mathrm{Dec}(G,L)} \int_{\mathcal{M}_{(G,[K])}(\boldsymbol{p})} W_G \overline{\omega}_{L\setminus L(D)},$$

$$S_m = \sum_{(G,[K]) \in \mathcal{SR}^m_{g,k,l}} \sum_{D \in \mathrm{Dec}(G,L)} \int_{\partial \mathcal{M}_{(G,[K])}(\boldsymbol{p})} W_G s^* \overline{\Phi}_{L\setminus L(D)},$$

where s is a nowhere-vanishing special canonical multisection. We will begin by showing that

$$S_m = A_{m+1} + S_{m+1}, \tag{41}$$

and that

$$p^{2L} 2^{\frac{1}{2}(g+k-1)} \langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle = A_0 + S_0. \tag{42}$$

For the first claim, consider S_m. Recall that for any G,

$$\partial \mathcal{M}_{(G,[K])} = \bigcup_{[e] \in [E(G)]} \partial_e \overline{\mathcal{M}}_{(G,[K])} = \bigcup_{[e] \in [E(G)]} \overline{\mathcal{M}}_{\partial_e (G,[K])}.$$

Since for different edges the boundary cells intersect in positive codimension, the integral over the union is just the sum over the edges e of the integrals over $\partial_e \overline{\mathcal{M}}_{(G,[K])}$.

For an edge e which is not a bridge or a boundary loop, by Corollary 5.50 we know that $\partial_e \overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p}) = -\partial_e \overline{\mathcal{M}}_{(G_e,[K_e])}(\boldsymbol{p})$, considered as oriented orbifolds, with the orientation induced as a boundary.

Now, $\mathrm{Dec}(G, L)$ and $\mathrm{Dec}(G_e, L)$ are the same sets, and it is easy to see that

$$W_G|_{\partial_e \overline{\mathcal{M}}_{(G,[K])}} = W_{G_e}|_{\partial_e \overline{\mathcal{M}}_{(G_e,[K_e])}}.$$

Thus, given a decoration D, and using the first item of Proposition 5.46,

$$\int_{\partial_e \overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p})} W_G s^* \overline{\Phi}_{L \setminus L(D)} = -\int_{\partial_e \overline{\mathcal{M}}_{(G_e,[K])}(\boldsymbol{p})} W_{G_e} s^* \overline{\Phi}_{L \setminus L(D)}.$$

For an effective loop e, the same argument, only using Corollary 5.51 instead of Corollary 5.50, and item (c) of Proposition 5.46 instead of item (a), shows that given a decoration D,

$$\int_{\partial_e \overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p})} W_G s^* \overline{\Phi}_{L \setminus L(D)} = -\int_{\partial_e \overline{\mathcal{M}}_{(G_e,[K])_e}(\boldsymbol{p})} W_{G_e} s^* \overline{\Phi}_{L \setminus L(D)}.$$

We should note that this is the second place that we use s being special canonical.

If e is a bridge or a boundary loop which is not effective, from Observation 6.9, for any decoration D,

$$\int_{\partial_e \mathcal{M}_{(G,[K])}(\boldsymbol{p})} W_G s^* \overline{\Phi}_{L \setminus L(D)} = 0.$$

Thus, we can write

$$S_m = \sum_{(G,[K])\in\mathcal{SR}^m_{g,k,l}} \sum_{D\in\mathrm{Dec}(G,L)} \sum_{[e]\in[\mathrm{Br}^{\mathrm{eff}}(G)]} \int_{\mathcal{M}_{\partial_e(G,[K])}(\boldsymbol{p})} W_G s^*\overline{\Phi}_{L\setminus L(D)}.$$

Applying Lemma 6.6, we obtain

$$S_m = \sum_{(G,[K])\in\mathcal{SR}^m_{g,k,l}} \sum_{D\in\mathrm{Dec}(G,L)} \sum_{[e]\in[\mathrm{Br}^{\mathrm{eff}}(G)]} \sum_{j\in(L\setminus L(D))_{i(e)}} \int_{\mathcal{M}_{\partial_e(G,[K])}(\boldsymbol{p})} W_G s^*\overline{\Phi}^j_{L\setminus L(D)}.$$

When e is an effective bridge, then $G' = \mathcal{B}_e(G,[K]) \in \mathcal{SR}^{m+1}_{g,k,l}$. We should note that this operation is also responsible for the appearance of ghost components, which result from contracting a boundary edge between two legal boundary tails. In addition, $j \in (L \setminus L(D))_{i(e)}$ induces a single decoration D' of G', which is defined by $(G,D) \in \mathcal{B}^{-1}(G',D')$ and $j \in L(D')$. Moreover, any $(G',[K']) \in \mathcal{SR}^{m+1}_{g,k,l}$ with $D' \in \mathrm{Dec}(G',L)$ is obtained in this way; see Observation 5.41. Hence, we can apply Lemma 6.7 and get

$$\begin{aligned} S_m = &\sum_{(G,[K])\in\mathcal{SR}^{m+1}_{g,k,l}} \sum_{D\in\mathrm{Dec}(G,L)} \int_{\mathcal{M}_{(G,[K])}(\boldsymbol{p})} W_G \overline{\omega}_{L\setminus L(D)} \\ &\qquad + \sum_{(G,[K])\in\mathcal{SR}^{m+1}_{g,k,l}} \sum_{D\in\mathrm{Dec}(G,L)} \int_{\partial\overline{\mathcal{M}}_{(G,[K])}(\boldsymbol{p})} W_G s^*\overline{\Phi}_{L\setminus L(D)} \\ = &A_{m+1} + S_{m+1}, \end{aligned}$$

as claimed.

For the second claim, using Lemma 4.47, we can write

$$\begin{aligned} &p^{2L} 2^{\frac{1}{2}(g+k-1)} \langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle \\ &\quad = \sum_{(G,[K])\in\mathcal{SR}^0_{g,k,l}} \int_{\mathcal{M}_{(G,[K])}(\boldsymbol{p})} \overline{\omega}_L \\ &\qquad\qquad + \sum_{(G,[K])\in\mathcal{SR}^0_{g,k,l}} \sum_{[e]\in[\mathrm{Br}(G)\cup\mathrm{Loop}(G)]} \int_{\mathcal{M}_{\partial_e(G,[K])}(\boldsymbol{p})} s^*\overline{\Phi}_L. \end{aligned}$$

Note that this is the nonnodal case, so all bridges and boundary loops are effective and the decorations are empty. The cancellation-in-pairs argument used above for the contribution of the integrals over edges which are neither boundary loops nor bridges shows, in particular, that

$$\sum_{(G,[K])\in\mathcal{SR}^0_{g,k,l}} \sum_{[e]\in[\mathrm{Br}(G)\cup\mathrm{Loop}(G)]} \int_{\mathcal{M}_{\partial_e(G,[K])}(\boldsymbol{p})} s^*\bar{\Phi}_L$$
$$= \sum_{(G,[K])\in\mathcal{SR}^0_{g,k,l}} \sum_{[e]\in[E(G)]} \int_{\mathcal{M}_{\partial_e(G,[K])}(\boldsymbol{p})} s^*\bar{\Phi}_L = S_0,$$

which, combined with the previous equation, gives (42).

Iterating (41) for $m \geq 0$ and using (42), we see that the left-hand side of equation (40) is $\sum_{m\geq 0} A_m$.

We now claim:

Proposition 6.11 *If G is a nodal graph such that on at least one boundary component there is an even total number of boundary marked points and legal nodes, then*

$$\int_{\mathcal{M}_{(G,[K])}(\boldsymbol{p})} W_G \bar{\omega}_{L\setminus L(D)} = 0.$$

The proof is given in Section 6.2; see Lemma 6.19. Thus,

$$\sum_{m\geq 0} A_m = \sum_{m\geq 0} \sum_{(G,[K])\in\mathcal{OSR}^m_{g,k,l}} \sum_{D\in\mathrm{Dec}(G,L)} \int_{\mathcal{M}_{(G,[K])}(\boldsymbol{p})} W_G \bar{\omega}_{L\setminus L(D)},$$

as claimed. □

Observation 6.12 *We have*

$$|\mathrm{Dec}(G,L)| = \binom{L_i}{\{m(e) \mid e\in E,\, i(e)=i\}}$$
$$= \prod_{i\in[l]} \frac{L_i!}{\left(\prod_{\{e\in E\mid i(e)=i\}} m(e)!\right)\left(L_i - \sum_{\{e\in E\mid i(e)=i\}} m(e)\right)!}.$$

Thus, with the above notation, we have

$$2^{\frac{1}{2}(g+k-1)} \prod_{i\in[l]} p_i^{2a_i} \langle \tau_{a_1}\cdots\tau_{a_l}\sigma^k\rangle$$
$$= \sum_{m\geq 0} \sum_{(G,[K])\in\mathcal{OSR}^m_{g,k,l}} \left(\prod_{i\in[l]} \binom{a_i}{\{m(e)\mid e\in E,\, i(e)=i\}}\right) \int_{\mathcal{M}_{(G,[K])}(\boldsymbol{p})} W_G \bar{\omega}_{L\setminus L(D)}$$
$$= \sum_{m\geq 0} \sum_{(G,[K])\in\mathcal{OSR}^m_{g,k,l}} \left(\prod_{i\in[l]} \frac{a_i!}{\left(a_i - \sum_{\{e\in E\mid i(e)=i\}} m(e)\right)!}\right) \int_{\mathcal{M}_{(G,[K])}(\boldsymbol{p})} \widetilde{W}_G \bar{\omega}_{L\setminus L(D)},$$

where $\widetilde{W}_G$ is defined in Definition 6.1, and $D \in D(G, L)$ are arbitrary decorations. Summing over all possible L and dividing by $d!$ we get

$$2^{\frac{1}{2}(g+k-1)} \sum_{\sum a_i = d} \prod_{i \in [l]} \frac{p_i^{2a_i}}{a_i!} \langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle = \sum_{m \geq 0} \sum_{(G,[K]) \in \mathcal{OSR}^m_{g,k,l}} \int_{\mathcal{M}_{(G,[K])}(\boldsymbol{p})} \widetilde{W}_G \frac{\overline{\omega}^{d-m}}{(d-m)!}. \tag{43}$$

Dimensional reasons give:

Observation 6.13 *Let L' be an l–set, and let $(G, [K]) \in \mathcal{OSR}^*_{g,k,l}$. Suppose that for some component $C \in C(G, [K])$,*

$$\dim(C) < \sum_{i \in I(C)} L'_i.$$

Then $\int_{\mathcal{M}_G} f \omega_{L'} = 0$ for any function f.

Now, $\overline{\omega} = \sum_{C \in C(G)} \overline{\omega}^C$, where $\overline{\omega}^C = \sum_{i \in I(C)} \overline{\omega}_i$. Thus, together with the observation, we get the following:

Corollary 6.14 *We have*

$$\widetilde{W}_G \frac{\overline{\omega}^{d-m}}{(d-m)!} = \prod_{C \in C(G)} \widetilde{W}_C \frac{(\overline{\omega}^C)^{\dim(C)}}{\dim(C)!}.$$

Thus,

$$\begin{aligned} \sum_{\sum a_i = d} \prod_{i \in [l]} \frac{p_i^{2a_i}}{a_i!} 2^{\frac{1}{2}(k-1)} \langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle &= \sum_{m \geq 0} \sum_{(G,[K]) \in \mathcal{OSR}^m_{g,k,l}} \int_{\mathcal{M}_{(G,[K])}(\boldsymbol{p})} \widetilde{W}_G \prod_{C \in C(G,[K])} \frac{(\overline{\omega}^C)^{\dim(C)}}{\dim(C)!} \\ &= \sum_{m \geq 0} \sum_{(G,[K]) \in \mathcal{OSR}^m_{g,k,l}} \prod_{C \in C(G,[K])} \int_{\mathcal{M}_C} \widetilde{W}_C \frac{(\overline{\omega}^C)^{\dim(C)}}{\dim(C)!}. \end{aligned} \tag{44}$$

In the above formula there may appear components C with $\dim(C) = 0$. These are precisely the ghost components and the genus 0 components with one internal tail and one legal boundary tail.

6.2 Power of 2

We aim now to gain a better understanding of the forms $\bigwedge dp_i \wedge (\bar{\omega}^d/d!)$ and $\mathfrak{o}_{(G,[K])}$ and their ratio.

Definition 6.15 For $(G,[K]) \in \mathcal{SR}^*_{g,k,l}$, define $s(G,[K])$ to be the sign of

$$\bigwedge dp_i \wedge \frac{\bar{\omega}^d}{d!} : \mathfrak{o}_{(G,[K])}.$$

For $G \in \mathcal{R}^*_{g,k,l}$, define

$$c_{\mathrm{spin}}(G) = \sum_{[K]\in[K(G)]} s(G,[K]).$$

Lemma 6.16 *For $G \in \mathcal{SR}^*_{g,k,l}$,*

$$\bigwedge dp_i \wedge \frac{\bar{\omega}^d}{d!} : \mathfrak{o}_{(G,[K])} = s(G,[K])c_{\mathrm{spin}}(G)2^{|V^I(G)|}.$$

In particular, $c_{\mathrm{spin}}(G) \geq 0$.

Proof Both the left-hand side and the right-hand side are multiplicative with respect to taking nonnodal components, by the first statement in Corollary 6.14 and the construction of $\mathfrak{o}_{(G,[K])}$. Thus, it is enough to prove the lemma for graphs in $\mathcal{SR}^{\mathbf{0}}_{g,k,l}$.

Recall that any class $[K]$ of Kasteleyn orientations is of size $2^{|V^I(G)|}$, by Lemma 5.8. In addition, by Lemma 5.29, the $\mathfrak{o}_{(G,K)}$ for different $K \in [K]$ are equal. Thus, the lemma is equivalent to the equality

$$\bigwedge dp_i \wedge \frac{\bar{\omega}^d}{d!} = \sum_{K\in K(G)} \mathfrak{o}_{(G,[K])}. \tag{45}$$

Recall that $\bar{\omega} = \sum_{i=1}^{l} \bar{\omega}_i$. Fix a good ordering n. To prove equation (45), it will be more comfortable to work with new variables $\ell_h, h \in H^I$, instead of $\ell_e, e \in E$. Set

$$H_{K,i} = \left\{h \in H_K \;\middle|\; \frac{h}{s_2} = i\right\}, \quad d_{K,i} = \frac{|H_{K,i}|-1}{2},$$
$$p_{K,i} = \sum_{h\in H_{K,i}} \ell_h, \qquad \bar{\omega}_{K,i} = \sum_{\substack{h_1,h_2\in H_{K,i} \\ n(h_1)<n(h_2)}} d\ell_{h_1} \wedge d_{\ell_{h_2}}.$$

Remark 6.17 Only $\bar{\omega}_{K,i}$ depends on the ordering n. For different orders the change in $\bar{\omega}_{K,i}$ is of the form $dp_{K,i} \wedge dx$, where x is a linear combination of $\{d\ell_h\}_h \in H_{K,i}$. Thus, for any a, the form $dp_{K,i} \wedge \bar{\omega}^a_{K,i}$ is independent of n.

Express each dp_i by $\sum_{h\in H_i} d\ell_h$, and express also each $\bar\omega_i$ in the $\{d\ell_h\}_{h\in H^I}$ basis as above. Our next aim is to show that

$$(46)\qquad \bigwedge dp_i \wedge \frac{\bar\omega^d}{d!} = \sum_{K\in K(G)} \bigwedge_{i\in[l]} dp_{K,i} \wedge \frac{\bar\omega_{K,i}^{d_{K,i}}}{d_{K,i}!} \pmod{I},$$

where I is the ideal $(d\ell_h - d\ell_{s_1h})_{h\in H^I}$. In order to show equation (46) we expand $\bigwedge dp_i \wedge (\bar\omega^d/d!)$ multilinearly, in terms of $\{d\ell_h\}_{h\in H^I}$, without cancellations. Any monomial which appears in this expression and contains exactly one of $d\ell_h, d\ell_{s_1h}$ for any $h\in H^I$ defines a unique Kasteleyn orientation K, defined by $K(h)=1$ if and only if $d\ell_h$ appears in the monomial. This is indeed a Kasteleyn orientation since any $h\in s_1H^B$ has $K(h)=1$, and for any $i\in[l]$, an odd number of variables of half-edges appear: one comes from dp_i, and the others come in pairs via powers of $\bar\omega_i$.

It is transparent that any Kasteleyn orientation $K\in K(G)$, is generated this way. Moreover, regrouping all terms which correspond to the same Kasteleyn orientation, and using the identity

$$\left(\sum_{i=1}^{2m+1} x_i\right) \wedge \frac{\left(\sum_{i<j} x_i\wedge x_j\right)^m}{m!} = x_1\wedge x_2\wedge\ldots\wedge x_{2m+1},$$

we get equation (46).

The "In particular" follows from the fact that $\bigwedge dp_i \wedge (\bar\omega^d/d!)$ and $s(G,[K])\mathfrak{o}_{(G,[K])}$ have the same sign. □

Proposition 6.18 *For $G\in\mathcal{SR}^0_{g,k,l}$ and $e\notin \mathrm{Br}(G)\cup\mathrm{Loop}(G)$,*

$$c_{\mathrm{spin}}(G) = c_{\mathrm{spin}}(G_e).$$

Proof It follows from Lemma 6.16 that

$$c_{\mathrm{spin}}(G) = \pm \sum_{[K']\in[K(G)]} \mathfrak{o}_{(G,[K'])} : \mathfrak{o}_{(G,[K])}$$

for any fixed $[K]\in[K(G)]$. If $K,K'\in K(G)$, then by the orientability of the moduli, Theorem 5.32, we see that

$$\mathfrak{o}_{(G,[K])} : \mathfrak{o}_{(G,[K'])} = \mathfrak{o}_{(G_e,[K_e])} : \mathfrak{o}_{(G_e,[K'_e])},$$

as $(G,[K]),(G_e,[K_e])$ and $(G,[K']),(G_e,[K'_e])$ parametrize adjacent cells. Thus, $c_{\mathrm{spin}}(G) = \pm c_{\mathrm{spin}}(G_e)$. But $c_{\mathrm{spin}}\ge 0$, hence the equality. □

Lemma 6.19 *If* $G \in \mathcal{R}^m_{g,k,l} \setminus \mathcal{OR}^m_{g,k,l}$, *then* $c_{\mathrm{spin}} = 0$.

Proof Again, as c_{spin} is multiplicative in nonnodal components, it is enough to consider the case of nonnodal graphs. Let $\partial\Sigma_b$ be a boundary with an even number of boundary marked points. Note that given a surface Σ and a boundary component $\partial\Sigma_b$, graded spin structures on Σ can be partitioned into pairs which differ exactly in the lifting of $\partial\Sigma_b$. Thus, we can partition $[K(G)]$ into pairs which differ exactly in the boundary conditions at $\partial\Sigma_b$. In combinatorial terms, for any pair $\{(G,[K_1])$ and $(G,[K_2])\}$ in the partition we can find $K_1 \in [K_1]$ and $K_2 \in [K_2]$ which agree everywhere, except on edges with exactly one vertex in $\partial\Sigma_b$, where they disagree. We shall show that $s(G,[K_1]) = -s(G,[K_2])$.

As a consequence of Proposition 6.18, $c_{\mathrm{spin}}(G,[K]) = c_{\mathrm{spin}}(G_e,[K_e])$ for $G \in \mathcal{R}^0_{g,k,l}$ and $e \notin \mathrm{Br}(G) \cup \mathrm{Loop}(G)$. By performing enough such Feynman moves at boundary edges of G, see Figure 14 moves (b) and (c), we may assume only one nonboundary edge emanates from $\partial\Sigma_b$. Let $2a$ denote the number of the boundary marked points on $\partial\Sigma_b$. Note that $\partial\Sigma_b$ is part of the boundary of a single face, say face 1. Let h and $s_1(h)$ be the internal half-edges which touch $\partial\Sigma_b$. Choose a good ordering n on G, so that $n(h) = 1$, $n(h_1) = 2, \dots, n(h_{2a+1}) = 2a+2$ and $n(s_1h) = 2a+3$, where $h_i \in H^I$ are the other half-edges on $\partial\Sigma_b$. This can always be done, possibly after interchanging h and s_1h. Choose any $K_1 \in [K_1]$ and $K_2 \in [K_2]$, which differ only in their values at h and s_1h. Thus, the sign difference between $\mathfrak{o}_{(G,[K_1])}$ and $\mathfrak{o}_{(G,[K_2])}$ is just $(-1)^{2a+1} = -1$, since we change only the location of the variable $d\ell_{h/s_1}$, by $2a+1$ spots. As claimed. $\square$

We can now prove Proposition 6.11.

Proof By Lemma 6.16, the proposition is equivalent to $c_{\mathrm{spin}}(G) = 0$. But $c_{\mathrm{spin}}(G) = \prod_{C \in C(G)} c_{\mathrm{spin}(C)}$, which is 0 by Lemma 6.19. $\square$

We can now also prove Corollary 5.51.

Proof As above, it is enough to prove it for smooth G. The case where e is a boundary loop is a special case of the graph considered in the proof of Lemma 6.19, and in particular we see that the orientation expressions for $(G,[K])$ and $(G_e,[K_e])$ are opposite. Recall that the map Ψ^{comb} preserves the edge-lengths of all edges, but changes the Kasteleyn orientation to $[K_e]$. By contracting these orientation expressions with the vector $-\partial/\partial\ell_e$, we see that the induced orientation on $\partial_e\overline{\mathcal{M}}_{(G_e,[K_e])}$ and the $(\Psi^{\mathrm{comb}})^*$–pushforward of the induced orientation on $\partial_e\overline{\mathcal{M}}_{(G,[K])}$ are opposite. $\square$

Lemma 6.20 *For $G \in \mathcal{OR}^0_{g,k,l}$, we have*

$$c_{\mathrm{spin}}(G) = 2^{\frac{1}{2}(g+b-1)},$$

where g is the genus of G, and b is the number of boundaries. For $G \in \mathcal{OR}^m_{g,k,l}$,

$$c_{\mathrm{spin}}(G) = \prod c_{\mathrm{spin}}(G_i),$$

where G_i are the smooth components of G.

Proof Again it is enough to consider nonnodal graphs. By Lemma 6.16, $c_{\mathrm{spin}}(G) \geq 0$. By Proposition 6.18 $c_{\mathrm{spin}}(G,[K]) = c_{\mathrm{spin}}(G_e,[K_e])$, whenever $G \in \mathcal{OSR}^0_{g,k,l}$ and $e \notin \mathrm{Br}(G) \cup \mathrm{Loop}(G)$. Thus, it is enough to calculate c_{spin} for the graph $\overline{G}$, where G is the graph constructed in Example 5.18; see Figure 17. We shall work with the notation of that example. We shall order the faces according to their labels, and we choose an ordering n of the edges of face 1 such that a_1 is the first edge. Choose a Kasteleyn orientation and write

$$\mathfrak{o}_G = W_1 \wedge W_2 \wedge \cdots \wedge W_{g_s} \wedge d\ell_{h_2} \wedge d\ell_{x_2} \wedge \cdots \wedge d\ell_{h_l} \wedge d\ell_{x_l} \wedge d\ell_{e_{1,0}} \wedge \cdots \wedge d\ell_{e_{1,k_1}} \\ \wedge R \wedge d\ell_{y_2} \wedge \cdots \wedge d\ell_{y_l},$$

where W_i is the wedge of $d\ell_{a_i}, d\ell_{b_i}, d\ell_{c_i}, d\ell_{d_i}, d\ell_{f_i}, d\ell_{g_i}$, according to the order induced by K, and R is the wedge of the remaining variables, according to the ordering. The ordering n, restricted to the half-edges which are involved in W_i, is

$$a_i, f_i, d_i, \overline{g}_i, c_i, \overline{f}_i, b_i, g_i.$$

There are four possibilities for $K(\overline{f}_i)$ and $K(\overline{g}_i)$. Let K_i^0 denote the set of possibilities with $K(\overline{f}_i)K(\overline{g}_i) = 0$. Let K_i^1 be the singleton made of the remaining possibility. One can check by hand that the form W_i is constant in K_i^0, and minus that constant in the fourth possibility.

The ordering restricted to the remaining edges is

$$b_{1,2}, e_{2,k_2+1}, b_{2,3}, e_{3,k_3+1}, \ldots, b_{b-1,b}, e_{b,0}, e_{b,1}, \ldots, e_{b,k_b}, \overline{b}_{b-1,b}, \\ e_{b-1,0}, e_{b-1,1}, \ldots, e_{b-1,k_{b-1}}, \overline{b}_{b-2,b-1}, e_{b-2,0}, \ldots, e_{2,k_2}\overline{b}_{1,2}.$$

The only freedom in K is in the values of $K(b_{j,j+1})$. The relative order of these edges is

$$b_{1,2}, b_{2,3}, \ldots, b_{b-1,b}, \overline{b}_{b-1,b}, \ldots, \overline{b}_{1,2}.$$

Observe that between $b_{j,j+1}$ and $\overline{b}_{j,j+1}$ in the ordering, there is an even number of half-edges. Thus, different assignments of $K(b_{j,j+1})$ do not change the orientation $\mathfrak{o}_G$. There are 2^{b-1} such assignments, where b is the number of boundary components.

To summarize, $s(G,[K])$ depends only on $\sum_i K(\overline{f_i})K(\overline{g}_i)$, which is just the parity of the graded spin structure (see Remark 5.19), and different parities give rise to different signs. By the calculation in Remark 5.19 we see that $c_{\mathrm{spin}}(G) = \pm 2^{\frac{1}{2}(g-b+1)+b-1}$, but as it cannot be negative we end with $c_{\mathrm{spin}}(G) = 2^{\frac{1}{2}(g+b-1)}$. $\square$

Remark 6.21 An analogous power of 2 appears in [25] when one wants to calculate the Laplace transform of the integral combinatorial formula. The method developed in this paper is also applicable to that calculation. It shows exactly where this power of 2 comes from, and how is it connected to spin structures. In fact, our c_{spin} can be thought as an open analog of the push down of the $r=2$–spin Witten's class to the spinless moduli; see [39].

Corollary 6.22 *For $G \in \mathcal{SR}^0_{g,k,l}$,*

$$\bigwedge dp_i \wedge \frac{\overline{\omega}^d}{d!} : \mathfrak{o}_{(G,[K])} = s(G,[K])2^{|V^I(G)|+\frac{1}{2}(g(G)+b(G)-1)}.$$

6.3 Laplace transform and the combinatorial formula

As in the closed case, a more compact formula may be obtained after performing a Laplace transform to Corollary 6.14.

Let λ_i be the variable dual to p_i and write, for $e = \{h_1, h_2 = s_1 h_1\}$,

$$\lambda(e) = \begin{cases} \dfrac{1}{\lambda_i+\lambda_j} & \text{if } i(h_1)=i \text{ and } i(h_2)=j, \\ \dfrac{1}{m(e)+1}\dbinom{2m(e)}{m(e)}\lambda_i^{-2m(e)-1} & \text{if } i(h_1)=i \text{ and } h_2 \in H^B. \end{cases}$$

We also define $\widetilde{\lambda}(e) = 1/\lambda(e)$ for an internal edge and $\widetilde{\lambda}(e) = \lambda_{i(e)}$ for a boundary edge of face i.

Applying the transform to the left-hand side of Corollary 6.14 gives

$$\begin{aligned}\int_{p_1,\dots,p_l>0} \bigwedge dp_i e^{-\sum \lambda_i p_i} \sum_{\sum a_i = d} \prod_{i\in[l]} \frac{p_i^{2a_i}}{a_i!} 2^{\frac{1}{2}(g+k-1)} \langle \tau_{a_1}\cdots\tau_{a_l}\sigma^k\rangle \\ = 2^{d+\frac{1}{2}(g+k-1)} \sum_{\sum a_i = d} \prod_{i\in[l]} \frac{(2a_i-1)!!}{\lambda_i^{2a_i+1}} \langle \tau_{a_1}\cdots\tau_{a_l}\sigma^k\rangle,\end{aligned}$$

where $d = \frac{1}{2}(k+2l+3g-3)$.

Transforming the right-hand side leaves us with

$$\sum_{m\geq 0}\sum_{G\in\mathcal{OSR}^m_{g,k,l}}\int_{p_1,\dots,p_l>0}\bigwedge dp_i e^{-\sum\lambda_i p_i}\prod_{C\in C(G,[K])}\int_{\mathcal{M}_C}\widetilde{W}_C\frac{(\overline{\omega}^C)^{\dim(C)}}{\dim(C)!}$$
$$=\sum_{m\geq 0}\sum_{G\in\mathcal{OSR}^m_{g,k,l}}\int_{p_1,\dots,p_l>0}\bigwedge dp_i e^{-\sum\tilde{\lambda}(e)\ell_e}\prod_{C\in C(G,[K])}\int_{\mathcal{M}_C}\widetilde{W}_C\frac{(\overline{\omega}^C)^{\dim(C)}}{\dim(C)!},$$

where we have used the fact that the perimeter of a face is the sum of its edges' lengths.

Recall that

$$\prod_{C\in C(G,[K])}\widetilde{W}_C=\prod_{e\in E^B(G)}\frac{\ell_e^{2m(e)}}{(m(e))!\,(m(e)+1)!}.$$

By Corollary 6.22, applied to $(G,[K])\in\mathcal{OSR}^0_{g,k,l}$, we have

$$\frac{\left(\bigwedge_{i\in[l]}dp_i\right)\dfrac{\overline{\omega}^d}{d!}}{\bigwedge_{e\in E(G)}d\ell_e}=s(G,[K])2^{|V^I(G)|+\frac{1}{2}(g(G)+b(G)-1)},$$

where the variables in the denominator are ordered by $\mathfrak{o}_{(G,[K])}$, and $|V^I|$, g and b are the number of internal vertices of G, its genus and the number of boundary components, respectively. In addition,

$$\sum_{[K]\in[K(G)]}s(G,[K])=c_{\mathrm{spin}}=2^{\frac{1}{2}(g+b-1)},$$

by Lemma 6.20. Moreover, since $\mathrm{Aut}(G)$ acts on $[K(G)]$, and is sign-preserving, we see that

$$\sum_{[K]\in[K(G)]}\frac{s(G,[K])}{|\mathrm{Aut}(G)|}=\sum_{[K]\in[K(G)]/\mathrm{Aut}(G)}\frac{s(G,[K])}{|\mathrm{Aut}(G,[K])|}.$$

Thus, for a fixed $G\in\mathcal{OR}^m_{g,k,l}$, summing over $\mathrm{for}^{-1}_{\mathrm{spin}}(G)$ using Observation 5.38, and recalling that $\mathcal{M}_{(G,[K])}\simeq\mathbb{R}^{E(G)}/|\mathrm{Aut}(G,[K])|$, we get

$$\sum_{[K]}\frac{1}{|\mathrm{Aut}(G,[K])|}\int_{p_1,\dots,p_l>0}\bigwedge dp_i e^{-\sum\tilde{\lambda}(e)\ell_e}\prod_{C\in C(G,[K])}\int_{\mathbb{R}^{E(C)}}\widetilde{W}_C\frac{(\overline{\omega}^C)^{\dim(C)}}{\dim(C)!}$$
$$=\frac{\prod_{C\in C(G)}c(C)}{|\mathrm{Aut}(G)|}\prod_{e\in E\setminus E^B}\int_0^\infty e^{-\tilde{\lambda}(e)\ell_e}\,d\ell_e\prod_{e\in E^B}\int_0^\infty e^{-\tilde{\lambda}(e)\ell_e}\frac{\ell_e^{2m(e)}}{m(e)!(m(e)+1)!}\,d\ell_e$$
$$=\frac{\prod_{C\in C(G)}c(C)}{|\mathrm{Aut}(G)|}\prod_{e\in E}\lambda(e),$$

where $c(C) = 2^{|V^I(C)|+g(C)+b(C)-1}$. Summing over all $G \in \mathcal{OR}^*_{g,k,l}$,

$$2^{d+\frac{1}{2}(g+k-1)} \sum_{\sum a_i = d} \prod_{i=1}^{l} \frac{(2a_i-1)!!}{\lambda_i^{2a_i+1}} \langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle = \sum_{G \in \mathcal{OR}^*_{g,k,l}} \frac{\prod_{C \in C(G)} c(C)}{|\mathrm{Aut}(G)|} \prod_{e \in E} \lambda(e).$$

This proves Theorem 1.5.

Open problem 3 The moduli space $\overline{\mathcal{M}}_{g,k,l}$ is disconnected, and is composed of components which parametrize different topologies, partitions of boundary markings along boundary components and graded structures. The boundary conditions of [35; 31] define in fact an intersection number on each such component, and their sum is what we denote in this work by $\langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle_g$. Using the techniques presented in this section one can actually calculate all these refined intersection numbers; see [2]. The intersection numbers $\langle \tau_{a_1} \cdots \tau_{a_l} \sigma^k \rangle_g$ are related to the KdV wave function, and therefore satisfy many recursion relations. A natural question is whether the refined numbers also satisfy interesting recursion relations, and whether they are related to an integrable hierarchy. The paper [2] proposes a conjecture in this direction.

Appendix Properties of the stratification

A.0.1 Proof of Proposition 4.25 Fix sets $\mathcal{I}$, $\mathcal{B}$ and $\mathcal{P}_0$. For a stable open ribbon graph G, write $\mathcal{M}_G = \mathbb{R}_+^{E(G)}/\mathrm{Aut}(G)$. Let $G_{g,\mathcal{B},(\mathcal{I},\mathcal{P}_0)}$ be the set of all such graphs with boundary markings $\mathcal{B}$, internal markings $\mathcal{I}$ and internal markings of perimeter zero $\mathcal{P}_0$. We will show that $\mathrm{comb}^{\mathbb{R}}$ maps $\overline{\mathcal{M}}^{\mathbb{R}}_{g,\mathcal{B},\mathcal{I}\cup\mathcal{P}_0}$ to $\coprod_{G_{g,\mathcal{B},(\mathcal{I},\mathcal{P}_0)}} \mathcal{M}_G(\boldsymbol{p})$ surjectively, and that it is one-to-one on smooth or effective loci.

Step 1 An antiholomorphic involution ϱ of a connected stable curve X is *separating* if X/ϱ is a connected orientable stable surface with boundary. X^{ϱ} is called the *real* locus. A *half* of X is a stable connected subsurface with boundary $\Sigma \subseteq X$ such that the composition $\Sigma \hookrightarrow X \to X/\varrho$ is a homeomorphism.

A *doubled* $(g,\mathcal{B},\mathcal{I}\cup\mathcal{P}_0)$*–surface* is a closed stable marked surface X with markings $\{x_i\}_{i\in\mathcal{B}}$ and $\{z_i,\bar{z}_i\}_{i\in\mathcal{I}\cup\mathcal{P}_0}$, together with a separating antiholomorphic involution ϱ and a preferred half Σ, satisfying

(a) $x_i \in X^{\varrho}$ for all i, and

(b) $z_i \in \mathrm{int}(\Sigma)$ for all i.

Observation A.1 *There is a natural one-to-one correspondence between open stable $(g, \mathcal{B}, \mathcal{I} \cup \mathcal{P}_0$–surfaces Σ and doubled $(g, \mathcal{B}, \mathcal{I} \cup \mathcal{P}_0)$–surfaces (X, ϱ, Σ), given by $\Sigma \to (D(\Sigma), \Sigma)$, where Σ is taken as a subset of $D(\Sigma)$.*

Note that all components of X^{ϱ} which are not isolated points are canonically oriented as boundaries of the distinguished half.

Step 2 Fix positive $\{p_i\}_{i \in \mathcal{I}}$. For convenience we denote by $\overline{\mathcal{I}}$ and $\overline{\mathcal{P}}_0$ the markings of $\bar{z}_i$ for $i \in \mathcal{I}, \mathcal{P}_0$. We now analyze the image of doubled surfaces (X, ϱ, Σ) under the (closed) map $\mathrm{comb}_{\boldsymbol{q}}$ defined on $\overline{\mathcal{M}}_{g,k+2l}$, where the perimeters $\boldsymbol{q}$ are defined so that the faces of z_i and $\bar{z}_i$ for $i \in \mathcal{I}$ have perimeter p_i, and the other points are boundary marked points or internal marked with perimeter 0. By the construction for closed surfaces, the image is a stable ribbon graph G in the sense of Definition 4.2, embedded in $\widetilde{X} = K_{\mathcal{B} \cup \mathcal{P}_0 \cup \overline{\mathcal{P}}_0}(X)$. Moreover, ϱ induces an involution on $\widetilde{X}$ and G, which we also denote by ϱ, and by Lemma 4.13 $\widetilde{X}^{\varrho} \subseteq G$. Faces and vertices marked by $\mathcal{I} \cup \mathcal{P}_0$ are in one distinguished half $\widetilde{\Sigma}$ of $\widetilde{X}$, where a half is defined analogously to the above.

Write $E^{\boldsymbol{B}}$ for ϱ–invariant edges. Let $H^{\boldsymbol{B}}$ be their halves which do not agree with the orientation induced by $\widetilde{\Sigma}$. Write $V^{\boldsymbol{B}}$ for ϱ–invariant vertices. Let $V^{\boldsymbol{I}}$ be vertices in $\mathrm{int}(\widetilde{\Sigma})$, let $H^{\boldsymbol{I}}$ be either half-edges in $s_1 H^{\boldsymbol{B}}$ or half-edges which intersect $\mathrm{int}(\widetilde{\Sigma})$, and let $E^{\boldsymbol{I}} = (H^{\boldsymbol{I}} \setminus s_1 H^{\boldsymbol{B}})/s_1$.

Observation A.2 *$H^{\boldsymbol{I}} \cup H^{\boldsymbol{B}}$ is invariant under s_1, and s_0 takes $H^{\boldsymbol{I}}$ to $H^{\boldsymbol{I}} \cup H^{\boldsymbol{B}}$.*

Indeed, if there were $h \in H^{\boldsymbol{I}}$ and $h' \notin H^{\boldsymbol{I}} \cup H^{\boldsymbol{B}}$ with $s_0 h = h'$, then there was a common face which contained h and $s_1 h'$. But then this face would intersect both $\mathrm{int}(\widetilde{\Sigma})$ and $\varrho(\mathrm{int}(\widetilde{\Sigma}))$, which is impossible.

Let v be a vertex, and consider its half-edges. The permutation s_0 acts on them, and also ϱ. Write B_v for the set of s_0–cycles which contain an element of $H^{\boldsymbol{B}}$, and write I_v for those cycles in $H^{\boldsymbol{I}}$. It is easy to see that no s_0–cycle contains more than two boundary edges. It follows from the observation that inside a cycle in B_v the half-edges are s_0–ordered as $h_1, \dots, h_{2r+2}$ so that

$$\begin{cases} h_1 \in s_1 H^{\boldsymbol{B}}, \\ h_i \in H^{\boldsymbol{I}} \setminus s_1 H^{\boldsymbol{B}} & \text{if } i \in [r+1] \setminus \{1\}, \\ h_i = \varrho(h_{i-r-1}) & \text{if } i \in [2r+2] \setminus [r+1]. \end{cases}$$

In particular, $h_{r+2} \in H^{\boldsymbol{B}}$ and $h_i \notin (H^{\boldsymbol{I}} \cup H^{\boldsymbol{B}})$ for $i \in [2r+2] \setminus [r+2]$. Define a permutation $\widetilde{s}_0$ of $H^{\boldsymbol{I}} \cup H^{\boldsymbol{B}}$ which is s_0 on $H^{\boldsymbol{I}}$, and otherwise, we are in the scenario just described, $\widetilde{s}_0 h_{r+2} = h_1$.

Define new marking assignments f^I, f^B and f^{P_0} as follows: f^I maps $i \in \mathcal{I}$ to the face containing z_i, f^B maps $i \in \mathcal{B}$ to the vertex x_i is mapped to, and f^{P_0} is defined similarly.

Recall Notation 4.17. Let $\widetilde{\mathcal{IT}}(g, I, B)$ be the set of isotopy types of smooth doubled (g, I, B)–surfaces. Write $\widetilde{\mathcal{IT}}(g, I) = \mathcal{IT}(g, I)$. Clearly there exists a canonical identification $\alpha\colon \widetilde{\mathcal{IT}}(g, I, B) \simeq \mathcal{IT}(g, I, B)$.

We can enrich the graph (G, ϱ) with a defect function d on $V^I \cup V^B$, defined as follows. Let $v \in V^I \cup V^B$ be a vertex, and consider its preimage X_v in X. If X_v is not a point, then it is a pointed nodal surface, doubled in case $v \in V^B$, and otherwise just a usual closed one, without $z_i, \bar{z}_i$ for $i \in \mathcal{I}$. Some of the special points of X_v correspond to nodes whose two halves belong to X_v. Smooth X_v along these nodes. There is a unique topological way to perform the smoothing process on a doubled surface, which is consistent with the choice of a half, and is such that the resulting surface is doubled. Define $d(v) \in \mathcal{IT}\big(g(v)I_v \cup (f^{P_0})^{-1}(v), B_v \cup (f^B)^{-1}(v)\big)$ to be the class of the smoothed X_v in the doubled case. Otherwise, $d(v)$ is the unique element in $\mathcal{IT}\big(g(v), I_v \cup (f^{P_0})^{-1}(v)\big)$.

The ribbon graph G, together with the involution ϱ, and the *doubled data*, which consists of the sets H^I, H^B, V^I, V^B and the maps d, f^I, f^B, f^{P_0} is called a *doubled ribbon graph*. We see that any doubled surface, together with perimeters as above, is associated with a doubled graph. Call this association Dcomb. It now follows from definitions that:

Observation A.3 *There is a canonical bijection* Half *between doubled* $(g, \mathcal{B}, (\mathcal{I}, \mathcal{P}_0))$*–metric ribbon graphs and open* $(g, \mathcal{B}, (\mathcal{I}, \mathcal{P}_0))$*–metric ribbon graphs.* $\mathrm{Half}(G)$ *is the graph spanned by* H^I, H^B, V^I, V^B*, with permutations* $\tilde{s}_0$ *and* s_1*, maps* f^I, f^B, f^{P_0}*, the same genus defect as* G *and topological defect* $\alpha(d)$*.*

$\mathrm{Half}(G)$ *is embedded in* $\tilde{\Sigma}$*, which, after defining the corresponding defects, is exactly* $K_{\mathcal{B},\mathcal{P}_0}\Sigma$*.*

Thus, by Observations A.1 and A.3, for any $\Sigma \in \overline{\mathcal{M}}^{\mathbb{R}}_{g,k,l}$ and perimeters $\boldsymbol{p}$, the symmetric JS differential indeed defines a stable open ribbon graph with perimeters $\boldsymbol{p}$ embedded in $K_{\mathcal{B},\mathcal{P}_0}\Sigma$.

Step 3 We now show that:

Proposition A.4 *The map*

$$\mathrm{comb}^{\mathbb{R}}\colon \overline{\mathcal{M}}^{\mathbb{R}}_{g,\mathcal{B},\mathcal{I}\cup\mathcal{P}_0} \times \mathbb{R}^{\mathcal{I}} \to \coprod_{G_{g,\mathcal{B},(\mathcal{I},\mathcal{P}_0)}} \mathcal{M}_G$$

is a surjection, and in the smooth case, or more generally when unmarked components are not adjacent and form a moduli of dimension zero, it is in fact a bijection onto its image.

This proposition is true in the closed case. By the above construction, it will be enough to show these properties for Dcomb. By the closed theory, from the doubled metric graph (G,ℓ) one can reconstruct the unique surface with extra structure $\widetilde{X}$ into which it embeds, including the complex structure on its marked components. Write $\boldsymbol{q}$ for the set of perimeters of faces of G. It is evident that the perimeters of faces i and $\bar{\imath}$ are the same. The involution on (G,ℓ) lifts to an involution on $\widetilde{X}$. For any singular point $v \in \widetilde{X}$ which corresponds to the vertex v of the graph, any s_0–cycle $\widetilde{v}$ of half-edges corresponds a new marked point labeled $\widetilde{v}$ in the normalization of $\widetilde{\Sigma}$. We define a surface X as follows. For a singular v with $v \in V^{\boldsymbol{B}}$, replace v by a doubled surface Σ_v in the isotopy class $d(v)$. For a singular $v \in V^{\boldsymbol{I}}$, replace v and $\varrho(v)$ by two conjugate closed surfaces Σ_v and $\overline{\Sigma}_v$, where Σ_v is in the class of $d(v)$. Note that Σ_v is not necessarily stable. Let $\Sigma_1,\ldots,\Sigma_r$ be the marked components of $\widetilde{\Sigma}$. Define

$$X = \mathrm{Stab}\Big(\Big(\coprod X_i \cup \coprod X_v\Big)/{\sim}\Big),$$

where the $\sim$ identifies a marked point in some Σ_v which corresponds to a s_0–cycle $\widetilde{v}$ with the corresponding point in some Σ_i. Stab is the stabilization map which contracts an unstable component to a point.

One can easily extend ϱ and the choice of a half to X, and $\mathrm{Dcomb}(X,\boldsymbol{q}) = (G,\ell)$, where $\boldsymbol{q}$ is the set of perimeters.

In the smooth or the more general case described in the statement, we have no freedom in the reconstruction of X.

References

[1] **A Alexandrov**, *Open intersection numbers, matrix models and MKP hierarchy*, J. High Energy Phys. (2015) art. id. 042 MR Zbl

[2] **A Alexandrov**, **A Buryak**, **R J Tessler**, *Refined open intersection numbers and the Kontsevich–Penner matrix model*, J. High Energy Phys. (2017) art. id. 123 MR Zbl

[3] **A Bawane**, **H Muraki**, **C Rim**, *Open KdV hierarchy and minimal gravity on disk*, Phys. Lett. B 783 (2018) 183–185 MR Zbl

[4] **R Bott**, **L W Tu**, *Differential forms in algebraic topology*, Graduate Texts in Math. 82, Springer (1982) MR Zbl

[5] **A Buryak**, *Equivalence of the open KdV and the open Virasoro equations for the moduli space of Riemann surfaces with boundary*, Lett. Math. Phys. 105 (2015) 1427–1448 MR Zbl

[6] **A Buryak**, *Open intersection numbers and the wave function of the KdV hierarchy*, Mosc. Math. J. 16 (2016) 27–44 MR Zbl

[7] **A Buryak**, **E Clader**, **R J Tessler**, *Open r–spin theory, II: The analogue of Witten's conjecture for r–spin disks*, preprint (2018) arXiv 1809.02536 To appear in J. Differential Geom.

[8] **A Buryak**, **E Clader**, **R J Tessler**, *Open r–spin theory, I: Foundations*, Int. Math. Res. Not. 2022 (2022) 10458–10532 MR Zbl

[9] **A Buryak**, **A Netser Zernik**, **R Pandharipande**, **R J Tessler**, *Stationary open $\mathbb{CP}^1$ descendent theory, II: The open GW/H correspondence*, in preparation

[10] **A Buryak**, **R J Tessler**, *Matrix models and a proof of the open analog of Witten's conjecture*, Comm. Math. Phys. 353 (2017) 1299–1328 MR Zbl

[11] **A Buryak**, **A N Zernik**, **R Pandharipande**, **R J Tessler**, *Open $\mathbb{CP}^1$ descendent theory, I: The stationary sector*, Adv. Math. 401 (2022) art. id. 108249 MR Zbl

[12] **D Cimasoni**, **N Reshetikhin**, *Dimers on surface graphs and spin structures, I*, Comm. Math. Phys. 275 (2007) 187–208 MR Zbl

[13] **P Deligne**, **D Mumford**, *The irreducibility of the space of curves of given genus*, Inst. Hautes Études Sci. Publ. Math. 36 (1969) 75–109 MR Zbl

[14] **R Dijkgraaf**, **H Verlinde**, **E Verlinde**, *Loop equations and Virasoro constraints in nonperturbative two-dimensional quantum gravity*, Nuclear Phys. B 348 (1991) 435–456 MR

[15] **R Dijkgraaf**, **E Witten**, *Developments in topological gravity*, preprint (2018) arXiv 1804.03275

[16] **K Fukaya**, **K Ono**, *Arnold conjecture and Gromov–Witten invariant for general symplectic manifolds*, from "The Arnoldfest" (E Bierstone, B Khesin, A Khovanskii, J E Marsden, editors), Fields Inst. Commun. 24, Amer. Math. Soc., Providence, RI (1999) 173–190 MR Zbl

[17] **J Harris**, **I Morrison**, *Moduli of curves*, Graduate Texts in Math. 187, Springer (1998) MR Zbl

[18] **T J Jarvis**, *Geometry of the moduli of higher spin curves*, Internat. J. Math. 11 (2000) 637–663 MR Zbl

[19] **T J Jarvis**, **T Kimura**, **A Vaintrob**, *Moduli spaces of higher spin curves and integrable hierarchies*, Compositio Math. 126 (2001) 157–212 MR Zbl

[20] **D Johnson**, *Spin structures and quadratic forms on surfaces*, J. London Math. Soc. 22 (1980) 365–373 MR Zbl

[21] **D Joyce**, *D–manifolds, d–orbifolds and derived differential geometry: a detailed summary*, preprint (2012) arXiv 1208.4948

[22] **D Joyce**, *On manifolds with corners*, from "Advances in geometric analysis" (S Janeczko, J Li, D H Phong, editors), Adv. Lect. Math. 21, International, Somerville, MA (2012) 225–258 MR Zbl

[23] **P W Kasteleyn**, *Dimer statistics and phase transitions*, J. Mathematical Phys. 4 (1963) 287–293 MR

[24] **S Katz**, **C-C M Liu**, *Enumerative geometry of stable maps with Lagrangian boundary conditions and multiple covers of the disc*, from "The interaction of finite-type and Gromov–Witten invariants" (D Auckly, J Bryan, editors), Geom. Topol. Monogr. 8, Geom. Topol. Publ., Coventry (2006) 1–47 MR Zbl

[25] **M Kontsevich**, *Intersection theory on the moduli space of curves and the matrix Airy function*, Comm. Math. Phys. 147 (1992) 1–23 MR Zbl

[26] **G Kuperberg**, *An exploration of the permanent-determinant method*, Electron. J. Combin. 5 (1998) art. id. 46 MR Zbl

[27] **C-C M Liu**, *Moduli of J–holomorphic curves with Lagrangian boundary conditions and open Gromov–Witten invariants for an S^1–equivariant pair*, J. Iran. Math. Soc. 1 (2020) 5–95 MR Zbl

[28] **E Looijenga**, *Cellular decompositions of compactified moduli spaces of pointed curves*, from "The moduli space of curves" (R Dijkgraaf, C Faber, G van der Geer, editors), Progr. Math. 129, Birkhäuser, Boston, MA (1995) 369–400 MR Zbl

[29] **M Mirzakhani**, *Weil–Petersson volumes and intersection theory on the moduli space of curves*, J. Amer. Math. Soc. 20 (2007) 1–23 MR Zbl

[30] **A Okounkov**, **R Pandharipande**, *Gromov–Witten theory, Hurwitz numbers, and matrix models*, from "Algebraic geometry, I" (D Abramovich, A Bertram, L Katzarkov, R Pandharipande, M Thaddeus, editors), Proc. Sympos. Pure Math. 80, Amer. Math. Soc., Providence, RI (2009) 325–414 MR Zbl

[31] **R Pandharipande**, **J P Solomon**, **R J Tessler**, *Intersection theory on moduli of disks, open KdV and Virasoro*, preprint (2014) arXiv 1409.2191

[32] **B Safnuk**, *Combinatorial models for moduli spaces of open Riemann surfaces*, preprint (2016) arXiv 1609.07226

[33] **B Safnuk**, *Topological recursion for open intersection numbers*, Commun. Number Theory Phys. 10 (2016) 833–857 MR Zbl

[34] **J P Solomon**, **R J Tessler**, *Graded spin surfaces and the open Arf invariant*, in preparation

[35] **J P Solomon**, **R J Tessler**, *Intersection theory on the moduli space of graded Riemann surfaces with boundary*, in preparation

[36] **J P Solomon**, **S B Tukachinsky**, *Differential forms on orbifolds with corners*, preprint (2020) arXiv 2011.10030

[37] **K Strebel**, *Quadratic differentials*, Ergebnisse der Math. (3) 5, Springer (1984) MR Zbl

[38] **E Witten**, *Two-dimensional gravity and intersection theory on moduli space*, from "Surveys in differential geometry" (H B Lawson, Jr, S-T Yau, editors), Lehigh Univ., Bethlehem, PA (1991) 243–310 MR Zbl

[39] **E Witten**, *Algebraic geometry associated with matrix models of two-dimensional gravity*, from "Topological methods in modern mathematics" (L R Goldberg, A V Phillips, editors), Publish or Perish, Houston, TX (1993) 235–269 MR Zbl

[40] **A N Zernik**, *Equivariant open Gromov–Witten theory of* $\mathbb{RP}^{2m} \hookrightarrow \mathbb{CP}^{2m}$, preprint (2017) arXiv 1709.09483

[41] **A N Zernik**, *Moduli of open stable maps to a homogeneous space*, preprint (2017) arXiv 1709.07402

[42] **D Zvonkine**, *Strebel differentials on stable curves and Kontsevich's proof of Witten's conjecture*, preprint (2002) arXiv math/0209071

Department of Mathematics, Weizmann Institute of Science
Rehovot, Israel

ran.tessler@weizmann.ac.il

Proposed: Jim Bryan — Received: 3 December 2018
Seconded: Richard P Thomas, Mark Gross — Revised: 25 December 2021

Geometry & Topology **27**:7 (2023) 2649–2693
DOI: 10.2140/gt.2023.27.2649
Published: 19 September 2023

Derived equivalences of hyperkähler varieties

LENNY TAELMAN

We show that the Looijenga–Lunts–Verbitsky Lie algebra acting on the cohomology of a hyperkähler variety is a derived invariant, and obtain from this a number of consequences for the action on cohomology of derived equivalences between hyperkähler varieties.

This includes a proof that derived equivalent hyperkähler varieties have isomorphic $\mathbb{Q}$–Hodge structures, the construction of a rational "Mukai lattice" functorial for derived equivalences, and the computation (up to index 2) of the image of the group of auto-equivalences on the cohomology of certain Hilbert squares of K3 surfaces.

14F05, 14J32

1 Introduction

1.1 Background

We briefly recall the background to our results. We refer to Huybrechts [24] for more details. For a smooth projective complex variety X, we denote by $\mathcal{D}X$ the bounded derived category of coherent sheaves on X. By a theorem of Orlov [37] any (exact, $\mathbb{C}$–linear) equivalence $\Phi\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ comes from a Fourier–Mukai kernel $\mathcal{P} \in \mathcal{D}(X_1 \times X_2)$, and convolution with the Mukai vector $v(\mathcal{P}) \in \mathrm{H}(X_1 \times X_2, \mathbb{Q})$ defines an isomorphism

$$\Phi^{\mathrm{H}}\colon \mathrm{H}(X_1, \mathbb{Q}) \xrightarrow{\sim} \mathrm{H}(X_2, \mathbb{Q})$$

between the total cohomology of X_1 and X_2. This isomorphism is not graded, and respects the Hodge structures only up to Tate twists. Nonetheless, Orlov has conjectured [38] that if X_1 and X_2 are derived equivalent, then for every i there exist (noncanonical) isomorphisms $\mathrm{H}^i(X_1, \mathbb{Q}) \cong \mathrm{H}^i(X_2, \mathbb{Q})$ of $\mathbb{Q}$–Hodge structures.

For every X we have a representation

$$\rho_X\colon \mathrm{Aut}(\mathcal{D}X) \to \mathrm{GL}(\mathrm{H}(X, \mathbb{Q})), \quad \Phi \mapsto \Phi^{\mathrm{H}}.$$

Its image is known for varieties with ample or antiample canonical class (in which case $\mathrm{Aut}(\mathcal{D}X)$ is small and well understood; see Bondal and Orlov [9]), for abelian varieties — see Golyshev, Lunts and Orlov [18] — and for K3 surfaces. To place our results in context, we recall the description of the image for K3 surfaces.

Let X be a K3 surface. Consider the *Mukai lattice*

$$\widetilde{\mathrm{H}}(X,\mathbb{Z}) := \mathrm{H}^0(X,\mathbb{Z}) \oplus \mathrm{H}^2(X,\mathbb{Z}(1)) \oplus \mathrm{H}^4(X,\mathbb{Z}(2)).$$

This is a Hodge structure of weight 0, and it comes equipped with a perfect bilinear form b of signature $(4,20)$. For convenience, we denote by α and β the natural generators of $\mathrm{H}^0(X,\mathbb{Z})$ and $\mathrm{H}^4(X,\mathbb{Z}(2))$ respectively, so that $\widetilde{\mathrm{H}}(X,\mathbb{Z}) = \mathbb{Z}\alpha \oplus \mathrm{H}^2(X,\mathbb{Z}(1)) \oplus \mathbb{Z}\beta$. The pairing b is the orthogonal sum of the intersection pairing on $\mathrm{H}^2(X,\mathbb{Z}(1))$ and the pairing on $\mathbb{Z}\alpha \oplus \mathbb{Z}\beta$ given by $b(\alpha,\alpha) = b(\beta,\beta) = 0$ and $b(\alpha,\beta) = -1$.

It was observed by Mukai [35] that if $\Phi\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ is a derived equivalence between K3 surfaces, then Φ^{H} restricts to an isomorphism $\Phi^{\widetilde{\mathrm{H}}}\colon \widetilde{\mathrm{H}}(X_1,\mathbb{Z}) \to \widetilde{\mathrm{H}}(X_2,\mathbb{Z})$ respecting the pairing and Hodge structures. Denote by $\mathrm{Aut}(\widetilde{\mathrm{H}}(X,\mathbb{Z}))$ the group of isometries of $\widetilde{\mathrm{H}}(X,\mathbb{Z})$ respecting the Hodge structure, and by $\mathrm{Aut}^+(\widetilde{\mathrm{H}}(X,\mathbb{Z}))$ the subgroup (of index 2) consisting of those isometries that respect the orientation on a four-dimensional positive definite subspace of $\widetilde{\mathrm{H}}(X,\mathbb{R})$.

Theorem 1.1 [22; 26; 35; 36; 39] *Let X be a* K3 *surface. Then the image of ρ_X is* $\mathrm{Aut}^+(\widetilde{\mathrm{H}}(X,\mathbb{Z}))$. □

In this paper, we prove Orlov's conjecture on $\mathbb{Q}$–Hodge structures for hyperkähler varieties, construct a rational version of the Mukai lattice for hyperkähler varieties, and compute (up to index 2) the image of ρ_X for certain Hilbert squares of K3 surfaces. The main tool in these results is the Looijenga–Lunts–Verbitsky Lie algebra.

1.2 The LLV Lie algebra and derived equivalences

Let X be a smooth projective complex variety. By the hard Lefschetz theorem, every ample class $\lambda \in \mathrm{NS}(X)$ determines a Lie algebra $\mathfrak{g}_\lambda \subset \mathrm{End}(\mathrm{H}(X,\mathbb{Q}))$ isomorphic to $\mathfrak{sl}_2$. More generally, this holds for every cohomology class $\lambda \in \mathrm{H}^2(X,\mathbb{Q})$ (algebraic or not) satisfying the conclusion of the hard Lefschetz theorem. Looijenga and Lunts [33] and Verbitsky [46] have studied the Lie algebra $\mathfrak{g}(X) \subset \mathrm{End}(\mathrm{H}(X,\mathbb{Q}))$ generated by the collection of the Lie algebras $\mathfrak{g}_\lambda$. We will refer to this as the LLV Lie algebra. See Section 2.1 for more details.

We say that X is *holomorphic symplectic* if it admits a nowhere degenerate holomorphic symplectic form $\sigma \in \mathrm{H}^0(X, \Omega_X^2)$.

Theorem A (Section 2.4) *Let X_1 and X_2 be holomorphic symplectic varieties. Then for every equivalence $\Phi\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ there exists a canonical isomorphism of rational Lie algebras*

$$\Phi^{\mathfrak{g}}\colon \mathfrak{g}(X_1) \xrightarrow{\sim} \mathfrak{g}(X_2)$$

with the property that the map $\Phi^{\mathrm{H}}\colon \mathrm{H}(X_1, \mathbb{Q}) \xrightarrow{\sim} \mathrm{H}(X_2, \mathbb{Q})$ is equivariant with respect to $\Phi^{\mathfrak{g}}$.

Note that $\mathfrak{g}(X)$ is defined in terms of the grading and the cup product on $\mathrm{H}(X, \mathbb{Q})$, neither of which are preserved under derived equivalences.

To prove Theorem A we introduce a complex Lie algebra $\mathfrak{g}'(X)$ whose definition is similar to the rational Lie algebra $\mathfrak{g}(X)$, but where the action of $\mathrm{H}^2(X, \mathbb{Q})$ on $\mathrm{H}(X, \mathbb{Q})$ is replaced with a natural action of the Hochschild cohomology group $\mathrm{HH}^2(X)$ on Hochschild homology $\mathrm{HH}_\bullet(X)$. Since Hochschild cohomology and its action on Hochschild homology is known to be invariant under derived equivalences, it follows that $\mathfrak{g}'(X)$ is a derived invariant. We show that if X is holomorphic symplectic, then the isomorphism $\mathrm{HH}_\bullet(X) \to \mathrm{H}(X, \mathbb{C})$ (coming from the Hochschild–Kostant–Rosenberg isomorphism) maps $\mathfrak{g}'(X)$ to $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$. This is closely related to Verbitsky's "mirror symmetry" for hyperkähler varieties [46; 47]. From this we deduce that the rational Lie algebra $\mathfrak{g}(X)$ is a derived invariant.

1.3 A rational Mukai lattice for hyperkähler varieties

A *hyperkähler* (or irreducible holomorphic symplectic) variety is a simply connected smooth projective variety X for which $\mathrm{H}^0(X, \Omega_X^2)$ is spanned by a nowhere degenerate form.

Let X be a hyperkähler variety. Consider the $\mathbb{Q}$–vector space

$$\widetilde{\mathrm{H}}(X, \mathbb{Q}) := \mathbb{Q}\alpha \oplus \mathrm{H}^2(X, \mathbb{Q}) \oplus \mathbb{Q}\beta$$

equipped with the bilinear form b which is the orthogonal sum of the Beauville–Bogomolov form on $\mathrm{H}^2(X, \mathbb{Q})$ and a hyperbolic plane $\mathbb{Q}\alpha \oplus \mathbb{Q}\beta$ with α and β isotropic and $b(\alpha, \beta) = -1$. By analogy with the case of a K3 surface, we will call $\widetilde{\mathrm{H}}(X, \mathbb{Q})$ the (rational) *Mukai lattice* of X. Looijenga and Lunts [33] and Verbitsky [46] have shown that the Lie algebra $\mathfrak{g}(X)$ can be canonically identified with $\mathfrak{so}(\widetilde{\mathrm{H}}(X, \mathbb{Q}))$;

see Section 3.1 for a precise statement. Moreover, Verbitsky [46] has shown that the subalgebra $\mathrm{SH}(X,\mathbb{Q})$ of $\mathrm{H}(X,\mathbb{Q})$ generated by $\mathrm{H}^2(X,\mathbb{Q})$ forms an irreducible sub-$\mathfrak{g}(X)$–module. Using this, we show that Theorem A implies:

Theorem B (Section 4.2) *Let X_1 and X_2 be hyperkähler varieties and*

$$\Phi\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$$

an equivalence. Then the induced isomorphism Φ^{H} restricts to an isomorphism $\Phi^{\mathrm{SH}}\colon \mathrm{SH}(X_1,\mathbb{Q}) \xrightarrow{\sim} \mathrm{SH}(X_2,\mathbb{Q})$.

Taking $X_1 = X_2 = X$ in Theorem B we obtain a homomorphism

$$\rho_X^{\mathrm{SH}}\colon \mathrm{Aut}(\mathcal{D}X) \to \mathrm{GL}(\mathrm{SH}(X,\mathbb{Q})).$$

The complex structure on a hyperkähler variety X induces a Hodge structure of weight 0 on $\widetilde{\mathrm{H}}(X,\mathbb{Q})$ given by

$$\widetilde{\mathrm{H}}(X,\mathbb{Q}) = \mathbb{Q}\alpha \oplus \mathrm{H}^2(X,\mathbb{Q}(1)) \oplus \mathbb{Q}\beta.$$

Denote by $\mathrm{Aut}\,\widetilde{\mathrm{H}}(X,\mathbb{Q})$ the group of Hodge isometries of $\widetilde{\mathrm{H}}(X,\mathbb{Q})$.

Theorem C (Section 4.2) *Let X be a hyperkähler variety of dimension $2d$ and second Betti number b_2. Assume that b_2 is odd or d is odd. Then ρ_X^{SH} factors over a map $\rho_X^{\widetilde{\mathrm{H}}}\colon \mathrm{Aut}(\mathcal{D}(X)) \to \mathrm{Aut}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$.*

See Sections 3.2 and 4.2 for an explicit description of the implicit map

$$\mathrm{Aut}(\widetilde{\mathrm{H}}(X,\mathbb{Q})) \to \mathrm{GL}(\mathrm{SH}(X,\mathbb{Q})).$$

Note that all known hyperkähler varieties satisfy the parity conditions in the theorem: there are two infinite series of deformation classes with odd b_2 (generalized Kummers and Hilbert schemes of points), and three exceptional deformation classes with odd d (K3, OG6, OG10).

1.4 Hodge structures of derived equivalent hyperkähler varieties

Another application of Theorem A is the following:

Theorem D (Section 5) *Let X_1 and X_2 be derived equivalent hyperkähler varieties. Then for every i the $\mathbb{Q}$–Hodge structures $\mathrm{H}^i(X_1,\mathbb{Q})$ and $\mathrm{H}^i(X_2,\mathbb{Q})$ are isomorphic.*

This confirms Orlov's conjecture for hyperkähler varieties. The proof is inspired by Soldatenkov [43].

1.5 Auto-equivalences of the Hilbert square of a K3 surface

In the second half of the paper we consider the problem of determining the image of ρ_X for certain hyperkähler varieties. An important difference with the first half of the paper is that *integral* structures (lattices, arithmetic subgroups, ...) will play an important role here.

As a first approximation to determining the image of ρ_X, we consider a variation of this problem which is deformation invariant. Let X be a smooth projective complex variety. If X' and X'' are smooth deformations of X (parametrized by paths in the base), and if $\Phi\colon \mathcal{D}X' \xrightarrow{\sim} \mathcal{D}X''$ is an equivalence, then we obtain an isomorphism as the composition

$$\mathrm{H}(X,\mathbb{Q}) \to \mathrm{H}(X',\mathbb{Q}) \xrightarrow{\Phi^{\mathrm{H}}} \mathrm{H}(X'',\mathbb{Q}) \to \mathrm{H}(X,\mathbb{Q}).$$

We define the *derived monodromy group* of X to be the subgroup $\mathrm{DMon}(X)$ of $\mathrm{GL}(\mathrm{H}(X,\mathbb{Q}))$ generated by all these isomorphisms. This group contains both the usual monodromy group of X and the image of $\rho_X : \mathrm{Aut}(\mathcal{D}X) \to \mathrm{GL}(\mathrm{H}(X,\mathbb{Q}))$.

If S is a K3 surface, then the result of Huybrechts, Macrì and Stellari [26] implies $\mathrm{DMon}(S) = \mathrm{O}^+(\widetilde{\mathrm{H}}(S,\mathbb{Z}))$, and that the image of ρ_S consists of those elements of $\mathrm{DMon}(S)$ that respect the Hodge structure on $\widetilde{\mathrm{H}}(S,\mathbb{Z})$. Similarly, for an abelian variety A, the results of [18] imply $\mathrm{DMon}(A) = \mathrm{Spin}(\mathrm{H}^1(A,\mathbb{Z}) \oplus \mathrm{H}^1(A^\vee,\mathbb{Z}))$, and that the image of ρ_A consists of those elements of $\mathrm{DMon}(A)$ that respect the Hodge structure on $\mathrm{H}^1(A,\mathbb{Z}) \oplus \mathrm{H}^1(A^\vee,\mathbb{Z})$.

Now let X be a hyperkähler variety of type $\mathrm{K3}^{[2]}$. We have $\mathrm{H}(X,\mathbb{Q}) = \mathrm{SH}(X,\mathbb{Q})$ and hence by Theorem C the action of $\mathrm{Aut}(\mathcal{D}X)$ on $\mathrm{H}(X,\mathbb{Q})$ factors over a subgroup $\mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ of $\mathrm{GL}(\mathrm{H}(X,\mathbb{Q}))$.

For an integral lattice $\Lambda \subset \widetilde{\mathrm{H}}(X,\mathbb{Q})$ we denote by $\mathrm{O}^+(\Lambda) \subset \mathrm{O}(\Lambda)$ the subgroup consisting of those Hodge isometries that respect the orientation of a positive 4–plane in $\Lambda_{\mathbb{R}}$.

Theorem E (Section 9.4) *Let X be a hyperkähler variety deformation equivalent to the Hilbert square of a* K3 *surface. There is an integral lattice $\Lambda \subset \widetilde{\mathrm{H}}(X,\mathbb{Q})$ such that*

$$\mathrm{O}^+(\Lambda) \subset \mathrm{DMon}(X) \subset \mathrm{O}(\Lambda)$$

inside $\mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$.

See Section 9.4 for a precise description of Λ. As an abstract lattice, Λ is isomorphic to $\mathrm{H}^2(X,\mathbb{Z}) \oplus U$, but its image in $\widetilde{\mathrm{H}}(X,\mathbb{Q})$ is *not* $\mathbb{Z}\alpha \oplus \mathrm{H}^2(X,\mathbb{Z}) \oplus \mathbb{Z}\beta$.

Crucial in the proof of Theorem E is the *derived McKay correspondence* due to Bridgeland, King and Reid [11] and Haiman [21]. It provides an ample supply of elements of $\mathrm{DMon}(X)$: every deformation of X to the Hilbert square $S^{[2]}$ of a K3 surface S induces an inclusion $\mathrm{DMon}(S) \to \mathrm{DMon}(X)$. As part of the proof, we explicitly compute this inclusion.

We denote by $\mathrm{Aut}(\Lambda)$ the group of isometries of $\Lambda \subset \widetilde{\mathrm{H}}(X, \mathbb{Q})$ that respect the Hodge structure on $\widetilde{\mathrm{H}}(X, \mathbb{Q})$. It follows from Theorem E that $\mathrm{im}(\rho_X)$ is contained in $\mathrm{Aut}(\Lambda)$ for every X which is deformation equivalent to the Hilbert square of a K3 surface. For some X we can show that the upper bound in the above corollary is close to being sharp. Denote by $\mathrm{Aut}^+(\Lambda) \subset \mathrm{Aut}(\Lambda)$ the subgroup consisting of those Hodge isometries that respect the orientation of a positive 4–plane in $\Lambda_{\mathbb{R}}$.

Theorem F (Section 10.2) *Let S be a complex* K3 *surface and $X = S^{[2]}$. Assume that* $\mathrm{NS}(X)$ *contains a hyperbolic plane. Then* $\mathrm{Aut}^+(\Lambda) \subset \mathrm{im}(\rho_X) \subset \mathrm{Aut}(\Lambda)$.

Remark 1.2 To determine $\mathrm{im}\,\rho_X$ up to index 2 for a general hyperkähler of type $\mathrm{K3}^{[2]}$ new constructions of derived equivalences will be needed.

Remark 1.3 Theorems E and F leave an ambiguity of index 2, related to orientations on a maximal positive subspace of $\widetilde{\mathrm{H}}(X, \mathbb{R})$. In the case of K3 surfaces, it was conjectured by Szendrői [44] that derived equivalences must respect such orientation, and this was proven by Huybrechts, Macrì, and Stellari [26]. Their method is based on deformation to generic (formal or analytic) K3 surfaces of Picard rank 0, and on a complete understanding of the space of stability conditions on those [25]. It is far from clear if such a strategy can be used to remove the index 2 ambiguity for hyperkähler varieties of type $\mathrm{K3}^{[2]}$.

Remark 1.4 That a lattice of signature $(4, b_2 - 2)$ should play a role in describing the image of ρ_X for hyperkähler varieties X was expected from the physics literature — see Dijkgraaf [16] — but it is not clear where the lattice should come from, nor what its precise description should be for general hyperkähler varieties. In the above results, the lattice Λ arises in a rather implicit way, and one may hope for a more concrete interpretation of its elements.

Remark 1.5 It is tempting to try to conjecture a description of the group $\mathrm{Aut}(\mathcal{D}X)$ in terms of an action on a space of stability conditions on X, generalizing Bridgeland's work on K3 surfaces [10]. However, there is a representation-theoretic obstruction against doing this naively. The central charge of a hypothetical stability condition on X

takes values in $\mathrm{H}(X,\mathbb{C})$, yet Theorems E and F suggest the central charge should take values in $\widetilde{\mathrm{H}}(X,\mathbb{C})$. If X is of type $\mathrm{K3}^{[2]}$, then $\mathrm{H}(X,\mathbb{C})$ and $\widetilde{\mathrm{H}}(X,\mathbb{C})$ are nonisomorphic irreducible $\mathrm{DMon}(X)$–modules, so this would require a modification of the notion of stability condition.

Acknowledgements

I am grateful to Nick Addington, Thorsten Beckmann, Dion Leijnse, Eyal Markman, Zoë Schroot, and the referees for many valuable comments on earlier versions of this paper. This project received funding from the European Research Council (ERC), grant 864145, and from the Dutch Research Council (NWO).

2 The LLV Lie algebra of a smooth projective variety

In this section we recall the construction of Looijenga and Lunts [33] and Verbitsky [46] of a Lie algebra acting naturally on the cohomology of algebraic varieties. For holomorphic symplectic varieties we show that this Lie algebra is a derived invariant.

2.1 The LLV Lie algebra

Let F be a field of characteristic zero and M be a $\mathbb{Z}$–graded F–vector space of finite F–dimension. Denote by h the endomorphism of M that is multiplication by n on M_n.

Let e be an endomorphism of M of degree 2. We say that *e has the hard Lefschetz property* if for every $n \geq 0$ the map $e^n\colon M_{-n} \to M_n$ is an isomorphism. This is equivalent to the existence of an $f \in \mathrm{End}(M)$ such that the relations

$$[h,e]=2e,\quad [h,f]=-2f,\quad [e,f]=h \tag{1}$$

hold in $\mathrm{End}(M)$. Thus, (e,h,f) forms an $\mathfrak{sl}_2$–triple and defines a Lie homomorphism $\mathfrak{sl}_2 \to \mathrm{End}(M)$.

Proposition 2.1 *Assume that e has the hard Lefschetz property. Then the element f satisfying* (1) *is unique, and if e and h lie in a semisimple sub-Lie algebra $\mathfrak{g} \subset \mathrm{End}(M)$, then so does f.*

Proof The action of $\mathrm{ad}\,e$ on $\mathrm{End}(M)$ has the hard Lefschetz property for the grading defined by $\mathrm{ad}\,h$. In particular,

$$(\mathrm{ad}\,e)^2\colon \mathrm{End}(M)_{-2} \xrightarrow{\sim} \mathrm{End}(M)_2$$

is an isomorphism. It sends f to $-2e$, so f is indeed uniquely determined.

If e and h lie in $\mathfrak{g}$, then $\mathfrak{g} \subset \mathrm{End}(M)$ is graded and the above map restricts to an injective map

$$(\operatorname{ad} e)^2 \colon \mathfrak{g}_{-2} \hookrightarrow \mathfrak{g}_2.$$

Since h is diagonalizable, it is contained in a Cartan subalgebra of $\mathfrak{g}$. The symmetry of the resulting root system implies that $\dim \mathfrak{g}_{-n} = \dim \mathfrak{g}_n$ for all n. In particular, the map $(\operatorname{ad} e)^2$ defines an isomorphism between $\mathfrak{g}_{-2}$ and $\mathfrak{g}_2$; thus f lies in $\mathfrak{g}$. $\square$

Let $\mathfrak{a}$ be an abelian Lie algebra and $e \colon \mathfrak{a} \to \mathfrak{gl}(M)$, defined by $a \mapsto e_a$, a Lie homomorphism. We say that *e has the hard Lefschetz property* if $e(\mathfrak{a}) \subset \mathfrak{gl}(M)_2$ and if there exists some $a \in \mathfrak{a}$ such that e_a has the hard Lefschetz property. Note that this is a Zariski open condition on $a \in \mathfrak{a}$.

If $e \colon \mathfrak{a} \to \mathfrak{gl}(M)$ has the hard Lefschetz property, then we denote by $\mathfrak{g}(\mathfrak{a}, M)$ the Lie algebra generated by the $\mathfrak{sl}_2$–triples (e_a, h, f_a) for $a \in \mathfrak{a}$ such that e_a has the hard Lefschetz property. We say that $(\mathfrak{a}, M)$ is a *Lefschetz module* if $\mathfrak{g}(\mathfrak{a}, M)$ is semisimple.

Now let X be a smooth projective complex variety of dimension d. Denote by $M := \mathrm{H}(X, \mathbb{Q})[d]$ the shifted total cohomology of X (with middle cohomology in degree 0). For a class $\lambda \in \mathrm{H}^2(X, \mathbb{Q})$, consider the endomorphism $e_\lambda \in \mathrm{End}(M)$ given by cup product with λ. If λ is ample, then e_λ has the hard Lefschetz property, so the map $e \colon \mathrm{H}^2(X, \mathbb{Q}) \to \mathfrak{gl}(M)$ has the hard Lefschetz property. We denote the corresponding Lie algebra by $\mathfrak{g}(X) := \mathfrak{g}(\mathrm{H}^2(X, \mathbb{Q}), M)$.

Proposition 2.2 [33, 1.6, 1.9] *$(\mathrm{H}^2(X, \mathbb{Q}), M)$ is a Lefschetz module.* $\square$

In other words, $\mathfrak{g}(X)$ is a semisimple Lie algebra over $\mathbb{Q}$.

2.2 Hochschild homology and cohomology

Let X be a smooth projective variety of dimension d with canonical bundle $\omega_X := \Omega_X^d$. Its *Hochschild cohomology* is defined as

$$\mathrm{HH}^n(X) := \mathrm{Ext}^n_{X \times X}(\Delta_* \mathcal{O}_X, \Delta_* \mathcal{O}_X)$$

and its *Hochschild homology* is defined as

$$\mathrm{HH}_n(X) := \mathrm{Ext}^{d-n}_{X \times X}(\Delta_* \mathcal{O}_X, \Delta_* \omega_X).$$

Composition of extensions defines maps

$$\mathrm{HH}^n \otimes \mathrm{HH}^m \to \mathrm{HH}^{n+m}, \quad \mathrm{HH}^n \otimes \mathrm{HH}_m \to \mathrm{HH}_{m-n},$$

making $\mathrm{HH}_\bullet(X)$ into a graded module over the graded ring $\mathrm{HH}^\bullet(X)$.

The Hochschild–Kostant–Rosenberg isomorphism (twisted by the square root of the Todd class as in [30; 15]) defines isomorphisms

$$I^n\colon \mathrm{HH}^n(X) \xrightarrow{\sim} \bigoplus_{i+j=n} \mathrm{H}^i\bigl(X, \textstyle\bigwedge^j T_X\bigr), \quad I_n\colon \mathrm{HH}_n(X) \xrightarrow{\sim} \displaystyle\bigoplus_{j-i=n} \mathrm{H}^i(X, \Omega_X^j).$$

Under these isomorphisms, multiplication in $\mathrm{HH}^\bullet(X)$ corresponds to the operation induced by the product in $\bigwedge^\bullet T_X$, and the action of $\mathrm{HH}^\bullet(X)$ on $\mathrm{HH}_\bullet(X)$ corresponds to the action induced by the contraction action of $\bigwedge^\bullet T_X$ on $\Omega_X^\bullet$; see [12; 13].

Together with the degeneration of the Hodge–de Rham spectral sequence, the isomorphism $I_\bullet$ defines an isomorphism

$$\mathrm{HH}_\bullet(X) \xrightarrow{\sim} \mathrm{H}(X, \mathbb{C}).$$

This map does not respect the grading; rather it maps HH_i to the i^{th} column of the Hodge diamond (normalized so that the 0^{th} column is the central column $\bigoplus_p \mathrm{H}^{p,p}$). Combining with the action of $\mathrm{HH}^\bullet$ on $\mathrm{HH}_\bullet$, we obtain an action of the ring $\mathrm{HH}^\bullet(X)$ on $\mathrm{H}(X, \mathbb{C})$.

Theorem 2.3 *Let $\Phi\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ be a derived equivalence between smooth projective complex varieties. Then we have natural graded isomorphisms*

$$\Phi^{\mathrm{HH}^\bullet}\colon \mathrm{HH}^\bullet(X_1) \xrightarrow{\sim} \mathrm{HH}^\bullet(X_2), \quad \Phi^{\mathrm{HH}_\bullet}\colon \mathrm{HH}_\bullet(X_1) \xrightarrow{\sim} \mathrm{HH}_\bullet(X_2),$$

compatible with the ring structure on $\mathrm{HH}^\bullet$ *and the module structure on* $\mathrm{HH}_\bullet$, *and such that the square*

$$\begin{tikzcd} \mathrm{HH}_\bullet(X_1) \arrow[r, "I"] \arrow[d, "\Phi^{\mathrm{HH}_\bullet}"] & \mathrm{H}(X_1, \mathbb{C}) \arrow[d, "\Phi^{\mathrm{H}}"] \\ \mathrm{HH}_\bullet(X_2) \arrow[r, "I"] & \mathrm{H}(X_2, \mathbb{C}) \end{tikzcd}$$

commutes.

Proof See [13; 34]. □

2.3 The Hochschild Lie algebra of a holomorphic symplectic variety

Now assume that X is holomorphic symplectic of dimension $2d$. That is, we assume that there exists a symplectic form $\sigma \in \mathrm{H}^0(X, \Omega_X^2)$. Note that this implies that a Zariski-dense collection of $\sigma \in \mathrm{H}^0(X, \Omega_X^2)$ will be nowhere degenerate.

Through the isomorphism $I\colon \mathrm{HH}_\bullet(X) \to \mathrm{H}(X, \mathbb{C})$, the vector space $\mathrm{H}(X, \mathbb{C})$ becomes a module under the ring $\mathrm{HH}^\bullet(X)$.

Lemma 2.4 $\mathrm{HH}^\bullet(X) \cong \mathrm{H}^\bullet(X,\mathbb{C})$ *as graded rings, and* $\mathrm{H}(X,\mathbb{C})$ *is free of rank one as an* $\mathrm{HH}^\bullet(X)$*–module.*

Proof A symplectic form σ defines an isomorphism $\Omega^1_X \xrightarrow{\sim} T_X$, and hence an isomorphism of algebras $\bigwedge^\bullet \Omega^1_X \xrightarrow{\sim} \bigwedge^\bullet T_X$. Combining this with the Hochschild–Kostant–Rosenberg isomorphism I and the degeneration of the Hodge–de Rham spectral sequence, we obtain a chain of isomorphisms of graded rings

$$\mathrm{HH}^\bullet(X) \xrightarrow{\sim} \mathrm{H}^\bullet\bigl(X, \textstyle\bigwedge^\bullet T_X\bigr) \xrightarrow{\sim} \mathrm{H}^\bullet(X,\Omega^\bullet_X) \xrightarrow{\sim} \mathrm{H}^\bullet(X,\mathbb{C}).$$

This proves the first assertion. For the second it suffices to observe that the module $\mathrm{HH}_\bullet(X,\mathbb{C})$ is generated by $\sigma^d \in \mathrm{HH}_{2d}(X) = \mathrm{H}^0(X,\Omega^{2d}_X)$. □

Consider the endomorphisms $h_p, h_q \in \mathrm{End}(\mathrm{H}(X,\mathbb{C}))$ given by

$$h_p = p-d, \quad h_q = q-d \quad \text{on } \mathrm{H}^{p,q}.$$

These define the Hodge bigrading on $\mathrm{H}(X,\mathbb{C})$, normalized to be symmetric along the central part $\mathrm{H}^{d,d}$. Note that $h = h_p + h_q$. The action of $\mathrm{HH}^n(X)$ on $\mathrm{H}(X,\mathbb{C})$ has degree n for the grading defined by $h' = h_q - h_p$.

Lemma 2.4 and hard Lefschetz imply:

Corollary 2.5 *For a Zariski-dense collection of* $\mu \in \mathrm{HH}^2(X)$*, the action by* μ*,*

$$e'_\mu \colon \mathrm{H}(X,\mathbb{C}) \to \mathrm{H}(X,\mathbb{C}),$$

has the hard Lefschetz property with respect to the grading defined by h'. □

In particular, for every such μ we have a complex subalgebra $\mathfrak{g}_\mu \subset \mathrm{End}(\mathrm{H}(X,\mathbb{C}))$ isomorphic to $\mathfrak{sl}_2$, and the collection of such algebras generates a Lie algebra which we denote by $\mathfrak{g}'(X) \subset \mathrm{End}(\mathrm{H}(X,\mathbb{C}))$. From Lemma 2.4 we also obtain:

Corollary 2.6 *The complex Lie algebras* $\mathfrak{g}'(X)$ *and* $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$ *are isomorphic.* □

In the next section, we will show something stronger: that $\mathfrak{g}'(X)$ and $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$ coincide as sub-Lie algebras of $\mathrm{End}(\mathrm{H}(X,\mathbb{C}))$. Theorem A then follows by combining this with the following proposition:

Proposition 2.7 *Assume that* X_1 *and* X_2 *are holomorphic symplectic varieties. Then for every equivalence* $\Phi \colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ *there exists a canonical isomorphism of complex Lie algebras*

$$\Phi^{\mathfrak{g}'} \colon \mathfrak{g}'(X_1) \xrightarrow{\sim} \mathfrak{g}'(X_2).$$

It has the property that the map $\Phi^{\mathrm{H}}\colon \mathrm{H}(X_1,\mathbb{C}) \xrightarrow{\sim} \mathrm{H}(X_2,\mathbb{C})$ *is equivariant with respect to* $\Phi^{\mathfrak{g}'}$.

Proof This follows immediately from Theorem 2.3. □

2.4 Comparison of the two Lie algebras and proof of Theorem A

The remainder of this section is devoted to the proof of the following:

Proposition 2.8 *If* X *is holomorphic symplectic, then* $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C} = \mathfrak{g}'(X)$ *as sub-Lie algebras of* $\mathrm{End}(\mathrm{H}(X,\mathbb{C}))$.

Let X be holomorphic symplectic. If $\mathcal{F}$ is a coherent $\mathcal{O}_X$–module then we will simply write $\mathrm{H}^i(\mathcal{F})$ for $\mathrm{H}^i(X,\mathcal{F})$. We have decompositions

$$\mathrm{H}^2(X,\mathbb{C}) = \mathrm{H}^2(\mathcal{O}_X) \oplus \mathrm{H}^1(\Omega^1_X) \oplus \mathrm{H}^0(\Omega^2_X)$$

and

$$\mathrm{HH}^2(X) = \mathrm{H}^2(\mathcal{O}_X) \oplus \mathrm{H}^1(T_X) \oplus \mathrm{H}^0\big(\textstyle\bigwedge^2 T_X\big).$$

We will use the same symbol λ to denote an element $\lambda \in \mathrm{H}^2(X,\mathbb{C})$ and the endomorphism of $\mathrm{End}(\mathrm{H}(X,\mathbb{C}))$ given by cup product with λ. Note that $\lambda \in \mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$ by construction. Similarly, we will use the same symbol for $\mu \in \mathrm{HH}^2(X)$ and the resulting $\mu \in \mathrm{End}(\mathrm{H}(X,\mathbb{C}))$, given by contraction with μ. We have $\mu \in \mathfrak{g}'(X)$.

For a symplectic form $\sigma \in \mathrm{H}^0(\Omega^2_X)$, we denote by $\check{\sigma} \in \mathrm{H}^0\big(\bigwedge^2 T_X\big)$ the image of the form $\sigma \in \mathrm{H}^0(\Omega^2_X)$ under the isomorphism $\Omega^2_X \to \bigwedge^2 T_X$ defined by σ. In suitable local coordinates, we have

$$\sigma = \mathrm{d}u_1 \wedge \mathrm{d}v_1 + \cdots + \mathrm{d}u_d \wedge \mathrm{d}v_d$$

and

$$\check{\sigma} = \frac{\partial}{\partial u_1} \wedge \frac{\partial}{\partial v_1} + \cdots + \frac{\partial}{\partial u_d} \wedge \frac{\partial}{\partial v_d}.$$

Lemma 2.9 *If* σ *is a nowhere degenerate symplectic form then* $(\sigma, h_p, \check{\sigma})$ *is an* $\mathfrak{sl}_2$*–triple in* $\mathrm{End}(\mathrm{H}(X,\mathbb{C}))$.

Proof Clearly σ has degree 2 and $\check{\sigma}$ has degree -2 for the grading given by h_p, so $[h_p,\sigma] = 2\sigma$ and $[h_p,\check{\sigma}] = -2\check{\sigma}$.

We need to show that $[\sigma,\check{\sigma}] = h_p$. This follows immediately from a local computation: in the above local coordinates, one verifies that on the standard basis of Ω^p the commutator $[\sigma,\check{\sigma}]$ acts as $p-d$. □

Note that the existence of one nowhere degenerate σ implies that a Zariski-dense collection of $\sigma \in \mathrm{H}^0(\Omega^2_X)$ is nowhere degenerate.

Lemma 2.10 *For a Zariski-dense collection $\alpha \in \mathrm{H}^2(X, \mathcal{O}_X)$, there is $\check{\alpha} \in \mathrm{End}(\mathrm{H}(X, \mathbb{C}))$ such that $(\alpha, h_q, \check{\alpha})$ is an $\mathfrak{sl}_2$–triple.*

Proof This follows from Lemma 2.9 and Hodge symmetry. □

Lemma 2.11 *For all $\tau \in \mathrm{H}^0\big(X, \bigwedge^2 T_X\big)$ the endomorphism τ lies in $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$.*

Proof It suffices to show that this holds for a Zariski-dense collection of τ; hence we may assume without loss of generality that $\tau = \check{\sigma}$ with σ and $\check{\sigma}$ as in Lemma 2.9. Let α and $\check{\alpha}$ be as in Lemma 2.10. Because σ and h_p commute with both α and h_q, we have that every element of the $\mathfrak{sl}_2$–triple $(\sigma, h_p, \check{\sigma})$ commutes with every element of the $\mathfrak{sl}_2$–triple $(\alpha, h_q, \check{\alpha})$. From this, it follows that

$$(\alpha + \sigma, h, \check{\alpha} + \check{\sigma}) \quad \text{and} \quad (\alpha - \sigma, h, \check{\alpha} - \check{\sigma})$$

are $\mathfrak{sl}_2$–triples. Since the elements $\alpha \pm \sigma$ lie in $\mathrm{H}^2(X, \mathbb{C})$, and apparently have the hard Lefschetz property, we conclude that the endomorphisms $\check{\alpha} \pm \check{\sigma}$ lie in $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$; hence also $\tau = \check{\sigma}$ lies in $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$. □

Corollary 2.12 *h_p and h_q lie in $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$.*

Proof By Lemma 2.9 we have $h_p = [\sigma, \check{\sigma}]$, which by Lemma 2.11 lies in $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$. Since $h_q = h - h_p$ we also have that h_q lies in $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$. □

Fix a $\tau \in \mathrm{H}^0\big(X, \bigwedge^2 T_X\big)$ that is nowhere degenerate as an alternating form on Ω^1_X. This defines isomorphisms $c_\tau \colon \Omega^1_X \to T_X$ and $c_\tau \colon \mathrm{H}^1(\Omega^1_X) \to \mathrm{H}^1(T_X)$ given by contracting sections of Ω^1_X with τ.

Lemma 2.13 *For all $\eta \in \mathrm{H}^1(\Omega^1_X)$, we have $[\tau, \eta] = c_\tau(\eta)$ in $\mathrm{End}(\mathrm{H}(X, \mathbb{C}))$.*

Proof This is again a local computation. If η is a local section of Ω^1_X, then a computation on a local basis shows $[\tau, \eta] = c_\tau(\eta)$ as maps $\Omega^p_X \to \Omega^{p-1}_X$. □

Corollary 2.14 *Every element η' of $\mathrm{H}^1(X, T_X)$ lies in $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$.*

Proof (See also [19, 4.5] for the case of a hyperkähler variety.) Every such η' is of the form $c_\tau(\eta)$ for a unique $\eta \in \mathrm{H}^1(\Omega^1_X)$, and hence the corollary follows from Lemmas 2.13 and 2.11 and the fact that η lies in $\mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$. □

We can now finish the comparison of the two Lie algebras.

Proof of Proposition 2.8 By Corollary 2.6 it suffices to show that $\mathfrak{g}'(X)$ is contained in $\mathfrak{g}(X)\otimes_{\mathbb{Q}}\mathbb{C}$. By Proposition 2.1 it suffices to show that h' is contained in $\mathfrak{g}(X)\otimes_{\mathbb{Q}}\mathbb{C}$, and that for almost every $a\in \mathrm{HH}^2(X)$ we have that the action of a on $\mathrm{H}(X,\mathbb{C})$ is contained in $\mathfrak{g}(X)\otimes_{\mathbb{Q}}\mathbb{C}$. This follows from Lemma 2.11, Corollaries 2.12 and 2.14, and the fact that the action of any $\alpha\in \mathrm{H}^2(\mathcal{O}_X)$ lies in $\mathfrak{g}(X)\otimes_{\mathbb{Q}}\mathbb{C}$. □

Together with Proposition 2.7, this proves Theorem A.

3 Rational cohomology of hyperkähler varieties

3.1 The BBF form and the LLV Lie algebra

Let X be a complex hyperkähler variety of dimension $2d$. We denote by

$$b=b_X\colon \mathrm{H}^2(X,\mathbb{Q})\times \mathrm{H}^2(X,\mathbb{Q})\to\mathbb{Q}$$

its Beauville–Bogomolov–Fujiki, and by c_X its Fujiki constant. These are related by

$$\int_X \lambda^{2d}=\frac{(2d)!}{2^d d!}c_X b(\lambda,\lambda)^d \tag{2}$$

for $\lambda\in \mathrm{H}^2(X,\mathbb{Q})$; see eg [41].

We extend b to a bilinear form on

$$\widetilde{\mathrm{H}}(X,\mathbb{Q}):=\mathbb{Q}\alpha\oplus \mathrm{H}^2(X,\mathbb{Q})\oplus\mathbb{Q}\beta,$$

by declaring α and β to be orthogonal to $\mathrm{H}^2(X,\mathbb{Q})$, and setting $b(\alpha,\beta)=-1$, $b(\alpha,\alpha)=0$ and $b(\beta,\beta)=0$. We equip $\widetilde{\mathrm{H}}(X,\mathbb{Q})$ with a grading satisfying $\deg\alpha=-2$ and $\deg\beta=2$, and for which $\mathrm{H}^2(X,\mathbb{Q})$ sits in degree 0. This induces a grading on the Lie algebra $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$.

For $\lambda\in \mathrm{H}^2(X,\mathbb{Q})$ we consider the endomorphism $e_\lambda\in\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ given by $e_\lambda(\alpha)=\lambda$, $e_\lambda(\mu)=b(\lambda,\mu)\beta$ for all $\mu\in \mathrm{H}^2(X,\mathbb{Q})$, and $e_\lambda(\beta)=0$.

Theorem 3.1 (Looijenga–Lunts, Verbitsky) *There is a unique isomorphism of graded Lie algebras*

$$\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))\xrightarrow{\sim}\mathfrak{g}(X)$$

that maps e_λ to e_λ for every $\lambda\in \mathrm{H}^2(X,\mathbb{Q})$.

Proof See [33, Proposition 4.5] or [46, Theorem 1.4] for the theorem over the real numbers. This readily descends to $\mathbb{Q}$; see [43, Proposition 2.9] for more details. □

The representation of $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ on $\mathrm{H}(X,\mathbb{Q})$ integrates to a representation of the group $\mathrm{Spin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ on $\mathrm{H}(X,\mathbb{Q})$. Let $\lambda \in \mathrm{H}^2(X,\mathbb{Q})$. Then e_λ is nilpotent, and hence $B_\lambda := \exp e_\lambda$ is an element of $\mathrm{Spin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$. It acts on $\widetilde{\mathrm{H}}(X,\mathbb{Q})$ by

$$B_\lambda(r\alpha + \mu + s\beta) = r\alpha + (\mu + r\lambda) + \bigl(s + b(\mu,\lambda) + r\tfrac{1}{2}b(\lambda,\lambda)\bigr)\beta \tag{3}$$

for all $r, s \in \mathbb{Q}$ and $\mu \in \mathrm{H}^2(X,\mathbb{Q})$. The action on the total cohomology of X is given by:

Proposition 3.2 *B_λ acts as multiplication by $\mathrm{ch}(\lambda)$ on $\mathrm{H}(X,\mathbb{Q})$.* □

In particular, if $\mathcal{L}$ is a line bundle on X and $\Phi\colon \mathcal{D}X \to \mathcal{D}X$ is the equivalence that maps $\mathcal{F}$ to $\mathcal{F} \otimes \mathcal{L}$, then $\Phi^{\mathrm{H}} = B_{c_1(\mathcal{L})}$.

3.2 The Verbitsky component of cohomology

Let X be a complex hyperkähler variety of dimension $2d$. We define the *even cohomology* of X as the graded $\mathbb{Q}$–algebra

$$\mathrm{H}^{\mathrm{ev}}(X,\mathbb{Q}) := \bigoplus_n \mathrm{H}^{2n}(X,\mathbb{Q}),$$

and the *Verbitsky component* of the cohomology of X as the sub-$\mathbb{Q}$–algebra $\mathrm{SH}(X,\mathbb{Q})$ of $\mathrm{H}^{\mathrm{ev}}(X,\mathbb{Q})$ generated by $\mathrm{H}^2(X,\mathbb{Q})$. Clearly, $\mathrm{SH}(X,\mathbb{Q})[2d]$ is a sub-Lefschetz module of $\mathrm{H}^{\mathrm{ev}}(X,\mathbb{Q})[2d]$ for $\mathrm{H}^2(X,\mathbb{Q})$.

Lemma 3.3 (Verbitsky [8; 45]) *The kernel of the $\mathbb{Q}$–algebra homomorphism*

$$\mathrm{Sym}^\bullet \mathrm{H}^2(X,\mathbb{Q}) \twoheadrightarrow \mathrm{SH}(X,\mathbb{Q})$$

is generated by the elements λ^{d+1} with $\lambda \in \mathrm{H}^2(X,\mathbb{Q})$ satisfying $b(\lambda,\lambda) = 0$. □

Lemma 3.4 (Verbitsky) *$\mathrm{SH}(X,\mathbb{Q})[2d]$ is an irreducible Lefschetz module.*

Proof It is the smallest sub-Lefschetz module of $\mathrm{H}^{\mathrm{ev}}(X,\mathbb{Q})[2d]$ having a nontrivial component of degree $-2d$. □

Verbitsky also describes the space $\mathrm{SH}(X,\mathbb{Q})$ explicitly. Below we normalize this description, and use it to compute the Mukai pairing on $\mathrm{SH}(X,\mathbb{Q})$.

Proposition 3.5 *There is a unique map*

$$\Psi\colon \mathrm{SH}(X,\mathbb{Q})[2d] \to \mathrm{Sym}^d \widetilde{\mathrm{H}}(X,\mathbb{Q})$$

satisfying

(i) Ψ *is morphism of Lefschetz modules*,

(ii) $\Psi(1) = \alpha^d/d!$.

Note that the Lefschetz module structure on $\mathrm{Sym}^d \widetilde{\mathrm{H}}(X,\mathbb{Q})$ is given by the Leibniz rule

$$e_\lambda(x_1\cdots x_d) := \sum_i x_1\cdots e_\lambda(x_i)\cdots x_d.$$

Proof Uniqueness is clear. For existence, consider the map

$$\widetilde{\Psi}\colon \mathrm{Sym}^\bullet \mathrm{H}^2(X,\mathbb{Q}) \to \mathrm{Sym}^d \widetilde{\mathrm{H}}(X,\mathbb{Q}),$$

given by

$$\lambda_1\cdots\lambda_n \mapsto e_{\lambda_1}\cdots e_{\lambda_n}(\alpha^d/d!).$$

This map is well defined since the e_{λ_i} commute. Moreover, the map is graded and satisfies $\widetilde{\Psi}(\lambda x) = e_\lambda \widetilde{\Psi}(x)$ for all $\lambda \in \mathrm{H}^2(X,\mathbb{Q})$ and $x \in \mathrm{Sym}^\bullet \mathrm{H}^2(X,\mathbb{Q})$. To show that $\widetilde{\Psi}$ induces a morphism of Lefschetz modules with the desired properties it now suffices to verify that it vanishes on the ideal generated by the λ^{d+1} for $\lambda \in \mathrm{H}^2(X,\mathbb{Q})$ satisfying $b(\lambda,\lambda)=0$. Equivalently, it suffices to show that for every $x \in \mathrm{Sym}^d \widetilde{\mathrm{H}}(X,\mathbb{Q})$ and for every $\lambda \in \mathrm{H}^2(X,\mathbb{Q})$ with $b(\lambda,\lambda)=0$ we have $e_\lambda^{d+1}(x)=0$.

Without loss of generality, we may assume that x is a monomial of the form

$$x = \alpha^i\beta^j\lambda_1\cdots\lambda_m, \quad i+j+m=d, \quad \lambda_i \in \mathrm{H}^2(X,\mathbb{Q}).$$

For degree reasons, we have $e_\lambda^k(\beta^j\lambda_1\cdots\lambda_m)=0$ for $k>m$. Moreover, it follows from $b(\lambda,\lambda)=0$ that $e_\lambda^k(\alpha^i)=0$ for $k>i$. Combining these, one concludes that $e_\lambda^{d+1}(x)=0$, which is what we had to prove. □

Lemma 3.6 $$\Psi(\mathrm{pt}_X) = \beta^d/c_X.$$

Proof Choose $\lambda \in \mathrm{H}^2(X,\mathbb{Q})$ with $b(\lambda,\lambda)\neq 0$. Then we have

$$\Psi(\lambda^{2d}) = e_\lambda^{2d}\left(\frac{\alpha^d}{d!}\right) = \frac{(2d)!}{2^d d!} b(\lambda,\lambda)^d \beta^d. \tag{4}$$

Dividing by (2) gives the claimed identity. □

Consider the contraction (or Laplacian) operator

$$\Delta\colon \operatorname{Sym}^d \widetilde{\mathrm{H}}(X,\mathbb{Q}) \to \operatorname{Sym}^{d-2} \widetilde{\mathrm{H}}(X,\mathbb{Q}),$$

given by

$$x_1 \dots x_d \mapsto \sum_{i<j} b(x_i, x_j) x_1 \cdots \hat{x}_i \cdots \hat{x}_j \cdots x_d.$$

This is a morphism of Lefschetz modules, or equivalently of $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$–modules.

Lemma 3.7 *The sequence of Lefschetz modules*

$$0 \to \mathrm{SH}(X,\mathbb{Q})[2d] \xrightarrow{\Psi} \operatorname{Sym}^d \widetilde{\mathrm{H}}(X,\mathbb{Q}) \xrightarrow{\Delta} \operatorname{Sym}^{d-2} \widetilde{\mathrm{H}}(X,\mathbb{Q}) \to 0$$

is exact.

Proof Since $\Delta\Psi(1) = 0$, we have $\Delta \circ \Psi = 0$. The map Δ is well known to be a surjective map of $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$–modules with irreducible kernel. Since Ψ is nonzero and $\mathrm{SH}(X,\mathbb{Q})$ is irreducible, it follows that the sequence is exact. □

The *Mukai pairing* [14] on $\mathrm{H}^{\mathrm{ev}}(X,\mathbb{Q})$ restricts to a pairing b_{SH} on $\mathrm{SH}(X,\mathbb{Q})$. It pairs elements of degree m with elements of degree $2d-m$, according to the formula

$$b_{\mathrm{SH}}(\lambda_1 \cdots \lambda_m,\ \mu_1 \cdots \mu_{2d-m}) = (-1)^m \int_X \lambda_1 \cdots \lambda_m \mu_1 \cdots \mu_{2d-m}.$$

Note that $b_{\mathrm{SH}}(e_\lambda x, y) + b_{\mathrm{SH}}(x, e_\lambda y) = 0$ for all $x, y \in \mathrm{SH}(X,\mathbb{Q})$ and $\lambda \in \mathrm{H}^2(X,\mathbb{Q})$, so b_{SH} is $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$–invariant.

The pairing on $\widetilde{\mathrm{H}}(X,\mathbb{Q})$ induces a pairing on $\operatorname{Sym}^d \widetilde{\mathrm{H}}(X,\mathbb{Q})$ defined by

$$b_{[d]}(x_1 \cdots x_d, y_1 \cdots y_d) := (-1)^d \sum_{\sigma \in \mathfrak{S}_d} \prod_i b(x_i, y_{\sigma i}).$$

By construction, $b_{[d]}$ is $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$–invariant. The map Ψ is almost an isometry, in the following sense:

Proposition 3.8 *For all $x, y \in \mathrm{SH}(X,\mathbb{Q})$,*

$$c_X b_{[d]}(\Psi x, \Psi y) = b_{\mathrm{SH}}(x, y).$$

Proof Both the Mukai form on $\mathrm{SH}(X,\mathbb{Q})[2d]$ and the pairing on $\operatorname{Sym}^d \widetilde{\mathrm{H}}(X,\mathbb{Q})$ are $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$–invariant. Since $\mathrm{SH}(X,\mathbb{Q})$ is an irreducible $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$–module, it suffices to verify the identity for some $x, y \in \mathrm{SH}(X,\mathbb{Q})$ with $b_{\mathrm{SH}}(x, y) \neq 0$.

Let $\lambda \in \mathrm{H}^2(X,\mathbb{Q})$ with $b(\lambda,\lambda) \neq 0$. We have

$$b_{\mathrm{SH}}(1,\lambda^{2d}) = \int_X \lambda^{2d} = \frac{(2d)!}{2^d d!} c_X b(\lambda,\lambda)^d \neq 0.$$

By (4),

$$\Psi(\lambda^{2d}) = \frac{(2d)!}{2^d d!} b(\lambda,\lambda)^d \beta^d,$$

and hence

$$c_X b_{[d]}(\Psi(1),\Psi(\lambda^{2d})) = \frac{c_X(2d)!}{2^d (d!)^2} b_{[d]}(\alpha^d,\beta^d) = \frac{c_X(2d)!}{2^d d!} b(\lambda,\lambda)^d,$$

which agrees with the above expression for $b_{\mathrm{SH}}(1,\lambda^{2d})$. □

Remark 3.9 If X is of type $\mathrm{K3}^{[d]}$ then $c_X = 1$ and Ψ is an isometry.

4 Action of derived equivalences on the Verbitsky component

In this section we prove Theorems B and C from the introduction.

4.1 A representation-theoretical construction

Let K be a field of characteristic different from 2, and let $V = (V,b)$ be a nondegenerate quadratic space over K. Let d be a positive integer and consider the space

$$S_{[d]}V := \ker(\mathrm{Sym}^d V \xrightarrow{\Delta} \mathrm{Sym}^{d-2} V).$$

The Lie algebra $\mathfrak{so}(V)$ acts faithfully on $S_{[d]}V$, inducing an inclusion

$$\mathfrak{so}(V) \subset \mathrm{End}(S_d V).$$

Consider the normalizer of $\mathfrak{so}(V)$ in $\mathrm{GL}(S_{[d]}V)$, that is, the group

$$N(V,d) := \{g \in \mathrm{GL}(S_{[d]}V) \mid g\,\mathfrak{so}(V)g^{-1} = \mathfrak{so}(V)\}.$$

Proposition 4.1 *Assume that K is separably closed. Then there is an exact sequence*

$$1 \to \{\pm 1\} \to \mathrm{O}(V) \times K^\times \to N(V,d) \to 1,$$

where the inclusion maps ϵ to (ϵ,ϵ^d) and the surjection maps (φ,λ) to $\lambda S_{[d]}(\varphi)$.

Proof The only nontrivial part is surjectivity of $\mathrm{O}(V) \times K^\times \to N(V,d)$. Denote by

$$\sigma\colon \mathrm{O}(V) \to N(V,d), \quad \varphi \mapsto S_{[d]}(\varphi),$$

the restriction of this map to the first component.

The representation $S_{[d]}V$ of $\mathfrak{so}(V)$ is irreducible, so by Schur's lemma the centralizer of $\mathfrak{so}(V)$ in $\mathrm{GL}(S_{[d]}V)$ is $K^\times$, and we have an exact sequence

$$1 \to K^\times \to N(V,d) \xrightarrow{\psi} \mathrm{Aut}(\mathfrak{so}(V)).$$

It therefore suffices to show that the image of ψ equals the image of $\psi \circ \sigma$.

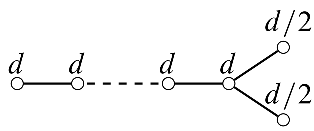
The adjoint group of $\mathfrak{so}(V)$ is $\mathrm{PSO}(V)$, and we have a short exact sequence

$$1 \to \mathrm{PSO}(V) \to \mathrm{Aut}(\mathfrak{so}(V)) \to \mathrm{Out}(\mathfrak{so}(V)) \to 1, \tag{5}$$

where $\mathrm{Out}(\mathfrak{so}(V))$ coincides with the group of symmetries of the Dynkin diagram.

If $\dim V = 2n+1$, then we have $\mathrm{PSO}(V) = \mathrm{SO}(V)$. The Dynkin diagram (of type B_n) has no nontrivial automorphisms, so $\mathrm{Aut}(\mathfrak{so}(V,b)) = \mathrm{SO}(V)$. The composition $\psi \circ \sigma$ maps $\mathrm{SO}(V)$ identically to $\mathrm{SO}(V)$, and we conclude that the image of ψ is the image of $\psi \circ \sigma$.

Now assume $\dim V = 2n$. Since K is algebraically closed, $\mathrm{PSO}(V) = \mathrm{SO}(V)/\{\pm 1\}$. The larger group $\mathrm{O}(V)/\{\pm 1\}$ embeds in $\mathrm{Aut}\,\mathfrak{so}(V)$, with elements of determinant -1 in $\mathrm{O}(V)$ inducing the reflection in the horizontal axis in the Dynkin diagram (of type D_n). For $n \neq 4$, this inclusion is an equality, while for $n = 4$ "triality" gives extra automorphisms. However, expressed on simple roots the highest weight of the representation $S_{[d]}V$ of $\mathfrak{so}(V)$ is



such that for $n = 4$ the extra automorphisms of $\mathfrak{so}(V)$ do not lift to automorphisms of $S_{[d]}V$. We conclude that the image of ψ is contained in $\mathrm{O}(V)/\{\pm 1\}$ and that the composition $\psi \circ \sigma$ is the natural map $\mathrm{O}(V) \to \mathrm{O}(V)/\{\pm 1\}$, so also in this case the image of ψ coincides with the image of $\psi \circ \sigma$. $\square$

Remark 4.2 The condition that K is algebraically closed is needed in the case of even $\dim V$. If K is not algebraically closed, then one still has the exact sequence (5), but one should be careful to define $\mathrm{PSO}(V)$ as the group of K–points of the algebraic group $\mathbf{PSO}(V)$ over K. In general, this group is bigger than $\mathrm{SO}(V)/\{\pm 1\}$). In particular, not every element of $N(V,d)$ can be lifted to $\mathrm{O}(V) \times K^\times$.

Proposition 4.3 *Let V_1 and V_2 be nondegenerate quadratic spaces over K. Assume that there is a linear isomorphism $f : S_{[d]}V_1 \to S_{[d]}V_2$ such that $f\,\mathfrak{so}(V_1)f^{-1} = \mathfrak{so}(V_2)$*

as subspaces of $\mathrm{End}(V_2)$. *Then there exists a* $\mu \in K^\times$ *and a similitude* $\varphi\colon V_1 \to V_2$ *such that* $f = \mu S_{[d]}(\varphi)$.

Proof Let $\overline{K}$ be a separable closure of K. Consider the $\mathrm{Gal}(\overline{K}/K)$–sets

$$S := \{\varphi\colon V_{1,\overline{K}} \to V_{2,\overline{K}} \mid \varphi \text{ is a similitude}\}$$

and

$$N := \{g\colon S_{[d]}V_{1,\overline{K}} \to S_{[d]}V_{2,\overline{K}} \mid g\,\mathfrak{so}(V_{1,\overline{K}})g^{-1} = \mathfrak{so}(V_{2,\overline{K}})\}$$

and the Galois-equivariant map

$$\xi\colon \overline{K}^\times \times S \to N, \quad (\mu,\varphi) \mapsto \mu S_{[d]}(\varphi).$$

The map ξ is surjective. Indeed, since over a separably closed field the quadratic spaces are isometric, we may assume without loss of generality that $V_1 = V_2$. Then $N = N(V_{1,\overline{K}}, d)$ and the surjectivity follows from Proposition 4.1 (it suffices even to consider isometries instead of similitudes).

The group $\overline{K}^\times$ acts on $\overline{K}^\times \times S$ by $\lambda(\mu,\varphi) := (\lambda^{-d}\mu, \lambda\varphi)$ and the fibers of ξ are principal homogenous spaces under this action.

The map f defines a Galois-invariant element $f \in N$, so its fiber $\xi^{-1}(f)$ carries a natural Galois action. By Hilbert 90, we have $\mathrm{H}^1(\mathrm{Gal}(\overline{K}/K, \overline{K}^\times) = \{1\}$, which implies that $\xi^{-1}(f)$ contains a Galois-invariant element (μ,φ). □

The bilinear form b on V induces a bilinear form $b_{[d]}$ on $S_{[d]}V$ defined as

$$b_{[d]}(x_1\cdots x_d, y_1\cdots y_d) := (-1)^d \sum_{\sigma\in S_n}\prod_i b(x_i, y_{\sigma i}),$$

Consider the group

$$G(V,d) := N(V,d) \cap \mathrm{O}(S_{[d]}, b_{[d]})$$

of isometries of $S_{[d]}V$ that preserve the subspace $\mathfrak{so}(V)$ of $\mathrm{End}\, S_{[d]}V$.

Proposition 4.4 *If d is odd, then the map*

$$\mathrm{O}(V) \to G(V,d), \quad \varphi \mapsto S_{[d]}(\varphi),$$

is an isomorphism. If d is even and $\dim V$ *is odd, then the map*

$$\mathrm{O}(V) \to G(V,d), \quad \varphi \mapsto \det(\varphi)S_{[d]}(\varphi),$$

is an isomorphism.

Proof Assume first that K is separably closed. The short exact sequence of Proposition 4.1 restricts to a short exact sequence

$$1 \to \{\pm 1\} \to \mathrm{O}(V) \times \{\pm 1\} \to G(V,d) \to 1,$$

from which one verifies directly that the given maps are isomorphisms. If K is not separably closed, then the result follows from taking Galois invariants. $\square$

Remark 4.5 If both d and $\dim V$ are even, one obtains

$$G(V_{\overline{K}}, d) \cong \mathrm{O}(V_{\overline{K}})/\{\pm 1\} \times \{\pm 1\}.$$

Note, however, that in general there are more Galois-invariant elements than just those in $\mathrm{O}(V)/\{\pm 1\}$. See also Remark 4.2.

4.2 The Verbitsky component

Theorem 4.6 *Let X_1 and X_2 be hyperkähler varieties and $\Phi\colon \mathcal{D}X_1 \to \mathcal{D}X_2$ an equivalence. Then the induced isomorphism $\Phi^{\mathrm{H}}\colon \mathrm{H}(X_1,\mathbb{Q}) \to \mathrm{H}(X_2,\mathbb{Q})$ restricts to an isomorphism $\Phi^{\mathrm{SH}}\colon \mathrm{SH}(X_1,\mathbb{Q}) \to \mathrm{SH}(X_2,\mathbb{Q})$. Moreover:*

(i) *Φ^{SH} is an isometry with respect to the Mukai pairings.*

(ii) *$\Phi^{\mathrm{SH}}\mathfrak{g}(X_1)(\Phi^{\mathrm{SH}})^{-1} = \mathfrak{g}(X_2)$ in $\mathrm{End}(\mathrm{SH}(X_2,\mathbb{Q}))$.*

Proof Note that $\mathrm{SH}(X,\mathbb{Q})$ can be characterized as the minimal sub-$\mathfrak{g}(X)$–module of $\mathrm{H}(X,\mathbb{Q})$ whose Hodge structure attains the maximal possible level (width). It then follows from Theorem A and from Lemma 3.4 that Φ^{H} restricts to an isomorphism

$$\Phi^{\mathrm{SH}}\colon \mathrm{SH}(X_1,\mathbb{Q}) \xrightarrow{\sim} \mathrm{SH}(X_2,\mathbb{Q})$$

respecting the Lie algebras $\mathfrak{g}(X_1)$ and $\mathfrak{g}(X_2)$. By [14], the map Φ^{H} respects the Mukai pairings, and the theorem follows. $\square$

Definition 4.7 For a complex hyperkähler variety we equip $\mathrm{SH}(X,\mathbb{Q})$ and $\widetilde{\mathrm{H}}(X,\mathbb{Q})$ with Hodge structures of weight 0, given by

$$\mathrm{SH}(X,\mathbb{Q}) \subset \mathrm{H}^{\mathrm{ev}}(X,\mathbb{Q}) = \bigoplus_n \mathrm{H}^{2n}(X,\mathbb{Q}(n))$$

and

$$\widetilde{\mathrm{H}}(X,\mathbb{Q}) = \mathbb{Q}\alpha \oplus \mathrm{H}^2(X,\mathbb{Q}(1)) \oplus \mathbb{Q}\beta.$$

Lemma 4.8 *Let X be a hyperkähler variety of dimension $2d$. Then the map*

$$\Psi\colon \mathrm{SH}(X,\mathbb{Q}) \to \mathrm{Sym}^d\, \widetilde{\mathrm{H}}(X,\mathbb{Q})$$

of Proposition 3.5 is a morphism of Hodge structures of weight 0.

Proof One verifies directly that the "action map"

$$\mathrm{H}^2(X,\mathbb{Q}(1))\otimes\widetilde{\mathrm{H}}(X,\mathbb{Q})\to\widetilde{\mathrm{H}}(X,\mathbb{Q}),$$

which maps (λ,x) to $e_\lambda(x)$ is a map of Hodge structures. From this it follows that the action map

$$\mathrm{H}^2(X,\mathbb{Q}(1))\otimes\mathrm{Sym}^d\,\widetilde{\mathrm{H}}(X,\mathbb{Q})\to\mathrm{Sym}^d\,\widetilde{\mathrm{H}}(X,\mathbb{Q})$$

is a map of Hodge structures, and that the map

$$\widetilde{\Psi}\colon\mathrm{Sym}^\bullet\,\mathrm{H}(X,\mathbb{Q}(1))\to\mathrm{Sym}^d\,\widetilde{\mathrm{H}}(X,\mathbb{Q})$$

from the proof of Proposition 3.5 is a morphism of Hodge structures.

Since multiplication in the cohomology of X preserves the Hodge structure, the quotient map $\mathrm{Sym}^\bullet\,\mathrm{H}(X,\mathbb{Q}(1))\to\mathrm{SH}(X,\mathbb{Q})$ is also a morphism of Hodge structures, and hence so is the map Ψ constructed in the proof of Proposition 3.5. □

Proposition 4.9 *Let X_1 and X_2 be derived equivalent hyperkähler varieties. Then there exists a Hodge similitude $\varphi\colon\widetilde{\mathrm{H}}(X_1,\mathbb{Q})\xrightarrow{\sim}\widetilde{\mathrm{H}}(X_2,\mathbb{Q})$ and a scalar $\lambda\in\mathbb{Q}^\times$ such that the square*

$$\begin{tikzcd}
\mathrm{SH}(X_1,\mathbb{Q}) \arrow[r, "\Phi^{\mathrm{SH}}"] \arrow[d, "\Psi"] & \mathrm{SH}(X_2,\mathbb{Q}) \arrow[d, "\Psi"] \\
\mathrm{Sym}^d\,\widetilde{\mathrm{H}}(X_1,\mathbb{Q}) \arrow[r, "\lambda\,\mathrm{Sym}^d(\varphi)"] & \mathrm{Sym}^d\,\widetilde{\mathrm{H}}(X_2,\mathbb{Q})
\end{tikzcd}$$

commutes.

Proof Recall from Lemma 3.7 that the image of Ψ is precisely $S_{[d]}\widetilde{\mathrm{H}}\subset\mathrm{Sym}^d\,\widetilde{\mathrm{H}}$. It then follows from Theorem 4.6 and Proposition 4.3 that there exists a similitude φ and a scalar λ that make the square commute.

It remains to check that φ respects the Hodge structures. The Hodge structure on $\widetilde{\mathrm{H}}(X_i,\mathbb{Q})$ is given by a morphism $h_i:\mathbb{C}^\times\to\mathrm{O}(\widetilde{\mathrm{H}}(X_i,\mathbb{R}))$, and the preceding lemma implies that the Hodge structure on $\mathrm{SH}(X_i,\mathbb{Q})$ is given by composing h_i with the injective map $\mathrm{O}(\widetilde{\mathrm{H}}(X_i,\mathbb{R}))\to\mathrm{GL}(\mathrm{SH}(X_i,\mathbb{R}))$. Since φ maps the Hodge structure on $\mathrm{SH}(X_1,\mathbb{Q})$ to the Hodge structure on $\mathrm{SH}(X_2,\mathbb{Q})$, we conclude that φ maps h_1 to h_2. □

Theorem 4.10 (d odd) *Assume that d is odd, and that X_1 and X_2 are deformation-equivalent hyperkähler varieties of dimension $2d$. Let $\Phi\colon\mathcal{D}X_1\xrightarrow{\sim}\mathcal{D}X_2$ be an* *equivalence. Then there is a unique Hodge isometry $\Phi^{\widetilde{\mathrm{H}}}$ making the square*

$$\begin{array}{ccc} \mathrm{SH}(X_1,\mathbb{Q}) & \xrightarrow{\Phi^{\mathrm{SH}}} & \mathrm{SH}(X_2,\mathbb{Q}) \\ \downarrow{\Psi} & & \downarrow{\Psi} \\ \mathrm{Sym}^d\,\widetilde{\mathrm{H}}(X_1,\mathbb{Q}) & \xrightarrow{\mathrm{Sym}^d(\Phi^{\widetilde{\mathrm{H}}})} & \mathrm{Sym}^d\,\widetilde{\mathrm{H}}(X_2,\mathbb{Q}) \end{array}$$

commute. The formation of $\Phi^{\widetilde{\mathrm{H}}}$ is functorial in Φ.

Proof Since X_1 and X_2 are deformation equivalent, we can choose an isometry $\varphi\colon \widetilde{\mathrm{H}}(X_1,\mathbb{Q}) \xrightarrow{\sim} \widetilde{\mathrm{H}}(X_2,\mathbb{Q})$. Moreover, X_1 and X_2 have the same Fujiki constant, so $\mathrm{Sym}^d\,\varphi$ restricts to an isometry between the images of Ψ. Then by Theorem 4.6 and Proposition 4.4, there is a unique isometry $\psi \in \mathrm{O}(\widetilde{\mathrm{H}}(X_2,\mathbb{Q}))$ such that $\Phi^{\widetilde{\mathrm{H}}} := \psi\varphi$ makes the square commute. Uniqueness forces its formation to be functorial.

That $\Phi^{\widetilde{\mathrm{H}}}$ respects the Hodge structures follows from the same argument as in the proof of Proposition 4.9. □

If d is even, then both existence and uniqueness of $\Phi^{\widetilde{\mathrm{H}}}$ in the statement of Theorem 4.10 fail. However, if we moreover assume that $b_2(X)$ is odd, then one can use the description of $G(V,d)$ from Proposition 4.4 to salvage this, at the cost of keeping track of a determinant character.

Define an *orientation* on X to be the choice of a generator of $\det \mathrm{H}^2(X,\mathbb{R})$, up to $\mathbb{R}^\times_{>0}$. Equivalently, an orientation is the choice of generator of $\det \widetilde{\mathrm{H}}(X,\mathbb{R})$ up to $\mathbb{R}^\times_{>0}$. Define the *sign* $\epsilon(\varphi)$ of a Hodge isometry $\varphi\colon \widetilde{\mathrm{H}}(X_1,\mathbb{Q}) \xrightarrow{\sim} \widetilde{\mathrm{H}}(X_2,\mathbb{Q})$ as $\epsilon(\varphi)=1$ if φ respects the orientations and $\epsilon(\varphi)=-1$ otherwise. A derived equivalence between oriented hyperkähler varieties is a derived equivalence between the underlying unoriented hyperkähler varieties.

Theorem 4.11 (d even) *Assume that d is even, and that $\Phi\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ is a derived equivalence between oriented hyperkähler varieties of dimension $2d$. Assume that X_1 and X_2 have odd b_2, and that the quadratic spaces $\mathrm{H}^2(X_1,\mathbb{Q})$ and $\mathrm{H}^2(X_2,\mathbb{Q})$ are isometric. Then there exists a unique Hodge isometry $\Phi^{\widetilde{\mathrm{H}}}$ making the square*

$$\begin{array}{ccc} \mathrm{SH}(X_1,\mathbb{Q}) & \xrightarrow{\epsilon(\Phi^{\widetilde{\mathrm{H}}})\Phi^{\mathrm{SH}}} & \mathrm{SH}(X_2,\mathbb{Q}) \\ \downarrow{\Psi} & & \downarrow{\Psi} \\ \mathrm{Sym}^d\,\widetilde{\mathrm{H}}(X_1,\mathbb{Q}) & \xrightarrow{\mathrm{Sym}^d(\Phi^{\widetilde{\mathrm{H}}})} & \mathrm{Sym}^d\,\widetilde{\mathrm{H}}(X_2,\mathbb{Q}) \end{array}$$

commute. Moreover, the formation of $\Phi^{\widetilde{\mathrm{H}}}$ is functorial for composition of derived equivalences between hyperkähler varieties equipped with orientations.

Proof The argument is quite similar to the proof of Theorem 4.10. Choose an isometry $\varphi\colon \widetilde{\mathrm{H}}(X_1,\mathbb{Q}) \xrightarrow{\sim} \widetilde{\mathrm{H}}(X_2,\mathbb{Q})$. Because the dimension of $\widetilde{\mathrm{H}}(X_i,\mathbb{Q})$ is odd, we may replace φ with $-\varphi$ if necessary to ensure that φ respects the orientations, and hence we may assume $\epsilon(\varphi)=1$. The map φ induces an isometry $\mathrm{Sym}^d\,\varphi$, which restricts to an isometry $\varphi^{\mathrm{SH}}\colon \mathrm{SH}(X_1,\mathbb{Q}) \to \mathrm{SH}(X_2,\mathbb{Q})$.

By Theorem 4.6, there is a $\psi \in G(\widetilde{\mathrm{H}}(X_2,\mathbb{Q}),d)$ such that $\Phi^{\mathrm{SH}} = \psi\circ\varphi^{\mathrm{SH}}$, and by Proposition 4.4, we have that $\psi = \det(\psi_0)S_{[d]}(\psi_0)$ for a unique $\psi_0 \in \mathrm{O}(\widetilde{\mathrm{H}}(X_2,\mathbb{Q}))$. Now take $\Phi^{\widetilde{\mathrm{H}}} := \psi_0\circ\varphi$. Then $\epsilon(\Phi^{\widetilde{\mathrm{H}}}) = \det(\psi_0)$ and $\mathrm{Sym}^d(\Phi^{\widetilde{\mathrm{H}}})$ lifts to the map $\det(\psi_0)^{-1}\psi\circ\varphi^{\mathrm{SH}} = \epsilon(\Phi^{\widetilde{\mathrm{H}}})\Phi^{\mathrm{SH}}$ as claimed.

Proposition 4.4 forces $\Phi^{\widetilde{\mathrm{H}}}$ to be unique, and this implies the functoriality for composition. Compatibility with Hodge structures follows from the same argument as in the proof of Proposition 4.9. □

Remark 4.12 If X_1 and X_2 are hyperkähler varieties belonging to one of the known families, and if $\Phi\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ is an equivalence, then the hypotheses of either Theorem 4.10 or Theorem 4.11 are satisfied. Indeed, X_1 and X_2 will have the same dimension $2d$ and because they have isomorphic LLV Lie algebra, they have the same second Betti number b_2. Going through the list of known families, one sees that this implies that X_1 and X_2 are deformation equivalent. In particular, they have isometric H^2. Finally, all known hyperkähler varieties of dimension $2d$ with d even have odd b_2.

Taking $X_1 = X_2$ in Theorems 4.10 and 4.11 yields Theorem C from the introduction:

Theorem 4.13 *Let X be a hyperkähler variety of dimension $2d$. Assume that either d is odd or that d is even and $b_2(X)$ is odd. Then the representation*

$$\rho^{\mathrm{SH}}\colon \mathrm{Aut}\,\mathcal{D}(X) \to \mathrm{GL}(\mathrm{SH}(X,\mathbb{Q}))$$

factors over a map $\rho^{\widetilde{\mathrm{H}}}\colon \mathrm{Aut}\,\mathcal{D}(X) \to \mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$.

Remark 4.14 For d odd, the implicit map $\mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q})) \to \mathrm{GL}(\mathrm{SH}(X,\mathbb{Q}))$ is the natural map coming from the isomorphism $\mathrm{SH}(X,\mathbb{Q}) \cong S_{[d]}\widetilde{\mathrm{H}}(X,\mathbb{Q}))$. For d even (and b_2 odd), it is the twist of the natural map with the determinant character

$$\mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q})) \to \{\pm 1\}.$$

5 Hodge structures

In this section we prove Theorem D from the introduction.

For a nondegenerate quadratic space V over $\mathbb{Q}$ we will make use of the algebraic groups $\mathbf{SO}(V)$, $\mathbf{Spin}(V)$, and $\mathbf{GSpin}(V)$ (sometimes denoted $\mathbf{CSpin}(V)$) over $\mathbb{Q}$. These groups sit in a commutative diagram with exact rows

$$\begin{tikzcd} 1 \arrow[r] & \mu_2 \arrow[r] \arrow[d] & \mathbf{Spin}(V) \arrow[r] \arrow[d] & \mathbf{SO}(V) \arrow[r] \arrow[d] & 1 \\ 1 \arrow[r] & \mathbf{G}_m \arrow[r] & \mathbf{GSpin}(V) \arrow[r] & \mathbf{SO}(V) \arrow[r] & 1 \end{tikzcd} \tag{6}$$

from which one deduces an exact sequence

$$1 \to \mu_2 \to \mathbf{G}_m \times \mathbf{Spin}(V) \to \mathbf{GSpin}(V) \to 1, \tag{7}$$

where the first map is the diagonal embedding $\epsilon \mapsto (\epsilon, \epsilon)$. Alternatively, one can use (7) as the definition of **GSpin**, and deduce the existence of the above commutative diagram.

We will write $\mathrm{SO}(V)$, $\mathrm{Spin}(V)$, and $\mathrm{GSpin}(V)$ for the groups of $\mathbb{Q}$–points of these algebraic groups. Note that the above exact sequences of algebraic groups need not induce exact sequences of groups of $\mathbb{Q}$–points, and the obstruction can be described in terms of Galois cohomology. The sequence for the Spin–cover of $\mathbf{SO}(V)$ induces an exact sequence

$$1 \to \{\pm 1\} \to \mathrm{Spin}(V) \to \mathrm{SO}(V) \to \mathrm{H}^1(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \{\pm 1\}) = \mathbb{Q}^\times/(\mathbb{Q}^\times)^2,$$

where the connecting homomorphism $\mathrm{SO}(V) \to \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ is the spinor norm. By Hilbert 90, we have $\mathrm{H}^1(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \overline{\mathbb{Q}}^\times) = \{1\}$ and the analogous sequence for the GSpin–cover does induce a short exact sequence

$$1 \to \mathbb{Q}^\times \to \mathrm{GSpin}(V) \to \mathrm{SO}(V) \to 1. \tag{8}$$

This will be used crucially in the proof of Theorem D.

Lemma 5.1 *Let X be a hyperkähler variety of dimension $2d$. There exists a unique action of $\mathbf{GSpin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ on $\mathrm{H}(X,\mathbb{Q})$ such that*

- (i) *the action of $\mathbf{Spin}(\widetilde{\mathrm{H}}(X,\mathbb{Q})) \subset \mathbf{GSpin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ integrates the action of $\mathfrak{g}(X) = \mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$;*
- (ii) *a section $\lambda \in \mathbf{G}_m \subset \mathbf{GSpin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ acts as λ^{i-2d} on $\mathrm{H}^i(X,\mathbb{Q})$.*

Proof The action of $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ integrates to an action of the simply connected algebraic group $\mathbf{Spin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$. This commutes with the action of $\mathbf{G}_m$ for which λ acts as λ^{i-2d} on $\mathrm{H}^i(X,\mathbb{Q})$, and we obtain an action of $\mathbf{G}_m \times \mathbf{Spin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ on $\mathrm{H}(X,\mathbb{Q})$. The lemma claims that this descends to an action of the quotient group $\mathbf{GSpin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$.

By (7) it suffices to verify that the kernel μ_2 acts trivially, ie that $-1 \in \mathrm{Spin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ acts as $(-1)^i$ on $\mathrm{H}^i(X,\mathbb{Q})$. Any $\mathfrak{sl}_2$–triple $(e_\lambda, h, f_\lambda)$ in $\mathfrak{g}(X)$ induces an algebraic subgroup $\mathbf{SL}_2 \subset \mathbf{Spin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ with the property that $\mathrm{diag}(\mu,\mu^{-1}) \in \mathrm{SL}_2(\mathbb{Q})$ acts as μ^i on $\mathrm{H}^{2d+i}(X,\mathbb{Q})$. It follows that $\mathrm{diag}(-1,-1)$ must be mapped to the nontrivial central element $-1 \in \mathrm{Spin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$, and that -1 acts as $(-1)^i$ on $\mathrm{H}^i(X,\mathbb{Q})$. □

Recall from Definition 4.7 that we have equipped $\widetilde{\mathrm{H}}(X,\mathbb{Q})$ and $\mathrm{H}^{\mathrm{ev}}(X,\mathbb{Q})$ with Hodge structures of weight 0. Similarly, we equip the odd cohomology of X with a Hodge structure of weight 1,

$$\mathrm{H}^{\mathrm{odd}}(X,\mathbb{Q}) := \bigoplus_i \mathrm{H}^{2i+1}(X,\mathbb{Q}(i)).$$

Lemma 5.2 *Let $g \in \mathrm{GSpin}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$. If the action of g on $\widetilde{\mathrm{H}}(X,\mathbb{Q})$ respects the Hodge structure, then so does its action on $\mathrm{H}^{\mathrm{ev}}(X,\mathbb{Q})$ and on $\mathrm{H}^{\mathrm{odd}}(X,\mathbb{Q})$.*

Proof This follows immediately from the fact that the Hodge structure is determined by the action of $h' \in \mathfrak{g}(X) \otimes_{\mathbb{Q}} \mathbb{C}$ (see Section 2.3), and from the faithfulness of the $\mathfrak{g}(X)$–module $\widetilde{\mathrm{H}}(X,\mathbb{Q})$. □

Theorem 5.3 *Let X_1 and X_2 be hyperkähler varieties, and let $\Phi\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ be an equivalence. Then for every i the $\mathbb{Q}$–Hodge structures $\mathrm{H}^i(X_1,\mathbb{Q})$ and $\mathrm{H}^i(X_2,\mathbb{Q})$ are isomorphic.*

Proof Consider the Lie algebra isomorphism $\Phi^{\mathfrak{g}}\colon \mathfrak{g}(X_1) \xrightarrow{\sim} \mathfrak{g}(X_2)$ from Theorem A. By Proposition 4.9, there exists a Hodge similitude $\phi\colon \widetilde{\mathrm{H}}(X_1,\mathbb{Q}) \xrightarrow{\sim} \widetilde{\mathrm{H}}(X_2,\mathbb{Q})$ such that the square

$$\begin{tikzcd}
\mathfrak{so}(\widetilde{\mathrm{H}}(X_1,\mathbb{Q})) \arrow[r, "\mathrm{Ad}(\phi)"] \arrow[d] & \mathfrak{so}(\widetilde{\mathrm{H}}(X_2,\mathbb{Q})) \arrow[d] \\
\mathfrak{g}(X_1) \arrow[r, "\Phi^{\mathfrak{g}}"] & \mathfrak{g}(X_2)
\end{tikzcd}$$

commutes. Here the vertical maps are the isomorphisms from Theorem 3.1.

The K3–type Hodge structure $\widetilde{\mathrm{H}}(X_2,\mathbb{Q})$ decomposes as $N \oplus T$, with N and T its algebraic and transcendental parts, respectively. The Hodge similitude ϕ maps the distinguished elements α_1 and β_1 of $\widetilde{\mathrm{H}}(X_1,\mathbb{Q})$ to N. By Witt cancellation, there exists a $\psi_N \in \mathrm{SO}(N)$ and $\lambda,\mu \in \mathbb{Q}^\times$ such that $\psi_N\phi(\alpha_1) = \lambda\alpha_2$ and $\psi_N\phi(\beta_1) = \mu\beta_2$. Extending by the identity, we find a Hodge isometry $\psi \in \mathrm{SO}(\widetilde{\mathrm{H}}(X_2,\mathbb{Q}))$ such that $\psi\phi\colon \widetilde{\mathrm{H}}(X_1,\mathbb{Q}) \xrightarrow{\sim} \widetilde{\mathrm{H}}(X_2,\mathbb{Q})$ is a *graded* Hodge similitude. In particular, the induced map $\psi\phi\colon \mathfrak{g}(X_1) \xrightarrow{\sim} \mathfrak{g}(X_2)$ is graded, and $\psi\phi$ maps the grading element $h_1 \in \mathfrak{g}(X_1)$ to the grading element $h_2 \in \mathfrak{g}(X_2)$.

By (8) the element ψ lifts to an element $\widetilde{\psi} \in \mathrm{GSpin}(\widetilde{\mathrm{H}}(X_2,\mathbb{Q}))$, which by Lemma 5.1 and Lemma 5.2 induces automorphisms of the Hodge structures $\mathrm{H}^{\mathrm{ev}}(X_2,\mathbb{Q})$ and $\mathrm{H}^{\mathrm{odd}}(X_2,\mathbb{Q})$. Now, by construction, the composition $\widetilde{\psi} \circ \Phi^{\mathrm{H}}$ defines isomorphisms

$$\widetilde{\psi} \circ \Phi^{\mathrm{H}}\colon \mathrm{H}^{\mathrm{ev}}(X_1,\mathbb{Q}) \xrightarrow{\sim} \mathrm{H}^{\mathrm{ev}}(X_2,\mathbb{Q}), \quad \widetilde{\psi} \circ \Phi^{\mathrm{H}}\colon \mathrm{H}^{\mathrm{odd}}(X_1,\mathbb{Q}) \xrightarrow{\sim} \mathrm{H}^{\mathrm{odd}}(X_2,\mathbb{Q}),$$

which respect both the grading and the Hodge structure, so they induce isomorphisms of Hodge structures $\mathrm{H}^i(X_1,\mathbb{Q}) \xrightarrow{\sim} \mathrm{H}^i(X_2,\mathbb{Q})$, for all i. □

6 Topological K–theory

6.1 Topological K–theory and the Mukai vector

We now briefly recall some basic properties of topological K–theory of projective algebraic varieties. See [1; 3; 4] for more details.

For every smooth and projective X over $\mathbb{C}$ we have a $\mathbb{Z}/2\mathbb{Z}$–graded abelian group

$$\mathrm{K}_{\mathrm{top}}(X) := \mathrm{K}^0_{\mathrm{top}}(X) \oplus \mathrm{K}^1_{\mathrm{top}}(X).$$

This is functorial for pullback and proper pushforward, and carries a product structure. The group $\mathrm{K}^0_{\mathrm{top}}(X)$ is the Grothendieck group of topological vector bundles on the differentiable manifold X^{an}. Pullback agrees with pullback of vector bundles, and the product structure agrees with the tensor product of vector bundles.

By [3, Section 1.10], the Chern character can be extended to odd degree, inducing a $\mathbb{Z}/2\mathbb{Z}$–graded map

$$v_X^{\mathrm{top}}\colon \mathrm{K}_{\mathrm{top}}(X) \to \mathrm{H}(X,\mathbb{Q}),$$

given by $v_X^{\mathrm{top}}(\mathcal{F}) = \sqrt{\mathrm{Td}_X} \cdot \mathrm{ch}(\mathcal{F})$. The image of v_X^{top} is a $\mathbb{Z}$–lattice of full rank.

There is a "forgetful" map $\mathrm{K}^0(X) \to \mathrm{K}_{\mathrm{top}}(X)$ from the Grothendieck group of algebraic vector bundles (or equivalently of the triangulated category $\mathcal{D}X$). This is compatible with pullback, multiplication, and proper pushforward. The Mukai vector

$$v_X \colon \mathrm{K}^0(X) \to \mathrm{H}(X, \mathbb{Q})$$

factors over v_X^{top}.

If $\mathcal{P}$ is an object in $\mathcal{D}(X \times Y)$ then convolution with its class in $\mathrm{K}^0_{\mathrm{top}}(X \times Y)$ defines a map $\Phi^{\mathrm{K}}_{\mathcal{P}} \colon \mathrm{K}_{\mathrm{top}}(X) \to \mathrm{K}_{\mathrm{top}}(Y)$, in such a way that the diagram

$$\begin{tikzcd}
\mathrm{K}^0(X) \arrow[r] \arrow[d, "\Phi_{\mathcal{P}}"] & \mathrm{K}_{\mathrm{top}}(X) \arrow[r, "v_X^{\mathrm{top}}"] \arrow[d, "\Phi^{\mathrm{K}}_{\mathcal{P}}"] & \mathrm{H}(X, \mathbb{Q}) \arrow[d, "\Phi^{\mathrm{H}}_{\mathcal{P}}"] \\
\mathrm{K}^0(Y) \arrow[r] & \mathrm{K}_{\mathrm{top}}(Y) \arrow[r, "v_Y^{\mathrm{top}}"] & \mathrm{H}(Y, \mathbb{Q}).
\end{tikzcd}$$

commutes.

6.2 Equivariant topological K–theory

The above formalism largely generalizes to an equivariant setting. Again, we briefly recall the most important properties; see [5; 6; 28; 42] for more details.

If X is a smooth projective complex variety equipped with an action of a finite group G, we denote by $\mathrm{K}^0_G(X)$ the Grothendieck group of G–equivariant algebraic vector bundles on X, or equivalently the Grothendieck group of the bounded derived category $\mathcal{D}_G X$ of G–equivariant coherent $\mathcal{O}_X$–modules. This is functorial for pullback along G–equivariant maps and pushforward along G–equivariant proper maps.

Similarly, we have the G–equivariant topological K–theory

$$\mathrm{K}_{\mathrm{top},G}(X) := \mathrm{K}^0_{\mathrm{top},G}(X) \oplus \mathrm{K}^1_{\mathrm{top},G}(X),$$

where $\mathrm{K}^0_{\mathrm{top},G}(X)$ is the Grothendieck group of topological G–equivariant vector bundles.

There is a natural map $\mathrm{K}^0_G(X) \to \mathrm{K}^0_{\mathrm{top},G}(X)$ compatible with pullback and tensor product. If $f \colon X \to Y$ is *proper* and G–equivariant, then we have a pushforward map $f_* \colon \mathrm{K}_{\mathrm{top},G}(X) \to \mathrm{K}_{\mathrm{top},G}(Y)$. There is a Riemann–Roch theorem [5; 28], stating that the square

$$\begin{tikzcd}
\mathrm{K}^0_G(X) \arrow[r] \arrow[d, "f_*"] & \mathrm{K}_{\mathrm{top},G}(X) \arrow[d, "f_*"] \\
\mathrm{K}^0_G(Y) \arrow[r] & \mathrm{K}_{\mathrm{top},G}(Y)
\end{tikzcd}$$

commutes.

Now assume that we have a finite group G acting on X, and a finite group H acting on Y. If $\mathcal{P}$ is an object in $\mathcal{D}_{G\times H}(X\times Y)$, then convolution with $\mathcal{P}$ induces a functor $\Phi_{\mathcal{P}}\colon \mathcal{D}_G X\to \mathcal{D}_H Y$, see [40] for more details. Similarly, convolution with the class of $\mathcal{P}$ in $\mathrm{K}^0_{\mathrm{top},G\times H}(X\times Y)$ induces a map $\Phi^{\mathrm{K}}_{\mathcal{P}}\colon \mathrm{K}_{\mathrm{top},G}(X)\to \mathrm{K}_{\mathrm{top},H}(Y)$. These satisfy the usual Fourier–Mukai calculus, and moreover they are compatible in the sense that the square

$$\begin{tikzcd}
\mathrm{K}^0_G(X) \arrow[r] \arrow[d, "\Phi_{\mathcal{P}}"] & \mathrm{K}_{\mathrm{top},G}(X) \arrow[d, "\Phi^{\mathrm{K}}_{\mathcal{P}}"] \\
\mathrm{K}^0_H(Y) \arrow[r] & \mathrm{K}_{\mathrm{top},H}(Y)
\end{tikzcd}$$

commutes.

7 Cohomology of the Hilbert square of a K3 surface

Let S be a K3 surface and $X=S^{[2]}$ its Hilbert square. In the coming few paragraphs we recall the structure of the cohomology of X in terms of the cohomology of S. See [7; 17; 23] for more details.

7.1 Line bundles on the Hilbert square

Let $G=\{1,\sigma\}$ be the group of order two, acting on $S\times S$ by permuting the factors. The Hilbert square X sits in a diagram

$$\begin{tikzcd}
 & Z \arrow[dl, "p"'] \arrow[dr, "q"] & \\
S\times S & & X
\end{tikzcd}$$

where $p\colon Z\to S\times S$ is the blow-up along the diagonal, and where $q\colon Z\to X$ is the quotient map for the natural action of G on Z. Denote by $R\subset Z$ the exceptional divisor of p. Then R equals the ramification locus of q. We have $q_*\mathcal{O}_Z=\mathcal{O}_X\oplus\mathcal{E}$ for some line bundle $\mathcal{E}$, and $q^*\mathcal{E}\cong\mathcal{O}_Z(-R)$.

If $\mathcal{L}$ is a line bundle on S then

$$\mathcal{L}_2 := (q_*p^*(\mathcal{L}\boxtimes\mathcal{L}))^G$$

is a line bundle on X. The map

$$\mathrm{Pic}(S)\oplus\mathbb{Z}\to\mathrm{Pic}(X),\quad (\mathcal{L},n)\mapsto\mathcal{L}_2\otimes\mathcal{E}^{\otimes n},$$

is an isomorphism.

7.2 Cohomology of the Hilbert square

There is an isomorphism

$$\mathrm{H}^2(S,\mathbb{Z})\oplus\mathbb{Z}\delta \xrightarrow{\sim} \mathrm{H}^2(X,\mathbb{Z})$$

with the property that $c_1(\mathcal{L})$ is mapped to $c_1(\mathcal{L}_2)$, and δ is mapped to $c_1(\mathcal{E})$. We will use this isomorphism to identify $\mathrm{H}^2(S,\mathbb{Z})\oplus\mathbb{Z}\delta$ with $\mathrm{H}^2(X,\mathbb{Z})$. The Beauville–Bogomolov form on $\mathrm{H}^2(X,\mathbb{Z})$ satisfies

$$b_X(\lambda,\lambda)=b_S(\lambda,\lambda),\quad b_X(\lambda,\delta)=0,\quad b_X(\delta,\delta)=-2$$

for all $\lambda\in\mathrm{H}^2(S,\mathbb{Z})$.

The cup product defines an isomorphism $\mathrm{Sym}^2\,\mathrm{H}^2(X,\mathbb{Q})\xrightarrow{\sim}\mathrm{H}^4(X,\mathbb{Q})$. By Poincaré duality, there is a unique $q_X\in\mathrm{H}^4(X,\mathbb{Q})$ representing the Beauville–Bogomolov form, in the sense that

$$\int_X q_X\lambda_1\lambda_2=b_X(\lambda_1,\lambda_2) \tag{9}$$

for all $\lambda_1,\lambda_2\in\mathrm{H}^2(X,\mathbb{Z})$. Multiplication by q_X defines an isomorphism $\mathrm{H}^2(X,\mathbb{Q})\to\mathrm{H}^6(X,\mathbb{Q})$, and, for all $\lambda_1,\lambda_2,\lambda_3\in\mathrm{H}^2(X,\mathbb{Q})$,

$$\lambda_1\lambda_2\lambda_3=b_X(\lambda_1,\lambda_2)q_X\lambda_3+b_X(\lambda_2,\lambda_3)q_X\lambda_1+b_X(\lambda_3,\lambda_1)q_X\lambda_2 \tag{10}$$

in $\mathrm{H}^6(X,\mathbb{Q})$. Finally, for all $\lambda\in\mathrm{H}^2(X,\mathbb{Q})$ the Fujiki relation

$$\int_X\lambda^4=3b_X(\lambda,\lambda)^2 \tag{11}$$

holds.

7.3 Todd class of the Hilbert square

Proposition 7.1 $\mathrm{Td}_X=1+\frac{5}{2}q_X+3[\mathrm{pt}]$.

Proof See also [23, Section 23.4]. Since the Todd class is invariant under the monodromy group of X, we necessarily have

$$\mathrm{Td}_X=1+sq_X+t[\mathrm{pt}]$$

for some $s,t\in\mathbb{Q}$. By Hirzebruch–Riemann–Roch, for every line bundle L on S with $c_1(L)=\lambda$,

$$\chi(X,L_2)=\int_X\mathrm{ch}(\lambda)\mathrm{Td}_X=\frac{1}{24}\int_X\lambda^4+\frac{s}{2}\int_X\lambda^2q_X+t.$$

By the relations (11) and (9), the right-hand side reduces to

$$\tfrac{1}{8}b(\lambda,\lambda)^2 + \tfrac{1}{2}sb(\lambda,\lambda) + t.$$

By [23, Section 23.4] or [17, 5.1], the left-hand side computes to

$$\chi(X, L_2) = \tfrac{1}{8}b(\lambda,\lambda)^2 + \tfrac{5}{4}b(\lambda,\lambda) + 3.$$

Comparing the two expressions yields the result. □

8 Derived McKay correspondence

8.1 The derived McKay correspondence

As in Section 7.1, we consider a K3 surface S, its Hilbert square $X = S^{[2]}$, the maps $p\colon Z \to S \times S$ and $q\colon Z \to X$, and the group $G = \{1, \sigma\}$ acting on $S \times S$ and Z.

The *derived McKay correspondence* [11] is the triangulated functor

$$\mathrm{BKR}\colon \mathcal{D}^b(X) \to \mathcal{D}^b_G(S \times S)$$

given as the composition

$$\mathrm{BKR}\colon \mathcal{D}X \xrightarrow{q^*} \mathcal{D}_G(Z) \xrightarrow{p_*} \mathcal{D}_G(S \times S),$$

where the first functor maps $\mathcal{F}$ to $q^*\mathcal{F}$ equipped with the trivial G–linearization. By [11, Theorem 1.1; 21, Theorem 5.1], the functor BKR is an equivalence of categories.

Its inverse has been described in [31, Section 4]. Denote by $j: Z \to S \times S \times X$ the G–equivariant closed immersion induced by p and q. The exceptional divisor $R \subset Z$ is G–invariant and hence defines a G–equivariant sheaf $\mathcal{O}(R)$, and a G–equivariant sheaf $\mathcal{Q} := j_*\mathcal{O}_Z(R)$ in $\mathcal{D}_G(S \times S \times X)$.

Proposition 8.1 *The inverse equivalence of* BKR *is given by the equivariant Fourier–Mukai transform with respect to $\mathcal{Q}$. It maps $\mathcal{F} \in \mathcal{D}_G(S \times S)$ to the object*

$$(q_* p^* \mathcal{F})^{\sigma=-1} \otimes \mathcal{E}^{-1}$$

of $\mathcal{D}(X)$.

Proof The first statement is [31, 4.1]. By the adjunction formula for $j: Z \to S \times S \to X$, this implies that $\mathcal{F}$ is mapped to $\big(q_*(p^*\mathcal{F} \otimes \mathcal{O}_Z(R))\big)^G \in \mathcal{D}(X)$. If we upgrade the line bundle $\mathcal{E}$ on X to a G–equivariant (for the trivial action on X) line bundle $\mathcal{E}_-$

by making σ act as -1, then $q^*\mathcal{E}_- \cong \mathcal{O}_Z(-R)$ as G–equivariant line bundles on Z. Applying the projection formula once more for the equivariant map q, we find

$$(q_*(p^*\mathcal{F} \otimes \mathcal{O}_Z(R)))^G \cong (q_* p^*\mathcal{F} \otimes \mathcal{E}_-^{-1})^G \cong (q_* p^*\mathcal{F})^{\sigma=-1} \otimes \mathcal{E}^{-1}. \qquad \square$$

Now let S_1 and S_2 be K3 surfaces with Hilbert squares X_1 and X_2. As was observed by Ploog [39], any equivalence $\Phi\colon \mathcal{D}S_1 \xrightarrow{\sim} \mathcal{D}S_2$ induces an equivalence

$$\mathcal{D}_G(S_1 \times S_2) \xrightarrow{\sim} \mathcal{D}_G(S_2 \times S_2),$$

and hence, via the derived McKay correspondence, an equivalence $\Phi^{[2]}\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$.

8.2 Topological K–theory of the Hilbert square

Theorem 8.2 *The composition*

$$\mathrm{BKR}_{\mathrm{top}}\colon \mathrm{K}_{\mathrm{top}}(X) \xrightarrow{q^*} \mathrm{K}_{\mathrm{top},G}(Z) \xrightarrow{p_*} \mathrm{K}_{\mathrm{top},G}(S \times S)$$

is an isomorphism.

Proof (See also [11, Section 10].) This is a purely formal consequence of the calculus of equivariant Fourier–Mukai transforms sketched in Section 6.2. The functor BKR and its inverse are given by kernels $\mathcal{P} \in \mathcal{D}_G(X \times S \times S)$ and $\mathcal{Q} \in \mathcal{D}_G(S \times S \times X)$. The map $\mathrm{BKR}_{\mathrm{top}}$ is given by convolution with the class of $\mathcal{P}$ in $\mathrm{K}^0_{\mathrm{top},G}(X \times S \times S)$. The identities in $\mathrm{K}^0(X \times X)$ and $\mathrm{K}^0_{G\times G}(S \times S \times S \times S)$ witnessing that $\mathcal{P}$ and $\mathcal{Q}$ are mutually inverse equivalences induce analogous identities in $\mathrm{K}^0_{\mathrm{top}}$. These show that convolution with the class of $\mathcal{Q}$ defines a two-sided inverse to $\mathrm{BKR}_{\mathrm{top}}$. $\square$

Consider the map

$$\psi^{\mathrm{K}}\colon \mathrm{K}^0_{\mathrm{top}}(X) \to \mathrm{K}^0_{\mathrm{top}}(S \times S)^G$$

obtained as the composition of $\mathrm{BKR}_{\mathrm{top}}$ and the forgetful map from $\mathrm{K}^0_{\mathrm{top},G}(S \times S)$ to $\mathrm{K}^0_{\mathrm{top}}(S \times S)$. Also, consider the map

$$\theta^{\mathrm{K}}\colon \mathrm{K}^0_{\mathrm{top}}(S) \to \mathrm{K}^0_{\mathrm{top}}(X), \quad [\mathcal{F}] \mapsto \mathrm{BKR}^{-1}_{\mathrm{top}}([\mathcal{F} \boxtimes \mathcal{F}, 1] - [\mathcal{F} \boxtimes \mathcal{F}, -1]),$$

where $[\mathcal{F} \boxtimes \mathcal{F}, \pm 1]$ denotes the class of the topological vector bundle $\mathcal{F} \boxtimes \mathcal{F}$ equipped with $\pm$ the natural G–linearization.

By construction, these maps are "functorial" in $\mathcal{D}S$, in the following sense:

Proposition 8.3 *If $\Phi\colon \mathcal{D}S_1 \xrightarrow{\sim} \mathcal{D}S_2$ is a derived equivalence between* K3 *surfaces, and $\Phi^{[2]}\colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ is the induced equivalence between their Hilbert squares, then the squares*

$$\begin{array}{ccc} \mathrm{K}^0_{\mathrm{top}}(X_1) & \xrightarrow{\psi^{\mathrm{K}}} & \mathrm{K}^0_{\mathrm{top}}(S_1\times S_1)^G \\ \downarrow{\scriptstyle \Phi^{[2],\mathrm{K}}} & & \downarrow{\scriptstyle \Phi^{\mathrm{K}}\otimes\Phi^{\mathrm{K}}} \\ \mathrm{K}^0_{\mathrm{top}}(X_2) & \xrightarrow{\psi^{\mathrm{K}}} & \mathrm{K}^0_{\mathrm{top}}(S_2\times S_2)^G \end{array} \qquad \begin{array}{ccc} \mathrm{K}^0_{\mathrm{top}}(S_1) & \xrightarrow{\theta^{\mathrm{K}}} & \mathrm{K}^0_{\mathrm{top}}(X_1) \\ \downarrow{\scriptstyle \Phi^{\mathrm{K}}} & & \downarrow{\scriptstyle \Phi^{[2],\mathrm{K}}} \\ \mathrm{K}^0_{\mathrm{top}}(S_2) & \xrightarrow{\theta^{\mathrm{K}}} & \mathrm{K}^0_{\mathrm{top}}(X_2) \end{array}$$

commute. □

Proposition 8.4 *The sequence*

$$0 \to \mathrm{K}^0_{\mathrm{top}}(S)\otimes_{\mathbb{Z}}\mathbb{Q} \xrightarrow{\theta^{\mathrm{K}}} \mathrm{K}^0_{\mathrm{top}}(X)\otimes_{\mathbb{Z}}\mathbb{Q} \xrightarrow{\psi^{\mathrm{K}}} \mathrm{K}^0_{\mathrm{top}}(S\times S)^G\otimes_{\mathbb{Z}}\mathbb{Q} \to 0$$

is exact.

Proof In the proof, we will implicitly identify $\mathrm{K}_{\mathrm{top},G}(S\times S)$ and $\mathrm{K}_{\mathrm{top}}(X)$.

Note that the map θ^{K} is additive. Indeed, let $\mathcal{F}_1$ and $\mathcal{F}_2$ be (topological) vector bundles on S. Then the cross term $\theta^{\mathrm{K}}[\mathcal{F}_1\oplus\mathcal{F}_2]-\theta^{\mathrm{K}}[\mathcal{F}_1]-\theta^{\mathrm{K}}[\mathcal{F}_2]$ computes to

$$\left[\mathcal{F}_1\boxtimes\mathcal{F}_2\oplus\mathcal{F}_2\boxtimes\mathcal{F}_1,\ \left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)\right]-\left[\mathcal{F}_1\boxtimes\mathcal{F}_2\oplus\mathcal{F}_2\boxtimes\mathcal{F}_1,\ \left(\begin{smallmatrix}0&-1\\-1&0\end{smallmatrix}\right)\right],$$

which vanishes because the matrices $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}0&-1\\-1&0\end{smallmatrix}\right)$ are conjugated over $\mathbb{Z}$.

Next we observe that $\psi^{\mathrm{K}}\colon \mathrm{K}^0_{\mathrm{top}}(X)\otimes_{\mathbb{Z}}\mathbb{Q}\to\mathrm{K}^0_{\mathrm{top}}(S\times S)^G\otimes_{\mathbb{Z}}\mathbb{Q}$ is surjective. Indeed, by the Künneth formula [2], the group $\mathrm{K}^0_{\mathrm{top}}(S\times S)^G\otimes_{\mathbb{Z}}\mathbb{Q}$ is generated by classes of the form $[\mathcal{F}_1\boxtimes\mathcal{F}_2\oplus\mathcal{F}_2\boxtimes\mathcal{F}_1]$, and these lie in the image of ψ^{K}.

Also, the composition $\psi^{\mathrm{K}}\theta^{\mathrm{K}}$ vanishes. Computing the $\mathbb{Q}$–dimensions one sees that it suffices to show that θ^{K} is injective to conclude that the sequence is exact.

Pulling back to the diagonal and taking invariants defines a map

$$\mathrm{K}^0_{\mathrm{top}}(S)\xrightarrow{\theta^{\mathrm{K}}}\mathrm{K}^0_{\mathrm{top},G}(S\times S)\xrightarrow{\Delta^*}\mathrm{K}^0_{\mathrm{top},G}(S)\xrightarrow{(-)^G}\mathrm{K}^0_{\mathrm{top}}(S).$$

This composition computes to

$$[\mathcal{F}]\mapsto[\mathrm{Sym}^2\mathcal{F}]-\left[\textstyle\bigwedge^2\mathcal{F}\right].$$

This coincides with the second Adams operation, which is injective on $\mathrm{K}^0_{\mathrm{top}}(S)\otimes_{\mathbb{Z}}\mathbb{Q}$, since it has eigenvalues 1, 2, and 4. We conclude that θ^{K} is injective, and the proposition follows. □

8.3 A computation in the cohomology of the Hilbert square

We now come to the technical heart of our computation of the derived monodromy of the Hilbert square of a K3 surface.

Consider the map $\theta^{\mathrm{H}}\colon \mathrm{H}(S,\mathbb{Q}) \to \mathrm{H}(X,\mathbb{Q})$ given by

$$\theta^{\mathrm{H}}(s+\lambda+t\,\mathrm{pt}_S) = (s\delta+\lambda\delta+tq_X\delta)\cdot e^{-\delta/2}, \tag{12}$$

for all $s,t\in\mathbb{Q}$ and $\lambda\in\mathrm{H}^2(S,\mathbb{Q})$. See Section 7.2 for the definition of $\delta\in\mathrm{H}^2(X,\mathbb{Q})$ and $q_X\in\mathrm{H}^4(X,\mathbb{Q})$.

Proposition 8.5 *The square*

$$\begin{array}{ccc} \mathrm{K}^0_{\mathrm{top}}(S) & \xrightarrow{\theta^{\mathrm{K}}} & \mathrm{K}^0_{\mathrm{top}}(X) \\ \big\downarrow{\scriptstyle v_S^{\mathrm{top}}} & & \big\downarrow{\scriptstyle v_X^{\mathrm{top}}} \\ \mathrm{H}(S,\mathbb{Q}) & \xrightarrow{\theta^{\mathrm{H}}} & \mathrm{H}(X,\mathbb{Q}) \end{array}$$

commutes.

Proof Since $\mathrm{K}^0_{\mathrm{top}}(S)\otimes_{\mathbb{Z}}\mathbb{Q}$ is additively generated by line bundles, it suffices to show

$$v_X^{\mathrm{top}}(\theta^{\mathrm{K}}(\mathcal{L})) = \big(\delta+\lambda\delta+\big(\tfrac{1}{2}b(\lambda,\lambda)+1\big)q_X\delta\big)\cdot e^{-\delta/2} \tag{13}$$

for a topological line bundle $\mathcal{L}$ with $\lambda=c_1(\mathcal{L})$. Deforming S if necessary, we may assume that $\mathcal{L}$ is algebraic.

Using Proposition 8.1 and the fact that the natural map

$$\mathcal{L}_2\otimes q_*\mathcal{O}_Z \to q_*p^*(\mathcal{L}\boxtimes\mathcal{L})$$

is an isomorphism of $\mathcal{O}_X$–modules, we find

$$\mathrm{BKR}^{-1}[\mathcal{L}\boxtimes\mathcal{L},1]=\mathcal{L}_2,\quad \mathrm{BKR}^{-1}[\mathcal{L}\boxtimes\mathcal{L},-1]=\mathcal{E}^{-1}\otimes\mathcal{L}_2.$$

We conclude that θ^{K} maps $\mathcal{L}$ to $[\mathcal{L}_2](1-[\mathcal{E}^{-1}])$ in $\mathrm{K}^0(X)$.

We compute its image under v_X. Using the formula for the Todd class from Proposition 7.1, we find

$$v_X(\theta^{\mathrm{K}}(\mathcal{L})) = \big(1+\tfrac{5}{4}q_X+\cdots\big)\exp(\lambda)(1-e^{-\delta}).$$

Since $1-e^{-\delta}$ has no term in degree 0, the degree 8 part of the square root of the Todd class is irrelevant, so we have

$$v_X(\theta^{\mathrm{K}}(\mathcal{L})) = \big(1+\tfrac{5}{4}q_X\big)\exp(\lambda)(1-e^{-\delta}).$$

By the Fujiki relation (11) from Section 7.2, we have $\lambda^3\delta = 0$, so the above can be rewritten as

$$v_X(\theta^{\mathrm{K}}(\mathcal{L})) = \bigl(1+\tfrac{5}{4}q_X\bigr)\cdot\bigl(\delta+\lambda\delta+\tfrac{1}{2}\lambda^2\delta\bigr)\cdot\frac{1-e^{-\delta}}{\delta}.$$

Since $q_X\delta\lambda = b(\delta,\lambda) = 0$, we can rewrite this further as

$$v_X(\theta^{\mathrm{K}}(\mathcal{L})) = \bigl(1+\tfrac{1}{4}q_X\bigr)\cdot\bigl(\delta+\lambda\delta+\bigl(\tfrac{1}{2}b(\lambda,\lambda)+1\bigr)q_X\delta\bigr)\cdot\frac{1-e^{-\delta}}{\delta}.$$

Comparing this with the right-hand side of (13), we see that it suffices to show

$$\bigl(1+\tfrac{1}{4}q_X\bigr)\cdot(1-e^{-\delta}) = \delta e^{-\delta/2}$$

in $\mathrm{H}(X,\mathbb{Q})$. This boils down to the identities

$$\tfrac{1}{6}\delta^3+\tfrac{1}{4}\delta q_X = \tfrac{1}{8}\delta^3,\quad \tfrac{1}{24}\delta^4+\tfrac{1}{8}\delta^2 q_X = \tfrac{1}{48}\delta^4$$

in $\mathrm{H}^6(X,\mathbb{Q})$ and $\mathrm{H}^8(X,\mathbb{Q})$, respectively. These follow easily from the relations (9), (10), and (11) in Section 7.2. □

9 Derived monodromy group of the Hilbert square of a K3 surface

9.1 Derived monodromy groups

Let X be a smooth projective complex variety. We call a *deformation* of X the data of a smooth projective variety X', a proper smooth family $\mathcal{X}\to B$, a path $\gamma\colon[0,1]\to\mathcal{X}$, and isomorphisms $X \xrightarrow{\sim} \mathcal{X}_{\gamma(0)}$ and $X' \xrightarrow{\sim} \mathcal{X}_{\gamma(1)}$. We will informally say that X' is a deformation of X, the other data being implicitly understood. Parallel transport along γ defines an isomorphism $\mathrm{H}(X,\mathbb{Q}) \xrightarrow{\sim} \mathrm{H}(X',\mathbb{Q})$.

If X' and X'' are deformations of X, and if $\phi\colon X'\to X''$ is an isomorphism of projective varieties, then we obtain a composite isomorphism

$$\mathrm{H}(X,\mathbb{Q}) \xrightarrow[\sim]{} \mathrm{H}(X',\mathbb{Q}) \xrightarrow[\sim]{\phi} \mathrm{H}(X'',\mathbb{Q}) \xrightarrow[\sim]{} \mathrm{H}(X,\mathbb{Q}).$$

We call such an isomorphism a *monodromy operator* for X, and denote by $\mathrm{Mon}(X)$ the subgroup of $\mathrm{GL}(\mathrm{H}(X,\mathbb{Q}))$ generated by all monodromy operators.

If X' and X'' are deformations of X, and if $\Phi\colon \mathcal{D}X' \xrightarrow{\sim} \mathcal{D}X''$ is an equivalence, then we obtain an isomorphism

$$\mathrm{H}(X,\mathbb{Q}) \xrightarrow[\sim]{} \mathrm{H}(X',\mathbb{Q}) \xrightarrow[\sim]{\Phi^{\mathrm{H}}} \mathrm{H}(X'',\mathbb{Q}) \xrightarrow[\sim]{} \mathrm{H}(X,\mathbb{Q}).$$

We call such an isomorphism a *derived monodromy operator* for X, and denote by $\mathrm{DMon}(X)$ the subgroup of $\mathrm{GL}(\mathrm{H}(X, \mathbb{Q}))$ generated by all derived monodromy operators.

By construction, the derived monodromy group is deformation invariant. It contains the usual monodromy group, and the image of ρ_X, and we have a commutative square of groups

$$\begin{tikzcd} \mathrm{Aut}(X) \arrow[r, hook] \arrow[d] & \mathrm{Aut}(\mathcal{D}X) \arrow[d, "\rho_X"] \\ \mathrm{Mon}(X) \arrow[r, hook] & \mathrm{DMon}(X) \end{tikzcd}$$

Remark 9.1 The above definition is somewhat ad hoc, and should be considered a poor man's derived monodromy group. This is sufficient for our purposes. A more mature definition should involve all noncommutative deformations of X.

Proposition 9.2 *If S is a* K3 *surface, then* $\mathrm{DMon}(S) = \mathrm{O}^+(\widetilde{\mathrm{H}}(S, \mathbb{Z}))$.

Proof Indeed, if $\Phi \colon \mathcal{D}S_1 \to \mathcal{D}S_2$ is an equivalence, then

$$\Phi^{\mathrm{H}} \colon \widetilde{\mathrm{H}}(S_1, \mathbb{Z}) \to \widetilde{\mathrm{H}}(S_2, \mathbb{Z})$$

preserves the Mukai form, as well as a natural orientation on four-dimensional positive subspaces; see [26, Section 4.5]. Also any deformation preserves the Mukai form and the natural orientation, so any derived monodromy operator will land in $\mathrm{O}^+(\widetilde{\mathrm{H}}(S, \mathbb{Z}))$.

The converse inclusion can be easily obtained from the Torelli theorem, together with the results of [22; 39] on derived auto-equivalences of K3 surfaces. Alternatively, one can use that the group $\mathrm{O}^+(\widetilde{\mathrm{H}}(S, \mathbb{Z}))$ is generated by reflections in -2–vectors δ. By the Torelli theorem, any such -2–vector will become algebraic on a suitable deformation S' of S, and by [32] there exists a spherical object $\mathcal{E}$ on S' with Mukai vector $v(\mathcal{E}) = \delta$. The spherical twist in $\mathcal{E}$ then shows that reflection in δ is indeed a derived monodromy operator. □

9.2 Action of DMon(S) on H($X, \mathbb{Q}$)

By the derived McKay correspondence, any derived equivalence $\Phi_S \colon \mathcal{D}S_1 \xrightarrow{\sim} \mathcal{D}S_2$ between K3 surfaces induces a derived equivalence $\Phi_X \colon \mathcal{D}X_1 \xrightarrow{\sim} \mathcal{D}X_2$ between the corresponding Hilbert squares. By Propositions 8.3 and 8.4, the induced map Φ_X^H only depends on Φ_S^H. Since any deformation of a K3 surface S induces a deformation of $X = S^{[2]}$, we conclude that we have a natural homomorphism

$$\mathrm{DMon}(S) \to \mathrm{DMon}(X),$$

and hence an action of $\mathrm{DMon}(S)$ on $\mathrm{H}(X, \mathbb{Q})$. In this subsection, we will explicitly compute this action. As a first approximation, we determine the $\mathrm{DMon}(S)$–module structure of $\mathrm{H}(X, \mathbb{Q})$, up to isomorphism.

Proposition 9.3 *We have* $\mathrm{H}(X, \mathbb{Q}) \cong \widetilde{\mathrm{H}}(S, \mathbb{Q}) \oplus \mathrm{Sym}^2 \widetilde{\mathrm{H}}(S, \mathbb{Q})$ *as representations of* $\mathrm{DMon}(S) = \mathrm{O}^+(\widetilde{\mathrm{H}}(S, \mathbb{Z}))$.

Proof This follows from Propositions 8.3 and 8.4. $\square$

Since $\mathfrak{g}(X)$ is a purely topological invariant, it is preserved under deformations. In particular, Theorem 4.13 implies that we have an inclusion $\mathrm{DMon}(X) \subset \mathrm{O}(\widetilde{\mathrm{H}}(X, \mathbb{Q}))$. We conclude there exists a unique map of algebraic groups h making the square

$$\begin{tikzcd} \mathrm{DMon}(S) \arrow[r] \arrow[d] & \mathrm{DMon}(X) \arrow[d] \\ \mathrm{O}(\widetilde{\mathrm{H}}(S, \mathbb{Q})) \arrow[r, "h"] & \mathrm{O}(\widetilde{\mathrm{H}}(X, \mathbb{Q})) \end{tikzcd} \tag{14}$$

commute.

Recall that in (3) we defined an isometry B_λ of $\widetilde{\mathrm{H}}(X, \mathbb{Q})$ for every $\lambda \in \mathrm{H}^2(X, \mathbb{Q})$.

Theorem 9.4 *The map h in the square* (14) *is given by*

$$g \mapsto \det(g) \cdot (B_{-\delta/2} \circ \iota(g) \circ B_{\delta/2}),$$

with $\iota\colon \mathrm{O}(\widetilde{\mathrm{H}}(S, \mathbb{Q})) \to \mathrm{O}(\widetilde{\mathrm{H}}(X, \mathbb{Q}))$ *the natural inclusion.*

The proof of this theorem will occupy the remainder of this section.

Consider the unique homomorphism of Lie algebras $\iota\colon \mathfrak{g}(S) \to \mathfrak{g}(X)$ that respects the grading and maps e_λ to e_λ for all $\lambda \in \mathrm{H}^2(S, \mathbb{Q}) \subset \mathrm{H}^2(X, \mathbb{Q})$. Under the isomorphism of Theorem 3.1 this corresponds to the map $\mathfrak{so}(\widetilde{\mathrm{H}}(S, \mathbb{Q})) \to \mathfrak{so}(\widetilde{\mathrm{H}}(X, \mathbb{Q}))$ induced by the inclusion of quadratic spaces $\widetilde{\mathrm{H}}(S, \mathbb{Q}) \subset \widetilde{\mathrm{H}}(X, \mathbb{Q})$.

Recall from Section 8.3 the map $\theta^{\mathrm{H}}\colon \mathrm{H}(S, \mathbb{Q}) \to \mathrm{H}(X, \mathbb{Q})$.

Lemma 9.5 *The map* $\theta^{\mathrm{H}}\colon \mathrm{H}(S,\mathbb{Q}) \to \mathrm{H}(X,\mathbb{Q})$ *is equivariant with respect to*

$$\theta^{\mathfrak{g}}\colon \mathfrak{g}(S) \to \mathfrak{g}(X), \quad x \mapsto B_{-\delta/2} \circ \iota(x) \circ B_{\delta/2}.$$

Proof We have $\theta^{\mathrm{H}} = e^{-\delta/2} \cdot \theta_0^{\mathrm{H}}$, with

$$\theta_0^{\mathrm{H}}(s + \lambda + t\,\mathrm{pt}_S) = s\delta + \lambda\delta + tq_X\delta.$$

The map θ_0^{H} respects the grading, and we claim that for every $\mu \in \mathrm{H}^2(S,\mathbb{Q})$ the diagram

$$\begin{tikzcd} \mathrm{H}(S,\mathbb{Q}) \arrow[r, "\theta_0^{\mathrm{H}}"] \arrow[d, "e_\mu"] & \mathrm{H}(X,\mathbb{Q}) \arrow[r, "e^{-\delta/2}"] \arrow[d, "e_\mu"] & \mathrm{H}(X,\mathbb{Q}) \arrow[d, "e^{-\delta/2}e_\mu e^{\delta/2}"] \\ \mathrm{H}(S,\mathbb{Q}) \arrow[r, "\theta_0^{\mathrm{H}}"] & \mathrm{H}(X,\mathbb{Q}) \arrow[r, "e^{-\delta/2}"] & \mathrm{H}(X,\mathbb{Q}) \end{tikzcd}$$

commutes. Indeed, we have

$$\begin{aligned} e_\mu(\theta_0^{\mathrm{H}}(s + \lambda + t\,\mathrm{pt}_S)) &= s\delta\mu + \lambda\delta\mu + tq_X\delta\mu, \\ \theta_0^{\mathrm{H}}(e_\mu(s + \lambda + t\,\mathrm{pt}_S)) &= s\delta\mu + b(\lambda,\mu)q_X\delta. \end{aligned}$$

One verifies easily that these agree, using the identities (10) and (9) from Section 7.2 and the fact that $b(\lambda,\delta) = b(\mu,\delta) = 0$. This shows that the left-hand square commutes. The right-hand square commutes trivially, so the outer rectangle commutes, which shows that $\theta^{\mathrm{H}} = e^{-\delta/2} \cdot \theta_0^{\mathrm{H}}$ is indeed equivariant with respect to $\theta^{\mathfrak{g}}$. □

Lemma 9.6 *There is an isomorphism*

$$\det(\widetilde{\mathrm{H}}(X,\mathbb{Q})) \otimes \mathrm{Sym}^2(\widetilde{\mathrm{H}}(X,\mathbb{Q})) \cong \mathrm{H}(X,\mathbb{Q}) \oplus \det(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$$

of representations of $G = \mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$.

Proof This follows from Lemma 3.7, Theorem 4.13 and Remark 4.14. □

We are now ready to prove the main result of this subsection.

Proof of Theorem 9.4 By Proposition 8.5, the map θ^{H} is equivariant for the action of $\mathrm{DMon}(S)$. Lemma 9.5 then implies that

$$h(g) = B_{-\delta/2} \circ \iota(g) \circ B_{\delta/2}$$

for all $g \in \mathrm{SO}(\widetilde{\mathrm{H}}(S,\mathbb{Q}))$. We have an orthogonal decomposition

$$\widetilde{\mathrm{H}}(X,\mathbb{Q}) = B_{-\delta/2}(\widetilde{\mathrm{H}}(S,\mathbb{Q})) \oplus C$$

with C of rank 1. Since $\mathrm{SO}(\widetilde{\mathrm{H}}(S,\mathbb{Q}))$ is normal in $\mathrm{O}(\widetilde{\mathrm{H}}(S,\mathbb{Q}))$, the action of $\mathrm{O}(\widetilde{\mathrm{H}}(S,\mathbb{Q}))$ (via h) must preserve this decomposition. With respect to this decomposition h must then be given by

$$h(g) = (B_{-\delta/2} \circ g\epsilon_1(g) \circ B_{\delta/2}) \oplus \epsilon_2(g),$$

where the $\epsilon_i(g)\colon \mathrm{O}(\widetilde{\mathrm{H}}(S,\mathbb{Q})) \to \{\pm 1\}$ are quadratic characters. This leaves four possibilities for h. One verifies that $\epsilon_1 = \epsilon_2 = \det g$ is the only possibility compatible with Proposition 9.3 and Lemma 9.6, and the theorem follows. □

9.3 A transitivity lemma

In this section we prove a lattice-theoretical lemma that will play an important role in the proofs of Theorems E and F.

Let $b\colon L \times L \to \mathbb{Z}$ be an even nondegenerate lattice. Let U be a hyperbolic plane with basis consisting of isotropic vectors α and β satisfying $b(\alpha,\beta) = -1$.

As before, to a $\lambda \in L$ we associate the isometry $B_\lambda \in \mathrm{O}(U \oplus L)$ defined as

$$B_\lambda(r\alpha + \mu + s\beta) = r\alpha + (\mu + r\lambda) + \big(s + b(\mu,\lambda) + r\tfrac{1}{2}b(\lambda,\lambda)\big)\beta$$

for all $r, s \in \mathbb{Z}$ and $\mu \in L$. Let γ be the isometry of $U \oplus L$ given by $\gamma(\alpha) = \beta$, $\gamma(\beta) = \alpha$, and $\gamma(\lambda) = -\lambda$ for all $\lambda \in L$.

Lemma 9.7 *Let L be an even lattice containing a hyperbolic plane. Let $G \subset \mathrm{O}(U \oplus L)$ be the subgroup generated by γ and by B_λ for all $\lambda \in L$. Then, for all $\delta \in U \oplus L$ with $\delta^2 = -2$ and for all $g \in \mathrm{O}(U \oplus L)$, there exists a $g' \in G$ such that $g'g$ fixes δ.*

Proof This follows from classical results of Eichler. A convenient modern source is [20, Section 3], whose notation we adopt. The isometry B_λ coincides with the Eichler transvection $t(\beta,-\lambda)$. The conjugate $\gamma B_\lambda \gamma^{-1}$ is the Eicher transvection $t(\alpha,\lambda)$. Hence G contains the subgroup $E_U(L) \subset \mathrm{O}(U \oplus L)$ of unimodular transvections with respect to U. By [20, Proposition 3.3], there exists a $g' \in E_U(L)$ mapping $g\delta$ to δ. □

9.4 Proof of Theorem E

Let X be a hyperkähler variety of type $\mathrm{K3}^{[2]}$. Let $\delta \in \mathrm{H}^2(X,\mathbb{Z})$ be any class satisfying $\delta^2 = -2$ and $b(\delta,\lambda) \in 2\mathbb{Z}$ for all $\lambda \in \mathrm{H}^2(X,\mathbb{Z})$. For example, if $X = S^{[2]}$, we may take $\delta = c_1(\mathcal{E})$ as in Section 7.2. Consider the integral lattice

$$\Lambda := B_{\delta/2}(\mathbb{Z}\alpha \oplus \mathrm{H}^2(X,\mathbb{Z}) \oplus \mathbb{Z}\beta) \subset \widetilde{\mathrm{H}}(X,\mathbb{Q}).$$

The subgroup $\Lambda \subset \widetilde{\mathrm{H}}(X,\mathbb{Q})$ does not depend on the choice of δ. In this section, we will prove Theorem E. More precisely, we will show:

Theorem 9.8 $$\mathrm{O}^+(\Lambda) \subset \mathrm{DMon}(X) \subset \mathrm{O}(\Lambda).$$

We start with the lower bound.

Proposition 9.9 $\mathrm{O}^+(\Lambda) \subset \mathrm{DMon}(X)$ *as subgroups of* $\mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$.

Proof Since the derived monodromy group is invariant under deformation, we may assume without loss of generality that $X = S^{[2]}$ for a K3 surface S and $\delta = c_1(\mathcal{E})$ as in Section 7.2.

The shift functor [1] on $\mathcal{D}X$ acts as -1 on $\mathrm{H}(X,\mathbb{Q})$, which coincides with the action of $-1 \in \mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$. In particular, $-1 \in \mathrm{O}^+(\Lambda)$ lies in $\mathrm{DMon}(X)$, so it suffices to show that $\mathrm{SO}^+(\Lambda)$ is contained in $\mathrm{DMon}(X)$.

Consider the isometry $\gamma \in \mathrm{O}^+(\widetilde{\mathrm{H}}(S,\mathbb{Q}))$ given by $\gamma(\alpha) = -\beta$, $\gamma(\beta) = -\alpha$, and $\gamma(\lambda) = \lambda$ for all $\lambda \in \mathrm{H}^2(S,\mathbb{Q})$. Then $\det(\gamma) = -1$ and by Theorem 9.4 its image $h(\gamma)$ interchanges $B_{\delta/2}\alpha$ and $B_{\delta/2}\beta$ and acts by -1 on $B_{\delta/2}\mathrm{H}^2(X,\mathbb{Z})$. Since γ lies in $\mathrm{DMon}(S) \subset \mathrm{O}(\widetilde{\mathrm{H}}(S,\mathbb{Q}))$, we have that $h(\gamma)$ lies in $\mathrm{DMon}(X) \subset \mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$.

Let $G \subset \mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ be the subgroup generated by $h(\gamma)$ and the isometries B_λ for $\lambda \in \mathrm{H}^2(X,\mathbb{Z})$. Clearly G is contained in $\mathrm{DMon}(X)$.

Let g be an element of $\mathrm{SO}^+(\Lambda)$, and consider the image $gB_{\delta/2}\delta$ of $B_{\delta/2}\delta$. By Lemma 9.7 there exists a $g' \in G \subset \mathrm{DMon}(X)$ such that $g'g$ fixes $B_{\delta/2}\delta$. But then $g'g$ acts on

$$(B_{\delta/2}\delta)^\perp = B_{\delta/2}(\mathbb{Z}\alpha \oplus \mathrm{H}^2(S,\mathbb{Z}) \oplus \mathbb{Z}\beta)$$

with determinant 1 and preserving the orientation of a maximal positive subspace. In particular, $g'g$ lies in the image of $\mathrm{DMon}(S) \to \mathrm{DMon}(X)$, and we conclude that g lies in $\mathrm{DMon}(X)$. □

The proof of the upper bound is now almost purely group-theoretical. Denote by $\mathrm{SO}^+(\Lambda)$ the intersection $\mathrm{O}^+(\Lambda) \cap \mathrm{SO}(\Lambda)$. This group coincides with the kernel of the spinor norm on $\mathrm{SO}(\Lambda)$.

Proposition 9.10 $\mathrm{SO}(\Lambda)$ *is the unique maximal arithmetic subgroup of* $\mathrm{SO}(\Lambda \otimes_{\mathbb{Z}} \mathbb{Q})$ *containing* $\mathrm{SO}^+(\Lambda)$.

Proof More generally, this holds for any even lattice Λ with the property that the quadratic form $q(x) = b(x,x)/2$ on the $\mathbb{Z}$–module Λ is semiregular [29, Section IV.3].

For such Λ, the group schemes $\mathbf{Spin}(\Lambda)$ and $\mathbf{SO}(\Lambda)$ are smooth over $\operatorname{Spec}\mathbb{Z}$; see eg [27]. In particular, for every prime p the subgroups $\mathrm{Spin}(\Lambda\otimes\mathbb{Z}_p)$ and $\mathrm{SO}(\Lambda\otimes\mathbb{Z}_p)$ of $\mathrm{Spin}(\Lambda\otimes\mathbb{Q}_p)$ and $\mathrm{SO}(\Lambda\otimes\mathbb{Q}_p)$, respectively, are maximal compact subgroups. It follows that the groups

$$\mathrm{Spin}(\Lambda) = \mathrm{Spin}(\Lambda\otimes\mathbb{Q}) \cap \prod_p \mathrm{Spin}(\Lambda\otimes\mathbb{Z}_p)$$

and

$$\mathrm{SO}(\Lambda) = \mathrm{SO}(\Lambda\otimes\mathbb{Q}) \cap \prod_p \mathrm{SO}(\Lambda\otimes\mathbb{Z}_p)$$

are maximal arithmetic subgroups of $\mathrm{Spin}(\Lambda\otimes\mathbb{Q})$ and $\mathrm{SO}(\Lambda\otimes\mathbb{Q})$, respectively.

The subgroup $\mathrm{SO}^+(\Lambda)\subset\mathrm{SO}(\Lambda)$ is the kernel of the spinor norm, and the short exact sequence $1\to\mu_2\to\mathbf{Spin}\to\mathbf{SO}\to 1$ of fppf sheaves on $\operatorname{Spec}\mathbb{Z}$ induces an exact sequence of groups

$$1\to\{\pm1\}\to\mathrm{Spin}(\Lambda)\to\mathrm{SO}^+(\Lambda)\to 1.$$

Let $\Gamma\subset\mathrm{SO}(\Lambda\otimes\mathbb{Q})$ be a maximal arithmetic subgroup containing $\mathrm{SO}^+(\Lambda)$. Let $\widetilde{\Gamma}$ be its inverse image in $\mathrm{Spin}(\Lambda\otimes\mathbb{Q})$, so that we have an exact sequence

$$1\to\{\pm1\}\to\widetilde{\Gamma}\to\Gamma\to\mathbb{Q}^\times/2.$$

Since the group $\widetilde{\Gamma}$ is arithmetic and contains $\mathrm{Spin}(\Lambda)$, we have $\widetilde{\Gamma}=\mathrm{Spin}(\Lambda)$. Moreover, Γ normalizes $\mathrm{SO}^+(\Lambda)=\ker(\Gamma\to\mathbb{Q}^\times/2)$, and, as the normalizer of an arithmetic subgroup of $\mathrm{SO}(\Lambda\otimes\mathbb{Q})$ is again arithmetic, Γ must equal the normalizer of $\mathrm{SO}^+(\Lambda)$. But then Γ contains $\mathrm{SO}(\Lambda)$, and we conclude $\Gamma=\mathrm{SO}(\Lambda)$. ◻

Corollary 9.11 $\mathrm{DMon}(X)\subset\mathrm{O}(\Lambda)$.

Proof $\mathrm{DMon}(X)$ preserves the integral lattice $\mathrm{K}_{\mathrm{top}}(X)$ in the representation $\mathrm{H}(X,\mathbb{Q})$ of $\mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$, and hence is contained in an arithmetic subgroup of

$$\mathrm{O}(\widetilde{\mathrm{H}}(X,\mathbb{Q})) = \mathrm{SO}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))\times\{\pm1\}.$$

By Proposition 9.9 it contains $\mathrm{SO}^+(\Lambda)\times\{\pm1\}$, so we conclude from the preceding proposition that $\mathrm{DMon}(X)$ must be contained in $\mathrm{O}(\Lambda)$. ◻

Together with Proposition 9.9 this proves Theorem 9.8.

10 The image of $\mathrm{Aut}(\mathcal{D}X)$ on $\mathrm{H}(X,\mathbb{Q})$

10.1 Upper bound for the image of ρ_X

We continue with the notation of the previous section. In particular, we denote by X a hyperkähler variety of type $\mathrm{K3}^{[2]}$, and by $\Lambda \subset \widetilde{\mathrm{H}}(X,\mathbb{Q})$ the lattice defined in Section 9.4. We equip $\widetilde{\mathrm{H}}(X,\mathbb{Q})$ with the weight 0 Hodge structure

$$\widetilde{\mathrm{H}}(X,\mathbb{Q}) = \mathbb{Q}\alpha \oplus \mathrm{H}^2(X,\mathbb{Q}(1)) \oplus \mathbb{Q}\beta.$$

We denote by $\mathrm{Aut}(\Lambda) \subset \mathrm{O}(\Lambda)$ the group of isometries of Λ that preserve this Hodge structure.

Proposition 10.1 $\mathrm{im}(\rho_X) \subset \mathrm{Aut}(\Lambda)$.

Proof By Theorem 9.8 we have $\mathrm{im}(\rho_X) \subset \mathrm{O}(\Lambda)$. The Hodge structure on

$$\mathrm{H}(X,\mathbb{Q}) = \bigoplus_{n=0}^{4} \mathrm{H}^{2n}(X,\mathbb{Q}(n))$$

induces a Hodge structure on $\mathfrak{g}(X) \subset \mathrm{End}(\mathrm{H}(X,\mathbb{Q}))$, which agrees with the Hodge structure on $\mathfrak{so}(\widetilde{\mathrm{H}}(X,\mathbb{Q}))$ induced by the Hodge structure on $\widetilde{\mathrm{H}}(X,\mathbb{Q})$. If

$$\Phi\colon \mathcal{D}X \xrightarrow{\sim} \mathcal{D}X$$

is an equivalence, then $\Phi^{\mathrm{H}}\colon \mathrm{H}(X,\mathbb{Q}) \xrightarrow{\sim} \mathrm{H}(X,\mathbb{Q})$ and $\Phi^{\mathfrak{g}}\colon \mathfrak{g}(X) \xrightarrow{\sim} \mathfrak{g}(X)$ are isomorphisms of $\mathbb{Q}$–Hodge structures, from which it follows that Φ^{H} must land in $\mathrm{Aut}(\Lambda) \subset \mathrm{O}(\Lambda)$. □

10.2 Lower bound for the image of ρ_X

We write $\mathrm{Aut}^+(\Lambda)$ for the index 2 subgroup $\mathrm{Aut}(\Lambda) \cap \mathrm{O}^+(\Lambda)$ of $\mathrm{Aut}(\Lambda)$.

Theorem 10.2 *Let S be a* K3 *surface and let X be the Hilbert square of S. Assume that* $\mathrm{NS}(X)$ *contains a hyperbolic plane. Then* $\mathrm{Aut}^+(\Lambda) \subset \mathrm{im}\,\rho_X \subset \mathrm{Aut}(\Lambda)$.

Proof In view of Proposition 10.1 we only need to show the lower bound. The argument for this is entirely parallel to the proof of Proposition 9.9. Recall that

$$\Lambda = B_{\delta/2}(\mathbb{Z}\alpha \oplus \mathrm{H}^2(S,\mathbb{Z}(1)) \oplus \mathbb{Z}\delta \oplus \mathbb{Z}\beta).$$

The shift functor $[1] \in \mathrm{Aut}(\mathcal{D}X)$ maps to $-1 \in \mathrm{Aut}^+(\Lambda)$, so it suffices to show that $\mathrm{Aut}^+(\Lambda) \cap \mathrm{SO}(\Lambda)$ is contained in $\mathrm{im}\,\rho_X$.

Let $\gamma_S \in \mathrm{Aut}(\mathcal{D}S)$ be the composition of the spherical twist in $\mathcal{O}_S$ with the shift [1]. On the Mukai lattice $\widetilde{\mathrm{H}}(S,\mathbb{Z}) = \mathbb{Z}\alpha \oplus \mathrm{H}^2(X,\mathbb{Z}(1)) \oplus \mathbb{Z}\beta$ this equivalence maps α to $-\beta$ and β to $-\alpha$ and is the identity on $\mathrm{H}^2(S,\mathbb{Z})$. Under the derived McKay correspondence this induces an autoequivalence $\gamma_X \in \mathrm{Aut}\,\mathcal{D}X$. By Theorem 9.4, the automorphism $\rho_X(\gamma_X) \in \mathrm{Aut}(\Lambda)$ interchanges $B_{\delta/2}\alpha$ and $B_{\delta/2}\beta$ and acts by -1 on $B_{\delta/2}\mathrm{H}^2(X,\mathbb{Z})$.

Denote by $G \subset \mathrm{Aut}(\Lambda)$ the subgroup generated by $\rho_X(\gamma_X)$ and the isometries $B_\lambda = \rho_X(- \otimes \mathcal{L})$ with $\mathcal{L}$ a line bundle of class $\lambda \in \mathrm{NS}(X)$. Clearly G is contained in the image of ρ_X. Note that G acts on the lattice

$$\Lambda_{\mathrm{alg}} := B_{\delta/2}(\mathbb{Z}\alpha \oplus \mathrm{NS}(X) \oplus \mathbb{Z}\beta)$$

and that by our assumption $\mathrm{NS}(X)$ contains a hyperbolic plane.

Let $g \in \mathrm{Aut}^+(\Lambda)$. By Lemma 9.7 applied to $L = \mathrm{NS}(X)$, there exists a $g' \in G$ such that $g'g$ fixes $B_{\delta/2}\delta$. But then $g'g$ acts on

$$(B_{\delta/2}\delta)^\perp = B_{\delta/2}(\mathbb{Z}\alpha \oplus \mathrm{H}^2(S,\mathbb{Z}) \oplus \mathbb{Z}\beta)$$

with determinant 1 and preserving the Hodge structure and the orientation of a maximal positive subspace. In particular, $g'g$ lies in the image of $\mathrm{Aut}(\mathcal{D}S)$, and we conclude that g lies in $\mathrm{im}\,\rho_X$. □

References

[1] **N Addington**, **R Thomas**, *Hodge theory and derived categories of cubic fourfolds*, Duke Math. J. 163 (2014) 1885–1927 MR Zbl

[2] **M F Atiyah**, *Vector bundles and the Künneth formula*, Topology 1 (1962) 245–248 MR Zbl

[3] **M F Atiyah**, **F Hirzebruch**, *Vector bundles and homogeneous spaces*, from "Differential geometry" (C B Allendoerfer, editor), Proc. Sympos. Pure Math. 3, Amer. Math. Soc., Providence, RI (1961) 7–38 MR Zbl

[4] **M F Atiyah**, **F Hirzebruch**, *The Riemann–Roch theorem for analytic embeddings*, Topology 1 (1962) 151–166 MR Zbl

[5] **M F Atiyah**, **G B Segal**, *The index of elliptic operators, II*, Ann. of Math. 87 (1968) 531–545 MR Zbl

[6] **P Baum**, **W Fulton**, **R MacPherson**, *Riemann–Roch and topological K–theory for singular varieties*, Acta Math. 143 (1979) 155–192 MR Zbl

[7] **A Beauville**, *Variétés Kähleriennes dont la première classe de Chern est nulle*, J. Differential Geom. 18 (1983) 755–782 MR Zbl

[8] **F A Bogomolov**, *On the cohomology ring of a simple hyperkähler manifold (on the results of Verbitsky)*, Geom. Funct. Anal. 6 (1996) 612–618 MR Zbl

[9] **A Bondal**, **D Orlov**, *Reconstruction of a variety from the derived category and groups of autoequivalences*, Compositio Math. 125 (2001) 327–344 MR Zbl

[10] **T Bridgeland**, *Stability conditions on $K3$ surfaces*, Duke Math. J. 141 (2008) 241–291 MR Zbl

[11] **T Bridgeland**, **A King**, **M Reid**, *The McKay correspondence as an equivalence of derived categories*, J. Amer. Math. Soc. 14 (2001) 535–554 MR Zbl

[12] **D Calaque**, **M Van den Bergh**, *Hochschild cohomology and Atiyah classes*, Adv. Math. 224 (2010) 1839–1889 MR Zbl

[13] **D Calaque**, **C A Rossi**, **M Van den Bergh**, *Căldăraru's conjecture and Tsygan's formality*, Ann. of Math. 176 (2012) 865–923 MR Zbl

[14] **A Caldararu**, *The Mukai pairing*, *I*: *The Hochschild structure*, preprint (2003) arXiv math/0308079

[15] **A Căldăraru**, *The Mukai pairing, II: The Hochschild–Kostant–Rosenberg isomorphism*, Adv. Math. 194 (2005) 34–66 MR Zbl

[16] **R Dijkgraaf**, *Instanton strings and hyper-Kähler geometry*, Nuclear Phys. B 543 (1999) 545–571 MR Zbl

[17] **G Ellingsrud**, **L Göttsche**, **M Lehn**, *On the cobordism class of the Hilbert scheme of a surface*, J. Algebraic Geom. 10 (2001) 81–100 MR Zbl

[18] **V Golyshev**, **V Lunts**, **D Orlov**, *Mirror symmetry for abelian varieties*, J. Algebraic Geom. 10 (2001) 433–496 MR Zbl

[19] **M Green**, **Y-J Kim**, **R Laza**, **C Robles**, *The LLV decomposition of hyper-Kähler cohomology*, preprint (2019) arXiv 1906.03432

[20] **V Gritsenko**, **K Hulek**, **G K Sankaran**, *Abelianisation of orthogonal groups and the fundamental group of modular varieties*, J. Algebra 322 (2009) 463–478 MR Zbl

[21] **M Haiman**, *Hilbert schemes, polygraphs and the Macdonald positivity conjecture*, J. Amer. Math. Soc. 14 (2001) 941–1006 MR Zbl

[22] **S Hosono**, **B H Lian**, **K Oguiso**, **S-T Yau**, *Autoequivalences of derived category of a $K3$ surface and monodromy transformations*, J. Algebraic Geom. 13 (2004) 513–545 MR Zbl

[23] **D Huybrechts**, *Compact hyperkähler manifolds*, from "Calabi–Yau manifolds and related geometries" (M Gross, D Huybrechts, D Joyce, editors), Springer (2003) 161–225 MR Zbl

[24] **D Huybrechts**, *Fourier–Mukai transforms in algebraic geometry*, Oxford Univ. Press (2006) MR Zbl

[25] **D Huybrechts**, **E Macrì**, **P Stellari**, *Stability conditions for generic $K3$ categories*, Compos. Math. 144 (2008) 134–162 MR Zbl

[26] **D Huybrechts**, **E Macrì**, **P Stellari**, *Derived equivalences of $K3$ surfaces and orientation*, Duke Math. J. 149 (2009) 461–507 MR Zbl

[27] **H Ikai**, *Spin groups over a commutative ring and the associated root data (odd rank case)*, Beiträge Algebra Geom. 46 (2005) 377–395 MR Zbl

[28] **R Joshua**, *Equivariant Riemann–Roch for G–quasi-projective varieties, I*, K–Theory 17 (1999) 1–35 MR Zbl

[29] **M-A Knus**, *Quadratic and Hermitian forms over rings*, Grundl. Math. Wissen. 294, Springer (1991) MR Zbl

[30] **M Kontsevich**, *Deformation quantization of Poisson manifolds*, Lett. Math. Phys. 66 (2003) 157–216 MR Zbl

[31] **A Krug**, *On derived autoequivalences of Hilbert schemes and generalized Kummer varieties*, Int. Math. Res. Not. 2015 (2015) 10680–10701 MR Zbl

[32] **S A Kuleshov**, *A theorem on the existence of exceptional bundles on surfaces of type $K3$*, Izv. Akad. Nauk SSSR Ser. Mat. 53 (1989) 363–378 MR Zbl In Russian; translated in Math. USSR-Izv. 34 (1990) 373–388

[33] **E Looijenga**, **V A Lunts**, *A Lie algebra attached to a projective variety*, Invent. Math. 129 (1997) 361–412 MR Zbl

[34] **E Macrì**, **P Stellari**, *Infinitesimal derived Torelli theorem for $K3$ surfaces*, Int. Math. Res. Not. 2009 (2009) 3190–3220 MR Zbl

[35] **S Mukai**, *On the moduli space of bundles on $K3$ surfaces, I*, from "Vector bundles on algebraic varieties", Tata Inst. Fund. Res. Stud. Math. 11, Tata Inst. Fund. Res., Bombay (1987) 341–413 MR Zbl

[36] **D O Orlov**, *Equivalences of derived categories and $K3$ surfaces*, J. Math. Sci. 84 (1997) 1361–1381 MR Zbl

[37] **D O Orlov**, *Derived categories of coherent sheaves and equivalences between them*, Uspekhi Mat. Nauk 58 (2003) 89–172 MR Zbl In Russian; translated in Russian Math. Surveys 58 (2003) 511–591

[38] **D O Orlov**, *Derived categories of coherent sheaves, and motives*, Uspekhi Mat. Nauk 60 (2005) 231–232 MR Zbl In Russian; translated in Russian Math. Surveys 60 (2005) 1242–1244

[39] **D Ploog**, *Groups of autoequivalences of derived categories of smooth projective varieties*, PhD thesis, Freie Universität Berlin (2005) Available at `https://www.mathematik.hu-berlin.de/~ploog/PAPERS/Ploog_phd.pdf`

[40] **D Ploog**, *Equivariant autoequivalences for finite group actions*, Adv. Math. 216 (2007) 62–74 MR Zbl

[41] **A Rapagnetta**, *On the Beauville form of the known irreducible symplectic varieties*, Math. Ann. 340 (2008) 77–95 MR Zbl

[42] **G Segal**, *Equivariant K–theory*, Inst. Hautes Études Sci. Publ. Math. 34 (1968) 129–151 MR Zbl

[43] **A Soldatenkov**, *On the Hodge structures of compact hyperkähler manifolds*, Math. Res. Lett. 28 (2021) 623–635 MR Zbl

[44] **B Szendrői**, *Diffeomorphisms and families of Fourier–Mukai transforms in mirror symmetry*, from "Applications of algebraic geometry to coding theory, physics and computation" (C Ciliberto, F Hirzebruch, R Miranda, M Teicher, editors), NATO Sci. Ser. II Math. Phys. Chem. 36, Kluwer, Dordrecht (2001) 317–337 MR Zbl

[45] **M S Verbitsky**, *Cohomology of compact hyperkaehler manifolds*, PhD thesis, Harvard University (1995) MR Available at `https://www.proquest.com/docview/304181888`

[46] **M Verbitsky**, *Cohomology of compact hyperkähler manifolds and its applications*, Geom. Funct. Anal. 6 (1996) 601–611 MR Zbl

[47] **M Verbitsky**, *Mirror symmetry for hyper-Kähler manifolds*, from "Mirror symmetry, III" (D H Phong, L Vinet, S-T Yau, editors), AMS/IP Stud. Adv. Math. 10, Amer. Math. Soc., Providence, RI (1999) 115–156 MR Zbl

Korteweg-de Vries Institute for Mathematics, University of Amsterdam
Amsterdam, Netherlands

`l.d.j.taelman@uva.nl`

Proposed: Richard P Thomas — Received: 17 March 2020
Seconded: Jim Bryan, Dan Abramovich — Revised: 2 November 2021

Geometry & Topology 27:7 (2023) 2695–2761
DOI: 10.2140/gt.2023.27.2695
Published: 19 September 2023

A new cohomology class on the moduli space of curves

PAUL NORBURY

We define a collection $\Theta_{g,n} \in H^{4g-4+2n}(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})$ for $2g-2+n>0$ of cohomology classes that restrict naturally to boundary divisors. We prove that the intersection numbers $\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \prod_{i=1}^n \psi_i^{m_i}$ can be recursively calculated. We conjecture that a generating function for these intersection numbers is a tau function of the KdV hierarchy. This is analogous to the conjecture of Witten proven by Kontsevich that a generating function for the intersection numbers $\int_{\overline{\mathcal{M}}_{g,n}} \prod_{i=1}^n \psi_i^{m_i}$ is a tau function of the KdV hierarchy.

14D23, 32G15, 53D45

1 Introduction

Let $\overline{\mathcal{M}}_{g,n}$ be the moduli space of genus g stable curves — curves with only nodal singularities and finite automorphism group — with n labelled points disjoint from nodes. Define $\psi_i = c_1(L_i) \in H^2(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})$ to be the first Chern class of the line bundle $L_i \to \overline{\mathcal{M}}_{g,n}$ with fibre above $[(C, p_1, \dots, p_n)]$ given by $T^*_{p_i}C$. Consider the natural maps given by the forgetful map which forgets the last point,

$$\overline{\mathcal{M}}_{g,n+1} \xrightarrow{\pi} \overline{\mathcal{M}}_{g,n}, \tag{1}$$

and the gluing maps which glue the last two points,

$$\begin{aligned}\overline{\mathcal{M}}_{g-1,n+2} &\xrightarrow{\phi_{\mathrm{irr}}} \overline{\mathcal{M}}_{g,n},\\ \overline{\mathcal{M}}_{h,|I|+1}\times\overline{\mathcal{M}}_{g-h,|J|+1} &\xrightarrow{\phi_{h,I}} \overline{\mathcal{M}}_{g,n}, \quad I\sqcup J=\{1,\dots,n\}.\end{aligned} \tag{2}$$

In this paper we construct cohomology classes $\Theta_{g,n}\in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$ for $g\geq 0$, $n\geq 0$ and $2g-2+n>0$ such that

(i) $\Theta_{g,n}\in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$ is of pure degree,

(ii) $\phi_{\mathrm{irr}}^*\Theta_{g,n}=\Theta_{g-1,n+2}$ and $\phi_{h,I}^*\Theta_{g,n}=\pi_1^*\Theta_{h,|I|+1}\cdot\pi_2^*\Theta_{g-h,|J|+1}$,

(iii) $\Theta_{g,n+1}=\psi_{n+1}\cdot\pi^*\Theta_{g,n}$,

(iv) $\Theta_{1,1}\neq 0$,

where π_i is projection onto the i^{th} factor of $\overline{\mathcal{M}}_{h,|I|+1}\times\overline{\mathcal{M}}_{g-h,|J|+1}$. We prove below that properties (i)–(iv) uniquely define intersection numbers of the classes $\Theta_{g,n}$ with the classes ψ_i and more generally with classes in the tautological ring $RH^*(\overline{\mathcal{M}}_{g,n})\subset H^{2*}(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$.

Remark 1.1 One can replace (ii) by the equivalent property

$$\phi_\Gamma^*\Theta_{g,n}=\Theta_\Gamma$$

for any stable graph Γ, defined in Section 3, of genus g with n external edges. Here

$$\phi_\Gamma:\overline{\mathcal{M}}_\Gamma=\prod_{v\in V(\Gamma)}\overline{\mathcal{M}}_{g(v),n(v)}\to\overline{\mathcal{M}}_{g,n},\quad \Theta_\Gamma=\prod_{v\in V(\Gamma)}\pi_v^*\Theta_{g(v),n(v)}\in H^*(\overline{\mathcal{M}}_\Gamma,\mathbb{Q}),$$

where π_v is projection onto the factor $\overline{\mathcal{M}}_{g(v),n(v)}$. This generalises (ii) from 1–edge stable graphs given by $\phi_{\Gamma_{\mathrm{irr}}}=\phi_{\mathrm{irr}}$ and $\phi_{\Gamma_{h,I}}=\phi_{h,I}$.

Remark 1.2 The sequence of classes $\Theta_{g,n}$ satisfies many properties of a cohomological field theory (CohFT). It is essentially a 1–dimensional CohFT with vanishing genus zero classes, not to be confused with Hodge classes which are trivial in genus zero but do not vanish there. The trivial cohomology class $1\in H^0(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$, which is a trivial example of a CohFT known as a topological field theory, satisfies conditions (i)–(ii), while the forgetful map property (iii) is replaced by $\Theta_{g,n+1}=\pi^*\Theta_{g,n}$.

Theorem 1.3 *There exists a class $\Theta_{g,n}$ satisfying* (i)–(iv) *and, furthermore, any such class satisfies the following properties*:

(I) $\Theta_{g,n}\in H^{4g-4+2n}(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$.

(II) $\Theta_{0,n} = 0$ *for all* n *and* $\phi_\Gamma^*\Theta_{g,n} = 0$ *for any* Γ *with a genus* 0 *vertex.*

(III) $\Theta_{g,n} \in H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})^{S_n}$, *ie it is symmetric under the* S_n *action.*

(IV) $\Theta_{1,1} = 3\psi_1$.

(V) *For any* $\eta \in RH^{g-1}(\overline{\mathcal{M}}_{g,n})$, *the intersection number* $\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n}\eta \in \mathbb{Q}$ *is uniquely determined by* (i)–(iii) *and* (IV).

The main content of Theorem 1.3 is the existence of $\Theta_{g,n}$, the rigidity property (IV) and the uniqueness property (V). The existence of $\Theta_{g,n}$ is constructed via the pushforward of a class over the moduli space of spin curves in Section 2. The rigidity property (IV) is proven in Section 3 by starting with $\Theta_{1,1} = \lambda\psi_1$ and determining constraints on λ to arrive at $\lambda = 3$, which does occur due to the construction of $\Theta_{g,n}$. The uniqueness result (V) involving classes in the tautological ring $RH^*(\overline{\mathcal{M}}_{g,n})$ is nonconstructive since it relies on the existence of nonexplicit tautological relations. The proofs of properties (I)–(III) are straightforward and presented in Section 3. Section 4 describes how the classes $\Theta_{g,n}$ naturally combine with any cohomological field theory.

Remark 1.4 Properties (i)–(iv) uniquely define the classes $\Theta_{g,n}$ for $g \le 4$ and all n, but it is not known if they uniquely define the classes $\Theta_{g,n}$ in general. Uniqueness would follow from injectivity of the pullback map to the boundary

$$RH^{2g-2}(\overline{\mathcal{M}}_g) \to RH^{2g-2}(\partial\overline{\mathcal{M}}_g),$$

which holds for $g = 2$, 3 and 4. It would show that $\Theta_g \in RH^{2g-2}(\overline{\mathcal{M}}_g)$ is uniquely determined from its restriction, and consequently $\Theta_{g,n}$ would coincide with the classes constructed in Section 2 for all $n \ge 0$.

The following conjecture allows one to recursively calculate all intersection numbers $\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \prod_{i=1}^n \psi_i^{m_i}$ via relations coming out of the KdV hierarchy. Such a recursive calculation would strengthen property (V) since intersections of $\Theta_{g,n}$ with ψ classes determine all tautological intersections with $\Theta_{g,n}$ algorithmically.

Conjecture 1.5 *The function*

$$Z^\Theta(\hbar, t_0, t_1, \dots) = \exp \sum_{g,n,\vec{k}} \frac{\hbar^{g-1}}{n!} \int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \cdot \prod_{j=1}^n \psi_j^{k_j} \prod t_{k_j}$$

is the Brézin–Gross–Witten tau function of the KdV hierarchy.

The Brézin–Gross–Witten KdV tau function Z^{BGW} was defined in [6; 30]. Conjecture 1.5 has been verified up to $g = 7$, ie the coefficients of the expansion of the logarithm of the Brézin–Gross–Witten tau function are given by intersection numbers of the classes $\Theta_{g,n}$ for $g \leq 7$ and all n. Progress towards Conjecture 1.5, including a purely combinatorial formulation that can be stated without reference to the moduli space of stable curves or the KdV hierarchy, is discussed in Section 6.

Acknowledgements I would like to thank Dimitri Zvonkine for his ongoing interest in this work, which benefited immensely from many conversations together. I would also like to thank Vincent Bouchard, Alessandro Chiodo, Alessandro Giacchetto, Oliver Leigh, Danilo Lewanksi, Rahul Pandharipande, Johannes Schmitt, Mehdi Tavakol, Ran Tessler, Ravi Vakil and Edward Witten for useful conversations, the referee for comments which improved the paper, and the Institut Henri Poincaré, where part of this work was carried out.

2 Existence

The existence of a cohomology class $\Theta_{g,n} \in H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})$ satisfying (i)–(iv) is proven here using the moduli space of stable twisted spin curves $\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$, which consists of pairs (Σ, θ) given by a twisted stable curve Σ equipped with an orbifold line bundle θ together with an isomorphism $\theta^{\otimes 2} \cong \omega_{\Sigma}^{\log}$. See precise definitions below. We first construct a cohomology class on $\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$ and then push it forward to a cohomology class on $\overline{\mathcal{M}}_{g,n}$.

A stable twisted curve, with group $\mathbb{Z}_2$, is a 1–dimensional orbifold, or stack, $\mathcal{C}$ such that generic points of $\mathcal{C}$ have trivial isotropy group and nontrivial orbifold points have isotropy group $\mathbb{Z}_2$. A stable twisted curve is equipped with a map which forgets the orbifold structure $\rho\colon \mathcal{C} \to C$, where C is a stable curve known as the coarse curve of $\mathcal{C}$. We say that $\mathcal{C}$ is smooth if its coarse curve C is smooth. Each nodal point of $\mathcal{C}$ (corresponding to a nodal point of C) has nontrivial isotropy group, the local picture at each node is $\{xy = 0\}/\mathbb{Z}_2$ with $\mathbb{Z}_2$ action given by $(-1) \cdot (x, y) = (-x, -y)$, and all other points of $\mathcal{C}$ with nontrivial isotropy group are labelled points of $\mathcal{C}$.

A line bundle L over $\mathcal{C}$ is a locally equivariant bundle over the local charts such that, at each nodal point, there is an equivariant isomorphism of fibres. Hence, each orbifold point p associates a representation of $\mathbb{Z}_2$ on $L|_p$ acting by multiplication by $\exp(2\pi i \lambda_p)$ for $\lambda_p = 0$ or $\frac{1}{2}$. One says L is *banded* at p by λ_p. The equivariant isomorphism at nodes guarantees that the representations agree on each local irreducible component at the node.

The canonical bundle $\omega_{\mathcal{C}}$ of $\mathcal{C}$ is generated by dz for any local coordinate z. At an orbifold point $x = z^2$, the canonical bundle $\omega_{\mathcal{C}}$ is generated by dz; hence, it is banded by $\frac{1}{2}$, ie $dz \mapsto -dz$ under $z \mapsto -z$. Over the coarse curve, ω_C is generated by $dx = 2z\,dz$. In other words, $\rho^*\omega_C \ncong \omega_{\mathcal{C}}$; however, $\omega_C \cong \rho_*\omega_{\mathcal{C}}$. Moreover, $\deg \omega_C = 2g-2$ and

$$\deg \omega_{\mathcal{C}} = 2g - 2 + \tfrac{1}{2}n.$$

For $\omega_{\mathcal{C}}^{\log} = \omega_{\mathcal{C}}(p_1, \dots, p_n)$, locally $dx/x = 2\,dz/z$, so $\rho^*\omega_C^{\log} \cong \omega_{\mathcal{C}}^{\log}$ and $\deg \omega_C^{\log} = 2g-2+n = \deg \omega_{\mathcal{C}}^{\log}$.

Following [1], define the moduli space of stable twisted spin curves by

$$\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}} = \{(\mathcal{C}, \theta, p_1, \dots, p_n, \phi) \mid \phi \colon \theta^2 \xrightarrow{\cong} \omega_{\mathcal{C}}^{\log}\}.$$

Here $\omega_{\mathcal{C}}^{\log}$ and θ are line bundles over the stable twisted curve $\mathcal{C}$ with labelled orbifold points p_j and $\deg\theta = g - 1 + \frac{1}{2}n$. The pair (θ, ϕ) is a *spin structure* on $\mathcal{C}$. The relation $\theta^2 \xrightarrow{\cong} \omega_{\mathcal{C}}^{\log}$ is possible because the representation associated to $\omega_{\mathcal{C}}^{\log}$ at p_i is trivial: $dz/z \to dz/z$, $z \mapsto -z$. The equivariant isomorphism of fibres over nodal points forces the balanced condition $\lambda_{p_+} = \lambda_{p_-}$ for $p_\pm$ corresponding to p on each irreducible component.

We can now define a vector bundle over $\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$ using the dual bundle $\theta^\vee$ on each stable twisted curve. Denote by $\mathcal{E}$ the universal spin structure on the universal stable twisted spin curve over $\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$. Given a map $S \to \overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$, $\mathcal{E}$ pulls back to θ, giving a family $(\mathcal{C}, \theta, p_1, \dots, p_n, \phi)$, where $\pi \colon \mathcal{C} \to S$ has stable twisted curve fibres, $p_i \colon S \to \mathcal{C}$ are sections with orbifold isotropy $\mathbb{Z}_2$, and $\phi \colon \theta^2 \xrightarrow{\cong} \omega_{\mathcal{C}/S}^{\log} = \omega_{\mathcal{C}/S}(p_1, \dots, p_n)$. Consider the pushforward sheaf $\pi_*\mathcal{E}^\vee$ over $\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$. We have

$$\deg \theta^\vee = 1 - g - \tfrac{1}{2}n < 0.$$

Furthermore, for any irreducible component $\mathcal{C}' \xrightarrow{i} \mathcal{C}$, the pole structure on sections of the log canonical bundle at nodes yields $i^*\omega_{\mathcal{C}/S}^{\log} = \omega_{\mathcal{C}'/S}^{\log}$. Hence, $\phi' \colon (\theta|_{\mathcal{C}'})^2 \xrightarrow{\cong} \omega_{\mathcal{C}'/S}^{\log}$, where $\phi' = i^* \circ \phi|_{\mathcal{C}'}$. Since the irreducible component $\mathcal{C}'$ is stable, its log canonical bundle has negative degree and

$$\deg \theta^\vee|_{\mathcal{C}'} < 0.$$

The negative degree of $\theta^\vee$ restricted to any irreducible component implies $R^0\pi_*\mathcal{E}^\vee = 0$ and the following definition makes sense:

Definition 2.1 Define a bundle $E_{g,n}=-R\pi_*\mathcal{E}^\vee$ over $\overline{\mathcal{M}}_{g,n}^{\text{spin}}$ with fibre $H^1(\mathcal{C},\theta^\vee)$.

Represent the band of θ at the labelled points by $\vec{\sigma}=(\sigma_1,\dots,\sigma_n)\in\{0,1\}^n$ so that, at each labelled point p_i, the representation of $\mathbb{Z}_2$ on $\theta|_{p_i}$ is given by multiplication by $\exp(2\pi i\lambda_{p_i})$ for $\lambda_{p_i}=\frac{1}{2}\sigma_i\in\{0,\frac{1}{2}\}$. The number of p_i with $\lambda_{p_i}=0$ is even due to evenness of the degree of the pushforward sheaf $|\theta|:=\rho_*\mathcal{O}_{\mathcal{C}}(\theta)$ on the coarse curve C [33]. In the smooth case, the boundary type of a spin structure is determined by an associated quadratic form, applied to each of the n boundary classes, which vanishes since it is a homological invariant, again implying that the number of p_i with $\lambda_{p_i}=0$ is even. The moduli space of stable twisted spin curves decomposes into components determined by the band $\vec{\sigma}$,

$$\overline{\mathcal{M}}_{g,n}^{\text{spin}}=\bigsqcup_{\sigma}\overline{\mathcal{M}}_{g,n,\vec{\sigma}}^{\text{spin}},$$

where $\overline{\mathcal{M}}_{g,n,\vec{\sigma}}^{\text{spin}}$ consists of those spin curves with θ banded by $\vec{\sigma}$, and the union is over the 2^{n-1} functions $\vec{\sigma}$ satisfying $|\vec{\sigma}|+n=\sum_{i=1}^n(\sigma_i+1)\in 2\mathbb{Z}$. Each component $\overline{\mathcal{M}}_{g,n,\vec{\sigma}}^{\text{spin}}$ is connected except when $|\vec{\sigma}|=n$, in which case there are two connected components determined by their Arf invariant, known as even and odd spin structures. This follows from the case of smooth spin curves proven in [42].

Restricted to $\overline{\mathcal{M}}_{g,n,\vec{\sigma}}^{\text{spin}}$, the bundle $E_{g,n}$ has rank

$$\operatorname{rank}E_{g,n}=2g-2+\tfrac{1}{2}(n+|\vec{\sigma}|) \tag{3}$$

by the following Riemann–Roch calculation. Orbifold Riemann–Roch takes into account the representation information

$$\begin{aligned}h^0(\mathcal{C},\theta^\vee)-h^1(\mathcal{C},\theta^\vee)&=1-g+\deg\theta^\vee-\sum_{i=1}^n\lambda_{p_i}=1-g+1-g-\tfrac{1}{2}n-\tfrac{1}{2}|\vec{\sigma}|\\&=2-2g-\tfrac{1}{2}(n+|\vec{\sigma}|).\end{aligned}$$

Alternatively, one can use the usual Riemann–Roch calculation on the pushforward of θ to the underlying coarse curve C as follows. The sheaf of local sections $\mathcal{O}_{\mathcal{C}}(L)$ of any line bundle L on $\mathcal{C}$ pushes forward to a sheaf $|L|:=\rho_*\mathcal{O}_{\mathcal{C}}(L)$ on C, which can be identified with the local sections of L invariant under the $\mathbb{Z}_2$ action. Away from nodal points, $|L|$ is locally free, and hence a line bundle. At nodal points, the pushforward $|L|$ is locally free when L is banded by the trivial representation, and $|L|$ is a torsion-free sheaf that is not locally free when L is banded by the nontrivial representation; see [25]. The pullback bundle is given by

$$\rho^*(|\theta^\vee|) = \theta^\vee \otimes \bigotimes_{i\in I} \mathcal{O}(-\sigma_i p_i)$$

since locally invariant sections must vanish when the representation is nontrivial. Hence, $\deg|\theta^\vee| = \deg\theta^\vee - \frac{1}{2}|\vec{\sigma}|$. Hence, Riemann–Roch on the coarse curve yields the same result as above: $h^0(C,|\theta^\vee|) - h^1(C,|\theta^\vee|) = 2-2g-\frac{1}{2}(n+|\vec{\sigma}|)$. It is proven in [25] that $H^i(\mathcal{C},\theta^\vee) = H^i(C,|\theta^\vee|)$, so the calculations agree.

We have $h^0(\mathcal{C},\theta^\vee)=0$ since $\deg\theta^\vee = 1-g-\frac{1}{2}n<0$, and the restriction of $\theta^\vee$ to any irreducible component C', say of type (g',n'), also has negative degree, $\deg\theta^\vee|_{C'} = 1-g'-\frac{1}{2}n'<0$. Hence, $h^1(\mathcal{C},\theta^\vee) = 2g-2+\frac{1}{2}(n+|\vec{\sigma}|)$. Thus, $H^1(\mathcal{C},\theta^\vee)$ gives fibres of a rank $2g-2+\frac{1}{2}(n+|\vec{\sigma}|)$ vector bundle.

The analogue of the boundary maps ϕ_{irr} and $\phi_{h,I}$ defined in (2) are multivalued maps defined as follows. Consider a node $p\in\mathcal{C}$ for $(\mathcal{C},\theta,p_1,\dots,p_n,\phi)\in\overline{\mathcal{M}}^{\mathrm{spin}}_{g,n}$. Denote the normalisation by $\nu\colon\widetilde{\mathcal{C}}\to\mathcal{C}$ with points $p_\pm\in\widetilde{\mathcal{C}}$ that map to the node, $p=\nu(p_\pm)$. When $\widetilde{\mathcal{C}}$ is not connected, the spin structure $\nu^*\theta$ decomposes into two spin structures θ_1 and θ_2. Any two spin structures θ_1 and θ_2 with bands at p_+ and p_- that agree can glue, but not uniquely, to give a spin structure on $\mathcal{C}$. This gives rise to a multivalued map, as described in [26, page 27], which uses the fibre product

$$\begin{tikzcd}
(\overline{\mathcal{M}}_{h,|I|+1}\times\overline{\mathcal{M}}_{g-h,|J|+1})\times_{\overline{\mathcal{M}}_{g,n}}\overline{\mathcal{M}}^{\mathrm{spin}}_{g,n} \arrow[r] \arrow[d] & \overline{\mathcal{M}}^{\mathrm{spin}}_{g,n} \arrow[d] \\
\overline{\mathcal{M}}_{h,|I|+1}\times\overline{\mathcal{M}}_{g-h,|J|+1} \arrow[r] & \overline{\mathcal{M}}_{g,n}
\end{tikzcd}$$

and is given by

$$\begin{tikzcd}
 & (\overline{\mathcal{M}}_{h,|I|+1}\times\overline{\mathcal{M}}_{g-h,|J|+1})\times_{\overline{\mathcal{M}}_{g,n}}\overline{\mathcal{M}}^{\mathrm{spin}}_{g,n} \arrow[dl, "\hat{\nu}"'] \arrow[dr, "\phi_{h,I}"] & \\
\overline{\mathcal{M}}^{\mathrm{spin}}_{h,|I|+1}\times\overline{\mathcal{M}}^{\mathrm{spin}}_{g-h,|J|+1} \arrow[rr, dashed] & & \overline{\mathcal{M}}^{\mathrm{spin}}_{g,n}
\end{tikzcd}$$

where $I\sqcup J=\{1,\dots,n\}$. The map $\hat{\nu}$ is given by the pullback of the spin structure obtained from $\overline{\mathcal{M}}^{\mathrm{spin}}_{g,n}$ to the normalisation defined by the points of $\overline{\mathcal{M}}_{h,|I|+1}$ and $\overline{\mathcal{M}}_{g-h,|J|+1}$. The broken arrow $\dashrightarrow$ represents the multiply defined map $\phi_{h,I}\circ\hat{\nu}^{-1}$. The multivalued map $\phi_{h,I}\circ\hat{\nu}^{-1}$ naturally restricts to components

$$\overline{\mathcal{M}}^{\mathrm{spin}}_{h,|I|+1,\sigma_1}\times\overline{\mathcal{M}}^{\mathrm{spin}}_{g-h,|J|+1,\sigma_2}\dashrightarrow\overline{\mathcal{M}}^{\mathrm{spin}}_{g,n,\vec{\sigma}},$$

where $\vec{\sigma}$ and I uniquely determine σ_1 and σ_2, since θ must be banded by $\lambda_p = 0$ at an even number of orbifold points, which uniquely determines the band $\lambda_{p_+} = \lambda_{p_-}$ at the separating node.

When $\widetilde{\mathcal{C}}$ is connected, a spin structure θ on $\mathcal{C}$ pulls back to a spin structure $\tilde{\theta} = \nu^*\theta$ on $\widetilde{\mathcal{C}}$. As above, any spin structure $\tilde{\theta}$ with bands at p_+ and p_- that agree glues nonuniquely, to give a spin structure on $\mathcal{C}$, and defines a multiply defined map which uses the fibre product

$$\begin{tikzcd} \overline{\mathcal{M}}_{g-1,n+2} \times_{\overline{\mathcal{M}}_{g,n}} \overline{\mathcal{M}}_{g,n}^{\text{spin}} \arrow[r] \arrow[d] & \overline{\mathcal{M}}_{g,n}^{\text{spin}} \arrow[d] \\ \overline{\mathcal{M}}_{g-1,n+2} \arrow[r] & \overline{\mathcal{M}}_{g,n} \end{tikzcd}$$

and is given by

$$\begin{tikzcd} & \overline{\mathcal{M}}_{g-1,n+2} \times_{\overline{\mathcal{M}}_{g,n}} \overline{\mathcal{M}}_{g,n}^{\text{spin}} \arrow[dl, "\hat{\nu}"'] \arrow[dr, "\phi_{\text{irr}}"] & \\ \overline{\mathcal{M}}_{g-1,n+2}^{\text{spin}} \arrow[rr, dashed] & & \overline{\mathcal{M}}_{g,n}^{\text{spin}} \end{tikzcd}$$

Again, $\phi_{\text{irr}} \circ \hat{\nu}^{-1}$ naturally restricts to components $\overline{\mathcal{M}}_{g-1,n+2,\vec{\sigma}'}^{\text{spin}} \dashrightarrow \overline{\mathcal{M}}_{g,n,\vec{\sigma}}^{\text{spin}}$, but, unlike the case of $\phi_{h,I} \circ \hat{\nu}^{-1}$ above, $\vec{\sigma}$ does not uniquely determine $\vec{\sigma}'$. The map $\hat{\nu}$ now depends on θ and there are two cases, corresponding to the decomposition of the fibre product $\overline{\mathcal{M}}_{g-1,n+2} \times_{\overline{\mathcal{M}}_{g,n}} \overline{\mathcal{M}}_{g,n,\vec{\sigma}}^{\text{spin}}$ into two components which depend on the behaviour of θ at the nodal point $p_\pm$. Either θ is banded by $\lambda_{p_\pm} = \frac{1}{2}$, or it is banded by $\lambda_{p_\pm} = 0$, corresponding to $\vec{\sigma}' = (\vec{\sigma}, 1, 1)$ and $\vec{\sigma}' = (\vec{\sigma}, 0, 0)$, respectively.

The bundle $E_{g,n}$ behaves naturally with respect to the boundary divisors.

Lemma 2.2 *On components where θ is banded by $\lambda_{p_\pm} = \frac{1}{2}$, at the node,*

$$\phi_{\text{irr}}^* E_{g,n} \cong \hat{\nu}^* E_{g-1,n+2}, \quad \phi_{h,I}^* E_{g,n} \cong \hat{\nu}^*(\pi_1^* E_{h,|I|+1} \oplus \pi_2^* E_{g-h,|J|+1}),$$

where π_i is projection from $\overline{\mathcal{M}}_{h,|I|+1}^{\text{spin}} \times \overline{\mathcal{M}}_{g-h,|J|+1}^{\text{spin}}$ onto the i^{th} factor for $i = 1, 2$.

Proof A spin structure $\tilde{\theta}$ on a connected normalisation $\widetilde{\mathcal{C}}$ has

$$\deg \tilde{\theta}^\vee = 1 - (g-1) - \tfrac{1}{2}(n+2) < 0$$

and also negative degree on all irreducible components; hence, $H^0(\widetilde{\mathcal{C}}, \tilde{\theta}^\vee) = 0$. By Riemann–Roch,

$$h^0(\widetilde{\mathcal{C}}, \tilde{\theta}^\vee) - h^1(\widetilde{\mathcal{C}}, \tilde{\theta}^\vee) = 1 - (g-1) + \deg \tilde{\theta}^\vee - \tfrac{1}{2}(n+2) = 2 - 2g - n.$$

Hence, $\dim H^1(\widetilde{\mathcal{C}},\tilde{\theta}^\vee)=\dim H^1(\mathcal{C},\theta^\vee)$ and the natural map

$$0\to H^1(\mathcal{C},\theta^\vee)\to H^1(\widetilde{\mathcal{C}},\tilde{\theta}^\vee)$$

is an isomorphism. In other words, $\phi_{\mathrm{irr}}^*E_{g,n}\cong\nu^*E_{g-1,n+2}$.

The argument is analogous when $\widetilde{\mathcal{C}}$ is not connected and $\lambda_{p_\pm}=\frac{1}{2}$. Again $\deg\theta_i^\vee<0$, and it has negative degree on all irreducible components; hence, $H^0(\mathcal{C},\theta_i^\vee)=0$ for $i=1,2$. By Riemann–Roch,

$$\dim H^1(\mathcal{C},\theta_1^\vee)+\dim H^1(\mathcal{C},\theta_2^\vee)=\dim H^1(\mathcal{C},\theta^\vee),$$

so the natural map

$$0\to H^1(\mathcal{C},\theta^\vee)\to H^1(\widetilde{\mathcal{C}}_1,\theta_1^\vee)\oplus H^1(\widetilde{\mathcal{C}}_2,\theta_2^\vee)$$

is an isomorphism. In other words, $\phi_{h,I}^*E_{g,n}\cong\hat{\nu}^*(\pi_1^*E_{h,|I|+1}\oplus\pi_2^*E_{g-h,|J|+1})$. □

The pullback of $E_{g,n}$ to boundary divisors with trivial isotropy at the node is described in the following lemma:

Lemma 2.3 *On components where θ is banded by $\lambda_{p_\pm}=0$, at the node,*

$$0\to\mathcal{O}_{X_{h,I}}\to\phi_{h,I}^*E_{g,n}\to\hat{\nu}^*(\pi_1^*E_{h,|I|+1}\oplus\pi_2^*E_{g-h,|J|+1})\to 0 \tag{4}$$

for $X_{h,I}=(\overline{\mathcal{M}}_{h,|I|+1}\times\overline{\mathcal{M}}_{g-h,|J|+1})\times_{\overline{\mathcal{M}}_{g,n}}\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$ and

$$0\to\mathcal{O}_{X_{\mathrm{irr}}}\to\phi_{\mathrm{irr}}^*E_{g,n}\to\hat{\nu}^*E_{g-1,n+2}\to 0 \tag{5}$$

for $X_{\mathrm{irr}}=\overline{\mathcal{M}}_{g-1,n+2}\times_{\overline{\mathcal{M}}_{g,n}}\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$.

Proof When the bundle θ is banded by $\lambda_{p_\pm}=0$, the map between sheaves of local holomorphic sections

$$\Gamma(U,\theta)\to\Gamma(\nu^{-1}U,\nu^*\theta)$$

is not surjective whenever $U\ni p$. The image consists of local sections that agree, under an identification of fibres, at p_+ and p_-. Hence we have an exact sequence

$$0\to\theta^\vee\to\nu_*\nu^*\theta^\vee\to\nu_*\nu^*\theta^\vee|_p\to 0, \tag{6}$$

where the quotient sends a local section $s\in\Gamma(\nu^{-1}U,\nu^*\theta^\vee)$ to $s(p_+)-s(p_-)$. Note that this difference of sections over different points makes sense since $X_{h,I}$ and X_{irr} come with a choice of isomorphism between the fibres over p_+ and p_-. The exact sequence (6) splits as follows. We can choose a representative ϕ upstairs of any element from the quotient space so that $\phi(p_+)=0$, ie $\Gamma(U,\theta^\vee)$ corresponds to elements of $\Gamma(\nu^{-1}U,\nu^*\theta^\vee)$ that vanish at p_+. This is achieved by adding the appropriate multiple of $s(p_+)-s(p_-)$ to a given $\phi\in\Gamma(\nu^{-1}U,\nu^*\theta^\vee)$. (Note that $\phi(p_-)$ is arbitrary. One could instead arrange $\phi(p_-)=0$ with $\phi(p_+)$ arbitrary.) In other words, we can identify $\theta^\vee$ with $\nu^*\theta^\vee(-p_+)$ in the complex

$$0\to\nu^*\theta^\vee(-p_+)\to\nu^*\theta^\vee\to\nu^*\theta^\vee|_{p_+}\to 0.$$

In a family $\pi\colon C\to S$, $R^0\pi_*(\nu^*\theta^\vee)=0=R^0\pi_*(\nu^*\theta^\vee(-p_+))$ since $\deg\nu^*\theta^\vee<0$, and it has negative degree on all irreducible components. Also $R^1\pi_*(\nu^*\theta^\vee|_{p_+})=0$ since p_+ has relative dimension 0. Thus,

$$0\to R^0\pi_*(\nu^*\theta^\vee|_{p_+})\to R^1\pi_*(\nu^*\theta^\vee(-p_+))\to R^1\pi_*(\nu^*\theta^\vee)\to 0. \tag{7}$$

We can identify the sequence (7) with the sequences (4) and (5) as follows. For the first term of (7), $\nu^*\theta^\vee|_{p_+}\cong\mathbb{C}$ canonically, since $\omega_C^{\log}|_{p_+}\cong\mathbb{C}$ canonically by the residue map; hence, $R^0\pi_*(\nu^*\theta^\vee|_{p_+})\cong:\mathbb{C}_S$. The second and third terms of (7) are identified with the corresponding terms of (4) by $\hat{\nu}^*(\pi_1^*E_{h,|I|+1}\oplus\pi_2^*E_{g-h,|J|+1})=R^1\pi_*(\nu^*\theta^\vee)$ and $\phi_{h,I}^*E_{g,n}=R^1\pi_*(\nu^*\theta^\vee(-p_+))$, and similarly with those of (5) by $\hat{\nu}^*E_{g-1,n+2}=R^1\pi_*(\nu^*\theta^\vee)$ and $\phi_{\mathrm{irr}}^*E_{g,n}=R^1\pi_*(\nu^*\theta^\vee(-p_+))$. □

Remark 2.4 In Lemma 2.2, the nodal band is $\lambda_{p_\pm}=\frac{1}{2}$ and so $\lambda_{p_+}+\lambda_{p_-}=1$. We see from Lemma 2.3 that $\lambda_{p_\pm}=0$ really wants one of $\lambda_{p_\pm}$ to be 1 to preserve $\lambda_{p_+}+\lambda_{p_-}=1$.

Definition 2.5 For $2g-2+n>0$, define the Chern class

$$\Omega_{g,n}:=c_{2g-2+n}(E_{g,n})\in H^{4g-4+2n}(\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}},\mathbb{Q}).$$

On the component $\overline{\mathcal{M}}_{g,n,\vec{\sigma}}^{\mathrm{spin}}$ of $\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$ for $|\vec{\sigma}|=n$, this defines the top Chern class, or Euler class. The Chern class vanishes on all other components because, by (3), the rank of $E_{g,n}=2g-2+\frac{1}{2}(|\vec{\sigma}|+n)<2g-2+n$ when $|\vec{\sigma}|<n$. Note that $\Omega_{0,n}=0$ for $n\ge 3$ because $\operatorname{rank}(E_{0,n})=n-2$ is greater than $\dim\overline{\mathcal{M}}_{0,n}^{\mathrm{spin}}=n-3$, so its top Chern class vanishes.

The cohomology classes $\Omega_{g,n}$ behave well with respect to inclusion of strata.

Lemma 2.6 *We have*

$$\phi_{\mathrm{irr}}^*\Omega_{g,n}=\hat{\nu}^*\Omega_{g-1,n+2},\quad \phi_{h,I}^*\Omega_{g,n}=\hat{\nu}^*(\pi_1^*\Omega_{h,|I|+1}\cdot\pi_2^*\Omega_{g-h,|J|+1}).$$

Proof When $|\vec{\sigma}| = n$ and θ is banded by $\frac{1}{2}$ at the nodal point, this is an immediate application of Lemma 2.2 and the naturality of $c_{2g-2+n} = c_{\mathrm{top}}$: we have

$$\phi_{\mathrm{irr}}^* c_{\mathrm{top}}(E_{g,n}) = \hat{\nu}^* c_{\mathrm{top}}(E_{g-1,n+2}),$$
$$\phi_{h,I}^* c_{\mathrm{top}}(E_{g,n}) = \hat{\nu}^* (\pi_1^* c_{\mathrm{top}}(E_{h,|I|+1}) \cdot \pi_2^* c_{\mathrm{top}}(E_{g-h,|J|+1})).$$

When $|\vec{\sigma}| = n$ and θ is banded by 0 at the nodal point, the nodal point is necessarily nonseparating and we must consider the restriction of $\Omega_{g,n}$ to the component $\overline{\mathcal{M}}^{\mathrm{spin}}_{g-1,n+2,\vec{\sigma}'}$ of $\overline{\mathcal{M}}^{\mathrm{spin}}_{g-1,n+2}$ with $|\vec{\sigma}'| = n$. On this component, we have the exact sequence of Lemma 2.3,

$$0 \to E_{g-1,n+2} \to \phi_{\mathrm{irr}}^* E_{g,n} \to \mathcal{O}_{\overline{\mathcal{M}}^{\mathrm{spin}}_{g-1,n+2,\vec{\sigma}'}} \to 0,$$

which implies $\phi_{\mathrm{irr}}^* c_{2g-2+n}(E_{g,n}) = c_{2g-3+n}(E_{g-1,n+2,\vec{\sigma}'}) \cdot c_1(\mathcal{O}_{\overline{\mathcal{M}}^{\mathrm{spin}}_{g-1,n+2,\vec{\sigma}'}}) = 0$. This vanishing result is a special case of the pullback by ϕ_{irr}^* since $\Omega_{g-1,n+2}$ vanishes on $\overline{\mathcal{M}}^{\mathrm{spin}}_{g-1,n+2,\vec{\sigma}'}$ for $|\vec{\sigma}'| = n$.

Finally, when $|\vec{\sigma}| < n$, this is simply because the pullback of the trivial class is trivial, since in each case the restriction to an irreducible component has at least one labelled point with band equal to 0, so that the right-hand side vanishes. □

The cohomology classes $\Omega_{g,n}$ also behave well with respect to the forgetful map

$$\pi : \overline{\mathcal{M}}^{\mathrm{spin}}_{g,n+1} \to \overline{\mathcal{M}}^{\mathrm{spin}}_{g,n}$$

which is defined on components with θ banded by $\frac{1}{2}$ at p_{n+1} as follows. Define

$$\pi(\mathcal{C}, \theta, p_1, \dots, p_{n+1}, \phi) = (\rho(\mathcal{C}), \rho_*\theta, p_1, \dots, p_n, \rho_*\phi),$$

where $\rho(\mathcal{C})$ forgets the orbifold structure at p_{n+1}. The pushforward sheaf $\rho_*\theta$ consists of local sections invariant under the $\mathbb{Z}_2$ action. Since the representation at p_{n+1} is given by multiplication by -1, any invariant local section must vanish at p_{n+1}. In terms of a local orbifold coordinate $x = z^2$, an invariant section is of the form $zf(x)s$ for s a generator of θ and its square

$$(zf(x)s)^2 = z^2 f(x)^2 s^2 = xf(x)^2 \frac{dx}{x} = f(x)^2\, dx$$

has no pole. In other words, its square is a section of $\omega_{\mathcal{C}}^{\log}$ with no pole at p_{n+1} and hence a section of $\omega_{\rho(\mathcal{C})}^{\log} = \omega_{\rho(\mathcal{C})}(p_1 + p_2 + \cdots + p_n)$. Furthermore, we have $\rho_*\theta = \rho_*\{\theta(-p_{n+1})\}$, $\rho^*\rho_*\theta = \theta(-p_{n+1})$ and $\deg \rho_*\theta = \deg \theta - \frac{1}{2}$. The forgetful map π is used to denote any family $\pi : \mathcal{C} \to S$ since $\overline{\mathcal{M}}^{\mathrm{spin}}_{g,n+1}$ is essentially the universal curve of $\overline{\mathcal{M}}^{\mathrm{spin}}_{g,n}$.

Tautological line bundles $L_{p_i} \to \overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}$ for $i = 1, \dots, n$ are defined analogously to those defined over $\overline{\mathcal{M}}_{g,n}$ as follows. Consider a family $\pi : \mathcal{C} \to S$ with sections $p_i : S \to \mathcal{C}$ for $i = 1, \dots, n$, and define

$$L_{p_i} := p_i^*(\omega_{\mathcal{C}/S}), \quad \psi_i = c_1(L_{p_i}) \in H^*(\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}}, \mathbb{Q}).$$

Lemma 2.7 $$\Omega_{g,n+1} = -\psi_{p_{n+1}} \cdot \pi^* \Omega_{g,n}.$$

Proof Over a family $\pi : \mathcal{C} \to S$, where $S \to \overline{\mathcal{M}}_{g,n+1}^{\mathrm{spin}}$ and $\theta \to \mathcal{C}$ is the universal spin structure (also denoted by $\mathcal{E}$), tensor the exact sequence of sheaves

$$0 \to \mathcal{O}_{\mathcal{C}}(-p_{n+1}) \to \mathcal{O}_{\mathcal{C}} \to \mathcal{O}_{\mathcal{C}}|_{p_{n+1}} \to 0$$

with $\theta^\vee(p_{n+1})$ to get

$$0 \to \theta^\vee \to \theta^\vee(p_{n+1}) \to \theta^\vee(p_{n+1})|_{p_{n+1}} \to 0.$$

This induces a long exact sequence, which simplifies to the short exact sequence

$$0 \to R^0\pi_*(\theta^\vee(p_{n+1})|_{p_{n+1}}) \to R^1\pi_*\theta^\vee \to R^1\pi_*(\theta^\vee(p_{n+1})) \to 0$$

due to the vanishing $R^0\pi_*(\theta^\vee(p_{n+1})) = 0 = R^1\pi_*(\theta^\vee(p_{n+1})|_{p_{n+1}})$. The first of these vanishing results uses the identification $\theta^\vee(p_{n+1}) = \pi^*\theta^\vee$ described below together with the vanishing $R^0\pi_*\theta^\vee = 0$ due to the negative degree on each irreducible component described earlier. The second of these vanishing results uses the simple dimension argument that $R^1\pi_*$ vanishes on the image of p_{n+1}, which has relative dimension 0.

Recall that the forgetful map $(\mathcal{C}, \theta, p_1, \dots, p_{n+1}, \phi) \mapsto (\pi(\mathcal{C}), \pi_*\theta, p_1, \dots, p_n, \pi_*\phi)$ pushes forward θ via π which forgets the orbifold structure at p_{n+1}. As described earlier, $\pi^*\pi_*\theta = \theta(-p_{n+1})$ since the pushforward gives the sheaf of locally invariant sections, which necessarily vanish as the isotropy group acts by multiplication by -1. Hence, $\theta^\vee(p_{n+1}) = \pi^*\theta^\vee$, which is used to calculate R^0 above, and also to give $R^1\pi_*(\theta^\vee(p_{n+1})) = R^1\pi_*(\pi^*\theta^\vee) = \pi^*R^1\pi_*(\theta^\vee)$. Thus, the last two terms of the short exact sequence become $E_{g,n+1} \to \pi^*E_{g,n}$.

For the first term of the short exact sequence, the residue map produces a canonical isomorphism

$$\pi_*\omega_{C/S}^{\log}|_{p_{n+1}} = \mathcal{O}_S.$$

Thus, $\pi_*(\theta|_{p_{n+1}})$ and $\pi_*(\theta^\vee|_{p_{n+1}})$ define line bundles over S with square $\mathcal{O}_S$ and hence trivial Chern class $c(\pi_*(\theta|_{p_{n+1}})) = 1 = c(\pi_*(\theta^\vee|_{p_{n+1}}))$. The first term of the short exact sequence $R^0\pi_*(\theta^\vee(p_{n+1})|_{p_{n+1}})$ defines a line bundle $\xi \to S$ with Chern class

$$c(\xi) = c\big(R^0\pi_*(\mathcal{O}_{\mathcal{C}}(p_{n+1})|_{p_{n+1}})\big)$$

that fits into the short exact sequence

$$0 \to \xi \to E_{g,n+1} \to \pi^* E_{g,n} \to 0.$$

The triviality of $\pi_*(\omega_{\mathcal{C}/S}^{\log}|_{p_{n+1}})$ implies

$$L_{p_{n+1}} = R^0\pi_*(\omega_{\mathcal{C}/S}|_{p_{n+1}}) = -R^0\pi_*(\mathcal{O}_{\mathcal{C}}(p_{n+1})|_{p_{n+1}});$$

hence,

$$c(\xi) = \frac{1}{c(L_{p_{n+1}})} = 1 - \psi_{p_{n+1}}.$$

The short exact sequence then gives $c_{2g-2+n+1}(E_{g,n+1}) = -\psi_{p_{n+1}} \cdot \pi^* c_{2g-2+n}(E_{g,n})$, as required. □

Definition 2.8 For $p\colon \overline{\mathcal{M}}_{g,n}^{\mathrm{spin}} \to \overline{\mathcal{M}}_{g,n}$, define

$$\Theta_{g,n} = (-1)^n 2^{g-1+n} p_* \Omega_{g,n} \in H^{4g-4+2n}(\overline{\mathcal{M}}_{g,n}, \mathbb{Q}).$$

Lemma 2.7 and the relation

$$\psi_{n+1} = \tfrac{1}{2} p^* \psi_{n+1}$$

proven in [26, Proposition 2.4.1], together with the factor of 2^n in the definition of $\Omega_{g,n}$, immediately gives property (iii) of $\Theta_{g,n}$,

$$\Theta_{g,n+1} = \psi_{n+1} \cdot \pi^* \Theta_{g,n}.$$

Property (iv) of $\Theta_{g,n}$ is given by the following calculation:

Proposition 2.9 $$\Theta_{1,1} = 3\psi_1 \in H^2(\overline{\mathcal{M}}_{1,1}, \mathbb{Q}).$$

Proof A one-pointed twisted elliptic curve $(\mathcal{E}, p)$ is a one-pointed elliptic curve (E, p) such that p has isotropy $\mathbb{Z}_2$. The degree of the divisor p in $\mathcal{E}$ is $\frac{1}{2}$ and the degree of every other point in $\mathcal{E}$ is 1. If dz is a holomorphic differential on E (where $E = \mathbb{C}/\Lambda$ and z is the identity function on the universal cover $\mathbb{C}$), then, locally near p, we have $z = t^2$, so $dz = 2t\,dt$ vanishes at p. In particular, the canonical divisor $(\omega_{\mathcal{E}}) = p$ has degree $\frac{1}{2}$ and $(\omega_{\mathcal{E}}^{\log}) = (\omega_{\mathcal{E}}(p)) = 2p$ has degree 1.

A spin structure on $\mathcal{E}$ is a degree $\frac{1}{2}$ line bundle $\mathcal{L}$ satisfying $\mathcal{L}^2 = \omega_{\mathcal{E}}^{\log}$. Line bundles on $\mathcal{E}$ correspond to divisors on $\mathcal{E}$ up to linear equivalence. Note that meromorphic functions on $\mathcal{E}$ are exactly the meromorphic functions on E. The four spin structures on $\mathcal{E}$ are given by the divisors $\theta_0 = p$ and $\theta_i = q_i - p$ for $i = 1, 2, 3$, where q_i is a nontrivial order 2 element in the group E with identity p. Clearly, $\theta_0^2=2p=\omega_{\mathcal{E}}^{\log}$. For $i=1,2,3$, $\theta_i^2=2q_i-2p\sim 2p$ since there is a meromorphic function $\wp(z)-\wp(q_i)$ on E with a double pole at p and a double zero at q_i. Its divisor on $\mathcal{E}$ is $2q_i-4p$, since p has isotropy $\mathbb{Z}_2$; hence, $2q_i-2p\sim 2p$.

Since $H^2(\overline{\mathcal{M}}_{1,1},\mathbb{Q})$ is generated by ψ_1, it is enough to calculate $\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}$. The Chern character of the pushforward bundle $E_{1,1}$ is calculated via the Grothendieck–Riemann–Roch theorem:

$$\mathrm{ch}(R\pi_*\mathcal{E}^\vee)=\pi_*(\mathrm{ch}(\mathcal{E}^\vee)\,\mathrm{Td}(\omega_\pi^\vee)).$$

In fact we need to use the orbifold Grothendieck–Riemann–Roch theorem [53]. The calculation we need is a variant of the calculation in [26, Theorem 6.3.3] which applies to $\mathcal{E}$ such that $\mathcal{E}^2=\omega_C^{\log}$ instead of $\mathcal{E}^\vee$. Importantly, this means that the Todd class has been worked out, and it remains to adjust the $\mathrm{ch}(\mathcal{E}^\vee)$ term. We get

$$\begin{aligned}\int_{\overline{\mathcal{M}}_{1,1}} p_*c_1(E_{1,1}) &= -\mathrm{ch}(R\pi_*\mathcal{E}^\vee)\\ &= -2\int_{\overline{\mathcal{M}}_{1,1}}\bigl[\tfrac{11}{24}\kappa_1+\tfrac{1}{24}\psi_1+\tfrac12\bigl(-\tfrac{1}{24}+\tfrac{1}{12}\bigr)(i_\Gamma)_*(1)\bigr]\\ &= -2\bigl(\tfrac{11}{24^2}+\tfrac{1}{24^2}+\tfrac12\cdot\tfrac{1}{24}\cdot\tfrac12\bigr)=-\tfrac{1}{16},\end{aligned}$$

which agrees with

$$-\int_{\overline{\mathcal{M}}_{1,1}}\tfrac32\psi_1=-\tfrac32\cdot\tfrac{1}{24}=-\tfrac{1}{16}.$$

Hence, $p_*c_1(E_{1,1})=-\frac32\psi_1$ and $\Theta_{1,1}=-2p_*c_1(E_{1,1})=3\psi_1$. One can also calculate this using Chiodo's formula [10], given by (41) in Section 5. □

Proposition 2.10 *The classes* $\Theta_{g,n}\in H^{4g-4+2n}(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$ *satisfy property* (ii).

Proof The two properties (ii) of $\Theta_{g,n}$ follow from the analogous properties for $\Omega_{g,n}$. This uses the relationship between compositions of pullbacks and pushforwards in the diagrams

$$\begin{array}{ccc} \overline{\mathcal{M}}^{\mathrm{spin}}_{g-1,n+2} & \xrightarrow{\phi_{\mathrm{irr}}\circ\hat{\nu}^{-1}} & \overline{\mathcal{M}}^{\mathrm{spin}}_{g,n}\\ \Big\downarrow p & & \Big\downarrow p\\ \overline{\mathcal{M}}_{g-1,n+2} & \xrightarrow{\phi_{\mathrm{irr}}} & \overline{\mathcal{M}}_{g,n}\end{array}\qquad \begin{array}{ccc} \overline{\mathcal{M}}^{\mathrm{spin}}_{h,|I|+1}\times\overline{\mathcal{M}}^{\mathrm{spin}}_{g-h,|J|+1} & \xrightarrow{\phi_{h,I}\circ\hat{\nu}^{-1}} & \overline{\mathcal{M}}^{\mathrm{spin}}_{g,n}\\ \Big\downarrow p & & \Big\downarrow p\\ \overline{\mathcal{M}}_{h,|I|+1}\times\overline{\mathcal{M}}_{g-h,|J|+1} & \xrightarrow{\phi_{h,I}} & \overline{\mathcal{M}}_{g,n}\end{array}$$

where the broken arrows signify multiply defined maps which are defined above using fibre products.

On cohomology, we have $\phi_{\mathrm{irr}}^* p_* = 2p_*\hat{\nu}_*\phi_{\mathrm{irr}}^*$ and $\phi_{h,I}^* p_* = 2p_*\hat{\nu}_*\phi_{h,I}^*$, where the factor of 2 is due to the degree of $\hat{\nu}$ ramification of p and the isotropy of the orbifold divisor; see [33, (39)]. Hence,

$$\begin{aligned}\phi_{\mathrm{irr}}^*\Theta_{g,n} &= \phi_{\mathrm{irr}}^* p_*(-1)^n 2^{g-1+n}\Omega_{g,n} = 2p_*\hat{\nu}_*\phi_{\mathrm{irr}}^*(-1)^n 2^{g-1+n}\Omega_{g,n}\\ &= p_*(-1)^{n+2}2^{g+n}\Omega_{g-1,n+2} = \Theta_{g-1,n+2}\end{aligned}$$

and, similarly, $\phi_{h,I}^*\Theta_{g,n} = \pi_1^*\Theta_{h,|I|+1}\cdot\pi_2^*\Theta_{g-h,|J|+1}$, which uses

$$2\cdot(-1)^n 2^{g-1+n} = (-1)^n 2^{g+n} = (-1)^{|I|+1}2^{h-1+|I|+1}(-1)^{|J|+1}2^{g-h-1+|J|+1}. \qquad \square$$

Remark 2.11 The construction of $\Omega_{g,n}$ should also follow from the cosection construction in [7] using the moduli space of spin curves with fields

$$\overline{\mathcal{M}}_{g,n}(\mathbb{Z}_2)^p = \{(C,\theta,\rho) \mid (C,\theta)\in\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}},\ \rho\in H^0(\mathcal{C},\theta)\}.$$

A cosection of the pullback of $E_{g,n}$ to $\overline{\mathcal{M}}_{g,n}(\mathbb{Z}_2)^p$ is given by ρ^{-3} since it pairs well with $H^1(\mathcal{C},\theta)$: we have $\rho^{-3}\in H^0(\mathcal{C},(\theta^\vee)^3)$ while $H^1(\mathcal{C},\theta)\cong H^0(\mathcal{C},\omega\otimes\theta^\vee)^\vee = H^0(\mathcal{C},(\theta^\vee)^3)^\vee$. Using the cosection ρ^{-3}, a virtual fundamental class is constructed in [7] that likely gives rise to $\Omega_{g,n}\in H^{4g-4+2n}(\overline{\mathcal{M}}_{g,n}^{\mathrm{spin}},\mathbb{Q})$. The virtual fundamental class is constructed away from the zero set of ρ.

3 Uniqueness

The degree property (I) of Theorem 1.3, $\Theta_{g,n}\in H^{4g-4+2n}(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$, proven below, implies the initial value

$$\Theta_{1,1} = \lambda\psi, \quad \lambda\in\mathbb{Q}.$$

It leads to uniqueness of intersection numbers $\int_{\overline{\mathcal{M}}_{g,n}}\Theta_{g,n}\prod_{i=1}^n\psi_i^{m_i}\prod_{j=1}^N\kappa_{\ell_j}$ via a reduction argument, and consequently property (V) of Theorem 1.3. The proofs in this section of properties (II), (III) and (V) apply for any $\lambda\neq 0$. We finish the section with a rigidity result given by Theorem 3.6, proving that necessarily $\lambda = 3$.

We first prove the following lemma, which will be needed later:

Lemma 3.1 *Properties* (i)–(iv) *imply that* $\Theta_{g,n}\neq 0$ *for* $g>0$ *and all* n.

Proof We have $\Theta_{1,1} = a$ or $\Theta_{1,1} = a\psi_1$ for $a\neq 0$ by (i) and (iv). Using the pullback property (iii) together with the equality $\psi_n\psi_i = \psi_n\pi^*\psi_i$ for $i<n$, we have $\Theta_{1,n} = a\psi_2\cdots\psi_n$ or $\Theta_{1,n} = a\psi_1\psi_2\cdots\psi_n$; hence, $(1+\psi_1)\Theta_{1,n} = a\psi_1\psi_2\cdots\psi_n$ and $\int_{\overline{\mathcal{M}}_{g,n}}(1+\psi_1)\Theta_{1,n} = \frac{1}{24}a(n-1)!$, proving $\Theta_{1,n}\neq 0$.

Now we proceed by induction on g. For the base case of $g=1$, we have $\Theta_{1,n}\neq 0$ for all $n>0$. Assume $\Theta_{h,n}\neq 0$ for $0<h<g$ and all n. For $g>1$, let Γ be the stable graph consisting of a genus $g-1$ vertex attached by a single edge to a genus 1 vertex with n labelled leaves (called *ordinary* leaves in Section 5.0.1). Then, by (ii),

$$\phi_\Gamma^*\Theta_{g,n}=\Theta_{g-1,1}\otimes\Theta_{1,n+1},$$

which is nonzero since $\Theta_{g-1,1}\neq 0$ by the inductive hypothesis and $\Theta_{1,n+1}\neq 0$ by the calculation above. $\square$

Proof of (I) Write

$$d(g,n)=\text{degree}(\Theta_{g,n}),$$

which exists by (i). Note that the degree here is half the cohomological degree, so $\Theta_{g,n}\in H^{2d(g,n)}(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$. Using (ii), $\phi_{\text{irr}}^*\Theta_{g,n}=\Theta_{g-1,n+2}$ implies that

$$d(g,n)=d(g-1,n+2)$$

since $\Theta_{g-1,n+2}\neq 0$ by Lemma 3.1. Hence, $d(g,n)=f(2g-2+n)$ is a function of $2g-2+n$. Similarly, using (ii), $\phi_{h,I}^*\Theta_{g,n}=\Theta_{h,|I|+1}\otimes\Theta_{g-h,|J|+1}$ implies that $f(a+b)=f(a)+f(b)=(a+b)f(1)$ since $\Theta_{h,|I|+1}\neq 0$ and $\Theta_{g-h,|J|+1}\neq 0$, again by Lemma 3.1. Hence,

$$d(g,n)=(2g-2+n)k$$

for an integer k. But $d(g,n)\leq 3g-3+n$ implies $k\leq 1$. When $k=0$, this gives $\deg\Theta_{g,n}=0$, which contradicts (iii) together with Lemma 3.1; hence, $k=1$ and $\deg\Theta_{g,n}=2g-2+n$. $\square$

Proof of (II) This is an immediate consequence of (I) since

$$\deg\Theta_{0,n}=n-2>n-3=\dim\overline{\mathcal{M}}_{0,n}$$

and hence $\Theta_{0,n}=0$. For any stable graph Γ with a genus 0 vertex, Remark 1.1 gives $\phi_\Gamma^*\Theta_{g,n}=\Theta_\Gamma=\prod_{v\in V(\Gamma)}\pi_v^*\Theta_{g(v),n(v)}=0$ since the genus 0 vertex contributes a factor of 0 to the product. $\square$

Proof of (III) Property (iii) implies that

$$\Theta_{g,n}=\prod_{i=1}^n\psi_i\cdot\pi^*\Theta_g,$$

where $\pi : \overline{\mathcal{M}}_{g,n} \to \overline{\mathcal{M}}_g$ is the forgetful map. Since $\pi^*\omega \in H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})^{S_n}$ for any class $\omega \in H^*(\overline{\mathcal{M}}_g, \mathbb{Q})$ and clearly $\prod_{i=1}^n \psi_i \in H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})^{S_n}$, we have $\Theta_{g,n} \in H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})^{S_n}$, as required. □

The proof of (V) follows from the special case of the intersection of $\Theta_{g,n}$ with a polynomial in κ and ψ classes.

Proposition 3.2 *For any $\Theta_{g,n}$ satisfying properties* (i)–(iii), *the intersection numbers*

$$\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \prod_{i=1}^n \psi_i^{m_i} \prod_{j=1}^N \kappa_{\ell_j} \tag{8}$$

are uniquely determined from the initial condition $\Theta_{1,1} = \lambda\psi_1$ for $\lambda \in \mathbb{Q}$.

Proof For $n > 0$, we will push forward the integral (8) via the forgetful map $\pi : \overline{\mathcal{M}}_{g,n} \to \overline{\mathcal{M}}_{g,n-1}$ as follows. Consider first the case when there are no κ classes. The presence of ψ_n in $\Theta_{g,n} = \psi_n \cdot \pi^*\Theta_{g,n-1}$ gives

$$\Theta_{g,n}\psi_k = \Theta_{g,n}\pi^*\psi_k, \quad k < n,$$

since $\psi_n\psi_k = \psi_n\pi^*\psi_k$ for $k < n$. Hence,

$$\begin{aligned}
\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \prod_{i=1}^n \psi_i^{m_i} &= \int_{\overline{\mathcal{M}}_{g,n}} \pi^*\Big(\Theta_{g,n-1} \prod_{i=1}^{n-1} \psi_i^{m_i}\Big)\psi_n^{m_n+1} \\
&= \int_{\overline{\mathcal{M}}_{g,n-1}} \pi_*\Big\{\pi^*\Big(\Theta_{g,n-1} \prod_{i=1}^{n-1} \psi_i^{m_i}\Big)\psi_n^{m_n+1}\Big\} \\
&= \int_{\overline{\mathcal{M}}_{g,n-1}} \Theta_{g,n-1} \prod_{i=1}^{n-1} \psi_i^{m_i}\kappa_{m_n},
\end{aligned}$$

so we have reduced an intersection number over $\overline{\mathcal{M}}_{g,n}$ to an intersection number over $\overline{\mathcal{M}}_{g,n-1}$. In the presence of κ classes, replace κ_{ℓ_j} by $\kappa_{\ell_j} = \pi^*\kappa_{\ell_j} + \psi_n^{\ell_j}$ and repeat the pushforward as above on all summands. By induction, we see that, for $g > 1$,

$$\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \prod_{i=1}^n \psi_i^{m_i} \prod_{j=1}^N \kappa_{\ell_j} = \int_{\overline{\mathcal{M}}_g} \Theta_g \cdot p(\kappa_1, \kappa_2, \dots, \kappa_{3g-3}),$$

ie the intersection number (8) reduces to an intersection number over $\overline{\mathcal{M}}_g$ of Θ_g times a polynomial in the κ classes. When $g = 1$, the right-hand side is instead $\int_{\overline{\mathcal{M}}_{1,1}} \Theta_{1,1} \cdot p$ for $p \in \mathbb{Q}$ a constant.

For $g > 1$, by (I), $\deg \Theta_g = 2g-2$, so we may assume the polynomial p consists only of terms of homogeneous degree $g-1$ (where $\deg \kappa_r = r$). But, by a result of Faber and Pandharipande [24, Proposition 2], which strengthens Looijenga's theorem [38], a homogeneous degree $g-1$ monomial in the κ classes is equal in the tautological ring to the sum of boundary terms, ie the sum of pushforwards of polynomials in ψ and κ classes by the maps $(\phi_\Gamma)_*$. Such relations arise from Pixton's relations and are described algorithmically in [11]. Now, property (ii) of Θ_g shows that the pullback of Θ_g to these boundary terms is $\Theta_{g',n'}$ for $g' < g$, so we have expressed (8) as a sum of integrals of $\Theta_{g',n'}$ against ψ and κ classes. By induction, one can reduce to the integral $\int_{\overline{\mathcal{M}}_{1,1}} \Theta_{1,1} = \frac{1}{24}\lambda$ and the proposition is proven. □

A consequence of Proposition 3.2 is property (V) of Theorem 1.3, stated as Corollary 3.3 below. Let us first recall the definition of tautological classes in $H^*(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$. Dual to any point $(C, p_1,\dots,p_n) \in \overline{\mathcal{M}}_{g,n}$ is its stable graph Γ with vertices $V(\Gamma)$ representing irreducible components of C, internal edges representing nodal singularities and a (labelled) external edge for each p_i. Each vertex is labelled by a genus $g(v)$ and has valency $n(v)$. The genus of a stable graph is $g(\Gamma) = b_1(\Gamma) + \sum_{v\in V(\Gamma)} g(v)$.

The *strata algebra* $S_{g,n}$ is a finite-dimensional vector space over $\mathbb{Q}$ with basis given by isomorphism classes of pairs (Γ,ω) for Γ a stable graph of genus g with n external edges and $\omega \in H^*(\overline{\mathcal{M}}_\Gamma,\mathbb{Q})$ a product of κ and ψ classes in each $\overline{\mathcal{M}}_{g(v),n(v)}$ for each vertex $v \in V(\Gamma)$. There is a natural map

$$q\colon S_{g,n} \to H^*(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$$

defined by the pushforward $q(\Gamma,\omega) = \phi_\Gamma^*(\omega) \in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$. The map q allows one to define a multiplication on $S_{g,n}$, essentially coming from intersection theory in $\overline{\mathcal{M}}_{g,n}$, which can be described purely graphically. The image $q(S_{g,n}) \subset H^*(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$ is the tautological ring $RH^*(\overline{\mathcal{M}}_{g,n})$ and an element of the kernel of q is a tautological relation. See [47, Section 0.3] for a detailed description of $S_{g,n}$.

Corollary 3.3 *For all* $\eta \in RH^*(\overline{\mathcal{M}}_{g,n})$, $\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n}\eta \in \mathbb{Q}$ *is uniquely determined by properties* (i)–(iii) *and* (IV).

Proof The tautological ring $RH^*(\overline{\mathcal{M}}_{g,n})$ consists of polynomials in the classes κ_i, ψ_i and boundary classes, which are pushforwards under $(\phi_\Gamma)_*$ of polynomials in κ_i and ψ_i. By the natural restriction property (ii) satisfied by $\Theta_{g,n}$, given a monomial in κ and ψ classes $\omega \in H^*(\overline{\mathcal{M}}_\Gamma, \mathbb{Q})$,

$$\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \cdot (\phi_\Gamma)_*(\omega) = \int_{\overline{\mathcal{M}}_\Gamma} \phi_\Gamma^*(\Theta_{g,n}) \cdot \omega = \int_{\overline{\mathcal{M}}_\Gamma} \Theta_\Gamma \cdot \omega = \frac{1}{|\mathrm{Aut}\,\Gamma|} \prod_{v \in \Gamma} w(v).$$

The final term is a product over the vertices of Γ of intersections Θ classes with monomials in κ and ψ classes $w(v) = \int_{\overline{\mathcal{M}}_{g(v),n(v)}} \Theta_{g(v),n(v)} \cdot \prod_{i=1}^{n(v)} P_v(\{\psi_i, \kappa_j\})$, which, by Proposition 3.2, are uniquely determined by (i)–(iii) and (IV). □

Remark 3.4 The intersection numbers $\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \prod_{i=1}^n \psi_i^{m_i} \prod_{j=1}^N \kappa_{\ell_j}$ can be calculated algorithmically from the intersection numbers $\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \prod_{i=1}^n \psi_i^{m_i}$ with no κ classes. This essentially reverses the reduction shown in the proof of Proposition 3.2. Explicitly, for $\pi : \overline{\mathcal{M}}_{g,n+N} \to \overline{\mathcal{M}}_{g,n}$ and $\boldsymbol{m} = (m_1, \ldots, m_N)$, define a polynomial in κ classes by

$$R_{\boldsymbol{m}}(\kappa_1, \kappa_2, \ldots) = \pi_*(\psi_{n+1}^{m_1+1} \cdots \psi_{n+N}^{m_N+1}),$$

so, for example, $R_{(m_1,m_2)} = \kappa_{m_1}\kappa_{m_1} + \kappa_{m_1+m_2}$. Then

$$\begin{aligned} \Theta_{g,n} \cdot R_{\boldsymbol{m}} &= \Theta_{g,n} \cdot \pi_*(\psi_{n+1}^{m_1+1} \cdots \psi_{n+N}^{m_N+1}) \\ &= \pi_*(\pi^*\Theta_{g,n} \cdot \psi_{n+1}^{m_1+1} \cdots \psi_{n+N}^{m_N+1}) \\ &= \pi_*(\Theta_{g,n+N} \cdot \psi_{n+1}^{m_1} \cdots \psi_{n+N}^{m_N}). \end{aligned} \tag{9}$$

The polynomials $R_{\boldsymbol{m}}(\kappa_1, \kappa_2, \ldots)$ generate all polynomials in the κ_i, so (9) can be used to remove any κ class.

The following example demonstrates Proposition 3.2 with an explicit genus 2 relation:

Example 3.5 A genus two relation proven by Mumford [41, (8.5)], relating κ_1 and the divisors defined by the double covers $\overline{\mathcal{M}}_{1,1} \times \overline{\mathcal{M}}_{1,1} \to \mathcal{M}_{\Gamma_1}$ and $\overline{\mathcal{M}}_{1,2} \to \mathcal{M}_{\Gamma_2}$ in $\overline{\mathcal{M}}_2$ labelled by stable graphs Γ_i, is given by

$$\kappa_1 - \tfrac{7}{5}[\mathcal{M}_{\Gamma_1}] - \tfrac{1}{5}[\mathcal{M}_{\Gamma_2}] = 0,$$

which induces the relation

$$\Theta_2 \cdot \kappa_1 - \tfrac{7}{5}\Theta_2 \cdot [\mathcal{M}_{\Gamma_1}] - \tfrac{1}{5}\Theta_2 \cdot [\mathcal{M}_{\Gamma_2}] = 0.$$

Property (ii) of $\Theta_{g,n}$ yields

$$\int_{\overline{\mathcal{M}}_2} \Theta_2 \cdot [\mathcal{M}_{\Gamma_1}] = \int_{\mathcal{M}_{\Gamma_1}} \phi_{\Gamma_1}^* \Theta_2 = \int_{\overline{\mathcal{M}}_{1,1}} \Theta_{1,1} \cdot \int_{\overline{\mathcal{M}}_{1,1}} \Theta_{1,1} \cdot \frac{1}{|\mathrm{Aut}(\Gamma_1)|},$$

$$\int_{\overline{\mathcal{M}}_2} \Theta_2 \cdot [\mathcal{M}_{\Gamma_2}] = \int_{\mathcal{M}_{\Gamma_2}} \phi_{\Gamma_2}^* \Theta_2 = \int_{\overline{\mathcal{M}}_{1,2}} \Theta_{1,2} \cdot \frac{1}{|\mathrm{Aut}(\Gamma_2)|};$$

hence, the relation on the level of intersection numbers is given by

$$\int_{\overline{\mathcal{M}}_2}\Theta_2\cdot\kappa_1-\tfrac{7}{5}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\frac{1}{|\mathrm{Aut}(\Gamma_1)|}-\tfrac{1}{5}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}\cdot\frac{1}{|\mathrm{Aut}(\Gamma_2)|}=0.$$

We have $\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}=\frac{1}{24}\lambda=\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}$ from (iii), and $|\mathrm{Aut}(\Gamma_1)|=2=|\mathrm{Aut}(\Gamma_2)|$. Hence,

$$\begin{aligned}\int_{\overline{\mathcal{M}}_2}\Theta_2\cdot\kappa_1&=\tfrac{7}{5}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\frac{1}{|\mathrm{Aut}(\Gamma_1)|}+\tfrac{1}{5}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}\cdot\frac{1}{|\mathrm{Aut}(\Gamma_2)|}\\&=\tfrac{7}{5}\cdot\left(\tfrac{1}{24}\lambda\right)^2\cdot\tfrac{1}{2}+\tfrac{1}{5}\cdot\tfrac{1}{24}\lambda\cdot\tfrac{1}{2}=\tfrac{1}{5760}(7\lambda^2+24\lambda).\end{aligned}$$

Until now, $\Theta_{1,1}=\lambda\psi_1$ for any nonzero $\lambda\in\mathbb{Q}$. The following theorem proves the rigidity condition (IV) that $\lambda=3$. The proof of the theorem relies on the fact that, for low genus and small n, the cohomology is tautological. This allows us to work in the tautological ring in order to construct $\Theta_{g,n}$ from properties (i)–(iv).

Theorem 3.6 *Let $\Theta_{g,n}\in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$ satisfy* (i)–(iv) *and set the initial condition to be $\Theta_{1,1}=\lambda\psi_1\neq 0$. Then $\lambda=3$.*

Proof The existence proof in Section 2 shows that $\lambda=3$ is possible but it does not exclude other values. The strategy of proof of this theorem is to attempt to construct classes, beginning with the initial condition $\Theta_{1,1}=\lambda\psi_1$. Importantly, condition (iii) determines $\Theta_{g,n}$ for all $n>0$ uniquely from Θ_g, so the main calculation occurs over $\overline{\mathcal{M}}_g$. We consider classes in $RH^{2g-2}(\overline{\mathcal{M}}_g)$ since, for small values of g, it is known that $H^{2*}(\overline{\mathcal{M}}_g,\mathbb{Q})=RH^*(\overline{\mathcal{M}}_g)$. The essential idea is as follows. A class $\Theta_g\in H^{2g-2}(\overline{\mathcal{M}}_g,\mathbb{Q})$ pulls back under boundary maps to $\Theta_{g-1,2}$ and $\Theta_{g-1,1}\otimes\Theta_{1,1}$. The relationship

$$\Theta_{g-1,2}=\psi_2\pi^*\Theta_{g-1,1}$$

constrains the class Θ_g. We find that Θ_2 exists (and hence also $\Theta_{2,n}$ exists for all n) for all $\lambda\in\mathbb{Q}$, but that Θ_3 (and $\Theta_{3,n}$) exists only for $\lambda=3$ or $\lambda=-\frac{11}{15}$. The existence of Θ_4 constrains λ further, allowing only $\lambda=3$.

***g* = 1** From $\Theta_{1,1}=\lambda\psi_1$, condition (iii) yields

$$\Theta_{1,n}=\lambda\psi_1\psi_2\cdots\psi_n$$

since $\psi_n\psi_j=\psi_n\pi^*\psi_j$ for any $j<n$.

***g* = 2** The cohomology group $H^4(\overline{\mathcal{M}}_2, \mathbb{Q})$ has basis $\{\kappa_1^2, \kappa_2\}$. Set $\Theta_2 = a_{11}\kappa_1^2 + a_2\kappa_2$ and deduce a_{11} and a_2 from restriction to $\mathcal{M}_{\Gamma_i} \subset \overline{\mathcal{M}}_2$ for $i = 1, 2$, defined in Example 3.5. Since $\kappa_2 \cdot \mathcal{M}_{\Gamma_1} = 0$, we deduce that $a_{11} = \frac{1}{2}\lambda^2$ and restriction to $\mathcal{M}_{\Gamma_2}$ then uniquely determines

$$\Theta_2 = \tfrac{1}{2}\lambda^2\kappa_1^2 + \bigl(\lambda - \tfrac{3}{2}\lambda^2\bigr)\kappa_2.$$

Commutativity of the boundary maps with the forgetful map shown in the diagrams

$$\begin{tikzcd}
\overline{\mathcal{M}}_{g-1,n+2} \arrow[r, "\phi_{\mathrm{irr}}"] \arrow[d, "\pi"'] & \overline{\mathcal{M}}_{g,n} \arrow[d, "\pi"'] & \overline{\mathcal{M}}_{h,|I|+1} \times \overline{\mathcal{M}}_{g-h,|J|+1} \arrow[r, "\phi_{h,I}"] \arrow[d, "\pi"'] & \overline{\mathcal{M}}_{g,n} \arrow[d, "\pi"'] \\
\overline{\mathcal{M}}_{g-1,2} \arrow[r, "\phi_{\mathrm{irr}}"] & \overline{\mathcal{M}}_g & \overline{\mathcal{M}}_{h,1} \times \overline{\mathcal{M}}_{g-h,1} \arrow[r, "\phi_h"] & \overline{\mathcal{M}}_g
\end{tikzcd}$$

implies that the classes $\Theta_{2,n} = \psi_1 \cdots \psi_n \pi^*\Theta_2$ restrict consistently to the boundary to give the correct genus 1 classes $\Theta_{1,n'}$ for all $\lambda \in \mathbb{Q}$.

***g* = 3** In genus 3, $H^{2*}(\overline{\mathcal{M}}_3, \mathbb{Q}) = RH^*(\overline{\mathcal{M}}_3)$ due to the calculation of the cohomology $H^*(\overline{\mathcal{M}}_3, \mathbb{Q})$, for example by using the calculation of $H^*(\overline{\mathcal{M}}_{3,1}, \mathbb{Q})$ in [28] together with the calculation of the tautological ring $RH^*(\overline{\mathcal{M}}_3)$ via Pixton's relations [47] implemented using the Sage package `admcycles` [12]. We have $\dim RH^4(\overline{\mathcal{M}}_3) = 7$ and we write Θ_3 as a general linear combination of basis vectors in $RH^4(\overline{\mathcal{M}}_3)$,

$$\Theta_3 = a_{1111}\kappa_1^4 + a_{112}\kappa_1^2\kappa_2 + a_{13}\kappa_1\kappa_3 + a_{22}\kappa_2^2 + a_4\kappa_4 + b_1B_1 + b_2B_2,$$

where $B_i \in RH^4(\overline{\mathcal{M}}_3)$ are given by

$$B_1 = \overset{\kappa_1}{\textcircled{1}} - \overset{\kappa_2}{\textcircled{2}} \quad \text{and} \quad B_2 = \textcircled{1} - \overset{\kappa_3}{\textcircled{2}}.$$

The pullback map

$$RH^4(\overline{\mathcal{M}}_3) \to RH^4(\overline{\mathcal{M}}_{2,2}) \oplus RH^3(\overline{\mathcal{M}}_{2,1}) \otimes RH^1(\overline{\mathcal{M}}_{1,1})$$

is injective (which implies that the map from $RH^4(\overline{\mathcal{M}}_3)$ to the boundary is injective). The restriction map

$$RH^4(\overline{\mathcal{M}}_3) \to RH^4(\overline{\mathcal{M}}_{2,2})$$

has 2–dimensional kernel and is surjective onto the S_2–invariant part of $RH^4(\overline{\mathcal{M}}_{2,2})$. Hence, the condition

$$\phi_{\mathrm{irr}}^*\Theta_3 = \Theta_{2,2} = \psi_1\psi_2\pi^*\Theta_2$$

determines Θ_3 up to parameters $s,t\in\mathbb{Q}$:

$$\begin{aligned}
a_{1111} &= s,\\
a_{112} &= \tfrac{11}{10}\lambda+\tfrac{17}{15}\lambda^2-18s+\tfrac{4}{3}t,\\
a_{13} &= -12\lambda-12\lambda^2+104s-13t,\\
a_{22} &= -\tfrac{33}{10}\lambda-\tfrac{29}{10}\lambda^2+27s-5t,\\
a_4 &= \tfrac{376}{5}\lambda+\tfrac{1933}{30}\lambda^2-426s+\tfrac{250}{3}t,\\
b_1 &= t,\\
b_2 &= \tfrac{2}{5}\lambda(3-\lambda).
\end{aligned}$$

The pullback map

$$RH^4(\overline{\mathcal{M}}_3)\to RH^3(\overline{\mathcal{M}}_{2,1})\otimes RH^1(\overline{\mathcal{M}}_{1,1})$$

has 3–dimensional image, and the condition

$$\phi_\Gamma^*\Theta_3=\Theta_{2,1}\otimes\Theta_{1,1}=(\psi_1\pi^*\Theta_2)\otimes(\lambda\psi_1)$$

is a linear system which cannot be satisfied for a general choice of the two parameters s and t defining Θ_3 for general λ, forcing λ to satisfy a polynomial relation. We find that

$$\begin{aligned}
a_{1111} &= \tfrac{5}{24}\lambda^3-\tfrac{19}{120}\lambda^2-\tfrac{11}{40}\lambda,\\
a_{112} &= \tfrac{5}{4}\lambda^3-\tfrac{147}{20}\lambda^2-\tfrac{99}{20}\lambda,\\
a_{13} &= \tfrac{403}{24}\lambda^3-\tfrac{209}{12}\lambda^2-\tfrac{239}{8}\lambda=a_{13}-\tfrac{3108}{53}b_1,\\
a_{22} &= -\tfrac{3867}{212}\lambda^3+\tfrac{99\,471}{2120}\lambda^2+\tfrac{22\,143}{530}\lambda=a_{22}+12b_1,\\
a_4 &= -\tfrac{115}{2}\lambda^3+\tfrac{1221}{20}\lambda^2+\tfrac{618}{5}\lambda,\\
b_1 &= \tfrac{1}{40}\lambda(\lambda-3)(15\lambda+11),\\
b_2 &= \tfrac{2}{5}\lambda(3-\lambda).
\end{aligned}$$

The expressions for a_{13} and a_{22} are consistent only when $b_1=0$; hence,

$$\lambda(\lambda-3)(15\lambda+11)=0.$$

***g* = 4** In genus 4, $H^{2*}(\overline{\mathcal{M}}_4,\mathbb{Q})=RH^*(\overline{\mathcal{M}}_4)$ is due to the calculation by Bergström and Tommasi [4] of the Hodge polynomial of $\overline{\mathcal{M}}_4$ together with the calculation of the tautological ring $RH^*(\overline{\mathcal{M}}_4)$ via Pixton's relations using `admcycles` [12]. We choose a general element $\Theta_4\in RH^6(\overline{\mathcal{M}}_4)$ which is a linear combination of basis vectors for the 32–dimensional space $RH^6(\overline{\mathcal{M}}_4)$. The pullback map of $RH^6(\overline{\mathcal{M}}_4)$ to the boundary can be shown to be injective using `admcycles`.

The main purpose of the $g=4$ calculation is to prove that $\lambda=-\frac{11}{15}$ is impossible, so we substitute $\lambda=-\frac{11}{15}$ into Θ_3 above to get

$$\Theta_3=\tfrac{2783}{81\,000}\kappa_1^4-\tfrac{11\,011}{13\,500}\kappa_1^2\kappa_2+\tfrac{59\,939}{10\,125}\kappa_1\kappa_3+\tfrac{16\,093}{9000}\kappa_2^2-\tfrac{474\,287}{13\,500}\kappa_4-\tfrac{1232}{1125}B_2.$$

As in the $g=3$ case above, we consider the pullback map

$$RH^6(\overline{\mathcal{M}}_4)\to RH^6(\overline{\mathcal{M}}_{3,2}),$$

which has a 6–dimensional kernel. The S_2–invariant part of $H^{12}(\overline{\mathcal{M}}_{3,2},\mathbb{Q})$ is proven in [3] to be 31–dimensional, and using `admcycles` it can be shown to be tautological. The condition $\phi_{\mathrm{irr}}^*\Theta_4=\Theta_{3,2}=\psi_1\psi_2\pi^*\Theta_3$ produces a system of 31 equations in 32 unknowns. Using `admcycles`, we find that $\Theta_{3,2}$ lies in the image of the pullback map, and constrains Θ_4 to depend linearly on six parameters. The pullback map composed with projection

$$RH^6(\overline{\mathcal{M}}_4)\to RH^5(\overline{\mathcal{M}}_{3,1})\otimes RH^1(\overline{\mathcal{M}}_{1,1})$$

uniquely determines the six parameters, and finally the resulting class Θ_4 is shown under the pullback map composed with projection

$$RH^6(\overline{\mathcal{M}}_4)\to RH^3(\overline{\mathcal{M}}_{2,1})\otimes RH^3(\overline{\mathcal{M}}_{2,1})$$

to disagree with $\Theta_{2,1}\otimes\Theta_{2,1}$. We conclude that $\lambda=-\frac{11}{15}$ is impossible, leaving $\lambda=3$. □

4 Cohomological field theories

The class $\Theta_{g,n}$ combines with known enumerative invariants, such as Gromov–Witten invariants, to give rise to new invariants. More generally, $\Theta_{g,n}$ pairs with any cohomological field theory, which is fundamentally related to the moduli space of curves $\overline{\mathcal{M}}_{g,n}$, retaining many of the properties of the cohomological field theory, and is in particular often calculable.

A *cohomological field theory* is a pair (H,η) composed of a finite-dimensional complex vector space H equipped with a symmetric, bilinear, nondegenerate form, or metric, η, and a sequence of S_n–equivariant maps. Many CohFTs are naturally defined on H defined over $\mathbb{Q}$; nevertheless, we use $\mathbb{C}$ in order to relate them to Frobenius manifolds, and to use normalised canonical coordinates, defined later,

$$\Omega_{g,n}\colon H^{\otimes n}\to H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C})$$

that satisfy compatibility conditions, from the inclusions of strata

$$\phi_{\mathrm{irr}}\colon \overline{\mathcal{M}}_{g-1,n+2} \to \overline{\mathcal{M}}_{g,n},$$
$$\phi_{h,I}\colon \overline{\mathcal{M}}_{h,|I|+1} \times \overline{\mathcal{M}}_{g-h,|J|+1} \to \overline{\mathcal{M}}_{g,n}, \quad I \sqcup J = \{1,\dots,n\},$$

given by

$$\phi_{\mathrm{irr}}^*\Omega_{g,n}(v_1 \otimes \cdots \otimes v_n) = \Omega_{g-1,n+2}(v_1 \otimes \cdots \otimes v_n \otimes \Delta), \tag{10}$$

$$\phi_{h,I}^*\Omega_{g,n}(v_1 \otimes \cdots \otimes v_n) = \Omega_{h,|I|+1} \otimes \Omega_{g-h,|J|+1}\Big(\bigotimes_{i\in I} v_i \otimes \Delta \otimes \bigotimes_{j\in J} v_j\Big), \tag{11}$$

where $\Delta \in H \otimes H$ is dual to $\eta \in H^* \otimes H^*$. When $n=0$, $\Omega_g := \Omega_{g,0} \in H^*(\overline{\mathcal{M}}_g,\mathbb{C})$. There exists a unit vector $\mathbb{1} \in H$ which satisfies

$$\Omega_{0,3}(\mathbb{1} \otimes v_1 \otimes v_2) = \eta(v_1,v_2).$$

The CohFT has *flat unit* if $\mathbb{1} \in H$ is compatible with the forgetful map $\pi\colon \overline{\mathcal{M}}_{g,n+1} \to \overline{\mathcal{M}}_{g,n}$ by

$$\Omega_{g,n+1}(\mathbb{1} \otimes v_1 \otimes \cdots \otimes v_n) = \pi^*\Omega_{g,n}(v_1 \otimes \cdots \otimes v_n) \tag{12}$$

for $2g-2+n>0$.

For a 1–dimensional CohFT, ie $\dim H = 1$, identify $\Omega_{g,n}$ with the image $\Omega_{g,n}(\mathbb{1}^{\otimes n})$, so we write $\Omega_{g,n} \in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C})$. A trivial example of a CohFT is $\Omega_{g,n} = 1 \in H^0(\overline{\mathcal{M}}_{g,n},\mathbb{C})$, which is a topological field theory, as we now describe.

A 2–dimensional topological field theory (TFT) is a vector space H and a sequence of symmetric linear maps

$$\Omega^0_{g,n}\colon H^{\otimes n} \to \mathbb{C}$$

for integers $g \ge 0$ and $n>0$ satisfying the following conditions. The map $\Omega^0_{0,2} = \eta$ defines a symmetric, bilinear, nondegenerate form η, and together with $\Omega^0_{0,3}$ it defines a product $\bullet$ on H via

$$\eta(v_1 \bullet v_2, v_3) = \Omega^0_{0,3}(v_1,v_2,v_3) \tag{13}$$

with identity element $\mathbb{1}$ given by the dual of $\Omega^0_{0,1} = \mathbb{1}^* = \eta(\mathbb{1},\cdot)$. It satisfies

$$\Omega^0_{g,n+1}(\mathbb{1} \otimes v_1 \otimes \cdots \otimes v_n) = \Omega^0_{g,n}(v_1 \otimes \cdots \otimes v_n)$$

and the gluing conditions

$$\Omega^0_{g,n}(v_1 \otimes \cdots \otimes v_n) = \Omega^0_{g-1,n+2}(v_1 \otimes \cdots \otimes v_n \otimes \Delta),$$
$$\Omega^0_{g,n}(v_1 \otimes \cdots \otimes v_n) = \Omega^0_{g_1,|I|+1} \otimes \Omega^0_{g_2,|J|+1}\Big(\bigotimes_{i\in I} v_i \otimes \Delta \otimes \bigotimes_{j\in J} v_j\Big)$$

for $g = g_1 + g_2$ and $I \sqcup J = \{1,\dots,n\}$.

Consider the natural isomorphism $H^0(\overline{\mathcal{M}}_{g,n},\mathbb{C})\cong\mathbb{C}$. The degree zero part of a CohFT $\Omega_{g,n}$ is a TFT

$$\Omega^0_{g,n}\colon H^{\otimes n}\xrightarrow{\Omega_{g,n}} H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C})\to H^0(\overline{\mathcal{M}}_{g,n},\mathbb{C}).$$

We often write $\Omega_{0,3}=\Omega^0_{0,3}$ interchangeably. Associated to $\Omega_{g,n}$ is the product (13) built from η and $\Omega_{0,3}$.

Remark 4.1 The classes $\Theta_{g,n}$ satisfy properties (10) and (11) of a 1–dimensional CohFT. In place of property (12), they satisfy

$$\Theta_{g,n+1}(\mathbb{1}\otimes v_1\otimes\cdots\otimes v_n)=\psi_{n+1}\cdot\pi^*\Theta_{g,n}(v_1\otimes\cdots\otimes v_n)$$

and $\Theta_{0,3}=0$.

The product defined in (13) is *semisimple* if it is diagonal $H\cong\mathbb{C}\oplus\mathbb{C}\oplus\cdots\oplus\mathbb{C}$, ie there is a canonical basis $\{u_1,\dots,u_N\}\subset H$ such that $u_i\cdot u_j=\delta_{ij}u_i$. The metric is then necessarily diagonal with respect to the same basis, $\eta(u_i,u_j)=\delta_{ij}\eta_i$ for some $\eta_i\in\mathbb{C}\setminus\{0\}$ for $i=1,\dots,N$. The Givental–Teleman theorem described in Section 5 gives a construction of semisimple CohFTs.

4.1 Cohomological field theories coupled to $\Theta_{g,n}$

Definition 4.2 For any CohFT Ω defined on (H,η), define $\Omega^\Theta=\{\Omega^\Theta_{g,n}\}$ to be the sequence of S_n–equivariant maps $\Omega^\Theta_{g,n}\colon H^{\otimes n}\to H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C})$ given by

$$\Omega^\Theta_{g,n}(v_1\otimes\cdots\otimes v_n):=\Theta_{g,n}\cdot\Omega_{g,n}(v_1\otimes\cdots\otimes v_n).$$

This is essentially the tensor product of CohFTs, albeit involving $\Theta_{g,n}$. The tensor products of CohFTs is obtained as above by cup product on $H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C})$, generalising Gromov–Witten invariants of target products and the Künneth formula $H^*(X_1\times X_2)\cong H^*X_1\otimes H^*X_2$.

Generalising Remark 4.1, $\Omega^\Theta_{g,n}$ satisfies properties (10) and (11) of a CohFT on (H,η). In place of property (12), it satisfies

$$\Omega^\Theta_{g,n+1}(\mathbb{1}\otimes v_1\otimes\cdots\otimes v_n)=\psi_{n+1}\cdot\pi^*\Omega^\Theta_{g,n}(v_1\otimes\cdots\otimes v_n)$$

and $\Omega^\Theta_{0,3}=0$.

Given a CohFT $\Omega=\{\Omega_{g,n}\}$, or a more general collection of classes such as $\Omega=\{\Omega^{\Theta}_{g,n}\}$, and a basis $\{e_1,\dots,e_N\}$ of H, the partition function of Ω is defined by

$$Z_\Omega(\hbar,\{t^\alpha_k\})=\exp\sum_{g,n,\vec{k}}\frac{\hbar^{g-1}}{n!}\int_{\overline{\mathcal{M}}_{g,n}}\Omega_{g,n}(e_{\alpha_1}\otimes\cdots\otimes e_{\alpha_n})\cdot\prod_{j=1}^n\psi_j^{k_j}\prod t^{\alpha_j}_{k_j}\tag{14}$$

for $\alpha_i\in\{1,\dots,N\}$ and $k_j\in\mathbb{N}$. For $\dim H=1$ and $\Omega_{g,n}=1\in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C})$, its partition function is $Z_\Omega(\hbar,\{t_k\})=Z^{\mathrm{KW}}(\hbar,\{t_k\})$, which is defined in Section 5.1.

For $\Omega_{g,n}=\Theta_{g,n}\in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C})$, $Z_\Omega(\hbar,\{t_k\})=Z^\Theta(\hbar,\{t_k\})$ gives its partition function. Property (iii) is realised by the homogeneity property

$$\frac{\partial}{\partial t_0}Z^\Theta(\hbar,t_0,t_1,\dots)=\sum_{i=0}^\infty(2i+1)t_i\frac{\partial}{\partial t_i}Z^\Theta(\hbar,t_0,t_1,\dots)+\tfrac18 Z^\Theta(\hbar,t_0,t_1,\dots),\tag{15}$$

proven in the following proposition:

Proposition 4.3 *The function $Z^\Theta(\hbar,t_0,t_1,\dots)$ is homogeneous of degree $-\frac18$ with respect to $\{q=1-t_0,t_1,t_2,\dots\}$ with $\deg q=1$ and $\deg t_i=2i+1$ for $i>0$. Equivalently, it satisfies the dilaton equation* (15).

Proof We have

$$\begin{aligned}\int_{\overline{\mathcal{M}}_{g,n+1}}\Theta_{g,n+1}\cdot\prod_{j=1}^n\psi_j^{k_j}&=\int_{\overline{\mathcal{M}}_{g,n+1}}\pi^*\Theta_{g,n}\cdot\psi_{n+1}\cdot\prod_{j=1}^n\psi_j^{k_j}\\&=\int_{\overline{\mathcal{M}}_{g,n+1}}\pi^*\Theta_{g,n}\cdot\psi_{n+1}\cdot\prod_{j=1}^n\pi^*\psi_j^{k_j}\\&=\int_{\overline{\mathcal{M}}_{g,n}}\Theta_{g,n}\cdot\prod_{j=1}^n\psi_j^{k_j}\cdot\pi_*\psi_{n+1}\\&=(2g-2+n)\int_{\overline{\mathcal{M}}_{g,n}}\Theta_{g,n}\cdot\prod_{j=1}^n\psi_j^{k_j},\end{aligned}$$

which uses $\psi_{n+1}\cdot\psi_j=\psi_{n+1}\cdot\pi^*\psi_j$ for $j=1,\dots,n$ and $\pi_*(\pi^*\omega\cdot\psi_{n+1})=\omega\cdot\pi_*\psi_{n+1}$. In terms of the partition function $Z^\Theta(\hbar,t_0,t_1,\dots)$, this is realised by (15). □

4.1.1 Gromov–Witten invariants Let X be a projective algebraic variety and consider $(C,x_1,\dots,x_n)$ a connected smooth curve of genus g with n distinct marked points. For $\beta\in H_2(X,\mathbb{Z})$, the moduli space of stable maps $\overline{\mathcal{M}}_{g,n}(X,\beta)$ is defined by

$$\overline{\mathcal{M}}_{g,n}(X,\beta)=\{(C,x_1,\dots,x_n)\xrightarrow{\pi}X\mid\pi_*[C]=\beta\}/\sim,$$

where π is a morphism from a connected nodal curve C containing distinct points $\{x_1,\dots,x_n\}$ that avoid the nodes. Any genus zero irreducible component of C with fewer than three distinguished points (nodal or marked), or genus one irreducible component of C with no distinguished point, must not be collapsed to a point. We quotient by isomorphisms of the domain C that fix each x_i. The moduli space of stable maps has irreducible components of different dimensions but it has a virtual class of dimension

$$\dim[\overline{\mathcal{M}}_{g,n}(X,\beta)]^{\mathrm{virt}} = (\dim X-3)(1-g)+\langle c_1(X),\beta\rangle+n. \tag{16}$$

For $i=1,\dots,n$, there exist evaluation maps

$$\mathrm{ev}_i\colon \overline{\mathcal{M}}_{g,n}(X,\beta)\to X, \quad \mathrm{ev}_i(\pi)=\pi(x_i), \tag{17}$$

and classes $\gamma\in H^*(X,\mathbb{Z})$ pull back to classes in $H^*(\overline{\mathcal{M}}_{g,n}(X,\beta),\mathbb{C})$ via

$$\mathrm{ev}_i^*\colon H^*(X,\mathbb{Z})\to H^*(\overline{\mathcal{M}}_{g,n}(X,\beta),\mathbb{C}). \tag{18}$$

The forgetful map $p\colon \overline{\mathcal{M}}_{g,n}(X,\beta)\to\overline{\mathcal{M}}_{g,n}$ maps a stable map to its domain curve followed by contraction of unstable components. The pushforward map p_* on cohomology defines a CohFT Ω_X on the even part of the cohomology $H=H^{\mathrm{even}}(X,\mathbb{C})$ (and a generalisation of a CohFT on $H^*(X,\mathbb{C})$) equipped with the symmetric, bilinear, nondegenerate form

$$\eta(\alpha,\beta)=\int_X\alpha\wedge\beta.$$

We have $(\Omega_X)_{g,n}\colon H^{\mathrm{even}}(X,\mathbb{C})^{\otimes n}\to H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C})$ defined by

$$(\Omega_X)_{g,n}(\alpha_1,\dots,\alpha_n)=\sum_\beta p_*\biggl(\prod_{i=1}^n\mathrm{ev}_i^*(\alpha_i)\cap[\overline{\mathcal{M}}_{g,n}(X,\beta)]^{\mathrm{virt}}\biggr)\in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C}).$$

Note that it is the dependence of $p=p(g,n,\beta)$ on β (which is suppressed) that allows $(\Omega_X)_{g,n}(\alpha_1,\dots,\alpha_n)$ to be composed of different-degree terms. The partition function of the CohFT Ω_X with respect to a chosen basis e_α of $H^{\mathrm{even}}(X;\mathbb{C})$ is

$$Z_{\Omega_X}(\hbar,\{t_k^\alpha\}) = \exp\sum_{\substack{g,n,\vec{k}\\ \vec{\alpha},\beta}}\frac{\hbar^{g-1}}{n!}\int_{\overline{\mathcal{M}}_{g,n}}p_*\biggl(\prod_{i=1}^n\mathrm{ev}_i^*(e_{\alpha_i})\cap[\overline{\mathcal{M}}_{g,n}(X,\beta)]^{\mathrm{virt}}\biggr)\prod_{j=1}^n\psi_j^{k_j}\prod t_{k_j}^{\alpha_j}.$$

It stores *ancestor* invariants. These are different from *descendant* invariants, which use, in place of $\psi_j=c_1(L_j)$, $\Psi_j=c_1(\mathcal{L}_j)$ for line bundles $\mathcal{L}_j\to\overline{\mathcal{M}}_{g,n}(X,\beta)$ defined similarly as the cotangent bundle over the i^{th} marked point on the domain curve.

Following Definition 4.2, we define Ω_X^Θ by

$$(\Omega_X^\Theta)_{g,n}(\alpha_1,\dots,\alpha_n)=\Theta_{g,n}\cdot\sum_\beta p_*\biggl(\prod_{i=1}^n \mathrm{ev}_i^*(\alpha_i)\biggr)\in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{C})$$

and

$$Z^\Theta_{\Omega_X}(\hbar,\{t_k^\alpha\})=\exp\sum_{\substack{g,n,\vec{k}\\ \vec{\alpha},\beta}}\frac{\hbar^{g-1}}{n!}\int_{\overline{\mathcal{M}}_{g,n}}\Theta_{g,n}\cdot p_*\biggl(\prod_{i=1}^n \mathrm{ev}_i^*(e_{\alpha_i})\biggr)\cdot\prod_{j=1}^n\psi_j^{k_j}\prod t_{k_j}^{\alpha_j}.$$

Let $\Theta_{g,n}^{\mathrm{PD}}\subset A_{g-1}(\overline{\mathcal{M}}_{g,n},\mathbb{C})$ be the $(g-1)$–dimensional Chow class given by the pushforward of the top Chern class of the bundle $E_{g,n}$ defined in Definition 2.1. The virtual dimension of the pullback of $\Theta_{g,n}^{\mathrm{PD}}$ is

$$\dim\{[\overline{\mathcal{M}}_{g,n}(X,\boldsymbol{d})]^{\mathrm{virt}}\cap p^{-1}(\Theta_{g,n}^{\mathrm{PD}})\}=(\dim X-1)(1-g)+\langle c_1(X),\beta\rangle. \tag{19}$$

Comparing the dimension formulas (16) and (19), we see that elliptic curves now take the place of Calabi–Yau 3–folds to give virtual dimension zero moduli spaces, independent of genus and degree. The invariants of a target curve X are trivial when the genus of X is greater than 1 and computable when $X=\mathbb{P}^1$ [44], producing results analogous to the usual Gromov–Witten invariants in [46]. For $c_1(X)=0$ and $\dim X>1$, the invariants vanish for $g>1$, while for $g=1$ it seems to predict an invariant associated to maps of elliptic curves to X.

4.1.2 Weil–Petersson volumes A fundamental example of a 1–dimensional CohFT is given by

$$\Omega_{g,n}=\exp(2\pi^2\kappa_1)\in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{R}).$$

Its partition function stores Weil–Petersson volumes

$$V_{g,n}=\frac{(2\pi^2)^{3g-3+n}}{(3g-3+n)!}\int_{\overline{\mathcal{M}}_{g,n}}\kappa_1^{3g-3+n}$$

and deformed Weil–Petersson volumes studied by Mirzakhani [39]. Weil–Petersson volumes of the subvariety of $\overline{\mathcal{M}}_{g,n}$ dual to $\Theta_{g,n}$ make sense even before we find such a subvariety. They are given by

$$V^\Theta_{g,n}=\frac{(2\pi^2)^{g-1}}{(g-1)!}\int_{\overline{\mathcal{M}}_{g,n}}\Theta_{g,n}\cdot\kappa_1^{g-1},$$

which are calculable since they are given by a translation of Z^{BGW}. If we include ψ classes, we get polynomials $V^\Theta_{g,n}(L_1,\dots,L_n)$ which give the deformed volumes analogous to Mirzakhani's volumes. In [43; 51], the polynomials $V^{\Theta}_{g,n}(L_1, \dots, L_n)$ are related to the volume of the moduli space of super-Riemann surfaces.

4.1.3 ELSV formula Another example of a 1–dimensional CohFT is given by

$$\Omega_{g,n} = c(E^\vee) = 1 - \lambda_1 + \cdots + (-1)^g \lambda_g \in H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{C}),$$

where $\lambda_i = c_i(E)$ is the i^{th} Chern class of the Hodge bundle $E \to \overline{\mathcal{M}}_{g,n}$ defined to have fibres $H^0(C, \omega_C)$ over a nodal curve C.

Hurwitz [31] studied the problem of connected curves Σ of genus g covering $\mathbb{P}^1$, branched over $r+1$ fixed points $\{p_1, p_2, \dots, p_r, p_{r+1}\}$ with arbitrary profile $\mu = (\mu_1, \dots, \mu_n)$ over p_{r+1}. Over the other r branch points, one specifies simple ramification, ie the partition $(2, 1, 1, \dots)$. The Riemann–Hurwitz formula determines the number r of simple branch points via $2 - 2g - n = |\mu| - r$.

Definition 4.4 Define the simple Hurwitz number $H_{g,\mu}$ to be the weighted count of genus g connected covers of $\mathbb{P}^1$ with ramification $\mu = (\mu_1, \dots, \mu_n)$ over ∞ and simple ramification elsewhere. Each cover π is counted with weight $1/|\mathrm{Aut}(\pi)|$.

Coefficients of the partition function of the CohFT $\Omega_{g,n} = c(E^\vee)$ appear naturally in the ELSV formula [20], which relates the Hurwitz numbers $H_{g,\mu}$ to the Hodge classes. The ELSV formula is

$$H_{g,\mu} = \frac{r(g,\mu)!}{|\mathrm{Aut}\,\mu|} \prod_{i=1}^{n} \frac{\mu_i^{\mu_i}}{\mu_i!} \int_{\overline{\mathcal{M}}_{g,n}} \frac{1 - \lambda_1 + \cdots + (-1)^g \lambda_g}{(1 - \mu_1 \psi_1) \cdots (1 - \mu_n \psi_n)},$$

where $\mu = (\mu_1, \dots, \mu_n)$ and $r(g,\mu) = 2g - 2 + n + |\mu|$.

Using $\Omega^{\Theta}_{g,n} = \Theta \cdot c(E^\vee)$, we can define an analogue of the ELSV formula,

$$H^{\Theta}_{g,\mu} = \frac{(2g - 2 + n + |\mu|)!}{|\mathrm{Aut}\,\mu|} \prod_{i=1}^{n} \frac{\mu_i^{\mu_i}}{\mu_i!} \int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \cdot \frac{1 - \lambda_1 + \cdots + (-1)^{g-1} \lambda_{g-1}}{(1 - \mu_1 \psi_1) \cdots (1 - \mu_n \psi_n)}.$$

It may be that $H^{\Theta}_{g,\mu}$ has an interpretation of enumerating a new type of Hurwitz covers. Note that it makes sense to set all $\mu_i = 0$, and, in particular, there are nontrivial primary invariants over $\overline{\mathcal{M}}_g$, unlike for simple Hurwitz numbers. An example calculation:

$$\int_{\overline{\mathcal{M}}_2} \Theta_2 \lambda_1 = \tfrac{1}{5} \cdot \tfrac{1}{8} \cdot \tfrac{1}{8} \cdot \tfrac{1}{2} + \tfrac{1}{10} \cdot \tfrac{1}{8} \cdot \tfrac{1}{2} = \tfrac{1}{128} \Longleftarrow \lambda_1 = \tfrac{1}{10}(2\delta_{1,1} + \delta_{\mathrm{irr}}).$$

4.1.4 The versal deformation space of the A_2 singularity The A_2 singularity has a 2–dimensional versal deformation space $M \cong \mathbb{C}^2 = \{(t_1, t_2)\}$ parametrising the family

$$W_t(z) = z^3 - t_2 z + t_1$$

that admits a semisimple Frobenius manifold structure. Dubrovin [15] associated a family of linear systems, defined in (20) below, depending on the canonical coordinates $(u_1, \dots, u_N)$ of any semisimple Frobenius manifold M. This produces a CohFT Ω^{A_2} defined on $\mathbb{C}^2$ from the A_2 singularity using Definition 5.2. More generally, to any point of a Frobenius manifold one can associate a cohomological field theory and, conversely, the genus zero part of a cohomological field theory defines a Frobenius manifold [15].

Recall that a Frobenius manifold is a complex manifold M equipped with an associative product on its tangent bundle compatible with a flat metric — a nondegenerate symmetric bilinear form — on the manifold. It is encoded by a single function $F(t_1, \dots, t_N)$, known as the *prepotential*, which satisfies a nonlinear partial differential equation, known as the Witten–Dijkgraaf–Verlinde–Verlinde equation,

$$F_{ijm}\eta^{mn}F_{k\ell n} = F_{i\ell m}\eta^{mn}F_{jkn}, \quad \eta_{ij} = F_{1ij}$$

where $\eta^{ik}\eta_{kj} = \delta_{ij}$, $F_i = \partial/\partial t_i F$, $\partial/\partial t_1 = \mathbb{1}$ corresponds to the flat unit vector field for the product, and $\{t_1, \dots, t_N\}$ are (flat) local coordinates on M. The Frobenius manifold is *conformal* if it comes equipped with an Euler vector field E which describes symmetries of the Frobenius manifold, neatly encoded by

$$E \cdot F(t_1, \dots, t_N) = c \cdot F(t_1, \dots, t_N) + \text{quadratic polynomial}, \quad c \in \mathbb{C}.$$

For a semisimple conformal Frobenius manifold, multiplication by the Euler vector field E produces an endomorphism U with eigenvalues $\{u_1, \dots, u_N\}$ known as *canonical coordinates* on M. They give rise to vector fields $\partial/\partial u_i$ with respect to which the metric η, product $\bullet$ and Euler vector field E are diagonal:

$$\frac{\partial}{\partial u_i} \bullet \frac{\partial}{\partial u_j} = \delta_{ij}\frac{\partial}{\partial u_i}, \quad \eta\Big(\frac{\partial}{\partial u_i}, \frac{\partial}{\partial u_j}\Big) = \delta_{ij}\Delta_i, \quad E = \sum u_i \frac{\partial}{\partial u_i}.$$

At any point of the Frobenius manifold, the endomorphism U, defined by multiplication by the Euler vector field E, and the endomorphism $V = [\Gamma, U]$, where $\Gamma_{ij} = \partial_{u_i}\Delta_j / 2\sqrt{\Delta_i \Delta_j}$ for $i \neq j$ are the so-called rotation coefficients of the metric η in the normalised canonical basis, produce the differential equation

$$\Big(\frac{d}{dz} - U - \frac{V}{z}\Big)Y = 0. \tag{20}$$

Choose a solution of (20) of the form $Y = R(z^{-1})e^{zU}$ and substitute $z \mapsto z^{-1}$ to get

$$0 = \left(\frac{d}{dz} + \frac{U}{z^2} + \frac{V}{z}\right)R(z)e^{U/z} = \left(\frac{d}{dz}R(z) + \frac{1}{z^2}[U, R(z)] + \frac{1}{z}VR(z)\right)e^{U/z}.$$

This associates an element $R(z) = \sum R_k z^k$ to each point of the Frobenius manifold. Teleman [52] defined the endomorphisms R_k of $H = T_pM$ recursively from $R_0 = I$ by

$$[R_{k+1}, U] = (k + V)R_k, \quad k = 0, 1, \ldots. \tag{21}$$

It is useful to consider three natural bases of the tangent space $H = T_pM \cong \mathbb{C}^N$ at any point p of a semisimple Frobenius manifold: the flat basis $\{\partial/\partial t_i\}$, which gives a constant metric η; the canonical basis $\{\partial/\partial u_i\}$, which gives a trivial product $\bullet$; and the normalised canonical basis $\{v_i\}$ for $v_i = \Delta_i^{-1/2}\partial/\partial u_i$, which gives a trivial metric η. (A different choice of square root of Δ_i would simply give a different choice of normalised canonical basis.) The transition matrix Ψ from flat coordinates to normalised canonical coordinates sends the metric η to the dot product, ie $\Psi^T\Psi = \eta$. The topological field theory structure on H induced from η and $\bullet$ is diagonal in the normalised canonical basis. It is given by

$$\Omega_{g,n}(v_i^{\otimes n}) = \Delta_i^{1-g-1/2n}$$

and vanishes on mixed products of v_i and v_j for $i \neq j$. In the normalised canonical basis, the unit vector is given by

$$\mathbb{1} = (\Delta_1^{1/2}, \ldots, \Delta_N^{1/2});$$

hence, it uniquely determines the topological field theory. We find the normalised canonical basis most useful for comparisons with topological recursion; see Section 5.2.1.

The Frobenius manifold structure on the versal deformation space M of the A_2 singularity was constructed in [15; 48]. The product on tangent spaces of the family $W_t(z) = z^3 - t_2z + t_1$ is induced from the isomorphism

$$T_tM \cong \mathbb{C}[z]/W_t'(z)$$

given by $\partial/\partial t_k \mapsto \partial W_t/\partial t_k = (-z)^{k-1}$, producing

$$\frac{\partial}{\partial t_1}\bullet\frac{\partial}{\partial t_1} = \frac{\partial}{\partial t_1}, \quad \frac{\partial}{\partial t_1}\bullet\frac{\partial}{\partial t_2} = \frac{\partial}{\partial t_2}, \quad \frac{\partial}{\partial t_2}\bullet\frac{\partial}{\partial t_2} = \tfrac{1}{3}t_2\frac{\partial}{\partial t_1}.$$

The metric is given by

$$\eta(p(z), q(z)) = -3\operatorname{Res}_{\infty}\frac{p(z)q(z)\,dz}{W_t'(z)}.$$

With respect to the basis $\{\partial/\partial t_1, \partial/\partial t_2\}$, it is constant and hence flat:

$$\eta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The Frobenius manifold structure on M is conformal. The unit and Euler vector fields are $\mathbb{1} = \partial/\partial t_1$ and $E = t_1\,\partial/\partial t_1 + \frac{2}{3}t_2\,\partial/\partial t_2$, which correspond respectively to the images of 1 and $W_t(z)$ in $\mathbb{C}[z]/W_t'(z)$.

The prepotential is produced via $\eta_{ij} = F_{1ij}$ and $\eta(\partial/\partial t_i \bullet \partial/\partial t_j, \partial/\partial t_k) = F_{ijk}$,

$$F(t_1, t_2) = \tfrac{1}{2}t_1^2 t_2 + \tfrac{1}{72}t_2^4,$$

and satisfies $E \cdot F(t_1, t_2) = \frac{8}{3}F(t_1, t_2)$. The canonical coordinates are

$$u_1 = t_1 + \frac{2}{3\sqrt{3}}t_2^{3/2}, \quad u_2 = t_1 - \frac{2}{3\sqrt{3}}t_2^{3/2}.$$

In the normalised canonical basis, the rotation coefficients $\Gamma_{12} = -i\frac{\sqrt{3}}{8}t_2^{-3/2} = \Gamma_{21}$ give rise to $V = [\Gamma, U] = i\frac{\sqrt{3}}{2}t_2^{-3/2}\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. In canonical coordinates we have

$$U = \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix}, \quad V = \frac{2i}{3(u_1 - u_2)}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{22}$$

The metric η applied to the vector fields $\partial/\partial u_i = \frac{1}{2}(\partial/\partial t_1 - (-1)^i(3/t_2)^{1/2}\,\partial/\partial t_2)$ is $\eta(\partial/\partial u_i, \partial/\partial u_j) = \delta_{ij}\Delta_i$, where $\Delta_1 = \frac{\sqrt{3}}{2}t_2^{-1/2} = -\Delta_2$. Restrict to the point of M with coordinates $(u_1, u_2) = (2, -2)$ or, equivalently, $(t_1, t_2) = (0, 3)$. Then $\Delta_1 = \frac{1}{2} = -\Delta_2$ determines the TFT and

$$U = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}, \quad V = \frac{1}{6}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

determines $R(z) \in L^{(2)}\mathrm{GL}(2, \mathbb{C})$ and $T(z) \in z^2\mathbb{C}^2[\![z]\!]$ via (21) to get

$$\begin{aligned} R(z) &= \sum_m \frac{(6m)!}{(6m-1)(3m)!(2m)!}\begin{pmatrix} -1 & (-1)^m 6mi \\ -6mi & (-1)^{m-1} \end{pmatrix}\left(\tfrac{1}{1728}z\right)^m, \\ T(z) &= z(\mathbb{1} - R^{-1}(z)(\mathbb{1})), \quad \text{where } \mathbb{1} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}. \end{aligned} \tag{23}$$

The triple $(R(z), T(z), \mathbb{1}) \in L^{(2)}\mathrm{GL}(N, \mathbb{C}) \times z^2\mathbb{C}^N[\![z]\!] \times \mathbb{C}^N$ in (23) produces the cohomological field theory Ω^{A_2} associated to the A_2 singularity at the point $(t_1, t_2) = (0, 3)$ via Definition 5.2 in the next section.

Remark 4.5 The matrix $R(z)$ defined in (23) — which uses the normalised canonical basis for H, so that η is the dot product — is related to the matrix $R(z)$ in [47] by conjugation by the transition matrix Ψ from flat coordinates to normalised canonical coordinates

$$R(z) = \Psi \sum_m \frac{(6m)!}{(3m)!(2m)!} \begin{pmatrix} (1+6m)/(1-6m) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^m \left(\tfrac{1}{1728} z\right)^m \Psi^{-1}$$

for

$$\Psi = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}.$$

5 Givental construction of cohomological field theories

Givental produced a construction of partition functions of cohomological field theories in [29]. He defined an action of the twisted loop group, and elements of $z^2\mathbb{C}^N[\![z]\!]$ known as translations, on partition functions of cohomological field theories and used this to build partition functions of semisimple cohomological field theories out of the basic building block $Z^{\mathrm{KW}}(\hbar, t_0, t_1, \dots)$ combined with the vector $\mathbb{1} \in \mathbb{C}^N$ which represents the topological field theory. This action was interpreted as an action on the actual cohomology classes in $H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{C})$, independently, by Katzarkov, Kontsevich and Pantev, and Kazarian and Teleman; see [47; 49].

The Givental action is defined on more general sequences of cohomology classes in $H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{C})$ such as the collection of classes $\Theta_{g,n}$ or $\Omega^\Theta_{g,n}$ defined from any CohFT $\Omega_{g,n}$ in Definition 4.2. If $\Omega_{g,n}$ is semisimple, the classes $\Omega^\Theta_{g,n}$ can be obtained by applying Givental's action to the collection $\Theta_{g,n}$.

5.0.1 The twisted loop group action The loop group $L\mathrm{GL}(N, \mathbb{C})$ is the group of formal series

$$R(z) = \sum_{k=0}^{\infty} R_k z^k,$$

where R_k are $N \times N$ matrices and $R_0 \in \mathrm{GL}(N, \mathbb{C})$. Define the *twisted loop group* $L^{(2)}\mathrm{GL}(N, \mathbb{C}) \subset L\mathrm{GL}(N, \mathbb{C})$ to be the subgroup of elements satisfying $R_0 = I$ and

$$R(z)R(-z)^T = I.$$

Elements of $L^{(2)}\mathrm{GL}(N, \mathbb{C})$ naturally arise out of solutions to the linear system (20) given by $(d/dz - U - V/z)Y = 0$, where $Y(z) \in \mathbb{C}^N$, $U = \mathrm{diag}(u_1, \dots, u_N)$ for u_i

distinct, and V is skew-symmetric. One can choose a solution of (20) which behaves asymptotically for $z \to \infty$ as

$$Y(z) = R(z^{-1})e^{zU}, \quad R(z) = I + R_1 z + R_2 z^2 + \cdots .$$

This defines a power series $R(z)$ with coefficients given by $N \times N$ matrices, which is easily shown to satisfy $R(z)R^T(-z) = I$; hence, $R(z) \in L^{(2)}\mathrm{GL}(N, \mathbb{C})$.

Givental [29] constructed an action on CohFTs using a triple

$$(R(z), T(z), \mathbb{1}) \in L^{(2)}\mathrm{GL}(N, \mathbb{C}) \times z^2\mathbb{C}^N[\![z]\!] \times \mathbb{C}^N$$

as follows. For a given stable graph Γ of genus g and with n external edges, we have

$$\phi_\Gamma \colon \overline{\mathcal{M}}_\Gamma = \prod_{v \in V(\Gamma)} \overline{\mathcal{M}}_{g(v),n(v)} \to \overline{\mathcal{M}}_{g,n}.$$

Given $(R(z), T(z), \mathbb{1}) \in L^{(2)}\mathrm{GL}(N, \mathbb{C}) \times z^2\mathbb{C}^N[\![z]\!] \times \mathbb{C}^N$, Givental's action is defined via weighted sums over stable graphs. For $R(z) \in L^{(2)}\mathrm{GL}(N, \mathbb{C})$, define

$$\mathcal{E}(z, w) = \frac{I - R^{-1}(z)R^{-1}(w)^T}{z + w} = \sum_{i,j \geq 0} \mathcal{E}_{ij} w^i z^j,$$

which has the power series expansion on the right since $R^{-1}(z)$ is also an element of the twisted loop group, so the numerator $I - R^{-1}(z)R^{-1}(w)^T$ vanishes at $w = -z$.

Definition 5.1 For a stable graph Γ denote by

$$V(\Gamma), \quad E(\Gamma), \quad H(\Gamma), \quad L(\Gamma) = L^*(\Gamma) \sqcup L^\bullet(\Gamma)$$

its sets of vertices, edges, half-edges and leaves. The disjoint splitting of $L(\Gamma)$ into ordinary leaves L^* and dilaton leaves $L^\bullet$ is part of the structure on Γ. The set of half-edges consists of leaves and oriented edges, so there is an injective map $L(\Gamma) \to H(\Gamma)$ and a multiply defined map $E(\Gamma) \to H(\Gamma)$ denoted by $E(\Gamma) \ni e \mapsto \{e^+, e^-\} \subset H(\Gamma)$. The map sending a half-edge to its vertex is given by $v \colon H(\Gamma) \to V(\Gamma)$. Decorate Γ by functions

$$\begin{aligned} &g \colon V(\Gamma) \to \mathbb{N}, \qquad \alpha \colon V(\Gamma) \to \{1, \dots, N\}, \\ &k \colon H(\Gamma) \to \mathbb{N} \qquad p \colon L^*(\Gamma) \xrightarrow{\cong} \{1, 2, \dots, n\}, \end{aligned}$$

such that $k|_{L^\bullet(\Gamma)} > 1$ and $n = |L^*(\Gamma)|$. We write $g_v = g(v)$, $\alpha_v = \alpha(v)$, $\alpha_\ell = \alpha(v(\ell))$, $p_\ell = p(\ell)$ and $k_\ell = k(\ell)$. The *genus* of Γ is $g(\Gamma) = b_1(\Gamma) + \sum_{v \in V(\Gamma)} g(v)$. We say Γ is *stable* if any vertex labelled by $g = 0$ is of valency ≥ 3 and there are no isolated vertices labelled by $g = 1$. We write n_v for the valency of the vertex v. Define $G_{g,n}$ to be the finite set of all stable, connected, genus g, decorated graphs with n ordinary leaves and at most $3g-3+n$ dilaton leaves.

Definition 5.2 [47; 49] Given a CohFT $\Omega' = \{\Omega'_{g,n}\}$ and

$$(R(z), T(z)) \in L^{(2)}\mathrm{GL}(N,\mathbb{C}) \times z^2\mathbb{C}^N[\![z]\!],$$

define $R \cdot T \cdot \Omega' = \Omega = \{\Omega_{g,n}\}$ by a weighted sum over stable graphs,

$$\Omega_{g,n} := \sum_{\Gamma \in G_{g,n}} \frac{1}{|\mathrm{Aut}(\Gamma)|} (\phi_\Gamma)_* \pi_* \prod_{v \in V(\Gamma)} w(v) \prod_{e \in E(\Gamma)} w(e) \prod_{\ell \in L(\Gamma)} w(\ell) \in H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{C}), \tag{24}$$

where π is the map that forgets dilaton leaves. Weights are defined as follows:

(i) **Vertex weight** $w(v) = \Omega'_{g(v),n_v}$ at each vertex v.

(ii) **Edge weight** $w(e) = \mathcal{E}(\psi'_e, \psi''_e)$ at each edge e.

(iii) **Leaf weight** $w(\ell) = \begin{cases} R^{-1}(\psi_{p(\ell)}) & \text{at each ordinary leaf } \ell \in L^*, \\ T(\psi_{p(\ell)}) & \text{at each dilaton leaf } \ell \in L^\bullet. \end{cases}$

We consider only the even part of $H^*(\overline{\mathcal{M}}_{g,n}, \mathbb{C})$, so (24) is independent of the order in which we take the product of cohomology classes. If $\{\Omega_{g,n}\}$ is a CohFT defined on $(\mathbb{C}, \eta)$ for $H \cong \mathbb{C}^N$, then the classes $\{\Omega_{g,n}\}$ in (24) satisfy the same restriction conditions and hence define a CohFT on $(\mathbb{C}, \eta)$ with the same degree zero, or topological field theory, terms as those of Ω'. If we choose $T(z) \equiv 0$, then the sum in (24), which is over stable graphs without dilaton leaves, defines the action of the twisted loop group on CohFTs. If we choose $R(z) \equiv I$, then (24) is a graphical realisation of the translation action of $T(z) \in z^2 H[\![z]\!]$ on a CohFT $\Omega'_{g,n}$ defined by

$$(T \cdot \Omega')_{g,n}(v_1 \otimes \cdots \otimes v_n) = \sum_{m \geq 0} \frac{1}{m!} \pi_* \Omega'_{g,n+m}(v_1 \otimes \cdots \otimes v_n \otimes T(\psi_{n+1}) \otimes \cdots \otimes T(\psi_{n+m})),$$

where $\pi \colon \overline{\mathcal{M}}_{g,n+m} \to \overline{\mathcal{M}}_{g,n}$ is the forgetful map. The sum over $m \in \mathbb{N}$ defining $(T \cdot \Omega')_{g,n}$ is finite since $T(z) \in z^2 H[\![z]\!]$, so $\dim \overline{\mathcal{M}}_{g,n+m} = 3g-3+n+m$ grows more slowly in m than the degree $2m$ coming from T, resulting in at most $3g-3+n$ terms. We can relax this condition and allow $T(z) \in zH[\![z]\!]$ if we control the growth of the degrees of all terms of $\Omega'_{g,n}$ in n to ensure $T(z)$ produces a finite sum. In particular, $\Theta_{g,n}$, and more generally $\Omega'^{\Theta}_{g,n}$ for any CohFT $\Omega'_{g,n}$, is annihilated by terms of degree $> g-1$; hence, the sum defining $(T\Omega')_{g,n}$ consists of at most $g-1$ terms when $T(z) \in zH[\![z]\!]$.

The tensor product $\Omega \mapsto \Omega^\Theta$ given in Definition 4.2 commutes with the action of R and commutes with the action of T up to rescaling. For a CohFT Ω, and $R(z) \in L^{(2)}\mathrm{GL}(N,\mathbb{C})$ and $T(z) \in z\mathbb{C}^N[\![z]\!]$,

$$(R\cdot\Omega)^\Theta = R\cdot\Omega^\Theta, \quad (zT)\cdot\Omega^\Theta = T\cdot\Omega^\Theta. \tag{25}$$

The first relation in (25) uses the restriction properties (ii) of $\Theta_{g,n}$ and the second of these uses the forgetful property (iii) of $\Theta_{g,n}$ to see

$$\begin{aligned}
\pi_*\Omega^\Theta_{g,n+m}\biggl(\bigotimes_{i=1}^n v_i \otimes \bigotimes_{i=1}^m T(\psi_{n+i})\biggr)
&= \pi_*\Omega_{g,n+m}\biggl(\bigotimes_{i=1}^n v_i \otimes \bigotimes_{i=1}^m T(\psi_{n+i})\Theta_{g,n+m}\biggr)\\
&= \Theta_{g,n}\pi_*\Omega_{g,n+m}\biggl(\bigotimes_{i=1}^n v_i \otimes \bigotimes_{i=1}^m T(\psi_{n+i})\prod_{i=1}^m \psi_{n+i}\biggr)\\
&= \Theta_{g,n}\pi_*\Omega_{g,n+m}\biggl(\bigotimes_{i=1}^n v_i \otimes \bigotimes_{i=1}^m \psi_{n+i}T(\psi_{n+i})\biggr)
\end{aligned}$$

and sum over m to get $T\cdot\Omega^\Theta = (zT)\cdot\Omega^\Theta$.

The Givental–Teleman theorem [29; 52] proves that the action defined in Definition 5.2 is transitive on semisimple CohFTs. In particular, a semisimple CohFT defined on a vector space of dimension N can be constructed via the Givental action on N copies of the trivial CohFT. Given a semisimple CohFT Ω, there exists

$$(R(z), T(z), \mathbb{1}) \in L^{(2)}\mathrm{GL}(N,\mathbb{C}) \times z^2\mathbb{C}^N[\![z]\!] \times \mathbb{C}^N$$

such that $\Omega_{g,n}$ is defined by the weighted sum over graphs (24) using $R(z)$, $T(z)$ and $\Omega'_{g,n}$ given by the topological field theory underlying $\Omega_{g,n}$. Note that a semisimple topological field theory of dimension N is equivalent to $\mathbb{1} \in \mathbb{C}^N$ which gives the unit vector in terms of a basis in which the product is diagonal and the metric η is the dot product, known as a *normalised canonical basis*.

On the level of partition functions, the construction of a semisimple CohFT from the trivial CohFT is realised via an action of quantised differential operators $\widehat{R}$ and $\widehat{T}$ on products of $Z^{\mathrm{KW}}(\hbar, t_0, t_1, \dots)$, a KdV tau function defined in the next section.

Definition 5.3 Define, for $R(z) = \exp\bigl(\sum_{\ell>0} r_\ell z^\ell\bigr) \in L^{(2)}\mathrm{GL}(N,\mathbb{C})$ and $T(z) = \sum_{k>0} T_k^\alpha z^k \in z\mathbb{C}^N[\![z]\!]$,

$$\widehat{R} := \exp\biggl\{\sum_{\ell=1}^{\infty}\sum_{\alpha,\beta}\biggl(\sum_{k=0}^{\infty}(r_k)^\alpha_\beta t_k^\beta \frac{\partial}{\partial t^\alpha_{k+\ell}} + \tfrac{1}{2}\hbar\sum_{m=0}^{\ell-1}(-1)^{m+1}(r_\ell)^\alpha_\beta \frac{\partial^2}{\partial t_m^\alpha \partial t^\beta_{\ell-m-1}}\biggr)\biggr\},$$

$$\widehat{T} := \exp\biggl(\sum_{\alpha=1}^{m}\sum_{k>0} T_k^\alpha \frac{\partial}{\partial t_k^\alpha}\biggr).$$

The partition function of (24) is given in [19; 29; 49] by

$$\begin{aligned}(26)\quad & Z_\Omega(\hbar,\{t_k^\alpha\}) \\ &= \widehat{R}\cdot\widehat{T}\cdot\hat{\mathbb{1}}\cdot Z^{\mathrm{KW}}(\hbar,\{t_k^1\})\cdots Z^{\mathrm{KW}}(\hbar,\{t_k^N\}) \\ &= \exp\biggl\{\sum_{g,n}\hbar^{g-1}\sum_{\Gamma\in G_{g,n}}\frac{1}{|\mathrm{Aut}(\Gamma)|}\prod_{v\in V(\Gamma)}\widehat{w}(v)\prod_{e\in E(\Gamma)}\widehat{w}(e)\prod_{\ell\in L(\Gamma)}\widehat{w}(\ell)\biggr\}.\end{aligned}$$

The operator $\hat{\mathbb{1}}$ rescales the variables $\widehat{\Delta}\cdot Z^{\mathrm{KW}}(\hbar,\{t_k^\alpha\}) = Z^{\mathrm{KW}}((\mathbb{1}^\alpha)^2\hbar,\{\mathbb{1}^\alpha t_k^\alpha\})$. Vertex weights $\widehat{w}(v)$ store products of Z^{KW} corresponding to the partition function of a topological field theory, edge weights $\widehat{w}(e)$ store coefficients of the series $\mathcal{E}(w,z)$, and leaf weights $\widehat{w}(\ell)$ store the variables t_k^α in a series weighted by coefficients of the series $R^{-1}(-z)$. We do not give explicit formulas for the weights — see [19; 29; 49] — and instead use an equivalent elegant formulation given by topological recursion, defined in Section 5.2.

A consequence of the relations (25) is the following proposition, which modifies the construction of a semisimple CohFT Ω to produce Ω^Θ:

Proposition 5.4 *Given a semisimple CohFT Ω defined via* (24) *using*

$$(R(z), T(z), \mathbb{1}) \in L^{(2)}\mathrm{GL}(N,\mathbb{C}) \times z^2\mathbb{C}^N[\![z]\!] \times \mathbb{C}^N,$$

the collection of classes Ω^Θ is defined via (24) *using*

$$\Bigl(R(z), \frac{1}{z}T(z), \mathbb{1}\Bigr) \in L^{(2)}\mathrm{GL}(N,\mathbb{C}) \times z\mathbb{C}^N[\![z]\!] \times \mathbb{C}^N$$

and

$$\Omega'_{g,n} = \Theta_{g,n} \otimes \Omega^{(0)}_{g,n} \colon H^{\otimes n} \to H^{4g-4+2n}(\overline{\mathcal{M}}_{g,n},\mathbb{C})$$

for $\Omega^{(0)}_{g,n}$ the degree 0 part of $\Omega_{g,n}$ determined by the vector $\mathbb{1} \in \mathbb{C}^N$. Its partition function $Z_{\Omega^\Theta}(\hbar,\{t_k^\alpha\})$ is obtained by replacing each copy of $Z^{\mathrm{KW}}(\hbar,\{t_k\})$ in (26) *by a copy of $Z^\Theta(\hbar,\{t_k\})$ and shifting the operator $\widehat{T}$.*

5.1 KdV tau functions

The KdV hierarchy is a sequence of partial differential equations beginning with the KdV equation,

$$U_{t_1} = UU_{t_0} + \tfrac{1}{12}\hbar U_{t_0t_0t_0}, \quad U(t_0,0,0,\dots) = f(t_0). \tag{27}$$

A tau function $Z(t_0,t_1,\dots)$ of the KdV hierarchy (equivalently the KP hierarchy in odd times $p_{2m+1} = t_m/(2m+1)!!$) gives rise to a solution $U = \hbar\,\partial^2(\log Z)/\partial t_0^2$ of the KdV hierarchy. The first equation in the hierarchy is the KdV equation (27), and later equations $U_{t_k} = P_k(U, U_{t_0}, U_{t_0t_0},\dots)$ for $k>1$ determine U uniquely from $U(t_0,0,0,\dots)$. See [40] for the full definition.

The Kontsevich–Witten tau function Z^{KW} is defined by the initial condition

$$U^{\mathrm{KW}}(t_0,0,0,\dots) = t_0$$

for $U^{\mathrm{KW}} = \hbar\partial^2(\log Z^{\mathrm{KW}})/\partial t_0^2$. The low-genus terms of $\log Z^{\mathrm{KW}}$ are

$$\log Z^{\mathrm{KW}}(\hbar,t_0,t_1,\dots) = \hbar^{-1}\big(\tfrac{1}{3!}t_0^3 + \tfrac{1}{3!}t_0^3t_1 + \tfrac{1}{4!}t_0^4t_2 + \cdots\big) + \tfrac{1}{24}t_1 + \cdots.$$

Theorem 5.5 (Witten and Kontsevich [36; 54])

$$Z^{\mathrm{KW}}(\hbar,t_0,t_1,\dots) = \exp\sum_{g,n}\hbar^{g-1}\frac{1}{n!}\sum_{\vec{k}\in\mathbb{N}^n}\int_{\overline{\mathcal{M}}_{g,n}}\prod_{i=1}^{n}\psi_i^{m_i}t_{m_i}$$

is a tau function of the KdV hierarchy.

The Brézin–Gross–Witten solution $U^{\mathrm{BGW}} = \hbar\partial^2(\log Z^{\mathrm{BGW}})/\partial t_0^2$ of the KdV hierarchy arises out of a unitary matrix model studied in [6; 30]. It is defined by the initial condition

$$U^{\mathrm{BGW}}(t_0,0,0,\dots) = \frac{\hbar}{8(1-t_0)^2}.$$

The low-genus g terms (= coefficient of $\hbar^{g-1}$) of $\log Z^{\mathrm{BGW}}$ are

$$\begin{aligned}\log Z^{\mathrm{BGW}} &= -\tfrac{1}{8}\log(1-t_0) + \hbar\cdot\tfrac{3}{128}\frac{t_1}{(1-t_0)^3} + \hbar\cdot^2\tfrac{15}{1024}\cdot\frac{t_2}{(1-t_0)^5} \\ &\qquad + \hbar^2\cdot\tfrac{63}{1024}\cdot\frac{t_1^2}{(1-t_0)^6} + O(\hbar^3) \\ &= \tfrac{1}{8}t_0 + \tfrac{1}{16}t_0^2 + \cdots + \hbar\big(\tfrac{3}{128}t_1 + \tfrac{9}{128}t_0t_1 + \cdots\big) \\ &\qquad + \hbar^2\big(\tfrac{15}{1024}t_2 + \tfrac{63}{1024}t_1^2 + \cdots\big).\end{aligned} \tag{28}$$

It satisfies the homogeneity property

$$\frac{\partial}{\partial t_0} Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots) = \sum_{i=0}^{\infty} (2i+1) t_i \frac{\partial}{\partial t_i} Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots) + \tfrac{1}{8} Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots),$$

which coincides with (15), satisfied by $Z^{\Theta}(\hbar, t_0, t_1, \dots)$. A proof of this homogeneity property for Z^{BGW} can be found in [2; 14].

The tau function $Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots)$ shares many properties of the famous Kontsevich-Witten tau function $Z^{\mathrm{KW}}(\hbar, t_0, t_1, \dots)$ introduced in [54]. An analogue of Theorem 5.5 is given by Conjecture 1.5, which postulates that the function

$$Z^{\Theta}(\hbar, t_0, t_1, \dots) = \exp \sum_{g,n,\vec{k}} \frac{\hbar^{g-1}}{n!} \int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \cdot \prod_{j=1}^{n} \psi_j^{k_j} \prod t_{k_j}$$

coincides with $Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots)$. The tau function Z^{BGW} appears in a generalisation of Givental's decomposition of CohFTs in [9].

Definition 5.6 Given a semisimple CohFT Ω with partition function $Z_{\Omega}(\hbar, \{t_k^{\alpha}\})$ constructed as a graphical sum, via (26),

$$Z_{\Omega}(\hbar, \{t_k^{\alpha}\}) = \hat{R} \cdot \hat{T} \cdot \hat{\mathbb{1}} \cdot Z^{\mathrm{KW}}(\hbar, \{t_k^1\}) \cdots Z^{\mathrm{KW}}(\hbar, \{t_k^N\}),$$

define

$$Z_{\Omega}^{\mathrm{BGW}}(\hbar, \{t_k^{\alpha}\}) = \hat{R} \cdot \hat{T}_0 \cdot \hat{\mathbb{1}} \cdot Z^{\mathrm{BGW}}(\hbar, \{t_k^1\}) \cdots Z^{\mathrm{BGW}}(\hbar, \{t_k^N\}),$$

where $T_0 = T/z(z)$.

The same shift $T_0 = \frac{1}{z} T(z)$ is used by $Z^{\mathrm{BGW}}(\hbar, \{t_k\})$ and $Z^{\Theta}(\hbar, \{t_k\})$ due to their common homogeneity property (15). One can also replace only some copies of $Z^{\mathrm{KW}}(\hbar, \{t_k\})$ in (26) by copies of $Z^{\mathrm{BGW}}(\hbar, \{t_k\})$ and shift components of $\hat{T}$. For example, in [13], the enumeration of bipartite dessins d'enfant is shown to have partition function

$$Z(\hbar, \{t_k^{\alpha}\}) = \hat{R} \cdot \hat{T} \cdot Z^{\mathrm{BGW}}\bigl(-\tfrac{1}{2}\hbar, i\bigl\{\tfrac{1}{\sqrt{2}} t_k^1\bigr\}\bigr) Z^{\mathrm{KW}}(32\hbar, \{4\sqrt{2} t_k^2\}) \tag{29}$$

for R and T determined by the curve $xy^2 + xy + 1 = 0$ as described in Section 5.2.

$$\begin{tikzcd}
\begin{matrix} R(z) \in L^{(2)}\mathrm{GL}(N,\mathbb{C}) \\ T(z) \in z^2\mathbb{C}^N[\![z]\!], \mathbb{1} \in \mathbb{C}^N \end{matrix} \arrow[r, "\text{Givental construction}"] \arrow[d, leftrightarrow] & Z_\Omega(\hbar, \{t_k^\alpha\}) \arrow[d, leftrightarrow] \\
S = (C, x, y, B) \arrow[r, "\text{topological recursion}"] & Z^S(\hbar, \{t_k^\alpha\})
\end{tikzcd}$$

Figure 1: Constructions of CohFT partition functions.

5.2 Topological recursion

Figure 1 summarises the contents of this section. The upper horizontal arrow in the figure represents Givental's construction of a partition function defined in (26) and Definition 5.2. Topological recursion is defined in Section 5.2 — it produces a partition function from a spectral curve $S = (C, x, y, B)$ consisting of a Riemann surface C equipped with meromorphic functions x and y and a bidifferential B. We begin with a description of the left vertical arrow, which represents the construction of an element $R(z) \in L^{(2)}\mathrm{GL}(N, \mathbb{C})$ from (C, x, B) in (30) and $T(z)$ and $\mathbb{1}$ from (C, x, y) in (36) and (35). We then define topological recursion in Section 5.2.1 and state the result of [18], that topological recursion encodes the graphical construction in (26) and gives equality of partition functions, represented by the right vertical arrow.

An element of the twisted loop group $R(z) \in L^{(2)}\mathrm{GL}(N, \mathbb{C})$ can be naturally constructed from a Riemann surface Σ equipped with a bidifferential $B(p_1, p_2)$ on $\Sigma \times \Sigma$ and a meromorphic function $x\colon \Sigma \to \mathbb{C}$ for N the number of zeros of dx. A basic example is the function $x = z^2$ on $\Sigma = \mathbb{C}$, which gives rise to the constant element $R(z) = 1 \in \mathrm{GL}(1, \mathbb{C})$. More generally, any function x that looks like this example locally, ie $x = s^2 + c$ for s a local coordinate around a zero of dx and $c \in \mathbb{C}$, gives $R(z) = I + R_1 z + \cdots \in L^{(2)}\mathrm{GL}(N, \mathbb{C})$, which is in some sense a deformation of $I \in \mathrm{GL}(N, \mathbb{C})$, or N copies of the basic example.

Definition 5.7 On any compact Riemann surface $(\Sigma, \{\mathcal{A}_i\}_{i=1,\dots,g})$ with a choice of $\mathcal{A}$–cycles, define a *fundamental normalised bidifferential of the second kind* $B(p, p')$ to be a symmetric tensor product of differentials on $\Sigma \times \Sigma$, uniquely defined by the properties that it has a double pole on the diagonal of zero residue, double residue equal to 1, no further singularities and normalised by $\int_{p \in \mathcal{A}_i} B(p, p') = 0$ for $i = 1, \dots, g$ [27].

On a rational curve, which is sufficient for the examples in this paper, B is the Cauchy kernel

$$B(z_1, z_2) = \frac{dz_1\, dz_2}{(z_1 - z_2)^2}.$$

The bidifferential $B(p, p')$ acts as a kernel for producing meromorphic differentials on the Riemann surface Σ via $\omega(p) = \int_\Lambda \lambda(p') B(p, p')$, where λ is a function defined along the contour $\Lambda \subset \Sigma$. Depending on the choice of (Λ, λ), ω can be a differential of the 1st kind (holomorphic), 2nd kind (zero residues) or 3rd kind (simple poles).

Definition 5.8 For (Σ, x) a Riemann surface equipped with a meromorphic function, define evaluation of any meromorphic differential ω at a simple zero $\mathcal{P}$ of dx by

$$\omega(\mathcal{P}) := \mathrm{Res}_{p=\mathcal{P}} \frac{\omega(p)}{\sqrt{2(x(p) - x(\mathcal{P}))}},$$

where we choose a branch of $\sqrt{x(p) - x(\mathcal{P})}$ once and for all at $\mathcal{P}$ to remove the ± 1 ambiguity.

A fundamental example of Definition 5.8 required here is $B(\mathcal{P}, p)$, which is a normalised (trivial $\mathcal{A}$–periods) differential of the second kind holomorphic on $\Sigma \backslash \mathcal{P}$ with a double pole at the simple zero $\mathcal{P}$ of dx.

In order to produce an element of the twisted loop group, Shramchenko [50] constructed a solution $Y(z)$ of the linear system (20) using $V = [\mathcal{B}, U]$ for $\mathcal{B}_{\alpha\beta} = B(\mathcal{P}_\alpha, \mathcal{P}_\beta)$ (defined for $\alpha \neq \beta$) given by

$$Y(z)^\alpha_\beta = -\frac{\sqrt{z}}{\sqrt{2\pi}} \int_{\Gamma_\beta} B(\mathcal{P}_\alpha, p) \cdot e^{-x(p)/z}.$$

The proof in [50] is indirect, showing that $Y(z)^i_j$ satisfies an associated set of PDEs in u_i and using the Rauch variational formula to calculate $\partial_{u_k} B(\mathcal{P}_\alpha, p)$. Instead, here we work directly with the associated element $R(z)$ of the twisted loop group.

Definition 5.9 Define the asymptotic series in the limit $z \to 0$ by

$$R^{-1}(z)^\alpha_\beta = -\frac{\sqrt{z}}{\sqrt{2\pi}} \int_{\Gamma_\beta} B(\mathcal{P}_\alpha, p) \cdot e^{(x(\mathcal{P}_\beta) - x(p))/z}, \tag{30}$$

where Γ_β is a path of steepest descent for $-x(p)/z$ containing $x(\mathcal{P}_\beta)$.

Note that the asymptotic expansion of the contour integral (30) for $z \to 0$ depends only the intersection of Γ_β with a neighbourhood of $p = \mathcal{P}_\beta$. When $\alpha = \beta$, the integrand has zero residue at $p=\mathcal{P}_\beta$, so we deform Γ_β to go around $\mathcal{P}_\beta$ to get a well-defined integral. Locally, this is the same as defining $\int_{\mathbb{R}} s^{-2}\exp(-s^2)\,ds=-2\sqrt{\pi}$ by integrating the analytic function $z^{-2}\exp(-z^2)$ along the real line in $\mathbb{C}$ deformed to avoid 0.

Lemma 5.10 [50] *The asymptotic series $R(z)$ defined in* (30) *satisfies the twisted loop group condition*

$$R(z)R^T(-z)=\mathrm{Id}. \tag{31}$$

Proof The proof here is taken from [16]. We have

$$\begin{aligned}\sum_{\alpha=1}^{N}\mathrm{Res}_{q=\mathcal{P}_\alpha}\frac{B(p,q)B(p',q)}{dx(q)} &= -\mathrm{Res}_{q=p}\frac{B(p,q)B(p',q)}{dx(q)}-\mathrm{Res}_{q=p'}\frac{B(p,q)B(p',q)}{dx(q)}\\ &= -d_p\biggl(\frac{B(p,p')}{dx(p)}\biggr)-d_{p'}\biggl(\frac{B(p,p')}{dx(p')}\biggr),\end{aligned} \tag{32}$$

where the first equality uses the fact that the only poles of the integrand occur at $\{p,p',\mathcal{P}_\alpha\mid\alpha=1,\dots,N\}$, and the second equality uses the Cauchy formula satisfied by the Bergman kernel. Define the Laplace transform of the Bergman kernel by

$$\check{B}^{\alpha,\beta}(z_1,z_2)=\frac{e^{x(\mathcal{P}_\alpha)/z_1+x(\mathcal{P}_\beta)/z_2}}{2\pi\sqrt{z_1z_2}}\int_{\Gamma_\alpha}\int_{\Gamma_\beta}B(p,p')e^{-x(p)/z_1-x(p')/z_2}.$$

The Laplace transform of the left-hand side of (32) is

$$\begin{aligned}&\frac{e^{x(\mathcal{P}_\alpha)/z_1+x(\mathcal{P}_\beta)/z_2}}{2\pi\sqrt{z_1z_2}}\int_{\Gamma_\alpha}\int_{\Gamma_\beta}e^{-x(p)/z_1-x(p')/z_2}\sum_{\gamma=1}^{N}\mathrm{Res}_{q=\mathcal{P}_\gamma}\frac{B(p,q)B(p',q)}{dx(q)}\\ &\qquad=\sum_{\gamma=1}^{N}\frac{e^{x(\mathcal{P}_\alpha)/z_1+x(\mathcal{P}_\beta)/z_2}}{2\pi\sqrt{z_1z_2}}\int_{\Gamma_\alpha}e^{-x(p)/z_1}B(p,\mathcal{P}_\gamma)\int_{\Gamma_\beta}e^{-x(p')/z_2}B(p',\mathcal{P}_\gamma)\\ &\qquad=\sum_{\gamma=1}^{N}\frac{[R^{-1}(z_1)]_\alpha^\gamma[R^{-1}(z_2)]_\beta^\gamma}{z_1z_2}.\end{aligned}$$

Since the Laplace transform satisfies

$$\int_{\Gamma_\alpha}d\biggl(\frac{\omega(p)}{dx(p)}\biggr)e^{-x(p)/z}=\frac{1}{z}\int_{\Gamma_\alpha}\omega(p)e^{-x(p)/z}$$

for any differential $\omega(p)$ by integration by parts, the Laplace transform of the right-hand side of (32) is

$$-\frac{e^{x(\mathcal{P}_\alpha)/z_1+x(\mathcal{P}_\beta)/z_2}}{2\pi\sqrt{z_1z_2}}\int_{\Gamma_\alpha}\int_{\Gamma_\beta}e^{-x(p)/z_1-x(p')/z_2}\left\{d_p\left(\frac{B(p,p')}{dx(p)}\right)+d_{p'}\left(\frac{B(p,p')}{dx(p')}\right)\right\}$$
$$=-\Big(\frac{1}{z_1}+\frac{1}{z_2}\Big)\check{B}^{\alpha,\beta}(z_1,z_2).$$

Putting the two sides together yields the result, due to Eynard [21],

$$\check{B}^{\alpha,\beta}(z_1,z_2)=-\frac{\sum_{\gamma=1}^N[R^{-1}(z_1)]_\alpha^\gamma[R^{-1}(z_2)]_\beta^k}{z_1+z_2}. \tag{33}$$

Equation (31) is an immediate consequence of (33) and the finiteness of $\check{B}^{\alpha,\beta}(z_1,z_2)$ at $z_2=-z_1$. □

5.2.1 Topological recursion Topological recursion is a procedure which takes as input a spectral curve, defined below, and produces a collection of symmetric tensor products of meromorphic 1–forms $\omega_{g,n}$ on C^n. The correlators store enumerative information in different ways. Periods of the correlators store top intersection numbers of tautological classes in the moduli space of stable curves $\overline{\mathcal{M}}_{g,n}$ and local expansions of the correlators can serve as generating functions for enumerative problems.

A spectral curve $S=(C,x,y,B)$ is a Riemann surface C equipped with two meromorphic functions $x,y\colon C\to\mathbb{C}$ and a bidifferential $B(p_1,p_2)$ defined in Definition 5.7, which is the Cauchy kernel in this paper. Topological recursion, as developed by Chekhov, Eynard and Orantin [8; 22], is a procedure that produces from a spectral curve $S=(C,x,y,B)$ a symmetric tensor product of meromorphic 1–forms $\omega_{g,n}$ on C^n for integers $g\ge 0$ and $n\ge 1$, which we refer to as *correlation differentials* or *correlators*. The correlation differentials $\omega_{g,n}$ are defined by

$$\omega_{0,1}(p_1)=-y(p_1)\,dx(p_1)\quad\text{and}\quad\omega_{0,2}(p_1,p_2)=B(p_1,p_2)$$

and, for $2g-2+n>0$, they are defined recursively via

$$\omega_{g,n}(p_1,p_L)$$
$$=\sum_{dx(\alpha)=0}\operatorname{Res}_{p=\alpha}K(p_1,p)$$
$$\cdot\Bigg[\omega_{g-1,n+1}(p,\hat{p},p_L)+\sum^{\circ}_{\substack{g_1+g_2=g\\ I\sqcup J=L}}\omega_{g_1,|I|+1}(p,p_I)\omega_{g_2,|J|+1}(\hat{p},p_J)\Bigg].$$

Here, we use the notation $L=\{2,3,\dots,n\}$ and $p_I=\{p_{i_1},p_{i_2},\dots,p_{i_k}\}$ for $I=\{i_1,i_2,\dots,i_k\}$. The outer summation is over the zeroes α of dx and $p\mapsto\hat{p}$ is the involution defined locally near α satisfying $x(\hat{p})=x(p)$ and $\hat{p}\neq p$. The symbol $\circ$ over the inner summation means that we exclude any term that involves $\omega_{0,1}$. Finally, the recursion kernel is given by

$$K(p_1,p)=-\frac{1}{2}\cdot\frac{\int_{\hat{p}}^{p}\omega_{0,2}(p_1,\cdot)}{[y(p)-y(\hat{p})]\,dx(p)}.$$

which is well defined in the vicinity of each zero of dx. It acts on differentials in p and produces differentials in p_1 since the quotient of a differential in p by the differential $dx(p)$ is a meromorphic function. For $2g-2+n>0$, each $\omega_{g,n}$ is a symmetric tensor product of meromorphic 1–forms on C^n with residueless poles at the zeros of dx and holomorphic elsewhere. A zero α of dx is *regular* if y is regular at α, and irregular if y has a simple pole at α. A spectral curve is irregular if it contains an irregular zero of dx. The order of the pole in each variable of $\omega_{g,n}$ at a regular (resp. irregular) zero of dx is $6g-4+2n$ (resp. $2g$). Define $\Phi(p)$ up to an additive constant by $d\Phi(p)=y(p)\,dx(p)$. For $2g-2+n>0$, the invariants satisfy the dilaton equation [22]

$$\sum_{\alpha}\mathrm{Res}_{p=\alpha}\,\Phi(p)\omega_{g,n+1}(p,p_1,\dots,p_n)=(2g-2+n)\omega_{g,n}(p_1,\dots,p_n),$$

where the sum is over the zeros α of dx. This enables the definition of the so-called *symplectic invariants*

$$F_g=\sum_{\alpha}\mathrm{Res}_{p=\alpha}\,\Phi(p)\omega_{g,1}(p).$$

The correlators $\omega_{g,n}$ are normalised differentials of the second kind in each variable since they have zero $\mathcal{A}$–periods, and poles only at the zeros $\mathcal{P}_\alpha$ of dx of zero residue. Their principal parts are skew-invariant under the local involution $p\mapsto\hat{p}$. A basis of such normalised differentials of the second kind is constructed from x and B in the following definition:

Definition 5.11 For a Riemann surface C equipped with a meromorphic function $x\colon C\to\mathbb{C}$ and bidifferential $B(p_1,p_2)$ define the auxiliary differentials on C as follows. For each zero $\mathcal{P}_\alpha$ of dx, define

$$V_0^\alpha(p)=B(\mathcal{P}_\alpha,p),\quad V_{k+1}^\alpha(p)=-d\left(\frac{V_k^\alpha(p)}{dx(p)}\right)\tag{34}$$

for $\alpha=1,\dots,N$ and $k=0,1,2,\dots$, where evaluation $B(\mathcal{P}_\alpha,p)$ at $\mathcal{P}_\alpha$ is given in Definition 5.8.

The correlators $\omega_{g,n}$ are polynomials in the auxiliary differentials $V_k^\alpha(p)$. To any spectral curve S, one can define a partition function Z^S by assembling the polynomials built out of the correlators $\omega_{g,n}$ [18; 21; 45].

Definition 5.12 $$Z^S(\hbar,\{u_k^\alpha\}) := \exp\sum_{g,n}\frac{\hbar^{g-1}}{n!}\omega_{g,n}^S\Big|_{V_k^\alpha(p_i)=u_k^\alpha}.$$

As usual, define F_g to be the contribution from $\omega_{g,n}$,

$$\log Z^S(\hbar,\{u_k^\alpha\}) = \sum_{g\geq 0}\hbar^{g-1}F_g^S(\{u_k^\alpha\}).$$

5.2.2 From topological recursion to Givental's construction The input data for Givental's construction is a triple $(R(z),T(z),\mathbb{1})\in L^{(2)}\mathrm{GL}(N,\mathbb{C})\times z^2\mathbb{C}^N[\![z]\!]\times\mathbb{C}^N$. Its output is a CohFT Ω, and its partition function $Z_\Omega(\hbar,\{t_k^\alpha\})$. The input data for topological recursion is a spectral curve $S=(C,x,y,B)$. Its output is the correlators $\omega_{g,n}$, which can be assembled into a partition function $Z^S(\hbar,\{t_k^\alpha\})$.

From a compact spectral curve define a triple

$$S=(C,x,y,B)\to(R(z),T(z),\mathbb{1})\in L^{(2)}\mathrm{GL}(N,\mathbb{C})\times z\mathbb{C}^N[\![z]\!]\times\mathbb{C}^N$$

by

$$(C,x,B)\mapsto R(z)\in L^{(2)}\mathrm{GL}(N,\mathbb{C})$$

via (30),

$$\mathbb{1}^i=\begin{cases}dy(\mathcal{P}_\alpha) & \text{if } \mathcal{P}_\alpha \text{ is regular,}\\ (y\,dx)(\mathcal{P}_\alpha) & \text{if } \mathcal{P}_\alpha \text{ is irregular,}\end{cases}\tag{35}$$

which is the unit in normalised canonical coordinates, and

$$T(z)^\alpha=\begin{cases}z\Big(\mathbb{1}^\alpha-\dfrac{1}{\sqrt{2\pi z}}\displaystyle\int_{\Gamma_\alpha}dy(p)\cdot e^{(x(\mathcal{P}_\alpha)-x(p))/z}\Big) & \text{if } \mathcal{P}_\alpha \text{ is regular,}\\ \mathbb{1}^\alpha-\dfrac{1}{\sqrt{2\pi z}}\displaystyle\int_{\Gamma_\alpha}y(p)\,dx(p)\cdot e^{(x(\mathcal{P}_\alpha)-x(p))/z} & \text{if } \mathcal{P}_\alpha \text{ is irregular.}\end{cases}\tag{36}$$

Note that

$$\lim_{z\to 0}\frac{1}{\sqrt{2\pi z}}\int_{\Gamma_\alpha}dy(p)\cdot e^{(x(\mathcal{P}_\alpha)-x(p))/z}=\begin{cases}dy(\mathcal{P}_\alpha) & \text{if } \mathcal{P}_\alpha \text{ is regular,}\\ (y\,dx)(\mathcal{P}_\alpha) & \text{if } \mathcal{P}_\alpha \text{ is irregular,}\end{cases}$$

which defines $\mathbb{1}$; hence, the right-hand side of (36) lives in $z^2\mathbb{C}^N[\![z]\!]$ (resp. $z\mathbb{C}^N[\![z]\!]$) when $\mathcal{P}_\alpha$ is regular (resp. irregular). If Ω is a CohFT with flat unit — see (12) in Section 4 — given by $\mathbb{1} \in \mathbb{C}^N$, then $\mathbb{1}$ determines the translation via $T(z) = z(\mathbb{1} - R^{-1}(z)\mathbb{1}) \in z^2\mathbb{C}^N[\![z]\!]$. In this special case, y satisfies

$$(37)\qquad (R^{-1}(z)\mathbb{1})^\alpha = \sum_{k=1}^{N} R^{-1}(z)_k^\alpha \cdot \Delta_k^{1/2} = \frac{1}{\sqrt{2\pi z}} \int_{\Gamma_\alpha} dy(p) \cdot e^{(x(\mathcal{P}_\alpha) - x(p))/z},$$

which uniquely determines y from its first-order data $\{dy(\mathcal{P}_\alpha)\}$ at each $\mathcal{P}_\alpha$.

The map $(C, x, y, B) \mapsto (R(z), T(z), \mathbb{1})$ produces the left vertical arrow in Figure 1 and its generalisation to irregular spectral curves, ie a correspondence between the input data, and via the graphical construction (26) this produces the same output $Z_\Omega(\hbar, \{t_k^\alpha\}) = Z^S(\hbar, \{t_k^\alpha\})$, which is the main result of [18], stated in the following theorem:

Theorem 5.13 [18] *Given a CohFT Ω built from*

$$R(z) \in L^{(2)}\mathrm{GL}(N, \mathbb{C}), \quad T(z) \in z^2\mathbb{C}^N[\![z]\!], \quad \mathbb{1} \in \mathbb{C}^N$$

via Definition 5.2, there exists a local spectral curve

$$S = (C, x, y, B) \mapsto (R(z), T(z), \mathbb{1})$$

on which x and B correspond to $R(z)$ via Definition 5.9 and y corresponds to $T(z)$ and $\mathbb{1}$ via (36) *and* (35), *giving the partition function of the CohFT*

$$Z_\Omega(\hbar, \{t_k^\alpha\}) = Z^S(\hbar, \{t_k^\alpha\}).$$

In general, the spectral curve S in Theorem 5.13 is a local spectral curve which is a collection of disk neighbourhoods of zeros of dx on which B and y are defined locally, although we only consider compact spectral curves S in this paper. Theorem 5.13 was proven only in the case $T(z) = z(\mathbb{1} - R^{-1}(z)\mathbb{1})$ in [18] but it has been generalised to allow any $T(z) \in z^2\mathbb{C}^N[\![z]\!]$; see [9; 37]. We will use the converse of Theorem 5.13, proven in [16], beginning instead from S. Theorem 5.13 was also generalised in [9] to show that the operators $\hat{\Psi}$, $\hat{R}$ and $\hat{T}$ acting on copies of Z^{BGW} analogous to (26) arises by applying topological recursion to an irregular spectral curve. Equivalently, periods of the correlators of an irregular spectral curve store linear combinations of coefficients of $\log Z^{\mathrm{BGW}}$. The appearance of Z^{BGW} is due to its relationship with topological recursion applied to the curve $x = \frac{1}{2}z^2$, $y = 1/z$ [14].

5.2.3 Spectral curve examples We demonstrate Theorem 5.13 with four key examples of rational spectral curves equipped with the bidifferential $B(p_1, p_2)$ given by the Cauchy kernel. The spectral curves in Examples 5.14 and 5.15, denoted by S_{Airy} and S_{Bes}, have partition functions Z^{KW} and Z^{BGW}, respectively. Any spectral curve at regular (resp. irregular) zeros of dx is locally isomorphic to S_{Airy} (resp. S_{Bes}). A consequence is that the tau functions Z^{KW} and Z^{BGW} are fundamental to the correlators produced from topological recursion. Moreover, the topological recursion partition function Z^S is constructed via (26), using a product of copies of Z^{KW} *and* copies of Z^{BGW}, as in (29), where R and T are obtained from the spectral curve as described in Section 5.2.2. The third example, given by Theorem 5.16, brings together Z^{KW} and Z^{Θ} and conjecturally Z^{BGW} in the limit. Proposition 5.4, which gives the relationship between the Givental construction of a semisimple CohFT Ω and its associated Ω^{BGW}, has an elegant consequence for spectral curves. This is demonstrated explicitly in the fourth example, which shows the relationship between the spectral curves of a CohFT Ω^{A_2} associated to the A_2 singularity and $(\Omega^{A_2})^{\text{BGW}}$.

Examples 5.14 and 5.15 below use the differentials

$$\xi_m(z) = (2m+1)!!\, z^{-(2m+2)}\, dz$$

defined by (34) for $x = \frac{1}{2}z^2$ with respect to a global rational parameter z for the curve $C \cong \mathbb{C}$.

Example 5.14 Topological recursion applied to the Airy curve

$$S_{\text{Airy}} = \left(\mathbb{C},\ x = \tfrac{1}{2}z^2,\ y = z,\ B = \frac{dz\, dz'}{(z-z')^2}\right)$$

produces correlators which are proven in [23] to store intersection numbers

$$\omega^{\text{Airy}}_{g,n} = \sum_{\vec{m}\in\mathbb{Z}^n_+} \int_{\overline{\mathcal{M}}_{g,n}} \prod_{i=1}^n \psi_i^{m_i} (2m_i+1)!!\, \frac{dz_i}{z_i^{2m_i+2}}$$

and the coefficient is nonzero only for $\sum_{i=1}^n m_i = 3g-3+n$. Hence,

$$\begin{aligned} Z^{\text{KW}}(\hbar, t_0, t_1, \dots) = Z^{S_{\text{Airy}}}(\hbar, t_0, t_1, \dots) &= \exp \sum_{g,n} \frac{\hbar^{g-1}}{n!} \omega^{\text{Airy}}_{g,n}\bigg|_{\xi_m(z_i)=t_m} \\ &= \exp \sum_{g,n,\vec{m}} \frac{\hbar^{g-1}}{n!} \int_{\overline{\mathcal{M}}_{g,n}} \prod_{i=1}^n (\psi_i^{m_i} t_{m_i}). \end{aligned}$$

Example 5.15 Topological recursion applied to the Bessel curve

$$S_{\text{Bes}} = \left(\mathbb{C},\ x = \tfrac{1}{2}z^2,\ y = \frac{1}{z},\ B = \frac{dz\, dz'}{(z-z')^2}\right)$$

produces correlators

$$\omega^{\mathrm{Bes}}_{g,n}=\sum_{\vec{k}\in\mathbb{Z}^n_+}b_g(m_1,\dots,m_n)\prod_{i=1}^n(2m_i+1)!!\,\frac{dz_i}{z_i^{2m_i+2}},$$

where $b_g(m_1,\dots,m_n)\neq 0$ only for $\sum_{i=1}^n m_i=g-1$. It is proven in [14] that

$$Z^{\mathrm{BGW}}(\hbar,t_0,t_1,\dots)=Z^{S_{\mathrm{Bes}}}(\hbar,t_0,t_1,\dots)=\exp\sum_{g,n}\frac{\hbar^{g-1}}{n!}\omega^{\mathrm{Bes}}_{g,n}\bigg|_{\xi_m(z_i)=t_m}.$$

For the next example, define differentials $\xi^\alpha_m(z,t)$, using $x=\frac{1}{2}z^2-t\cdot\log z$, by

$$\begin{aligned}\xi^0_{-1}(z,t)&=t^{-1/2}z\,dz,\\ \xi^1_{-1}(z,t)&=dz,\\ \xi^\alpha_{m+1}(z,t)&=-d\left(\frac{\xi^\sigma_m(z,t)}{dx(z)}\right),\quad \sigma=0,1,\ m=-1,0,1,2,\dots.\end{aligned}\tag{38}$$

For $m\geq 0$, these are linear combinations of the $V^i_m(p)$ defined in (34). The following theorem uses the Chern polynomial

$$c(E^{\vec{\sigma}}_{g,n},t)=1+t\cdot c_1(E^{\vec{\sigma}}_{g,n})+t^2\cdot c_2(E^{\vec{\sigma}}_{g,n})+\cdots\in H^*(\overline{\mathcal{M}}^{\mathrm{spin}}_{g,n,\vec{\sigma}},\mathbb{Q}),\quad \vec{\sigma}\in\{0,1\}^n.$$

Theorem 5.16 [37] *Topological recursion applied to the spectral curve*

$$x=\tfrac{1}{2}z^2-t\cdot\log z,\quad y=z^{-1},\quad B=\frac{dz\,dz'}{(z-z')^2}\tag{39}$$

produces correlators $\omega_{g,n}$ *satisfying*

$$\begin{aligned}&\omega_{g,n}(t,z_1,\dots,z_n)\\ &\qquad=\sum_{\vec{\sigma},\vec{m}}(-1)^n t^{2g-2+n}2^{1-g}\int_{\overline{\mathcal{M}}_{g,n}}p_*c\Big(E^{\vec{\sigma}}_{g,n},\frac{2}{t}\Big)\prod_{i=1}^n\psi_i^{m_i}\xi^{\sigma_i}_{m_i}(z_i,t).\end{aligned}\tag{40}$$

Proof Theorem 5.16 is a specialisation of a theorem in [37] which applies to a generalisation of the moduli space of spin curves to the moduli space of r–spin curves

$$\overline{\mathcal{M}}^{1/r}_{g,n}=\{(\mathcal{C},\theta,p_1,\dots,p_n,\phi)\mid\phi\colon\theta^r\xrightarrow{\cong}\omega^{\log}_{\mathcal{C}}\}.$$

For any $s\in\mathbb{Z}$, there is a line bundle $\mathcal{E}$ on the universal r–spin curve over $\overline{\mathcal{M}}^{1/r}_{g,n}$ with fibres given by the universal r^{th} root of $(\omega^{\log}_{\mathcal{C}})^s$. Its derived pushforward $R^*\pi_*\mathcal{E}$ defines a virtual bundle over $\overline{\mathcal{M}}^{1/r}_{g,n}$. For example, when $s=1$ and $r=1$, $-R^*\pi_*\mathcal{E}$

is the Hodge bundle, and, when $s=-1$ and $r=2$, $-R^*\pi_*\mathcal{E}=E_{g,n}$ coincides with Definition 2.1 (where $\mathcal{E}^\vee$ has now become $\mathcal{E}$ due to $s=-1$.) Note that [37] considers r^{th} roots of $(\omega_C^{\log})^s\big(-\sum_{i=1}^n \sigma_i p_i\big)$ for C the underlying coarse curve of $\mathcal{C}$ with forgetful map $\rho\colon \mathcal{C}\to C$. The r^{th} roots in [37] coincide with the pushforward $|\theta|=\rho_*\theta$, which is the locally free sheaf of $\mathbb{Z}_2$–invariant sections of the pushforward sheaf of θ, and the isotropy representation at p_i determines σ_i as described in Section 2. For $r=2$, ie $\theta^2\cong\omega_{\mathcal{C}}^{\log}$, at any point p_i banded by $\frac{1}{2}$ the pushforward locally satisfies $|\theta|^2=\omega_C(2p_i)=\omega_C^{\log}(p_i)$; hence, $(|\theta|^\vee)^2=(\omega_C^{\log})^{-1}(-p_i)$, which corresponds to $\sigma_i=1$. At any point p_i banded by 0, the pushforward does not change local degree and corresponds to $\sigma_i=0$.

The Chern character of the virtual bundle $-R^*\pi_*\mathcal{E}$ is given by Chiodo's generalisation of Mumford's formula for the Chern character of the Hodge bundle. For $\sigma\in\{0,1,\dots,r-1\}$, let $j_\sigma\colon \mathrm{Sing}_\sigma\to\overline{\mathcal{M}}_{g,n}^{1/r}$ be the map from the singular set of the universal spin curve banded by σ/r, where now the local isotropy is $\mathbb{Z}_n$. Let $B_m(x)$ be the m^{th} Bernoulli polynomial. Chiodo [10] proved

$$\operatorname{ch}(R^*\pi_*\mathcal{E})=\sum_{m\ge 0}\bigg(\frac{B_{m+1}(s/r)}{(m+1)!}\kappa_m-\sum_{i=1}^n\frac{B_{m+1}(m_i/r)}{(m+1)!}\psi_i^m \\ +\tfrac{1}{2}r\sum_{\sigma=0}^{r-1}\frac{B_{m+1}(\sigma/r)}{(m+1)!}(j_\sigma)_*\frac{\psi_+^m+(-1)^{m-1}\psi_-^m}{\psi_++\psi_-}\bigg). \tag{41}$$

The total Chern class of a virtual bundle $c(E-F):=c(E)/c(F)$ can be calculated from its Chern character and in this case is given by

$$c(-R^*\pi_*\mathcal{E})=\exp\bigg(\sum_{m=1}^\infty(-1)^m(m-1)!\operatorname{ch}_m(R^*\pi_*\mathcal{E})\bigg).$$

The components of $\overline{\mathcal{M}}_{g,n}^{1/r}$ are given by $\overline{\mathcal{M}}_{g,n,\vec\sigma}^{1/r}$ for $\vec\sigma\in\mathbb{Z}_r^n$. The pushforward of the restriction of $c(-R^*\pi_*\mathcal{E})$ to a component is known as the Chiodo class

$$C_{g,n}(r,s;\vec\sigma):=p_*c(-R^*\pi_*\mathcal{E}|_{\overline{\mathcal{M}}_{g,n,\vec\sigma}^{1/r}})\in H^*(\overline{\mathcal{M}}_{g,n},\mathbb{Q}).$$

The sum of this pushforward over all components of $\overline{\mathcal{M}}_{g,n}^{1/r}$ is expressed as a weighted sum over stable graphs in [32] which encodes a twisted loop group action as described in Section 5, with edge and vertex weights proven in [37, Theorem 4.5] to exactly match the edge and vertex weights arising from the spectral curve

$$\hat x=z^r-\log z,\quad \hat y=\frac{r^{1+s/r}}{s}z^s,\quad B=\frac{dz\,dz'}{(z-z')^2}.$$

In particular, the term $\exp\bigl(-\sum_m (B_{m+1}(s/r)/m(m+1))\kappa_m\bigr)$ which arises from the $\sum_m (B_{m+1}(s/r)/(m+1)!)\kappa_m$ terms in Chiodo's formula exactly matches the local expansion of dy. More precisely, by [37, Lemma 4.1],

$$\frac{1}{\sqrt{2\pi\hbar}}\int_{\Gamma_\alpha} dy(p)\cdot e^{(x(\mathcal{P}_\alpha)-x(p))/\hbar} \sim dy(\mathcal{P}_\alpha)\exp\biggl(-\sum_m \frac{B_{m+1}(s/r)}{m(m+1)}(-\hbar)^m\biggr), \tag{42}$$

where $\sim$ means the asymptotic expansion in the limit $\hbar \to 0$.

Hence, topological recursion applied to this spectral curve produces correlators with expansion in terms of the local coordinate $e^{-\hat{x}_i} = e^{-\hat{x}(z_i)} = z_i e^{-z_i^r}$ around $z_i = 0$,

$$\begin{aligned} &\hat{\omega}_{g,n}(z_1,\dots,z_n) \\ &\qquad \sim \sum_{\vec{k}\in\mathbb{Z}_+^n}\prod_{i=1}^n c(k_i) r^{(k_i)_r/r} d(e^{-k_i\hat{x}_i}) \int_{\overline{\mathcal{M}}_{g,n}} \frac{C_{g,n}(r,s;(-\vec{k})_r)}{\prod_{i=1}^n (1-(k_i/r)\psi_i)}, \end{aligned} \tag{43}$$

where $\sim$ means expansion in a local coordinate, $(-\vec{k})_r \in \{0,\dots,r-1\}^n$ the residue class of $-\vec{k}$ modulo r, and

$$c(k) = \frac{k^{\lfloor k/r\rfloor}}{\lfloor k/r\rfloor!}.$$

We have used $\hat{x} = z^r - \log z$ and $y = (r^{1+s/r}/s)z^s$ here, rather than $\hat{x} = -z^r + \log z$ and $y = z^s$ as used in [37], because the convention for the kernel $K(p_1, p)$ used here differs by sign from [37], and also to remove a factor of $(r^{1+s/r}/s)^{2-2g-n}$ from the correlators. Chiodo's formula and the asymptotic expansion (42) are true for any $s \in \mathbb{Z}$; hence, (43) holds for any $s \in \mathbb{Z}$, although it is stated only for $s \geq 0$ in [37].

In [37], $(-\vec{k})_r \in \{1,\dots,r\}^n$; however, replacing $k_i = r$ by $k_i = 0$ leaves the Chiodo class invariant since it does not change the component, but rather it twists the universal bundle $\mathcal{E}$ over the component, resulting in adding a direct summand of a trivial bundle to the virtual bundle $-R^*\pi_*\mathcal{E}$ which does not affect the total Chern class. The invariance of the total Chern class, or equivalently the positive-degree terms of the Chern character, can also be seen in Chiodo's formula via properties of the Bernoulli polynomials.

We will use (43) in the case $r = 2$. Define

$$\hat{\xi}^0_{-1} = 2z\,dz, \quad \hat{\xi}^1_{-1} = dz, \quad \hat{\xi}^\sigma_m(z) = -d\biggl(\frac{\hat{\xi}^\sigma_{m-1}(z)}{d\hat{x}(z)}\biggr), \quad \sigma\in\{0,1\},\ m\in\{0,1,2,\dots\},$$

which have local expansion at $z = 0$ given by

$$\hat{\xi}^\sigma_m(z) \sim \sum_{\substack{k\in\mathbb{Z}_+\\ k\equiv\sigma\ (\mathrm{mod}\ 2)}} k^m c(k) d(e^{-k\hat{x}}).$$

Each ψ_i in the denominator of the right-hand side of (43) produces monomials $\bigl(\frac{1}{2}k_i\psi_i\bigr)^{m_i}$; hence, (43) with $r=2$ becomes

$$\widehat{\omega}_{g,n}(z_1,\dots,z_n)=\sum_{\vec{\sigma},\vec{m}}\int_{\overline{\mathcal{M}}_{g,n}} C_{g,n}(2,s;\vec{\sigma})\prod_{i=1}^n \psi_i^{m_i}\hat{\xi}_{m_i}^{\sigma_i}(z_i)2^{\sigma/2_i-m_i}.$$

Change $(\hat{x},\hat{y})\mapsto(x,y)$ by

$$x=t\hat{x}\Bigl(\frac{z}{\sqrt{2t}}\Bigr)-\tfrac{1}{2}t\log(2t)=\tfrac{1}{2}z^2-t\cdot\log z,\quad y=\tfrac{1}{2}st^{s/2}\hat{y}\Bigl(\frac{z}{\sqrt{2t}}\Bigr)=z^s.$$

The differentials defined in (38) using x are given by

$$\xi_m^\sigma(z,t)=t^{-m-1/2}2^{\sigma/2}\hat{\xi}_m^\sigma\Bigl(\frac{z}{\sqrt{2t}}\Bigr).$$

Hence,

$$\begin{aligned}
&\omega_{g,n}(t,z_1,\dots,z_n)\\
&\quad=\bigl(\tfrac{1}{2}st^{s/2+1}\bigr)^{2-2g-n}\widehat{\omega}_{g,n}\Bigl(\frac{z_1}{\sqrt{2t}},\dots,\frac{z_n}{\sqrt{2t}}\Bigr)\\
&\quad=\bigl(\tfrac{1}{2}st^{s/2+1}\bigr)^{2-2g-n}\sum_{\vec{\sigma},\vec{m}}\int_{\overline{\mathcal{M}}_{g,n}} C_{g,n}(2,s;\vec{\sigma})\prod_{i=1}^n \psi_i^{m_i}\hat{\xi}_{m_i}^{\sigma_i}\Bigl(\frac{z_i}{\sqrt{2t}}\Bigr)2^{\sigma/2_i-m_i}\\
&\quad=\bigl(\tfrac{1}{2}st^{s/2+1}\bigr)^{2-2g-n}\sum_{\vec{\sigma},\vec{m}}\int_{\overline{\mathcal{M}}_{g,n}} C_{g,n}(2,s;\vec{\sigma})\prod_{i=1}^n t^{m_i+1/2}\psi_i^{m_i}\xi_{m_i}^{\sigma_i}(z_i)2^{-m_i}\\
&\quad=\bigl(\tfrac{1}{2}st^{s/2+1}\bigr)^{2-2g-n}t^{n/2}\sum_{\vec{\sigma},\vec{m}}\int_{\overline{\mathcal{M}}_{g,n}} C_{g,n}(2,s;\vec{\sigma})\prod_{i=1}^n \bigl(\tfrac{1}{2}t\bigr)^{m_i}\psi_i^{m_i}\xi_{m_i}^{\sigma_i}(z_i)\\
&\quad=\sum_{\vec{\sigma},\vec{m}} t^{(1-s)(2g-2+n)/2}2^{1-g}s^{2-2g-n}\int_{\overline{\mathcal{M}}_{g,n}} C_{g,n}\Bigl(2,s;\vec{\sigma},\frac{2}{t}\Bigr)\prod_{i=1}^n \psi_i^{m_i}\xi_{m_i}^{\sigma_i}(z_i,t),
\end{aligned}$$

where the last equality uses $\bigl(\frac{1}{2}t\bigr)^{\sum m_i}=\bigl(\frac{1}{2}t\bigr)^{3g-3+n-\deg}$ for the degree operator $\deg c_k(E_{g,n}^{\vec{\sigma}})=k$ then $\bigl(\frac{1}{2}t\bigr)^{-\deg}$ is absorbed into the Chern polynomial. Set $s=-1$ to get the desired result. □

The classes $\Theta_{g,n}$ arise in the limit

$$\lim_{t\to 0}\omega_{g,n}(t,z_1,\dots,z_n)=\sum_{\vec{m}}\int_{\overline{\mathcal{M}}_{g,n}}\Theta_{g,n}\prod_{i=1}^n \psi_i^{m_i}\xi_{m_i}(z)$$

for $\xi_m(z)=(2m+1)!!\,z^{-(2m+2)}\,dz$. We explain the relationship of this limit with Conjecture 1.5 in Proposition 6.1.

5.2.4 A_2 singularity In this section we calculate the spectral curves of the CohFT Ω^{A_2} and $(\Omega^{A_2})^{\Theta}$. We begin with a general result relating the spectral curve of any semisimple CohFT Ω with the spectral curve of Ω^{BGW}.

Proposition 5.17 *Given a semisimple CohFT Ω with partition function $Z_\Omega(\hbar,\{t_k^\alpha\})$ encoded by the spectral curve*

$$S=(C,x,y,B)$$

via Theorem 5.13, $Z_\Omega^{\mathrm{BGW}}(\hbar,\{t_k^\alpha\})$ is encoded by the spectral curve

$$\widehat{S}=\Big(C,x,\hat{y}=\frac{dy}{dx},B\Big).$$

Proof Note that the spectral curves S and $\widehat{S}$ share the same (C,x,B) and hence produce the same operator $\widehat{R}(z)$ used in the construction of both Z_Ω and Z_Ω^{BGW}.

Proposition 5.4 shows that a shift in the translation operator $T(z)\mapsto T(z)/z$ combined with replacing each copy of $Z^{\mathrm{KW}}(\hbar,\{t_k\})$ in (26) by a copy of $Z^{\Theta}(\hbar,\{t_k\})$ produces the partition function of Ω^{Θ}. It relied upon the homogeneity property (15) satisfied by $Z^{\Theta}(\hbar,\{t_k\})$. But $Z^{\mathrm{BGW}}(\hbar,\{t_k\})$ also satisfies (15); hence, an identical argument to that in Proposition 4.3 proves that, for a semisimple CohFT Ω, the partition function $Z_\Omega^{\mathrm{BGW}}(\hbar,\{t_k^\alpha\})$ is obtained by replacing each copy of $Z^{\mathrm{KW}}(\hbar,\{t_k\})$ in (26) by a copy of $Z^{\mathrm{BGW}}(\hbar,\{t_k\})$ and replacing the translation operator by $T(z)\mapsto T(z)/z$.

Given an irregular spectral curve, it is proven in [9] that its partition function is obtained from (26) with translation operator given by (36). Given a semisimple CohFT Ω encoded by the regular spectral curve $S=(C,x,y,B)$, define $\hat{y}=dy/dx$. Then, since $dy=\hat{y}\,dx$, the translation operator shifts by $T(z)^\alpha\mapsto T(z)^\alpha/z$, which proves that Ω^{BGW} is encoded by the spectral curve $\widehat{S}=(C,x,\hat{y}=dy/dx,B)$. □

Define the spectral curves

$$\begin{aligned} S_{A_2}&=\Big(\mathbb{C},\,x=z^3-3z,\,y=z\sqrt{-3},\,B=\frac{dz\,dz'}{(z-z')^2}\Big),\\ S_{A_2}^{\mathrm{BGW}}&=\Big(\mathbb{C},\,x=z^3-3z,\,\hat{y}=\frac{\sqrt{-3}}{3z^2-3},\,B=\frac{dz\,dz'}{(z-z')^2}\Big).\end{aligned}\tag{44}$$

The partition functions associated to $S=S_{A_2}$ defined in 4.1.4 and $S=S_{A_2}^{\Theta}$ are built out of correlators $\omega_{g,n}^S$ by

$$Z^S(\hbar,\{t_k^\alpha\})=\exp\sum_{g,n}\frac{\hbar^{g-1}}{n!}\omega_{g,n}^S\Big|_{\xi_k^\alpha(z_i)=t_k^\alpha}$$

using the differentials $\xi_k^\alpha(z)$ defined on $\mathbb{C}$ by

$$\xi_0^\alpha = \frac{dz}{(1-z)^2} - \frac{(-1)^\alpha dz}{(1+z)^2}, \quad \xi_{k+1}^\alpha(p) = d\left(\frac{\xi_k^\alpha(p)}{dx(p)}\right), \quad \alpha \in \{1,2\},\ k \in \mathbb{N}. \tag{45}$$

These are linear combinations of the $V_k^i(p)$ defined in (34) with $x = z^3 - 3z$. The $V_k^i(p)$ correspond to normalised canonical coordinates while the $\xi_k^\alpha(p)$ correspond to flat coordinates. We have

$$Z_{\Omega^{A_2}} = Z^{S_{A_2}}, \quad Z_{(\Omega^{A_2})^\Theta} = Z^{S_{A_2}^\Theta}.$$

The equality $Z_{A_2} = Z^{S_{A_2}}$ was proven in [17]; hence, $Z_{(\Omega^{A_2})^\Theta} = Z^{S_{A_2}^\Theta}$ by Proposition 5.17. We verify this by giving the local expansions of B and $\hat{y}$ for S_{A_2}, which helps to deal with different normalisations in the references. Choose a local coordinate t around $z = -1 = \mathcal{P}_1$ so that $x(t) = \frac{1}{2}t^2 + 2$. Then

$$\begin{aligned}
B(\mathcal{P}_1, t) &= \frac{-i}{\sqrt{6}} \frac{dz}{(z+1)^2} = dt\big(t^{-2} - \tfrac{1}{144} + \tfrac{35}{41\,472}t^2 + \cdots + \text{odd terms}\big), \\
B(\mathcal{P}_2, t) &= \frac{1}{\sqrt{6}} \frac{dz}{(z-1)^2} = dt\big(-\tfrac{i}{24} + \tfrac{35i}{3456}t^2 + \cdots + \text{odd terms}\big).
\end{aligned}$$

Around $z = 1 = \mathcal{P}_2$, the local expansions of $B(\mathcal{P}_\alpha, z)$ are the same as those above, up to sign. The odd terms are annihilated by the Laplace transform, and we get

$$\begin{aligned}
R^{-1}(z)_\alpha^\alpha &= -\frac{\sqrt{z}}{\sqrt{2\pi}} \int_{\Gamma_\alpha} B(\mathcal{P}_\alpha, t) \cdot e^{(-t^2/2)/z} = 1 - (-1)^\alpha \tfrac{1}{144}z - \tfrac{35}{41\,472}z^2 + \cdots, \\
R^{-1}(z)_{3-\alpha}^\alpha &= -\frac{\sqrt{z}}{\sqrt{2\pi}} \int_{\Gamma_\alpha} B(\mathcal{P}_{3-\alpha}, t) \cdot e^{(-t^2/2)/z} = \tfrac{i}{24}z + (-1)^\alpha \tfrac{35i}{3456}z^2 + \cdots.
\end{aligned}$$

Hence, $R^{-1}(z) = I - R_1 z + (R_1^2 - R_2)z^2 + \cdots = I - R_1^T z + R_2^T z^2 + \cdots$ gives

$$R_1 = \frac{1}{144}\begin{pmatrix} -1 & -6i \\ -6i & 1 \end{pmatrix}, \quad R_2 = \frac{35}{41\,472}\begin{pmatrix} -1 & 12i \\ -12i & -1 \end{pmatrix},$$

which determines all other R_k via (21) and agrees with (23) for Ω^{A_2}.

The topological field theory is defined by $\{dy(\mathcal{P}_\alpha)\}$ for $i = 1, 2$. The translation operator $T(z)$ is determined by the (Laplace transform of the) local expansion of y given by (36). Moreover, Ω^{A_2} has flat unit, so in this case the odd expansions of dy is determined by $R^{-1}(z)\mathbb{1}$ via (37), and hence uniquely determined by the terms $dy(\mathcal{P}_\alpha)$ for $\alpha = 1, 2$. This is visible on the spectral curve by the fact that the poles of dy are dominated by the poles of dx, ie dy/dx has poles only at the zeros $\mathcal{P}_1$ and $\mathcal{P}_2$ of dx, and hence, by the Cauchy formula, dy satisfies

$$d\Big(\frac{dy}{dx}(p)\Big) = -\sum_{\alpha=1}^{N} \operatorname{Res}_{p'=\mathcal{P}_\alpha} \frac{dy}{dx}(p')B(p',p), \tag{46}$$

which is proven in [17] to imply (37). Thus, it remains to show that y defines the correct topological field theory, representing $\mathbb{1}$ in normalised canonical coordinates. The local expansion of $dy = \sqrt{-3}\,dz$ around $\mathcal{P}_1 = -1$ in the local coordinate $x(t) = \frac{1}{2}t^2 + 2$ is

$$dy = \sqrt{-3}\,dz = \Big(\frac{1}{\sqrt{2}} - \frac{5}{144\sqrt{2}}t^2 + \frac{385}{124\,416\,\sqrt{2}}t^4 + \cdots + \text{odd terms}\Big)\,dt$$

and around $\mathcal{P}_2 = 1$ replace t by it. Hence, the Laplace transform is

$$\begin{aligned}\Big\{\frac{1}{\sqrt{2\pi z}}\int_{\Gamma_\alpha} dy(p)\cdot e^{((x(\mathcal{P}_k)-x(p)))/z}\Big\} \\ &= R^{-1}(z)\mathbb{1} \\ &= \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix} + \frac{5}{144\sqrt{2}}\begin{pmatrix}-1\\ i\end{pmatrix}z + \frac{385}{41\,472\,\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}z^2 + \cdots.\end{aligned}$$

Note that $dy(\mathcal{P}_1) = \frac{1}{\sqrt{2}} = \sqrt{\mathbb{1}^1}$ and $dy(\mathcal{P}_2) = \frac{i}{\sqrt{2}} = \sqrt{\mathbb{1}^2}$ gives the unit $\mathbb{1}$, and hence the TFT. Thus, $S_{A_2} \mapsto (R(z), T(z), \mathbb{1})$ for Ω^{A_2} as required.

6 Progress towards a proof of Conjecture 1.5

A consequence of the homogeneity property (15) satisfied by both partition functions $Z^{\Theta}(\hbar, t_0, t_1, \dots)$ and $Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots)$ is that, for $g > 1$, the coefficient of $\hbar^{g-1}$ of the logarithm of the partition function, ie its genus g part, is a *finite* sum of rational functions. They are both of the form

$$\log Z(\hbar, t_0, t_1, \dots) = -\tfrac{1}{8}\log(1-t_0) + \sum_{g=2}^{\infty} \hbar^{g-1} \sum_{\mu \vdash g-1} \frac{c_\mu t_\mu}{(1-t_0)^{2g-2+n}},$$

where $t_\mu := \prod t_{\mu_i}$ for a partition $\mu = (\mu_1, \dots, \mu_n)$. Hence, for each g, one needs only match the finite set of coefficients c_μ, parametrised by partitions μ of $g-1$, of $\log Z^{\Theta}(\hbar, t_0, t_1, \dots)$ with those of $\log Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots)$, to determine equality.

The initial value of $\int_{\overline{\mathcal{M}}_{1,1}} \Theta_{1,1} = \frac{1}{8}$ together with (15) produces all genus 1 terms of $\log Z^{\Theta}$, and the calculation $\int_{\overline{\mathcal{M}}_{2,1}} \Theta_{2,1}\cdot\psi_1 = \frac{3}{128}$ from Example 3.5 together with (15) produces all genus 2 terms, giving

$$\log Z^{\Theta} = -\tfrac{1}{8}\log(1-t_0) + \hbar\cdot\tfrac{3}{128}\cdot\frac{t_1}{(1-t_0)^3} + O(\hbar^2).$$

Further calculations, such as the genus 3 calculation in the appendix and calculations up to $g=7$ and $n=6$ using `admcycles` [12], prove

$$\log Z^{\Theta}(\hbar, t_0, t_1, \dots) = \log Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots) + O(\hbar^8). \tag{47}$$

Conjecture 1.5 is reduced to a purely combinatorial or analytic problem in the following proposition. Recall the spectral curve (39) given by

$$x = \tfrac{1}{2}z^2 - t\cdot\log z, \quad y = z^{-1}, \quad B = \frac{dz\,dz'}{(z-z')^2}$$

with correlators $\omega_{g,n}(t, z_1, \dots, z_n)$.

Proposition 6.1 *Conjecture 1.5 is equivalent to*

$$\lim_{t\to 0}\omega_{g,n}(t, z_1, \dots, z_n) = \omega^{\mathrm{Bes}}_{g,n}(z_1, \dots, z_n). \tag{48}$$

Proof By Theorem 5.16,

$$\omega_{g,n}(t, z_1, \dots, z_n) = \sum_{\vec{\sigma},\vec{m}}(-1)^n t^{2g-2+n}2^{1-g}\int_{\overline{\mathcal{M}}_{g,n}} p_*c\Big(E^{\vec{\sigma}}_{g,n}, \frac{2}{t}\Big)\prod_{i=1}^n \psi_i^{m_i}\xi^{\sigma_i}_{m_i}(z_i, t),$$

which is regular in t since

$$\operatorname{rank} E^{\vec{\sigma}}_{g,n} = 2g-2+\tfrac{1}{2}(n+|\vec{\sigma}|),$$

so the Chern polynomial has degree at most $2g-2+n$ in t^{-1}. Hence, for $|\vec{\sigma}|=n$,

$$\lim_{t\to 0}(-1)^n t^{2g-2+n}2^{1-g}p_*c\Big(E^{\vec{\sigma}}_{g,n}, \frac{2}{t}\Big) = (-1)^n 2^{g-1+n}p_*c_{2g-2+n}(E^{\vec{\sigma}}_{g,n}) = \Theta_{g,n},$$

while, for $|\vec{\sigma}|<n$, $\operatorname{rank} E^{\vec{\sigma}}_{g,n} < 2g-2+n$, so

$$\lim_{t\to 0}(-1)^n t^{2g-2+n}2^{1-g}p_*c\Big(E^{\vec{\sigma}}_{g,n}, \frac{2}{t}\Big) = 0.$$

Thus, the $t\to 0$ limit exists to give

$$\lim_{t\to 0}\sum_{\vec{\sigma},\vec{m}}\int_{\overline{\mathcal{M}}_{g,n}}(-1)^n t^{2g-2+n}2^{1-g}p_*c\Big(E^{\vec{\sigma}}_{g,n}, \frac{2}{t}\Big)\prod_{i=1}^n\psi_i^{m_i}\xi^{\sigma_i}_{m_i}$$
$$= \sum_{\vec{m}}\int_{\overline{\mathcal{M}}_{g,n}}\Theta_{g,n}\prod_{i=1}^n\psi_i^{m_i}\xi_{m_i}(z)$$

for

$$\xi_m(z) = \lim_{t\to 0} \xi_m^1(z,t) = (2m+1)!!z^{-(2m+2)}\,dz.$$

Also, $\lim_{t\to 0} \xi_m^0(z,t) = 0$ for $m \geq 0$. The $t \to 0$ limit of the spectral curve (39) gives the Bessel spectral curve of Example 5.15 with correlators proven in [14] to be given by

$$\omega_{g,n}^{\mathrm{Bes}}(z_1,\ldots,z_n) = \sum_{\vec{m}} \frac{\partial^n F^{\mathrm{BGW}}(\hbar,\{t_k\})}{\partial t_{m_1}\cdots\partial t_{m_n}} \prod_{i=1}^n \xi_{m_i}(z).$$

Hence, the conjectured limit (48) yields

$$\sum_{\vec{m}} \int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \prod_{i=1}^n \psi_i^{m_i} \xi_{m_i}(z) = \sum_{\vec{m}} \frac{\partial^n F^{\mathrm{BGW}}(\hbar,\{t_k\})}{\partial t_{m_1}\cdots\partial t_{m_n}} \prod_{i=1}^n \xi_{m_i}(z),$$

which is equivalent to Conjecture 1.5. □

The subtlety of the limit (48), which is known up to $g = 7$ for all n by the verification of Conjecture 1.5 in these cases, can be seen as follows. The correlators are regular in t; for example,

$$\omega_{0,3}(t,z_1,z_2,z_3) = O(t) \implies \lim_{t\to 0} \omega_{0,3}(t,z_1,z_2,z_3) = 0.$$

However, the coefficients in the recursion can be irregular in t, ie blow up as $t \to 0$. For example, we next introduce the parameter a to keep track of the contribution of $\omega_{0,3}(t,z_1,z_2,z_3)$ and can set $a = 1$ at the end in this calculation of $\omega_{1,2}(t,z_1,z_2)$:

$$\begin{aligned}\omega_{1,2}&(t,z_1,z_2)\\ &= \sum_{dx(\alpha)=0} \mathrm{Res}_{z=\alpha} K(z_1,z)\big[a\cdot\omega_{0,3}(t,z,\sigma_\alpha(z),z_2) + \omega_{0,2}(z,z_2)\omega_{1,1}(t,\sigma_\alpha(z))\\ &\qquad\qquad\qquad\qquad + \omega_{0,2}(\sigma_\alpha(z),z_2)\omega_{1,1}(t,z)\big],\end{aligned}$$

$$\lim_{t\to 0} \omega_{1,2}(t,z_1,z_2) = \tfrac{1}{1080}(74a+61)\,\frac{dz_1\,dz_2}{z_1^2 z_2^2}.$$

This gives the expected limit of $\omega_{1,2}^{\mathrm{Bes}}(z_1,z_2)$ when $a = 1$, and shows the dependence of $\lim_{t\to 0}\omega_{1,2}(t,z_1,z_2)$ on $\omega_{0,3}(t,z_1,z_2,z_3)$ due to coefficients in the recursion which are irregular in t.

6.1 Pixton relations

A collection of relations in the tautological ring $RH^*(\overline{\mathcal{M}}_{g,n})$ was conjectured by Pixton and proven in [47] using the CohFT Ω^{A_2}. Such tautological relations can be used to produce topological recursion relations for CohFTs such as Gromov–Witten invariants. Similarly, the intersections of $\Theta_{g,n}$ with Pixton's relations produce topological recursion relations satisfied by the intersection numbers $\int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \prod_{i=1}^{n} \psi_i^{m_i}$.

The key idea behind the proof of Pixton's relations in [47] is a degree bound on the cohomology classes

$$\deg \Omega_{g,n}^{A_2} \leq \tfrac{1}{3}(g-1+n) < 3g-3+n$$

combined with Givental's construction of $\Omega_{g,n}^{A_2}$ in Definition 5.2 from the triple $(R(z), T(z), \mathbb{1}) \in L^{(2)}\mathrm{GL}(N,\mathbb{C}) \times z^2\mathbb{C}^N[\![z]\!] \times \mathbb{C}^N$ obtained from the Frobenius manifold structure on the versal deformation space of the A_2 singularity; see Section 4.1.4. Givental's construction produces $\Omega_{g,n}^{A_2}$, although it does not know about the degree bound and produces classes in the degrees where $\Omega_{g,n}^{A_2}$ vanishes. This leads to sums of tautological classes representing the zero class, ie relations given by the degree $d > \frac{1}{3}(g-1+n)$ part of the sum over stable graphs in (24) of the form

$$\Omega_{g,n}^{A_2} = \sum_{\Gamma \in G_{g,n}} \frac{1}{|\mathrm{Aut}(\Gamma)|} (\phi_\Gamma)_* \omega_\Gamma^{R,T,\mathbb{1}}.$$

Since Ω^{A_2} has flat unit, the pushforward classes in (24) produce κ polynomials; hence, only graphs without dilaton leaves in the sum are required and the classes $\omega_\Gamma^{R,T,\mathbb{1}}$ consist of products of ψ and κ classes associated to each vertex of Γ. The main result of [47] is the construction of elements $R_{g,A}^d \in S_{g,n}$ for $A = (a_1, \dots, a_n)$ with $a_\alpha \in \{0,1\}$ satisfying $q(R_{g,A}^d) = 0$ which push forward to tautological relations in $H^{2d}(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})$. They are defined by $R_{g,A}^d$, the degree d part of $\Omega_{g,n}^{A_2}(v_A)$ for a basis $\{v_0, v_1\}$. The element $R_2^1 \in H^2(\overline{\mathcal{M}}_2, \mathbb{Q})$ is given in Example 3.5.

When $n \leq g-1$ and $g > 1$, we have $d = g-1 > \frac{1}{3}(g-1+n)$; hence, there exist nontrivial relations $R_{g,A}^{g-1}$. This produces the sum over graphs

$$\Theta_{g,n} \cdot R_{g,A}^{g-1} = 0,$$

which defines a relation for each A between intersection numbers of ψ classes with $\Theta_{g,n}$, ie coefficients of $Z^\Theta(\hbar, \{t_k\})$. This uses $\Theta_{g,n} \cdot (\phi_\Gamma)_* = (\phi_\Gamma)_* \Theta_\Gamma$ together with Remark 3.4 to replace κ classes by ψ classes. We saw this in Example 3.5, arising from the genus two Pixton relation

$$\int_{\overline{\mathcal{M}}_{2,1}} \Theta_{2,1} \cdot \psi_1 - \tfrac{7}{10} \cdot \int_{\overline{\mathcal{M}}_{1,1}} \Theta_{1,1} \cdot \int_{\overline{\mathcal{M}}_{1,1}} \Theta_{1,1} - \tfrac{1}{10} \cdot \int_{\overline{\mathcal{M}}_{1,2}} \Theta_{1,2} = 0, \tag{49}$$

which determines $\int_{\overline{\mathcal{M}}_{2,1}} \Theta_2 \cdot \psi_1^2$ from $\int_{\overline{\mathcal{M}}_{1,1}} \Theta_{1,1}$ and $\int_{\overline{\mathcal{M}}_{1,2}} \Theta_{1,2}$. Similarly, the appendix uses genus three relations to deduce $\int_{\overline{\mathcal{M}}_{3,2}} \Theta_{3,2} \cdot \psi_1^2$ and $\int_{\overline{\mathcal{M}}_{3,2}} \Theta_{3,2} \cdot \psi_1 \psi_2$ from lower-genus coefficients of $Z^{\Theta}(\hbar, t_0, t_1, \dots)$.

The following theorem proves that the coefficients of $Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots)$ also satisfy (49), and more generally an infinite set of relations satisfied by coefficients of $Z^{\Theta}(\hbar, t_0, t_1, \dots)$ arising from Pixton relations:

Theorem 6.2 *Pixton relations produce infinitely many nontrivial relations satisfied by the coefficients of both $Z^{\Theta}(\hbar, t_0, t_1, \dots)$ and $Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots)$.*

Proof For each $g > 1$, n and $\lfloor \frac{1}{2}(n+1) \rfloor$ possible $A \in \{0,1\}^n$ (due to symmetry and vanishing of half for parity reasons), $R_{g,A}^{g-1} = 0$ defines a nontrivial Pixton relation. For each of these choices of g, n and A, due to the restriction and pullback properties of $\Theta_{g,n}$ as explained above, $\Theta_{g,n} \cdot R_{g,A}^{g-1} = 0$ defines a relation between coefficients of $Z^{\Theta}(\hbar, \{t_k\})$, such as (49).

The main goal is to prove that the corresponding coefficients of $Z^{\mathrm{BGW}}(\hbar, \{t_k\})$ also satisfy this infinite set of relations. To do this, we study the partition function $Z^{\mathrm{BGW}}_{\Omega^{A_2}}$, defined in Definition 5.6 via the spectral curve $S^{\mathrm{BGW}}_{A_2}$ defined in (44). The relations between coefficients of $Z^{\mathrm{BGW}}(\hbar, \{t_k\})$ will be stored in the spectral curve. This will produce identical relations satisfied by both the coefficients of Z^{BGW} and Z^{Θ}. To summarise, we have vanishing of certain coefficients of $Z^{\Theta}_{A_2}(\hbar, \{t_k^{\alpha}\})$ due to the cohomological viewpoint shown in the upper row in Figure 1, and vanishing of corresponding coefficients of $Z^{\mathrm{BGW}}_{\Omega^{A_2}}(\hbar, \{t_k^{\alpha}\})$ due to Givental's construction neatly encoded by topological recursion shown in the lower row in Figure 1.

Pixton relations induce relations between intersection numbers of ψ and κ classes or ψ classes alone, ie coefficients of $Z^{\mathrm{KW}}(\hbar, \{t_k\})$. These relations are realised by unexpected vanishing of coefficients of the partition function $Z_{A_2}(\hbar, \{t_k^{\alpha}\})$. Similarly, unexpected vanishing of coefficients of the partition function $Z^{\mathrm{BGW}}_{A_2}(\hbar, \{t_k^{\alpha}\})$ correspond to relations between coefficients of $Z^{\mathrm{BGW}}(\hbar, \{t_k\})$.

The coefficients of $\log Z^{\mathrm{BGW}}_{A_2}(\hbar, \{t_k^{\alpha}\})$ are obtained from the correlators $\omega_{g,n}^{\mathrm{BGW},A_2}$ of $S^{\mathrm{BGW}}_{A_2}$ by

$$\text{(50)} \quad \frac{\partial^n}{\partial t_{k_1}^{\alpha_1} \cdots \partial t_{k_n}^{\alpha_n}} (F^{\mathrm{BGW}}_{A_2})_g(\{t_k^{\alpha}\}) \bigg|_{t_k^{\alpha}=0} = \mathrm{Res}_{z_1=\infty} \cdots \mathrm{Res}_{z_n=\infty} \prod_{i=1}^{n} p_{\alpha_i,k_i}(z_i) \omega_{g,n}^{\mathrm{BGW},A_2}(z_1, \dots, z_n)$$

for polynomials $p_{\alpha,k}(z)=\sqrt{-3}((-1)^\alpha/\alpha)z^{3k+\alpha}+\text{lower-order terms}$ for $\alpha\in\{1,2\}$ and $k\in\mathbb{N}$ chosen so that the residues are dual to the differentials ξ_k^α defined in (45). The lower-order terms (and the top coefficient) will not be important here because we will only consider vanishing of (50) arising from high-enough-order vanishing of $\omega_{g,n}^{A_2}(z_1,\dots,z_n)$ at $z_i=\infty$, so that the integrand in (50) is holomorphic at $z_i=\infty$. Equation (50) is a special case of the more general phenomena, proven in [16], that periods of $\omega_{g,n}$ are dual to insertions of vectors in a CohFT. Thus, we have shown that *relations between coefficients of $Z^{\mathrm{BGW}}(\hbar,\{t_k\})$ induced from Pixton relations are detected by high-order vanishing of $\omega_{g,n}^{\mathrm{BGW},A_2}(z_1,\dots,z_n)$ at $z_i=\infty$*. The same is true for high-order vanishing $\omega_{g,n}^{A_2}(z_1,\dots,z_n)$ at $z_i=\infty$, which is shown by

$$\omega_{2,1}^{A_2}(z)=\tfrac{35}{243}\cdot\frac{z(11z^4+14z^2+2)}{(z^2-1)^{10}}\,dz \\ \implies\quad \operatorname{Res}_{z=\infty} z^m\omega_{2,1}^{A_2}(z)=0,\quad m\in\{0,1,\dots,12\}.$$

Hence, (50) vanishes for $k_1=0,1,2,3$ and $\alpha_1\equiv k_1 \bmod 2$, which gives the relations between intersection numbers, or coefficients of $Z^{\mathrm{KW}}(\hbar,\{t_k\})$,

$$\int_{\overline{\mathcal{M}}_{2,1}} R^d_{2,\bar d}\psi_1^{4-d}=0,\quad d=1,2,3,4, \tag{51}$$

where $R^d_{2,\bar d}$ is a nontrivial Pixton relation, for $\bar d\equiv d \bmod 2$, between cohomology classes in $H^{2d}(\overline{\mathcal{M}}_{2,1},\mathbb{Q})$ proven in [47], such as $R^2_{2,0}=\psi_1^2+\text{boundary terms}=0$.

Lemma 6.3 *We have*

$$\sum_{i=1}^n \operatorname{ord}_{z_i=\infty}\omega_{g,n}^{\mathrm{BGW},A_2}(z_1,\dots,z_n)\ge 2g-2,$$

where $\operatorname{ord}_{z=\infty}\eta(z)$ is the order of vanishing of the differential at $z=\infty$.

Proof We can make the rational differential

$$\omega_{g,n}^{A_2}(z_1,\dots,z_n)=\frac{p_{g,n}(z_1,\dots,z_n)}{\prod_{i=1}^n(z_i^2-1)^{2g}}\,dz_1\cdots dz_n$$

homogeneous by applying topological recursion to $x(z)=z^3-3Q^2z$ and $y=\sqrt{-3}/x'(z)$ which are homogeneous in z and Q. Then $\omega_{g,n}^{A_2}(Q,z_1,\dots,z_n)$ is homogeneous in z and Q of degree $2-2g-n$:

$$\omega_{g,n}^{A_2}(Q,z_1,\dots,z_n)=\lambda^{2-2g-n}\omega_{g,n}^{A_2}(\lambda Q,\lambda z_1,\dots,\lambda z_n).$$

The degree of homogeneity uses the fact that

$$(z,Q)\mapsto(\lambda z,\lambda Q)\implies y\,dx\mapsto\lambda y\,dx\implies \omega_{g,n}\mapsto\lambda^{2-2g-n}\omega_{g,n}$$

because $y\,dx$ appears in the kernel $K(p_1, p)$ with homogeneous degree -1, which easily leads to degree $2-2g-n$ for $\omega_{g,n}$. The degree $2-2g-n$ homogeneity of

$$\omega_{g,n}^{A_2}(Q, z_1, \dots, z_n) = \frac{p_{g,n}(Q, z_1, \dots, z_n)}{\prod_{i=1}^n (z_i^2 - Q^2)^{2g}}\, dz_1 \cdots dz_n$$

implies that $\deg p_{g,n}(Q, z_1, \dots, z_n) = 4gn - n + 2 - 2g - n$. But we also know that $\omega_{g,n}^{A_2}(Q, z_1, \dots, z_n)$ is well defined as $Q \to 0$ — the limit becomes $\omega_{g,n}$ of the spectral curve $x(z) = z^3$ and $y = \sqrt{-3}/x'(z)$ using the topological recursion defined by Bouchard and Eynard [5] — so $\deg p_{g,n}(z_1, \dots, z_n) \le 4gn - n + 2 - 2g - n$. Note that dz_i is homogeneous of degree 1 but has a pole of order 2 at $z_i = \infty$; hence,

$$\sum_{i=1}^n \operatorname{ord}_{z_i=\infty} \omega_{g,n}^{\mathrm{BGW},A_2}(z_1, \dots, z_n) = 4gn - \deg p_{g,n}(z_1, \dots, z_n) - 2n \ge 2g - 2. \quad \square$$

Primary invariants of a partition function are those coefficients of $\prod_{i=1}^n t_{k_i}^{\alpha_i}$ with all $k_i = 0$. They correspond to intersections in $\overline{\mathcal{M}}_{g,n}$ with no ψ classes. The primary invariants of $Z_{A_2}^{\Theta}(\hbar, \{t_k^\alpha\})$ vanish for $n < 2g - 2$. This uses $\deg \Omega_{g,n}^{A_2} \le \frac{1}{3}(g - 1 + n)$, so $\deg \Omega_{g,n}^{A_2} \cdot \Theta_{g,n} \le \frac{1}{3}(g - 1 + n) + 2g - 2 + n < 3g - 3 + n$ when $n < 2g - 2$. These vanishing coefficients correspond to the relations $\Theta_{g,n} \cdot R_{g,A}^{g-1} = 0$, which, as discussed above, give relations between coefficients of $Z^{\Theta}(\hbar, \{t_k\})$.

The primary coefficients of $Z_{A_2}^{\mathrm{BGW}}(\hbar, \{t_k^\alpha\})$ correspond to

$$\operatorname{Res}_{z_1=\infty} \cdots \operatorname{Res}_{z_n=\infty} \prod_{i=1}^n z_i^{\epsilon_i}\, \omega_{g,n}^{A_2}(z_1, \dots, z_n)$$

for $\epsilon_i = 1$ or 2. Different choices of ϵ_i give different relations (except half which vanish for parity reasons). By Lemma 6.3, $\sum_{i=1}^n \operatorname{ord}_{z_i=\infty} \omega_{g,n}^{\mathrm{BGW},A_2}(z_1, \dots, z_n) \ge 2g - 2$, so, for $n < 2g - 2$, there exists an i such that $\operatorname{ord}_{z_i=\infty} \omega_{g,n}^{\mathrm{BGW},A_2}(z_1, \dots, z_n) \ge 2$. Hence, $z_i^{\epsilon_i}\, \omega_{g,n}^{A_2}(z_1, \dots, z_n)$ is holomorphic at $z_i = \infty$, so

$$\operatorname{Res}_{z_i=\infty} z_i^{\epsilon_i}\, \omega_{g,n}^{A_2}(z_1, \dots, z_n) = 0$$

and we have

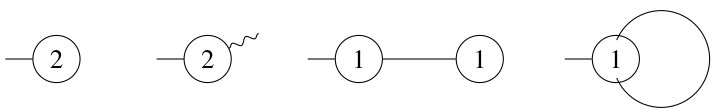
$$n < 2g - 2 \implies \operatorname{Res}_{z_1=\infty} \cdots \operatorname{Res}_{z_n=\infty} \prod_{i=1}^n z_i^{\epsilon_i}\, \omega_{g,n}^{A_2}(z_1, \dots, z_n) = 0. \tag{52}$$

Hence, the primary coefficients of $Z_{A_2}^{\mathrm{BGW}}(\hbar, \{t_k^\alpha\})$ vanish for $n < 2g - 2$, yielding a common set of relations satisfied by both the coefficients of $Z^{\Theta}(\hbar, t_0, t_1, \dots)$ and $Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots)$. $\square$

An example of a genus 2 relation produced by Theorem 6.2 is

$$\omega_{2,1}^{\mathrm{BGW},A_2}(z) = \frac{-5z^2-1}{16\sqrt{-3}(z-1)^4(z+1)^4}\,dz.$$

It immediately follows that $\operatorname{Res}_{z=\infty} \frac{\sqrt{-3}}{2} z \cdot \omega_{2,1}(z) = 0$, which signifies a relation between coefficients of $Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots)$. We will write the relations using $\Theta_{g,n}$; however, the relations are between coefficients of $Z^{\mathrm{BGW}}(\hbar, t_0, t_1, \dots)$ and what we are showing here is that these coefficients satisfy the same relations as intersection numbers involving $\Theta_{g,n}$, or, equivalently, coefficients of $Z^{\Theta}(\hbar, t_0, t_1, \dots)$. The graphical expansion encoded by both Givental's construction and topological recursion is given by



(plus graphs containing genus 0 vertices on which $\Theta_{2,1}$ vanishes), which contributes

$$2^2 \cdot \tfrac{60}{1728} \cdot \int_{\overline{\mathcal{M}}_{2,1}} \Theta_{2,1} \cdot \psi_1 + 2^2 \cdot \tfrac{-60}{1728} \cdot \int_{\overline{\mathcal{M}}_{2,1}} \Theta_{2,1} \cdot \kappa_1 \\ + 2^2 \cdot \tfrac{84}{1728} \cdot \int_{\overline{\mathcal{M}}_{1,2}} \Theta_{1,2} \cdot \int_{\overline{\mathcal{M}}_{1,1}} \Theta_{1,1} + \tfrac{2}{2} \cdot \tfrac{84-60}{1728} \cdot \int_{\overline{\mathcal{M}}_{1,3}} \Theta_{1,3},$$

which agrees with the expansion in weighted graphs of $\operatorname{Res}_{z=\infty} \frac{\sqrt{-3}}{2} z \cdot \omega_{2,1}(z) = 0$ given by

$$\tfrac{5}{1536} - \tfrac{15}{1536} + \tfrac{7}{2304} + \tfrac{1}{288} = 0.$$

Appendix Calculations

Here we show explicitly the equality $Z^{\mathrm{BGW}} = Z^{\Theta}$ up to genus 3. The coefficients of the Brézin–Gross–Witten tau function are calculated recursively since it is a tau function of the KdV hierarchy. It has low genus g (= coefficient of $\hbar^{g-1}$) terms given by

$$\log Z^{\mathrm{BGW}} = -\tfrac{1}{8}\log(1-t_0) + \hbar \cdot \tfrac{3}{128} \cdot \frac{t_1}{(1-t_0)^3} + \hbar^2 \cdot \tfrac{15}{1024} \cdot \frac{t_2}{(1-t_0)^5} \\ + \hbar^2 \cdot \tfrac{63}{1024} \cdot \frac{t_1^2}{(1-t_0)^6} + O(\hbar^3) \\ = \tfrac{1}{8}t_0 + \tfrac{1}{16}t_0^2 + \cdots + \hbar\big(\tfrac{3}{128}t_1 + \tfrac{9}{128}t_0t_1 + \cdots\big) + \hbar^2\big(\tfrac{15}{1024}t_2 + \tfrac{63}{1024}t_1^2 + \cdots\big).$$

The intersection numbers of $\Theta_{g,n}$ stored in

$$\log Z^{\Theta}(\hbar, t_0, t_1, \dots) = \sum_{g,n,\vec{k}} \frac{\hbar^{g-1}}{n!} \int_{\overline{\mathcal{M}}_{g,n}} \Theta_{g,n} \cdot \prod_{j=1}^{n} \psi_j^{k_j} \prod t_{k_j}$$

are calculated recursively via relations among tautological classes in $H^*(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$. The calculation of these intersection numbers up to genus 2 can be found throughout the text. We assemble them here for convenience, then present the genus 3 calculations.

$\boldsymbol{g=0}$ Theorem 1.3(II) gives $\Theta_{0,n}=0$, which agrees with the vanishing of all genus 0 terms in Z^{BGW}.

$\boldsymbol{g=1}$ Proposition 2.9 gives $\Theta_{1,1}=3\psi_1$; hence, $\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}=\frac{1}{8}$. We use this together with the dilaton equation to get $\int_{\overline{\mathcal{M}}_{1,n}}\Theta_{1,n}=\frac{1}{8}(n-1)!$. This agrees with $-\frac{1}{8}\log(1-t_0)$ in $\log Z^{\mathrm{BGW}}$.

$\boldsymbol{g=2}$ Using Mumford's relation [41], κ_1 is the sum of boundary terms in $\overline{\mathcal{M}}_2$, which coincides with a genus 2 Pixton relation; Example 3.5 produced the genus 2 intersection numbers from the genus 1 intersection numbers:

$$\begin{aligned}\int_{\overline{\mathcal{M}}_2}\Theta_2\cdot\kappa_1&=\tfrac{7}{5}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\frac{1}{|\mathrm{Aut}(\Gamma_1)|}+\tfrac{1}{5}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}\cdot\frac{1}{|\mathrm{Aut}(\Gamma_2)|}\\&=\tfrac{7}{5}\cdot\tfrac{1}{8}\cdot\tfrac{1}{8}\cdot\tfrac{1}{2}+\tfrac{1}{5}\cdot\tfrac{1}{8}\cdot\tfrac{1}{2}=\tfrac{3}{128}.\end{aligned}$$

Note that $\int_{\overline{\mathcal{M}}_{2,1}}\Theta_{2,1}\cdot\psi_1=\int_{\overline{\mathcal{M}}_{2,1}}\pi^*\Theta_2\cdot\psi_1^2=\int_{\overline{\mathcal{M}}_2}\Theta_2\cdot\kappa_1$. Using the dilaton equation, we then get $\int_{\overline{\mathcal{M}}_{2,n}}\Theta_{2,n}\cdot\psi_1=\frac{3}{256}(n+1)!$, which agrees with the $\hbar\cdot\frac{3}{128}t_1/(1-t_0)^3$ term in $\log Z^{\mathrm{BGW}}$.

$\boldsymbol{g=3}$ There are two independent genus 3 Pixton relations expressing κ_2 and κ_1^2 as sums of boundary terms in $\overline{\mathcal{M}}_3$. The relations correspond to sums over stable graphs in $\overline{\mathcal{M}}_3$; hence, they contain many terms. In place of these, we use the equivalent relations discovered earlier in [34; 35], which push forward to relations in $\overline{\mathcal{M}}_3$. In $\overline{\mathcal{M}}_{3,1}$, we can write ψ_1^3 as a sum of boundary terms, which yields

$$\begin{aligned}&\int_{\overline{\mathcal{M}}_{3,1}}\Theta_{3,1}\cdot\psi_1^2\\&=\int_{\overline{\mathcal{M}}_{3,1}}\pi^*\Theta_3\cdot\psi_1^3\\&=\tfrac{41}{21}\cdot\int_{\overline{\mathcal{M}}_{2,1}}\Theta_{2,1}\cdot\psi_1\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}+\tfrac{5}{42}\cdot\int_{\overline{\mathcal{M}}_{2,2}}\Theta_{2,2}\cdot\psi_1\\&\quad-\tfrac{1}{105}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,3}}\Theta_{1,3}\cdot\frac{1}{|\mathrm{Aut}|}+\tfrac{11}{70}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}\cdot\frac{1}{|\mathrm{Aut}|}\\&\quad-\tfrac{4}{35}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}-\tfrac{1}{105}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,3}}\Theta_{1,3}\cdot\frac{1}{|\mathrm{Aut}|}\\&\quad-\tfrac{1}{1260}\cdot\int_{\overline{\mathcal{M}}_{1,4}}\Theta_{1,4}\cdot\frac{1}{|\mathrm{Aut}|}\end{aligned}$$

$$= \tfrac{41}{21}\cdot\tfrac{3}{128}\cdot\tfrac{1}{8}+\tfrac{5}{42}\cdot\tfrac{9}{128}-\tfrac{1}{105}\cdot\tfrac{1}{8}\cdot\tfrac{2}{8}\cdot\tfrac{1}{2}+\tfrac{11}{70}\cdot\tfrac{1}{8}\cdot\tfrac{1}{8}\cdot\tfrac{1}{2}-\tfrac{4}{35}\cdot\tfrac{1}{8}\cdot\tfrac{1}{8}\cdot\tfrac{1}{8}-\tfrac{1}{105}\cdot\tfrac{1}{8}\cdot\tfrac{2}{8}\cdot\tfrac{1}{2}-\tfrac{1}{1260}\cdot\tfrac{6}{8}\cdot\tfrac{1}{4}$$
$$= \tfrac{15}{1024}.$$

In $\overline{\mathcal{M}}_{3,2}$, we can write $\psi_1^2\psi_2-\psi_1\psi_2^2$ as a sum of boundary terms, which yields

$$\begin{aligned}
&7\int_{\overline{\mathcal{M}}_{3,2}}\Theta_{3,2}\cdot(\psi_1^2-\psi_1\psi_2)\\
&=7\int_{\overline{\mathcal{M}}_{3,2}}\pi^*\Theta_{3,1}\cdot(\psi_1^2\psi_2-\psi_1\psi_2^2)\\
&=-\tfrac{16}{3}\cdot\int_{\overline{\mathcal{M}}_{2,2}}\Theta_{2,2}\cdot\psi_2\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}-5\int_{\overline{\mathcal{M}}_{2,2}}\Theta_{2,2}\cdot\psi_1\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\\
&\qquad-\tfrac{40}{3}\cdot\int_{\overline{\mathcal{M}}_{2,1}}\Theta_{2,1}\cdot\psi_1\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}-\tfrac{1}{6}\cdot\int_{\overline{\mathcal{M}}_{2,3}}\Theta_{2,3}\cdot\psi_1-\int_{\overline{\mathcal{M}}_{2,3}}\Theta_{2,3}\cdot\psi_1\cdot\frac{1}{|\mathrm{Aut}|}\\
&\qquad-\tfrac{1}{15}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,4}}\Theta_{1,4}\cdot\frac{1}{|\mathrm{Aut}|}-\tfrac{9}{10}\cdot\int_{\overline{\mathcal{M}}_{1,3}}\Theta_{1,3}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}\\
&\qquad-\tfrac{1}{15}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,4}}\Theta_{1,4}\cdot\frac{1}{|\mathrm{Aut}|}+\tfrac{4}{15}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}\cdot\int_{\overline{\mathcal{M}}_{1,3}}\Theta_{1,3}\cdot\frac{1}{|\mathrm{Aut}|}\\
&\qquad-\tfrac{4}{5}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,3}}\Theta_{1,3}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\\
&\qquad+\tfrac{16}{5}\cdot\int_{\overline{\mathcal{M}}_{1,1}}\Theta_{1,1}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}\cdot\int_{\overline{\mathcal{M}}_{1,2}}\Theta_{1,2}-\tfrac{1}{180}\cdot\int_{\overline{\mathcal{M}}_{1,5}}\Theta_{1,5}\cdot\frac{1}{|\mathrm{Aut}|}\\
&=-\tfrac{16}{3}\cdot\tfrac{9}{128}\cdot\tfrac{1}{8}-5\tfrac{9}{128}\cdot\tfrac{1}{8}-\tfrac{40}{3}\cdot\tfrac{3}{128}\cdot\tfrac{1}{8}-\tfrac{1}{6}\cdot\tfrac{36}{128}-\tfrac{36}{128}\cdot\tfrac{1}{2}\\
&\qquad-\tfrac{1}{15}\cdot\tfrac{1}{8}\cdot\tfrac{6}{8}\cdot\tfrac{1}{2}-\tfrac{9}{10}\cdot\tfrac{2}{8}\cdot\tfrac{1}{8}-\tfrac{1}{15}\cdot\tfrac{1}{8}\cdot\tfrac{6}{8}\cdot\tfrac{1}{2}+\tfrac{4}{15}\cdot\tfrac{1}{8}\cdot\tfrac{2}{8}\cdot\tfrac{1}{2}-\tfrac{4}{5}\cdot\tfrac{1}{8}\cdot\tfrac{2}{8}\cdot\tfrac{1}{8}+\tfrac{16}{5}\cdot\tfrac{1}{8}\cdot\tfrac{1}{8}\cdot\tfrac{1}{8}-\tfrac{1}{180}\cdot\tfrac{24}{8}\cdot\tfrac{1}{4}\\
&=-\tfrac{357}{1024}.
\end{aligned}$$

Hence,

$$\int_{\overline{\mathcal{M}}_{3,2}}\Theta_{3,2}\cdot\psi_1\psi_2=\int_{\overline{\mathcal{M}}_{3,2}}\Theta_{3,2}\cdot\psi_1^2+\tfrac{1}{7}\tfrac{357}{1024}=\tfrac{75}{1024}+\tfrac{51}{1024}=\tfrac{63}{512},$$

where $\int_{\overline{\mathcal{M}}_{3,2}}\Theta_{3,2}\cdot\psi_1^2=\frac{75}{1024}$ is obtained from $\int_{\overline{\mathcal{M}}_{3,1}}\Theta_{3,1}\cdot\psi_1^2=\frac{15}{1024}$ via the dilaton equation. The dilaton equation then yields

$$\int_{\overline{\mathcal{M}}_{3,n}}\Theta_{3,n}\cdot\psi_1^2=\tfrac{75}{1024}\cdot\tfrac{1}{5!}(n+3)!\quad\text{and}\quad\int_{\overline{\mathcal{M}}_{3,n}}\Theta_{3,n}\cdot\psi_1\psi_2=\tfrac{63}{512}\cdot\tfrac{1}{5!}(n+3)!,$$

which agree with the terms $\hbar^2\cdot\frac{15}{1024}t_2/(1-t_0)^5+\hbar^2\cdot\frac{63}{1024}t_1^2/(1-t_0)^6$ in $\log Z^{\mathrm{BGW}}$.

References

[1] **D Abramovich**, **T J Jarvis**, *Moduli of twisted spin curves*, Proc. Amer. Math. Soc. 131 (2003) 685–699 MR Zbl

[2] **A Alexandrov**, *Cut-and-join description of generalized Brezin–Gross–Witten model*, Adv. Theor. Math. Phys. 22 (2018) 1347–1399 MR Zbl

[3] **J Bergström**, *Cohomology of moduli spaces of curves of genus three via point counts*, J. Reine Angew. Math. 622 (2008) 155–187 MR Zbl

[4] **J Bergström**, **O Tommasi**, *The rational cohomology of* $\overline{\mathcal{M}}_4$, Math. Ann. 338 (2007) 207–239 MR Zbl

[5] **V Bouchard**, **B Eynard**, *Think globally, compute locally*, J. High Energy Phys. (2013) art. id. 143 MR Zbl

[6] **E Brézin**, **D J Gross**, *The external field problem in the large* N *limit of QCD*, Phys. Lett. B 97 (1980) 120–124 MR

[7] **H-L Chang**, **J Li**, **W-P Li**, *Witten's top Chern class via cosection localization*, Invent. Math. 200 (2015) 1015–1063 MR Zbl

[8] **L Chekhov**, **B Eynard**, *Hermitian matrix model free energy: Feynman graph technique for all genera*, J. High Energy Phys. (2006) art. id. 014 MR Zbl

[9] **L Chekhov**, **P Norbury**, *Topological recursion with hard edges*, Internat. J. Math. 30 (2019) art. id. 1950014 MR Zbl

[10] **A Chiodo**, *Towards an enumerative geometry of the moduli space of twisted curves and* r*th roots*, Compos. Math. 144 (2008) 1461–1496 MR Zbl

[11] **E Clader**, **S Grushevsky**, **F Janda**, **D Zakharov**, *Powers of the theta divisor and relations in the tautological ring*, Int. Math. Res. Not. 2018 (2018) 7725–7754 MR Zbl

[12] **V Delecroix**, **J Schmitt**, **J van Zelm**, *admcycles: a Sage package for calculations in the tautological ring of the moduli space of stable curves*, J. Softw. Algebra Geom. 11 (2021) 89–112 Zbl

[13] **N Do**, **P Norbury**, *Topological recursion for irregular spectral curves*, J. Lond. Math. Soc. 97 (2018) 398–426 MR Zbl

[14] **N Do**, **P Norbury**, *Topological recursion on the Bessel curve*, Commun. Number Theory Phys. 12 (2018) 53–73 MR Zbl

[15] **B Dubrovin**, *Geometry of* 2*D topological field theories*, from "Integrable systems and quantum groups" (M Francaviglia, S Greco, editors), Lecture Notes in Math. 1620, Springer (1996) 120–348 MR Zbl

[16] **P Dunin-Barkowski**, **P Norbury**, **N Orantin**, **A Popolitov**, **S Shadrin**, *Primary invariants of Hurwitz Frobenius manifolds*, from "Topological recursion and its influence in analysis, geometry, and topology" (C-C M Liu, M Mulase, editors), Proc. Sympos. Pure Math. 100, Amer. Math. Soc., Providence, RI (2018) 297–331 MR Zbl

[17] **P Dunin-Barkowski**, **P Norbury**, **N Orantin**, **A Popolitov**, **S Shadrin**, *Dubrovin's superpotential as a global spectral curve*, J. Inst. Math. Jussieu 18 (2019) 449–497 MR Zbl

[18] **P Dunin-Barkowski**, **N Orantin**, **S Shadrin**, **L Spitz**, *Identification of the Givental formula with the spectral curve topological recursion procedure*, Comm. Math. Phys. 328 (2014) 669–700 MR Zbl

[19] **P Dunin-Barkowski**, **S Shadrin**, **L Spitz**, *Givental graphs and inversion symmetry*, Lett. Math. Phys. 103 (2013) 533–557 MR Zbl

[20] **T Ekedahl**, **S Lando**, **M Shapiro**, **A Vainshtein**, *Hurwitz numbers and intersections on moduli spaces of curves*, Invent. Math. 146 (2001) 297–327 MR Zbl

[21] **B Eynard**, *Invariants of spectral curves and intersection theory of moduli spaces of complex curves*, Commun. Number Theory Phys. 8 (2014) 541–588 MR Zbl

[22] **B Eynard**, **N Orantin**, *Invariants of algebraic curves and topological expansion*, Commun. Number Theory Phys. 1 (2007) 347–452 MR Zbl

[23] **B Eynard**, **N Orantin**, *Topological recursion in enumerative geometry and random matrices*, J. Phys. A 42 (2009) art. id. 293001 MR Zbl

[24] **C Faber**, **R Pandharipande**, *Relative maps and tautological classes*, J. Eur. Math. Soc. 7 (2005) 13–49 MR Zbl

[25] **H Fan**, **T Jarvis**, **Y Ruan**, *Quantum singularity theory for $A_{(r-1)}$ and r–spin theory*, Ann. Inst. Fourier (Grenoble) 61 (2011) 2781–2802 MR Zbl

[26] **H Fan**, **T Jarvis**, **Y Ruan**, *The Witten equation, mirror symmetry, and quantum singularity theory*, Ann. of Math. 178 (2013) 1–106 MR Zbl

[27] **J D Fay**, *Theta functions on Riemann surfaces*, Lecture Notes in Math. 352, Springer (1973) MR Zbl

[28] **E Getzler**, **E Looijenga**, *The Hodge polynomial of $\overline{\mathcal{M}}_{3,1}$*, preprint (1999) arXiv math/9910174

[29] **A B Givental**, *Gromov–Witten invariants and quantization of quadratic Hamiltonians*, Mosc. Math. J. 1 (2001) 551–568 MR Zbl

[30] **D Gross**, **E Witten**, *Possible third-order phase transition in the large-N lattice gauge theory*, Phys. Rev. D (1980) 446–453

[31] **A Hurwitz**, *Ueber Riemann'sche Flächen mit gegebenen Verzweigungspunkten*, Math. Ann. 39 (1891) 1–60 MR Zbl

[32] **F Janda**, **R Pandharipande**, **A Pixton**, **D Zvonkine**, *Double ramification cycles on the moduli spaces of curves*, Publ. Math. Inst. Hautes Études Sci. 125 (2017) 221–266 MR Zbl

[33] **T J Jarvis**, **T Kimura**, **A Vaintrob**, *Moduli spaces of higher spin curves and integrable hierarchies*, Compositio Math. 126 (2001) 157–212 MR Zbl

[34] **T Kimura**, **X Liu**, *A genus-*3 *topological recursion relation*, Comm. Math. Phys. 262 (2006) 645–661 MR Zbl

[35] **T Kimura**, **X Liu**, *Topological recursion relations on* $\overline{\mathcal{M}}_{3,2}$, Sci. China Math. 58 (2015) 1909–1922 MR Zbl

[36] **M Kontsevich**, *Intersection theory on the moduli space of curves and the matrix Airy function*, Comm. Math. Phys. 147 (1992) 1–23 MR Zbl

[37] **D Lewanski**, **A Popolitov**, **S Shadrin**, **D Zvonkine**, *Chiodo formulas for the* r*–th roots and topological recursion*, Lett. Math. Phys. 107 (2017) 901–919 MR Zbl

[38] **E Looijenga**, *On the tautological ring of* $\mathcal{M}_g$, Invent. Math. 121 (1995) 411–419 MR Zbl

[39] **M Mirzakhani**, *Simple geodesics and Weil–Petersson volumes of moduli spaces of bordered Riemann surfaces*, Invent. Math. 167 (2007) 179–222 MR Zbl

[40] **T Miwa**, **M Jimbo**, **E Date**, *Solitons*: *differential equations*, *symmetries and infinite-dimensional algebras*, Cambridge Tracts in Mathematics 135, Cambridge Univ. Press (2000) MR Zbl

[41] **D Mumford**, *Towards an enumerative geometry of the moduli space of curves*, from "Arithmetic and geometry, II" (M Artin, J Tate, editors), Progr. Math. 36, Birkhäuser, Boston, MA (1983) 271–328 MR Zbl

[42] **S M Natanzon**, *Moduli of Riemann surfaces, real algebraic curves, and their superanalogs*, Translations of Mathematical Monographs 225, Amer. Math. Soc., Providence, RI (2004) MR Zbl

[43] **P Norbury**, *Enumerative geometry via the moduli space of super Riemann surfaces*, preprint (2020) arXiv 2005.04378

[44] **P Norbury**, *Gromov–Witten invariants of* $\mathbb{P}^1$ *coupled to a KdV tau function*, Adv. Math. 399 (2022) art. id. 108227 MR Zbl

[45] **P Norbury**, **N Scott**, *Polynomials representing Eynard–Orantin invariants*, Q. J. Math. 64 (2013) 515–546 MR Zbl

[46] **P Norbury**, **N Scott**, *Gromov–Witten invariants of* $\mathbb{P}^1$ *and Eynard–Orantin invariants*, Geom. Topol. 18 (2014) 1865–1910 MR Zbl

[47] **R Pandharipande**, **A Pixton**, **D Zvonkine**, *Relations on* $\overline{\mathcal{M}}_{g,n}$ *via* 3*–spin structures*, J. Amer. Math. Soc. 28 (2015) 279–309 MR Zbl

[48] **K Saito**, *Period mapping associated to a primitive form*, Publ. Res. Inst. Math. Sci. 19 (1983) 1231–1264 MR Zbl

[49] **S Shadrin**, *BCOV theory via Givental group action on cohomological fields theories*, Mosc. Math. J. 9 (2009) 411–429 MR Zbl

[50] **V Shramchenko**, *Riemann–Hilbert problem associated to Frobenius manifold structures on Hurwitz spaces: irregular singularity*, Duke Math. J. 144 (2008) 1–52 MR Zbl

[51] **D Stanford**, **E Witten**, *JT gravity and the ensembles of random matrix theory*, Adv. Theor. Math. Phys. 24 (2020) 1475–1680 MR Zbl

[52] **C Teleman**, *The structure of 2D semi-simple field theories*, Invent. Math. 188 (2012) 525–588 MR Zbl

[53] **B Toen**, *Théorèmes de Riemann–Roch pour les champs de Deligne–Mumford*, *K*–Theory 18 (1999) 33–76 MR Zbl

[54] **E Witten**, *Two-dimensional gravity and intersection theory on moduli space*, from "Surveys in differential geometry" (H B Lawson, Jr, S-T Yau, editors), Lehigh Univ., Bethlehem, PA (1991) 243–310 MR Zbl

School of Mathematics and Statistics, University of Melbourne
Melbourne VIC, Australia

norbury@unimelb.edu.au

Proposed: Haynes R Miller — Received: 2 April 2020
Seconded: Jim Bryan, Mark Gross — Revised: 22 September 2021

Geometry & Topology **27**:7 (2023) 2763–2831
DOI: 10.2140/gt.2023.27.2763
Published: 19 September 2023

The 2–primary Hurewicz image of tmf

MARK BEHRENS
MARK MAHOWALD
J D QUIGLEY

We determine the image of the 2–primary tmf Hurewicz homomorphism, where tmf is the spectrum of topological modular forms. We do this by lifting elements of tmf_* to the homotopy groups of the generalized Moore spectrum $M(8, v_1^8)$ using a modified form of the Adams spectral sequence and the tmf resolution, and then proving the existence of a v_2^{32}–self-map on $M(8, v_1^8)$ to generate 192–periodic families in the stable homotopy groups of spheres.

55Q45; 55Q51, 55T15

1 Introduction

The Hurewicz theorem implies that the Hurewicz homomorphism

$$h\colon \pi_*(S^n) \to \widetilde{H}_*(S^n; \mathbb{Z})$$

is an isomorphism for $* = n$, implying the well-known result that the 0^{th} stable stem is given by

$$\pi_0^s \cong \mathbb{Z}.$$

Adams [1] studied the Hurewicz homomorphism for real K–theory

$$h_{\mathrm{KO}}\colon \pi_*^s \to \pi_*\mathrm{KO} = \mathrm{KO}^{-*}(\mathrm{pt}).$$

The computation of the real K–theory of a point (the homotopy groups of the spectrum KO representing real K–theory) is a consequence of the Bott periodicity theorem [11]: these groups are given by the following 8–fold periodic pattern:

n mod 8	0	1	2	3	4	5	6	7
$\pi_n\mathrm{KO}$	$\mathbb{Z}$	$\mathbb{Z}/2$	$\mathbb{Z}/2$	0	$\mathbb{Z}$	0	0	0

The map h_{KO} is an isomorphism in degree 0, and Adams showed that h_{KO} is surjective in degrees $* \equiv 1, 2 \bmod 8$. He did this by constructing what is now known as a v_1–self-map

$$v_1^4\colon \Sigma^8 M(2) \to M(2),$$

where $M(2)$ denotes the mod 2 Moore spectrum, and considering the projections

$$\mu_{8j+1+\epsilon} \in \pi_{8j+1+\epsilon}^s$$

of the elements

$$\eta^\epsilon \cdot v_1^{4j}\widetilde{\eta} \in \pi_{8j+2+\epsilon}M(2) \tag{1.1}$$

to the top cell of $M(2)$. Here $\widetilde{\eta}$ denotes a lift of $\eta \in \pi_1^s$ to the top cell of $M(2)$ and $\epsilon \in \{0, 1\}$. Because we have

$$\pi_*^s \otimes \mathbb{Q} = 0$$

for $* > 0$, the homomorphism h_{KO} is necessarily trivial in positive degrees $* \equiv 0 \bmod 4$.

Goerss, Hopkins and Miller constructed the spectrum tmf of topological modular forms [16] as a higher analog of the real K–theory spectrum.[1] The homotopy groups of tmf are 576–periodic. The goal of this paper is to determine the image of the 2–local tmf–Hurewicz homomorphism

$$h_{\mathrm{tmf}}\colon \pi_*^s \to \pi_*\mathrm{tmf}_{(2)}.$$

The 3–primary Hurewicz image has recently been determined by Belmont and Shimomura [9]. Since $\pi_*\mathrm{tmf}_{(p)}$ has no torsion for $p \geq 5$, the p–primary tmf–Hurewicz image is trivial in positive degrees for these primes. *Henceforth, everything in this paper is implicitly* 2*–local.*

2–Locally, the homotopy groups of tmf are merely 192–periodic. These homotopy groups were originally computed by Hopkins and Mahowald [19] (see also Bauer [3])

[1] Here, tmf denotes *connective* topological modular forms.

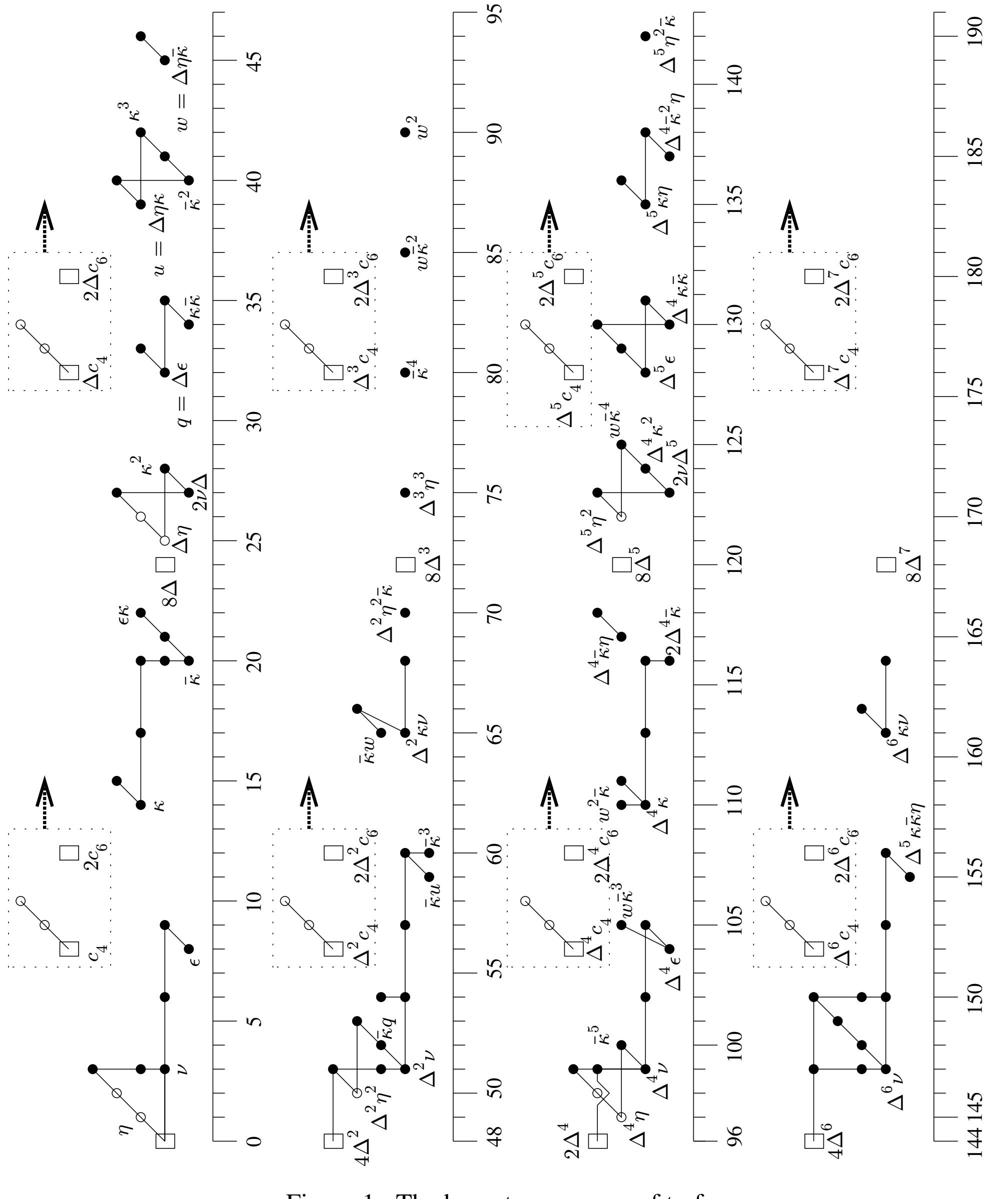


Figure 1: The homotopy groups of tmf.

using the descent spectral sequence

$$\mathrm{Ext}^{s,t}_{\Gamma^{\mathrm{ell}}}(A^{\mathrm{ell}}, A^{\mathrm{ell}}) \Rightarrow \pi_{t-s}(\mathrm{tmf}),$$

where $(A^{\mathrm{ell}}, \Gamma^{\mathrm{ell}})$ is the elliptic curve Hopf algebroid. These homotopy groups are displayed in Figure 1. In this figure:

- A series of i black dots joined by vertical lines corresponds to a factor of $\mathbb{Z}/2^i$ which is annihilated by some power of c_4.
- An open circle corresponds to a factor of $\mathbb{Z}/2$ which is not annihilated by a power of c_4.
- A box indicates a factor of $\mathbb{Z}_{(2)}$ which is not annihilated by a power of c_4.
- The nonvertical lines indicate multiplication by η and ν.
- A pattern with a dotted box around it and an arrow emanating from the right face indicates this pattern continues indefinitely to the right by c_4–multiplication (ie tensor the pattern with $\mathbb{Z}_{(2)}[c_4]$).
- The vertical arrangement of the chart is arbitrary.

The homotopy groups $\pi_*\mathrm{tmf}$ are given by tensoring the pattern depicted in Figure 1 with $\mathbb{Z}_{(2)}[\Delta^8]$, where $\Delta^8 \in \pi_{192}\mathrm{tmf}$. Our choice of names for generators in Figure 1 is motivated by the fact that the elements

$$\eta, \quad \nu, \quad \epsilon, \quad \kappa, \quad \overline{\kappa}, \quad q, \quad u, \quad w$$

in the stable stems map to the corresponding elements in $\pi_*\mathrm{tmf}$ under the tmf–Hurewicz homomorphism. The other indecomposable multiplicative generators are named based on the names of elements which detect them in the E_2–term of the descent spectral sequence. There is thus some ambiguity in the naming of some of these elements coming from the filtration associated to the descent spectral sequence.

For definiteness we fix $c_4 \in \pi_8\mathrm{tmf}$ to be the unique element detected by c_4 in the descent spectral sequence of Adams filtration 4. Note that the c_4–torsion in $\pi_*\mathrm{tmf}$ does not have c_4–exponent 1. Indeed, on c_4–torsion classes, multiplication by c_4 is equal to multiplication by ϵ — see Bruner and Rognes [14, Section 9.5] — so, for example, $c_4\kappa = \epsilon\kappa \neq 0$. However, all c_4–torsion has c_4–exponent 2; see loc. cit. and Behrens, Hill, Hopkins and Mahowald [7, Proposition 6.1].

The main theorem of this paper is the following:

Theorem 1.2 *The* tmf*–Hurewicz image is the subgroup of* $\pi_*\mathrm{tmf}$ *generated by*

(1) *all the elements of* $\pi_{\leq 3}(\mathrm{tmf})$,
(2) *the elements* $c_4^i\eta$ *and* $c_4^i\eta^2$,
(3) *all the elements of* $\pi_*\mathrm{tmf}$ *annihilated by a power of* c_4 *except those in* $\pi_{24k+3}\mathrm{tmf}$.

Remark 1.3 The reader will note from Figure 1 that the subgroup of $\pi_*(\mathrm{tmf})$ generated by the elements of type (3) above form a self-dual pattern centered in dimension 85. This is discussed in [14, Chapter 10].

Besides representing an advance in our understanding of v_2–periodic homotopy at the prime 2, Theorem 1.2 also has applications to smooth structures on spheres, as explained in [7]. Specifically, Hill, Hopkins and the first two authors consider the following question:

Question 1.4 In which dimensions n do there exist exotic smooth structures on the n–sphere?

Such spheres with exotic smooth structures are called exotic spheres. The work of Kervaire and Milnor [26] relates the existence of exotic spheres to the triviality of the Kervaire homomorphism

$$\pi^s_{4k+2} \to \mathbb{Z}/2$$

and the nontriviality of the cokernel of the J–homomorphism

$$J \colon \pi_n \mathrm{SO} \to \pi^s_n.$$

Specifically, they prove that exotic spheres exist in dimensions n for which:

$\boldsymbol{n = 4k}$ $n \geq 8$ and there exists a nontrivial element of coker J.

$\boldsymbol{n = 4k+1}$ There exists a nontrivial element of coker J, or there does not exist an element of Kervaire invariant 1 in dimension $n+1$.

$\boldsymbol{n = 4k+2}$ There exists a nontrivial element of coker J with Kervaire invariant 0.

$\boldsymbol{n = 4k+3}$ $n \geq 7$.

Combining this with the work of Moise [35], Browder [12], Barratt, Jones and Mahowald [2], Hill, Hopkins and Ravenel [18], and Wang and Xu [36], Question 1.4 has been answered completely for n odd:

> *The only odd dimensions n for which there do not exist exotic spheres are $n = 1, 3, 5$ and 61.*

For n even, the case of $n = 4$ is unresolved. For other even n, by the previous discussion, the question boils down to the existence of nontrivial elements of coker J (with Kervaire invariant 0). It is shown in [7]:

> *The only even dimensions $4 \neq n < 140$ for which there do not exist exotic spheres are $n = 2, 6, 12$ and 56.*

In the case of $n = 8k + 2 \geq 10$, Adams' elements μ_{8k+2} with nontrivial KO–Hurewicz image are not in the image of J and have trivial Kervaire invariant. It thus follows that:

There exist exotic spheres in all dimensions $n = 8k + 2 \geq 10$.

As is explained in [7], many of the 192–periodic families of elements of Theorem 1.2 also are not in the image of J and have trivial Kervaire invariant. Theorem 1.2 therefore has the following corollary:[2]

Corollary 1.5 *There exist exotic spheres in the following congruence classes of even dimensions* $n \geq 8$ *modulo* 192:

$$2, 6, 8, 10, 14, 18, 20, 22, 26, 28, 32, 34, 40, 42, 46, 50, 52, 54, 58, 60, 66, 68,$$
$$70, 74, 80, 82, 90, 98, 100, 102, 104, 106, 110, 114, 116, 118, 122, 124, 128,$$
$$130, 136, 138, 142, 146, 148, 150, 154, 156, 162, 164, 170, 178, 186.$$

(This accounts for over half of the even dimensions.)

We will prove Theorem 1.2 by first showing (Theorem 6.1) that the subgroup of $\pi_*\mathrm{tmf}$ described by Theorem 1.2 is contained in the Hurewicz image. This will be a relatively straightforward consequence of some v_1–periodic computations. The elements of Theorem 1.2(1) are already established to be in the Hurewicz image by the preceding discussion, and the elements of (2) are in the Hurewicz image because they are the images of the elements μ_{8i+j}. We are left to show that the elements of type (3) lift to π_*^s. This is the main task of this paper.

In [14], Bruner and Rognes give a systematic and careful study of the Adams spectral sequence for tmf, and in particular they have independently established the Hurewicz image in many low-dimensional cases. Specifically, they prove Theorem 1.2 for degrees $* \leq 101$ and also show that $w\bar{\kappa}^3$, $w^2\bar{\kappa}$, $w\bar{\kappa}^4$, $2\Delta^4\kappa\bar{\kappa}$ and $4\Delta^6\nu^2$ (in dimensions 105, 110, 125, 130 and 150) are in the Hurewicz image. Also, they use a different technique (Anderson duality) to prove that the Hurewicz image is contained in the subgroup of tmf_* described in Theorem 1.2.

Our strategy to lift elements from $\pi_*\mathrm{tmf}$ to π_*^s is to use the methods of [7]. We summarize that strategy here. We recall the following from [7, Proposition 6.1]:

[2] In fact, the v_2^{32}–self-map of Theorem 1.8 which is used to construct the periodic families of Theorem 1.2 also immediately implies the existence of some elements not in the image of the J–homomorphism which are in the kernel of the tmf–Hurewicz homomorphism, such as the beta elements $\beta_{32k/8}$. However, we will not concern ourselves here with the few additional dimensions such considerations add to the list of Corollary 1.5.

Proposition 1.6 [7] *Every c_4–torsion element $x \in \pi_*\mathrm{tmf}$ is 8–torsion and c_4^2–torsion.*

Let $M(2^i)$ denote the cofiber of 2^i, and let $M(2^i, v_1^j)$ denote the cofiber of a v_1–self-map (see Davis and Mahowald [15, Proposition 2.3])

$$v_1^j : \Sigma^{2j} M(2^i) \to M(2^i).$$

Corollary 1.7 *Every c_4–torsion element $x \in \pi_*(\mathrm{tmf})$ lifts to an element*

$$\tilde{x} \in \mathrm{tmf}_{*+18} M(8, v_1^8)$$

so that the projection to the top cell maps $\tilde{x}$ to x.

Given a c_4–torsion element $x \in \pi_{<192}(\mathrm{tmf})$, Proposition 1.6 implies it lifts to an element

$$\tilde{x} \in \mathrm{tmf}_* M(8, v_1^8)$$

so that the projection to the top cell maps $\tilde{x}$ to x. We will then show that $\tilde{x}$ lifts to an element

$$\tilde{y} \in \pi_* M(8, v_1^8).$$

Then the image

$$y \in \pi_*^s$$

given by projecting $\tilde{y}$ to the top cell is an element whose image under the tmf–Hurewicz homomorphism is x.

Every c_4–torsion element $x' \in \pi_{\geq 192}\mathrm{tmf}$ is of the form $v_2^{32k} x$ for $x \in \pi_{<192}\mathrm{tmf}$. We will prove the following theorem:

Theorem 1.8 *There exists a v_2^{32}–self-map*

$$v_2^{32} : \Sigma^{192} M(8, v_1^8) \to M(8, v_1^8).$$

If $\tilde{x} \in \mathrm{tmf}_* M(8, v_1^8)$ is a lift of x, and $\tilde{y} \in \pi_* M(8, v_1^8)$ is a lift of $\tilde{x}$, as in the discussion above, then the resulting element

$$v_2^{32k} \tilde{y} \in \pi_* M(8, v_1^8),$$

obtained by composing with the k–fold iterate of the v_2^{32}–self-map, projects to an element $y' \in \pi_*^s$ which maps to x' under the tmf–Hurewicz homomorphism.

As in [7], the analysis above rests on a systematic analysis of the homotopy groups $\pi_* M(8, v_1^8)$. This will be based on computations using the *modified Adams spectral sequence* (*MASS*). The E_2–term of the modified Adams spectral sequence will be analyzed in a region near its vanishing line by means of another spectral sequence, the *algebraic tmf resolution*.

The work of [7] was hampered by the fact that all of the algebraic tmf resolution computations were performed on the level of the E_1–term of the algebraic tmf resolution. In this paper, we will show that the weight spectral sequence, used in the context of bo resolutions by Lellmann and Mahowald [28] and Beaudry, Behrens, Bhattacharya, Culver and Xu [4], can be used to analyze the E_2–term of the algebraic tmf resolution, greatly simplifying the computations.

Conventions

- Homology will be implicitly taken with mod 2 coefficients.
- We let A_* denote the dual Steenrod algebra, $A/\!/A(2)_*$ denote the dual of the Hopf algebra quotient $A/\!/A(2)$, and, for an A_*–comodule M (or more generally an object of the stable homotopy category of A_*–comodules; see Hovey [21]), we let
$$\mathrm{Ext}^{s,t}_{A_*}(M)$$
denote the group $\mathrm{Ext}^{s,t}_{A_*}(\mathbb{F}_2, M)$.
- Given a Hopf algebroid (B, Γ) and a comodule M, we will let $C^*_\Gamma(M)$ denote the associated normalized cobar complex.
- For a spectrum E, we let E_* denote its homotopy groups $\pi_* E$.

Outline of the paper

In Section 2, we recall the modified Adams spectral sequence (MASS), which takes the form
$${}^{\mathrm{mass}}E_2^{*,*} = \mathrm{Ext}_{A_*}(H_*X \otimes H(8, v_1^8)) \Rightarrow \pi_*(X \wedge M(8, v_1^8))$$
for a certain object $H(8, v_1^8)$ in the stable homotopy category of A_*–comodules. We recall how the E_2–term of the MASS can be studied using the algebraic tmf resolution, which is a spectral sequence that takes the form
$${}^{\mathrm{tmf}}_{\mathrm{alg}}E_1(M)^{*,*,*} \Rightarrow \mathrm{Ext}^{*,*}_{A_*}(M)$$
for any M in the stable category of A_*–comodules. We then recall how the E_1–term of the algebraic tmf resolution decomposes as a sum of Ext groups involving tensor

powers of bo Brown–Gitler comodules, and also summarize an inductive method to compute these Ext groups.

In Section 3, we study the d_1–differential in the algebraic tmf resolution for $\mathbb{F}_2$, and introduce a tool, the weight spectral sequence (WSS)

$$ {}^{\mathrm{tmf}}_{\mathrm{alg}}E_1 = {}^{\mathrm{wss}}E_0 \Rightarrow {}^{\mathrm{tmf}}_{\mathrm{alg}}E_2, $$

which serves as an analog of the May spectral sequence and converges to the E_2–term of the algebraic tmf resolution. The E_0–page of the v_0–localized weight spectral sequence is identified with the cobar complex of a primitively generated Hopf algebra, and this allows us to give "names" to the v_0–torsion-free classes of ${}^{\mathrm{tmf}}_{\mathrm{alg}}E_1$. We include many charts of summands of ${}^{\mathrm{tmf}}_{\mathrm{alg}}E_1(\mathbb{F}_2)$ corresponding to tensor powers of bo Brown–Gitler comodules which illustrate this naming convention, and provide the essential data for the rest of the computations in this paper. Finally, we study the g–local WSS[3] using recent work of Bhattacharya, Bobkova and Thomas [10], and show that many classes are killed in the g–local WSS by d_1–differentials. This is the key fact we will use to systematically remove obstructions for lifting classes from tmf_*X to π_*X.

In Section 4 we study the structure of the MASS for $M(8, v_1^8)$. We recall the structure of the MASS for $\mathrm{tmf}_*M(8, v_1^8)$, and we explain how to adapt the Ext charts of Section 3 to give the corresponding computations of ${}^{\mathrm{tmf}}_{\mathrm{alg}}E_1(H(8, v_1^8))$. We then explain how to translate the computations of the g–localized algebraic tmf resolution of Section 3 to the case of $H(8, v_1^8)$.

Section 5 is dedicated to the proof of Theorem 1.8. We recall the work of Davis, Mahowald and Rezk, who discovered topological attaching maps between the first two bo Brown–Gitler spectra which constitute $\mathrm{tmf} \wedge \mathrm{tmf}$, which give extra differentials in the Adams spectral sequence of $\mathrm{tmf} \wedge \mathrm{tmf}$ that kill some g–torsion-free classes. We then prove a technical lemma (Lemma 5.5) which lifts differentials from the MASS for $\mathrm{tmf}^s \wedge M(8, v_1^8)$ to the MASS for $M(8, v_1^8)$. We prove Theorem 1.8 by listing all elements in ${}^{\mathrm{tmf}}_{\mathrm{alg}}E_1(H(8, v_1^8))$ which could detect a nontrivial differential $d_r(v_2^{32})$ in the MASS for $M(8, v_1^8)$, and then we systematically eliminate these possibilities. Most of these classes are g–torsion-free, and are eliminated in the WSS or by using Lemma 5.5.

In Section 6, we explain how v_1–periodic computations give an upper bound on the Hurewicz image.

[3]Here, $g \in \mathrm{Ext}^{4,24}_{A_*}(\mathbb{F}_2)$ is the element corresponding to the element $h_{2,1}^4$ in the May spectral sequence which detects $\bar{\kappa}$ in the Adams spectral sequence for the sphere.

Section 7 is devoted to showing this upper bound is sharp, by producing lifts of the remaining elements of $\pi_*\mathrm{tmf}$ to the sphere. We begin by identifying multiplicative generators of the Hurewicz image in dimensions less than 192, so that it suffices for us to lift these. We then lift these elements by producing elements in the MASS for $M(8, v_1^8)$ which we show are permanent cycles, and detect elements of $\pi_* M(8, v_1^8)$ which project to the desired elements on the top cell. These elements are then propagated to v_2^{32}–periodic families using the self-map, thus proving Theorem 1.2 in all dimensions.

Acknowledgments

We are grateful to Bob Bruner and John Rognes for generously sharing their results on their study of the Adams spectral sequence of tmf, and also to Rognes for pointing out a redundancy in Section 7. This project would have not been possible without the Ext computational software developed by Bob Bruner and Amelia Perry, and the detailed computations of the Adams spectral sequence of the sphere by Isaksen, Wang and Xu. The authors are especially grateful to Bob Bruner for providing them with a module definition file for $\overline{A/\!\!/A(2)}$. Behrens would also like to express his appreciation to Agnès Beaudry, Prasit Bhattacharya, Dominic Culver, Kyle Ormsby, Nat Stapleton, Vesna Stojanoska and Zhouli Xu, whose previous collaborative work on the tmf resolution was essential for the results of this paper, as well as to Mike Hill and Mike Hopkins, whose collaboration with Behrens and Mahowald was the genesis of this paper. Quigley also wishes to thank his coauthors for the opportunity to contribute to this project. Finally, the authors wish to thank the referees for important comments and corrections. Behrens was supported by NSF grants DMS-1050466, DMS-1452111, DMS-1547292, DMS-1611786 and DMS-2005476 over the course of this work. Quigley was partially supported by NSF grant DMS-1547292.

2 Preliminaries

The techniques and methods of this paper closely follow those of [7]. In this section we recall some spectral sequences used in that paper.

The modified Adams spectral sequence

Our computations of $\pi_* M(8, v_1^8)$ and $\mathrm{tmf}_* M(8, v_1^8)$ will be performed using the modified Adams spectral sequence (MASS). We refer the reader to [7, Section 6] for a complete account of the construction of the MASS and summarize the form it takes here.

Let St_{A_*} denote Hovey's stable homotopy category of A_*–comodules [21]. For objects M and N of St_{A_*}, we define the group

$$\mathrm{Ext}^{s,t}_{A_*}(M,N)=\mathrm{St}_{A_*}(\Sigma^t M, N[s])$$

as a group of maps in the stable homotopy category. Here $\Sigma^t M$ denotes the t–fold shift with respect to the internal grading of M, and $N[s]$ denotes the s–fold shift with respect to the triangulated structure of St_{A_*}. This reduces to the usual definition of Ext_{A_*} when M and N are A_*–comodules.

Define $H(8)$ to be the cofiber of the map

(2.1) $$\Sigma^3\mathbb{F}_2[-3]\xrightarrow{h_0^3}\mathbb{F}_2$$

in the stable homotopy category of A_*–comodules. Define $H(8,v_1^8)\in \mathrm{St}_{A_*}$ to be the cofiber

(2.2) $$\Sigma^{24}H(8)[-8]\xrightarrow{v_1^8}H(8)\to H(8,v_1^8).$$

For a spectrum X, the MASS takes the form

$$^{\mathrm{mass}}E_2^{s,t}(M(8,v_1^8)\wedge X)=\mathrm{Ext}^{s,t}_{A_*}(H(8,v_1^8)\otimes H_*X)\Rightarrow \pi_{t-s}M(8,v_1^8)\wedge X.$$

Recall the following from [7, Proposition 7.1]:

Proposition 2.3 *$M(8,v_1^8)$ is a weak homotopy ring spectrum.*[4]

It follows that, if X is a ring spectrum, the MASS above is a spectral sequence of (nonassociative) algebras.

We recall the following key theorem of Mathew:

Theorem 2.4 (Mathew [34]) *We have*

$$H_*\mathrm{tmf}\cong A/\!\!/A(2)_*$$

as an algebra in A_*–comodules.

Taking $X=\mathrm{tmf}\wedge Y$ for some Y and applying a change-of-rings theorem, the MASS takes the form

$$^{\mathrm{mass}}E_2^{s,t}(\mathrm{tmf}\wedge M(8,v_1^8)\wedge Y)=\mathrm{Ext}^{s,t}_{A(2)_*}(H(8,v_1^8)\otimes H_*Y)\Rightarrow \mathrm{tmf}_{t-s}(M(8,v_1^8)\wedge Y).$$

[4] By this, we mean a spectrum with a possibly nonassociative product and a two-sided unit in the stable homotopy category.

The algebraic tmf resolution

The E_2–page of the MASS for $M(8, v_1^8)$ will be analyzed using an algebraic analog of the tmf resolution (as in [7, Section 6]).

The (topological) tmf resolution of a space X is the Adams spectral sequence based on the spectrum tmf:

$$ {}^{\mathrm{tmf}}E_1^{s,t} = \pi_t \mathrm{tmf} \wedge \overline{\mathrm{tmf}}^s \wedge X \Rightarrow \pi_{t-s} X. $$

Here, $\overline{\mathrm{tmf}}$ is the cofiber of the unit

$$ S \to \mathrm{tmf} \to \overline{\mathrm{tmf}} $$

and $\overline{\mathrm{tmf}}^s = \overline{\mathrm{tmf}}^{\wedge s}$ denotes its s–fold smash power.

The *algebraic* tmf resolution is an algebraic analog. Namely, let M be an object of the stable homotopy category of A_*–comodules and let $\overline{A/\!/A(2)}_*$ denote the cokernel of the unit

$$ 0 \to \mathbb{F}_2 \to A/\!/A(2)_* \to \overline{A/\!/A(2)}_* \to 0 $$

(note that $H_*\overline{\mathrm{tmf}} = \overline{A/\!/A(2)}_*$). The *algebraic* tmf *resolution* of M is a spectral sequence of the form

$$ {}^{\mathrm{tmf}}_{\mathrm{alg}}E_1^{s,t,n}(M) = \mathrm{Ext}^{s,t}_{A(2)_*}(\overline{A/\!/A(2)}_*^{\otimes n} \otimes M) \Rightarrow \mathrm{Ext}^{s+n,t}_{A_*}(M). $$

bo Brown–Gitler comodules

We recall some material on bo Brown–Gitler comodules. These are A_*–comodules which are the homology of the bo Brown–Gitler spectra constructed by [17]. Mahowald used integral Brown–Gitler spectra to analyze the bo resolution [30]. The bo Brown–Gitler comodules play a similar role in the algebraic tmf resolution [6; 31; 15; 8; 7].

Endow the mod 2 homology of the connective real K–theory spectrum

$$ H_*(\mathrm{bo}) \cong A/\!/A(1)_* = \mathbb{F}_2[\zeta_1^4, \zeta_2^2, \zeta_3, \dots] $$

with a multiplicative grading by declaring the *weight* of ζ_i to be

(2.5) $$ \mathrm{wt}(\zeta_i) = 2^{i-1}. $$

The i^{th} bo *Brown–Gitler* comodule is the subcomodule

$$ \underline{\mathrm{bo}}_i = F_{4i} A/\!/A(1)_* \subset A/\!/A(1)_* $$

spanned by monomials of weight less than or equal to $4i$. It is isomorphic as an A_*–comodule to the homology of the i^{th} bo Brown–Gitler spectrum bo_i.

The analysis of the E_1–page of the algebraic tmf resolution is simplified via the decomposition of $A(2)_*$–comodules

$$\overline{A/\!\!/A(2)}_* \cong \bigoplus_{i>0} \Sigma^{8i}\underline{\mathrm{bo}}_i$$

of [6, Corollary 5.5]. We therefore have a decomposition of the E_1–page of the algebraic tmf resolution for M given by

$$\text{(2.6)}\qquad {}^{\mathrm{tmf}}_{\mathrm{alg}}E_1^{s,t,n}(M) \cong \bigoplus_{i_1,\dots,i_n>0} \mathrm{Ext}^{s,t}_{A(2)_*}(\Sigma^{8(i_1+\cdots+i_n)}\underline{\mathrm{bo}}_{i_1}\otimes\cdots\otimes\underline{\mathrm{bo}}_{i_n}\otimes M).$$

For any M, the computation of

$$\mathrm{Ext}^{s,t}_{A(2)_*}(\Sigma^{8(i_1+\cdots+i_n)}\underline{\mathrm{bo}}_{i_1}\otimes\cdots\otimes\underline{\mathrm{bo}}_{i_n}\otimes M)$$

can be inductively determined from $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes k}\otimes M)$ by means of a set of exact sequences of $A(2)_*$–comodules, which relate the $\underline{\mathrm{bo}}_i$ [6, Section 7] (see also [8]),

$$\text{(2.7)}\qquad 0\to\Sigma^{8j}\underline{\mathrm{bo}}_j\to\underline{\mathrm{bo}}_{2j}\to A(2)/\!\!/A(1)_*\otimes\underline{\mathrm{tmf}}_{j-1}\to\Sigma^{8j+9}\underline{\mathrm{bo}}_{j-1}\to 0,$$

$$\text{(2.8)}\qquad 0\to\Sigma^{8j}\underline{\mathrm{bo}}_j\otimes\underline{\mathrm{bo}}_1\to\underline{\mathrm{bo}}_{2j+1}\to A(2)/\!\!/A(1)_*\otimes\underline{\mathrm{tmf}}_{j-1}\to 0.$$

Here $\underline{\mathrm{tmf}}_j$ is the j^{th} tmf–Brown–Gitler comodule — it is the subcomodule of

$$H_*(\mathrm{tmf})\cong A/\!\!/A(2)_* = \mathbb{F}_2[\zeta_1^8,\zeta_2^4,\zeta_3^2,\zeta_4,\dots]$$

spanned by monomials of weight less than or equal to $8j$.[5]

The exact sequences (2.7) and (2.8) can be reexpressed as resolutions in the stable homotopy category of $A(2)_*$–comodules

$$\underline{\mathrm{bo}}_{2j}\to A(2)/\!\!/A(1)_*\otimes\underline{\mathrm{tmf}}_{j-1}\to\Sigma^{8j+9}\underline{\mathrm{bo}}_{j-1}\to\Sigma^{8j}\underline{\mathrm{bo}}_j[2],$$
$$\underline{\mathrm{bo}}_{2j+1}\to A(2)/\!\!/A(1)_*\otimes\underline{\mathrm{tmf}}_{j-1}\to\Sigma^{8j}\underline{\mathrm{bo}}_j\otimes\underline{\mathrm{bo}}_1[1],$$

[5] Technically speaking, as is addressed in [6, Section 7], the comodules $A(2)/\!\!/A(1)_*\otimes\underline{\mathrm{tmf}}_{j-1}$ in the above exact sequences have to be given a slightly different $A(2)_*$–comodule structure from the standard one arising from the tensor product. However, this different comodule structure ends up being Ext–isomorphic to the standard one. As we are only interested in Ext groups, the reader can safely ignore this subtlety.

which give rise to spectral sequences

$$(2.9)\quad \begin{gathered} E_1^{n,s,t} = \left\{ \begin{array}{ll} \mathrm{Ext}^{s,t}_{A(1)_*}(\underline{\mathrm{tmf}}_{j-1} \otimes M), & n = 0, \\ \mathrm{Ext}^{s,t}_{A(2)_*}(\Sigma^{8j+9}\underline{\mathrm{bo}}_{j-1} \otimes M[-1]), & n = 1, \\ \mathrm{Ext}^{s,t}_{A(2)_*}(\Sigma^{8j}\underline{\mathrm{bo}}_{j} \otimes M), & n = 2, \\ 0, & n > 2 \end{array} \right\} \Rightarrow \mathrm{Ext}^{s,t}_{A(2)_*}(\underline{\mathrm{bo}}_{2j} \otimes M), \\ E_1^{n,s,t} = \left\{ \begin{array}{ll} \mathrm{Ext}^{s,t}_{A(1)_*}(\underline{\mathrm{tmf}}_{j-1} \otimes M), & n = 0, \\ \mathrm{Ext}^{s,t}_{A(2)_*}(\Sigma^{8j}\underline{\mathrm{bo}}_{j} \otimes \underline{\mathrm{bo}}_1 \otimes M), & n = 1, \\ 0, & n > 1 \end{array} \right\} \Rightarrow \mathrm{Ext}^{s,t}_{A(2)_*}(\underline{\mathrm{bo}}_{2j+1} \otimes M). \end{gathered}$$

These spectral sequences have been observed to collapse in low degrees (see [8]) but it is not known if they collapse in general. They inductively build $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_i \otimes M)$ out of $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes k} \otimes M)$ and $\mathrm{Ext}_{A(1)_*}(\underline{\mathrm{tmf}}_j \otimes M)$.

3 Analysis of the algebraic tmf resolution

In this section we will compute the d_1–differential in the algebraic tmf resolution, and will introduce a tool, the *weight spectral sequence* (*WSS*), which is a variant of the May spectral sequence that converges to the E_2–page of the algebraic tmf resolution.

The d_1–differential in the algebraic tmf resolution

Our approach to understanding the d_1–differential in the algebraic tmf resolution will be to compute it on v_0–torsion-free classes, and then infer its effect on v_0–torsion classes by means of linearity over $\mathrm{Ext}_{A_*}(\mathbb{F}_2)$.

Consider the algebraic $\mathrm{BP}\langle 2\rangle$ and algebraic BP resolutions

$$\begin{aligned} {}^{\mathrm{BP}\langle 2\rangle}_{\mathrm{alg}}E^{s,t,n} &= \mathrm{Ext}^{s,t}_{E[2]_*}(\overline{A/\!/E[2]}_*^{\otimes n}) \Rightarrow \mathrm{Ext}^{s+n,t}_{A_*}(\mathbb{F}_2), \\ {}^{\mathrm{BP}}_{\mathrm{alg}}E^{s,t,n} &= \quad \mathrm{Ext}^{s,t}_{E_*}(\overline{A/\!/E}_*^{\otimes n}) \quad \Rightarrow \mathrm{Ext}^{s+n,t}_{A_*}(\mathbb{F}_2). \end{aligned}$$

Here $E[2] = E[Q_0, Q_1, Q_2]$ and $E = E[Q_0, Q_1, Q_2, \dots]$ denote subalgebras of the Steenrod algebra, where Q_i are the Milnor generators dual to $\xi_{i+1} \in A_*$.

The d_1–differential in the algebraic tmf resolution may be studied by means of the zigzag

$$(3.1)\qquad {}^{\mathrm{tmf}}_{\mathrm{alg}}E_1^{*,*,*} \to {}^{\mathrm{BP}\langle 2\rangle}_{\mathrm{alg}}E_1^{*,*,*} \leftarrow {}^{\mathrm{BP}}_{\mathrm{alg}}E_1^{*,*,*}.$$

Note that

$${}^{\mathrm{BP}}_{\mathrm{alg}}E_1^{*,*,n} \cong \mathbb{F}_2[v_0, v_1, v_2, \dots] \otimes \overline{\mathbb{F}_2[\zeta_1^2, \zeta_2^2, \dots]}^{\otimes n},$$

where $\overline{\mathbb{F}_2[\zeta_1^2, \zeta_2^2, \dots]}$ denotes the cokernel of the unit

$$\mathbb{F}_2 \to \mathbb{F}_2[\zeta_1^2, \zeta_2^2, \dots].$$

The Adams spectral sequences

$${}^{\mathrm{BP}}_{\mathrm{alg}}E_1^{n,*,*} = {}^{\mathrm{ass}}_{*,*}E_2(\mathrm{BP} \wedge \overline{\mathrm{BP}}^n) \Rightarrow C^n_{\mathrm{BP}_*\mathrm{BP}}(\mathrm{BP}_*)$$

collapse, where $C^*_{\mathrm{BP}_*\mathrm{BP}}$ is the normalized cobar complex for $\mathrm{BP}_*\mathrm{BP}$, and

$$\zeta_i^2 \in \overline{A/\!\!/E_*} \quad \text{detects} \quad t_i \in \mathrm{BP}_*\mathrm{BP}.$$

We conclude:

Lemma 3.2 *The d_1–differential in the algebraic* BP *resolution is the associated graded of the differential in the cobar complex for* $\mathrm{BP}_*\mathrm{BP}$ *with respect to Adams filtration.*

The weight spectral sequence

Endow the normalized cobar complex

$$C^*(A_*, A_*, \mathbb{F}_2)$$

with a decreasing filtration by weight by defining

$$\mathrm{wt}(a_0[a_1 \,|\cdots|\, a_s]) = \mathrm{wt}(a_1) + \cdots + \mathrm{wt}(a_s).$$

Applying $\mathrm{Ext}_{A_*}(\mathbb{F}_2, -)$ to the resulting filtered A_*–comodule produces a variant of the May spectral sequence, which we will call the *modified May spectral sequence* (MMSS),[6]

(3.3)
$${}^{\mathrm{mmss}}E_0^{w,s,t} = C^*_{E^0A_*}(\mathbb{F}_2) \Rightarrow \mathrm{Ext}^{s,t}_{A_*}(\mathbb{F}_2).$$

Since E^0A_* is primitively generated, we have

$${}^{\mathrm{mmss}}E_1^{*,*} = \mathbb{F}_2[h_{i,j} : i \geq 1,\ j \geq 0].$$

The map $\mathrm{tmf} \to H$ induces an inclusion

$$\Phi\colon H_*(\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^n) \hookrightarrow H_*(H \wedge \overline{H}^n) \cong C^n(A_*, A_*, \mathbb{F}_2).$$

Under this inclusion, the weight filtration restricts to a decreasing filtration on

$$H_*(\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^n) \cong A/\!\!/A(2)_* \otimes \overline{A/\!\!/A(2)}_*^{\otimes n}$$

[6]The authors of [29] construct a similar modified May spectral sequence, but with a slightly different filtration.

by A_*–subcomodules. Because the weights of all of the generators of $A/\!/A(2)_*$ are divisible by 8, we actually work with weights divided by 8. Applying $\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2, -)$ and taking cohomology, we get the *weight spectral sequence* (WSS)

$$ {}^{\mathrm{wss}}E_0^{w,n,s,t} = \bigoplus_{i_1+\cdots+i_n=w} \mathrm{Ext}^{s,t}_{A(2)_*}(\underline{\mathrm{bo}}_{i_1} \otimes \cdots \otimes \underline{\mathrm{bo}}_{i_n}) \Rightarrow {}^{\mathrm{tmf}}_{\mathrm{alg}}E_2^{n,s,t}. $$

The WSS serves as an analog of the May spectral sequence for the algebraic tmf resolution.

The map Φ above induces a map of spectral sequences

(3.4)
$$\begin{tikzcd} {}^{\mathrm{wss}}E_0^{w,n,0,t} \arrow[r, Rightarrow] \arrow[d, "\Phi_*"'] & {}^{\mathrm{tmf}}_{\mathrm{alg}}E_0^{n,0,t} \arrow[d, "\Phi_*"] \\ {}^{\mathrm{mmss}}E_0^{8w,n,t} \arrow[r] & \mathrm{Ext}^{n,t}_{A_*}(\mathbb{F}_2) \end{tikzcd}$$

The v_0–localized algebraic tmf resolution

Observe that we have

(3.5)
$$ v_0^{-1}\,\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2) = \mathbb{F}_2[v_0^{\pm}, v_1^4, v_2^2]. $$

Note that $c_4, c_6 \in (\mathrm{tmf}_*)_{\mathbb{Q}}$ are detected in the v_0–localized ASS by v_1^4 and $v_0^3 v_2^2$, respectively.

We recall from [8] that

(3.6)
$$ v_0^{-1}\,\mathrm{Ext}^{*,*}_{A(2)_*}(A/\!/A(2)_*) = \mathbb{F}_2[v_0^{\pm}, v_1^4, v_2^2][\zeta_1^8, \zeta_2^4] $$

and that there is an isomorphism

(3.7)
$$ v_0^{-1}\,\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_i) \cong \mathbb{F}_2[v_0^{\pm}, v_1^4, v_2^2]\{\zeta_1^{8i'}\zeta_2^{4i''}\}_{i=i'+i''}. $$

We will now compute the localized E_1–page $v_0^{-1}\,{}^{\mathrm{wss}}E_1$. The following is immediate from the computation of the cobar differential (modulo terms of higher Adams filtration) on the elements ζ_1^8 and ζ_2^4, using (3.6), (3.7) and (3.1):

Proposition 3.8 *There is an isomorphism of differential graded algebras*

$$ v_0^{-1}\,{}^{\mathrm{wss}}E_0^{*,n,*,*} \cong \mathbb{F}_2[v_0^{\pm}, v_1^4, v_2^2] \otimes C^n_{\mathbb{F}_2[\zeta_1^8, \zeta_2^4]}, $$

where $\mathbb{F}_2[\zeta_1^8, \zeta_2^4]$ is regarded as a primitively generated Hopf algebra.

Corollary 3.9 *There is an isomorphism*

$$ v_0^{-1}\,{}^{\mathrm{wss}}E_1 = \mathbb{F}_2[v_0^{\pm}, v_1^4, v_2^2] \otimes \mathbb{F}_2[h_{1,3}, h_{1,4}, \ldots, h_{2,2}, h_{2,3}, \ldots]. $$

Charts

For the convenience of the reader we include some charts of $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^k)$ for $0 \le k \le 3$ as well as $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2)$.

$\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$ (see Figure 2) All the elements are $c_4 = v_1^4$–periodic and v_2^8–periodic. Exactly one v_1^4–multiple of each element is displayed with the $\bullet$ replaced by a $\circ$. Observe the wedge pattern beginning in $t-s = 35$. This pattern is infinite, propagated horizontally by $h_{2,1}$–multiplication and vertically by v_1–multiplication. Here $h_{2,1}$ is the name of the generator in the May spectral sequence of bidegree $(t-s,s) = (5,1)$, and $h_{2,1}^4 = g$.

$\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes k})$ for $k = 1, 2, 3$ (Figure 3) Every element is v_2^8–periodic. However, unlike $\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$, not every element of these Ext groups is v_1^4–periodic. Rather, it is the case that an element $x \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes k})$ either satisfies $v_1^4 x = 0$ or is v_1^4–periodic. The v_1^4–periodic elements fit into families which look like shifted and truncated copies of $\mathrm{Ext}_{A(1)_*}(\mathbb{F}_2)$ and are labeled with a $\circ$. We have only included the beginning of these v_1^4–periodic patterns in the chart. The other generators are labeled with a $\bullet$. A $\square$ indicates a polynomial algebra $\mathbb{F}_2[h_{2,1}]$.

$\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2)$ (Figure 4) Via the spectral sequence (2.9), this Ext chart is assembled out of $\mathrm{Ext}_{A(1)_*}(\mathbb{F}_2)$, $\mathrm{Ext}_{A(2)_*}(\Sigma^8 \underline{\mathrm{bo}}_1)$ and $\mathrm{Ext}_{A(2)_*}(\Sigma^{17}\mathbb{F}_2[-1])$.

$h_{2,1}$–towers

Our computations of the MASS for $M(8, v_1^8)$ will rely on a detailed understanding of this spectral sequence near its vanishing line. Since $M(8, v_1^8)$ is a type 2 complex, the Hopkins–Smith periodicity theorem [20] implies that the E_∞–page of this MASS has a vanishing line of slope $1/|v_2| = \frac{1}{6}$. However, $g = h_{2,1}^4$ is not nilpotent in the modified Ext groups $\mathrm{Ext}_{A_*}(H(8, v_1^8))$, and $h_{2,1}$–multiplication has slope $\frac{1}{5}$. The goal of this subsection is to show that many of the $h_{2,1}$–towers in the E_1–page of the algebraic tmf resolution actually kill each other off by the E_2–page of the algebraic tmf resolution. We will then identify specific $h_{2,1}$–periodic elements of $\mathrm{Ext}_{A_*}(\mathbb{F}_2)$ that some of these remaining $h_{2,1}$–towers detect.

Consider the quotient Hopf algebra $C_* := \mathbb{F}_2[\zeta_2]/(\zeta_2^4)$ of $A(2)_*$, with

$$\mathrm{Ext}_{C_*}^{*,*}(\mathbb{F}_2) = \mathbb{F}_2[v_1, h_{2,1}].$$

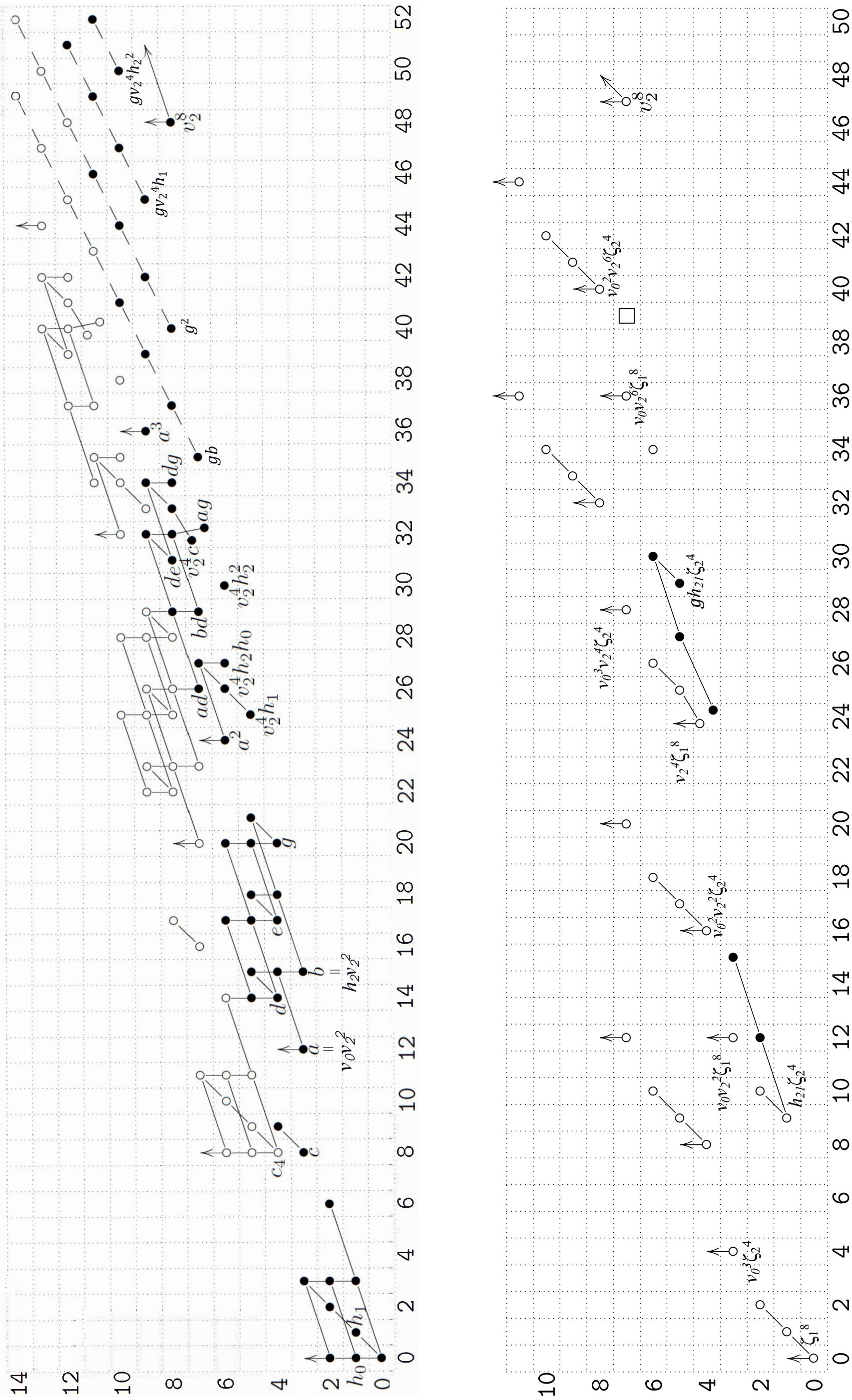


Figure 2: $\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$ (left) and $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1)$ (right).

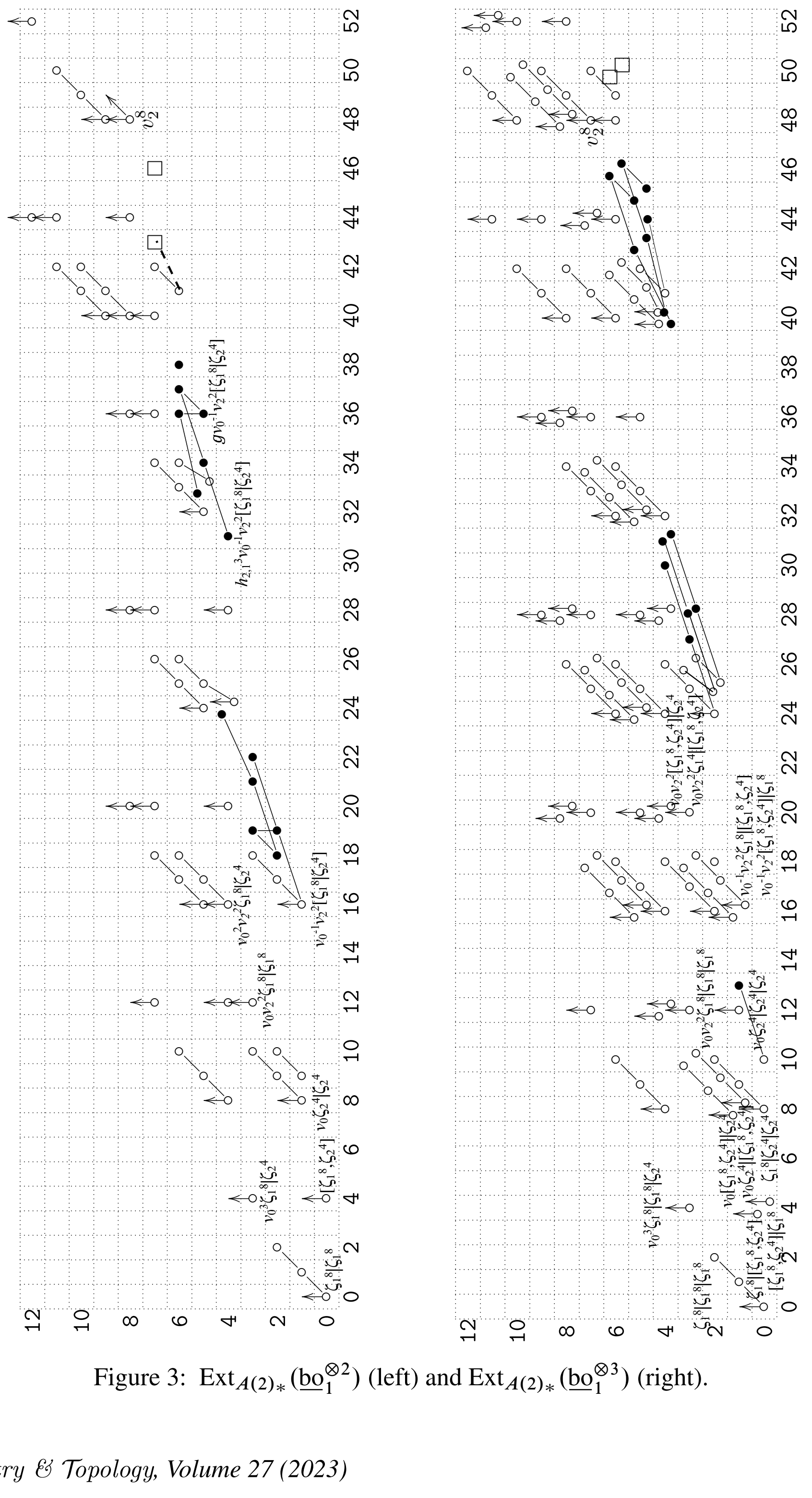


Figure 3: $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2})$ (left) and $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 3})$ (right).

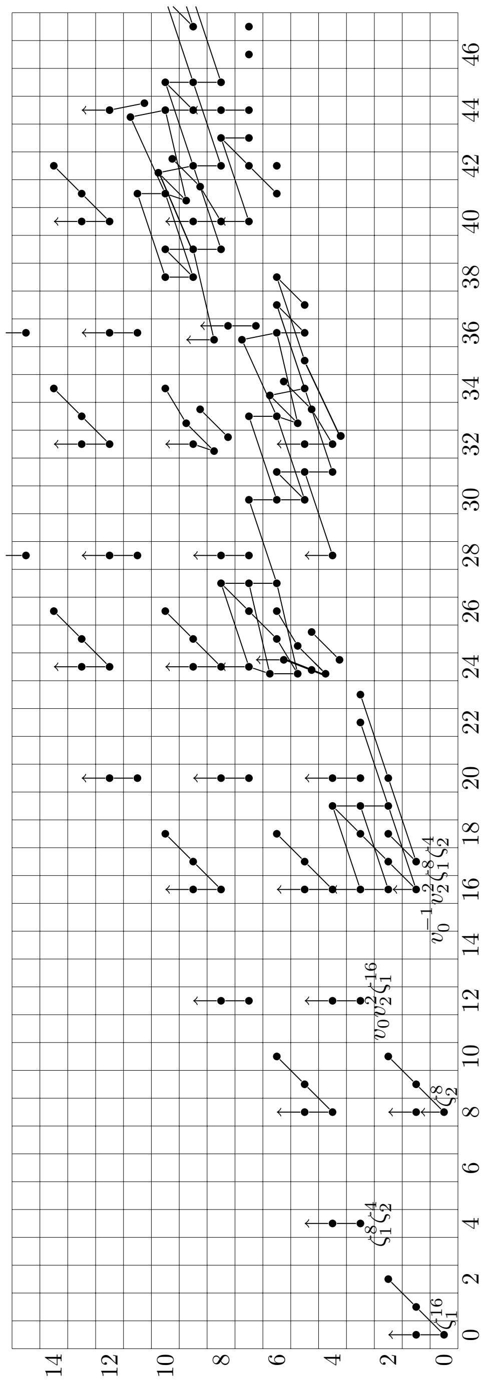


Figure 4: $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2)$.

Lemma 3.10 *Let $C(v_2^8)$ be the cofiber of the map*

$$v_2^8\colon \Sigma^{56}\mathbb{F}_2[-8] \to \mathbb{F}_2$$

in the stable homotopy category $\mathrm{St}_{A(2)_*}$. For any $M \in \mathrm{St}_{A(2)_*}$ there is an isomorphism

$$g^{-1}\,\mathrm{Ext}_{A(2)_*}(M \otimes C(v_2^8)) \cong h_{2,1}^{-1}\,\mathrm{Ext}_{C_*}(M).$$

Proof Since the element $v_2^8 \in \mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$ maps to zero in $\mathrm{Ext}_{C_*}(\mathbb{F}_2)$, it follows that there is a factorization

$$\begin{tikzcd} \mathbb{F}_2 \arrow[r] \arrow[d] & A(2)/\!/C_* \\ C(v_2^8) \arrow[ur, dotted] & \end{tikzcd}$$

in $\mathrm{St}_{A(2)_*}$. Explicit computation reveals

$$g^{-1}\,\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2) = \mathbb{F}_2[v_2^8, v_1, h_{2,1}^{\pm}]$$

and it follows that the map

$$g^{-1}C(v_2^8) \to g^{-1}A(2)/\!/C_*$$

induces an isomorphism on $\mathrm{Ext}_{A(2)_*}$, and is hence an equivalence. The result follows. □

Corollary 3.11 *For any $M \in \mathrm{St}_{A(2)_*}$, there is a v_2^8–Bockstein spectral sequence*

$$h_{2,1}^{-1}\,\mathrm{Ext}_{C_*}(M) \otimes \mathbb{F}_2[v_2^8] \Rightarrow g^{-1}\,\mathrm{Ext}_{A(2)_*}(M).$$

Bhattacharya, Bobkova and Thomas [10] computed the P_2^1–Margolis homology of the tmf resolution, and in the process computed the structure of $A/\!/A(2)_*^{\otimes n}$ as C_*–comodules. From this one can read off the Ext groups

$$h_{2,1}^{-1}\,\mathrm{Ext}_{C_*}(A/\!/A(2)_*^{\otimes n}),$$

which in turn determines the g–local algebraic tmf resolution by Corollary 3.11 (the spectral sequence in this corollary will collapse in the cases we consider it).

To state the results of [10], we will need to introduce some notation. The coaction of $\mathbb{F}_2[\zeta_2]/\zeta_2^4$ is encoded in the dual action of the algebra $E[Q_1, P_2^1]$ on $A/\!/A(2)_*^{\otimes n}$. Define elements

$$x_{i,j} = 1 \otimes \cdots \otimes 1 \otimes \underbrace{\zeta_{i+3}}_{j} \otimes 1 \otimes \cdots \otimes 1, \quad t_{i,j} = 1 \otimes \cdots \otimes 1 \otimes \underbrace{\zeta_{i+1}^4}_{j} \otimes 1 \otimes \cdots \otimes 1$$

in $A/\!/A(2)_*^{\otimes n}$. The weight filtration on $A/\!/A(2)_*$ induces a *multiweight filtration* on $A/\!/A(2)_*^{\otimes n}$ indexed by n–tuples of weights. The generators $x_{i,j}$ and $t_{i,j}$ have

multiweight
$$(0, \dots, 0, \underbrace{2^{i+2}}_{j}, 0, \dots, 0).$$
For sets of multi-indices
$$I = \{(i_1, j_1), \dots, (i_k, j_k)\}, \quad I' = \{(i'_1, j'_1), \dots, (i'_{k'}, j'_{k'})\}$$
with $I \cap I' = \varnothing$, let
$$x_I t_{I'} \in A/\!/A(2)_*$$
denote the corresponding monomial. The action of the algebra $E[Q_1, P_2^1]$ on the $\mathbb{F}_2$–submodule of $A/\!/A(2)_*^{\otimes n}$ spanned by such monomials is given by
$$Q_1(x_I t_{I'}) = \sum_{\ell} x_{I-\{(i_\ell, j_\ell)\}} t_{I' \cup \{(i_\ell, j_\ell)\}},$$
$$P_2^1(x_I t_{I'}) = \sum_{\ell < \ell'} x_{I-\{(i_\ell, j_\ell), (i_{\ell'}, j_{\ell'})\}} t_{I' \cup \{(i_\ell, j_\ell), (i_{\ell'}, j_{\ell'})\}}.$$
For an *ordered* set
$$J = ((i_1, j_1), \dots, (i_k, j_k))$$
of multi-indices, let
$$|J| := k$$
denote the number of pairs of indices it contains. Define linearly independent sets of elements
$$\mathcal{T}_J \subset A/\!/A(2)_*^{\otimes n}$$
inductively as follows. Define
$$\mathcal{T}_{(i,j)} = \{x_{i,j}\}.$$
For J as above with $|J|$ odd, define
$$\mathcal{T}_{J,(i,j)} = \{z \cdot x_{i,j}\}_{z \in \mathcal{T}_J},$$
$$\mathcal{T}_{J,(i,j),(i',j')} = \{Q_1(z \cdot x_{i,j}) x_{i',j'}\}_{z \in \mathcal{T}_J} \cup \{Q_1(z \cdot x_{i',j'}) x_{i,j}\}_{z \in \mathcal{T}_J}.$$
Let
$$N_J \subset A/\!/A(2)_*^{\otimes n}$$
denote the $\mathbb{F}_2$–subspace with basis
$$Q_1 \mathcal{T}_J := \{Q_1(z)\}_{z \in \mathcal{T}_J}.$$
While the set $\mathcal{T}_J$ depends on the ordering of J, the subspace N_J does not.

Finally, for a set of pairs of indices
$$J = \{(i_1, j_1), \dots, (i_k, j_k)\}$$

as before, define

$$x_J t_J := x_{i_1,j_1} t_{i_1,j_1} \cdots x_{i_k,j_k} t_{i_k,j_k}.$$

The following is the main theorem of [10]:[7]

Theorem 3.12 (Bhattacharya, Bobkova and Thomas) *As modules over $\mathbb{F}_2[h_{2,1}^{\pm}, v_1]$, we have*

$$\begin{aligned} &h_{2,1}^{-1} \operatorname{Ext}^{*,*}_{E[Q_1,P_2^1]}(A/\!/A(2)_*^{\otimes n}) \\ &\quad = \mathbb{F}_2[h_{2,1}^{\pm}] \otimes \Bigl(\mathbb{F}_2[v_1]\{x_{J'} t_{J'}\}_{J'} \oplus \bigoplus_{|J| \text{ odd}} N_J\{x_{J'} t_{J'}\}_{J \cap J' = \varnothing} \\ &\qquad\qquad \oplus \bigoplus_{|J| \neq 0 \text{ even}} \mathbb{F}_2[v_1]/v_1^2 \otimes N_J\{x_{J'} t_{J'}\}_{J \cap J' = \varnothing} \Bigr), \end{aligned}$$

where J and J' range over the subsets of

$$\{(i,j) : 1 \le i,\ 1 \le j \le n\}$$

and v_1 acts trivially on N_J for $|J|$ odd. The summand

$$h_{2,1}^{-1} \operatorname{Ext}^{*,*}_{E[Q_1,P_2^1]}(\underline{\mathrm{bo}}_{i_1} \otimes \cdots \otimes \underline{\mathrm{bo}}_{i_n})$$

is spanned by those monomials of multiweight $(8i_1, \ldots, 8i_n)$.

In light of Lemma 3.10 and Corollary 3.11, we may refer to elements of the g–local algebraic tmf resolution as $v_2^{8j} z$, where z is an element of the $h_{2,1}$–localized Ext groups described in the theorem above.

Lemma 3.13 *The WSS d_0–differential on the element*

$$x_{1,1} t_{1,1} \in g^{-1} \operatorname{Ext}^{*,*}_{A(2)_*}(\underline{\mathrm{bo}}_2)$$

is given by

$$d_0^{\mathrm{wss}}(x_{1,1} t_{1,1}) = Q_1(x_{1,1} x_{1,2}) \in \operatorname{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2}).$$

Proof We use the map of spectral sequences

$${}^{\mathrm{wss}}E_0 \to g^{-1}\,{}^{\mathrm{wss}}E_0.$$

[7] The main theorem of [10] is a computation of P_2^1–Margolis homology, but the actual content of the paper is a decomposition of $A/\!/A(2)_*$ in the stable module category of $E[Q_1, P_2^1]$.

By explicit computation of $g^{-1}\operatorname{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2)$, under the map

$$\operatorname{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2) \to g^{-1}\operatorname{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2)$$

we have

$$v_0^{-1}v_2^2\zeta_1^8\zeta_2^4 \mapsto h_{2,1}x_{1,1}t_{1,1}.$$

In the WSS, we have

$$d_0^{\mathrm{wss}}(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4) = v_0^{-1}v_2^2[\zeta_1^8,\zeta_2^4]. \tag{3.14}$$

Again, by explicit computation of g–local Ext groups, under the map

$$\operatorname{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2}) \to g^{-1}\operatorname{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2})$$

we have

$$v_0^{-1}v_2^2[\zeta_1^8,\zeta_2^4] \mapsto h_{2,1}Q_1(x_{1,1}x_{1,2}). \qquad \square$$

Proposition 3.15 *In $g^{-1\,\mathrm{wss}}E_0$, all of the $h_{2,1}$–towers coming from $\operatorname{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes k})$ for $k \geq 2$ either support nontrivial d_0–differentials or are the target of d_0–differentials.*

Proof By Lemma 3.10 and Theorem 3.12, the $h_{2,1}$–towers coming from

$$\operatorname{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes k})$$

are supported by the elements $\mathcal{T}_{\{(1,1),\dots,(1,k)\}}$. By Lemma 3.13, the WSS d_0 induces a surjection for $k=2$,

$$d_0^{\mathrm{wss}}\colon \mathbb{F}_2[h_{2,1}^{\pm},v_1,v_2^8]\{x_{1,1}t_{1,1}\} \twoheadrightarrow \mathbb{F}_2[h_{2,1}^{\pm},v_1,v_2^8]/v_1^2 \otimes N_{\{(1,1),(1,2)\}}.$$

For $k>2$, observe that

$$\mathcal{T}_{(1,1),\dots,(1,k)} = Q_1(x_{1,1}x_{1,2})\mathcal{T}_{(1,3),\dots,(1,k)} \cup Q_1(x_{1,2}x_{1,3})\mathcal{T}_{\{(1,1),(1,4),\dots,(1,k)\}}.$$

For $k>2$ even, the WSS d_0 gives isomorphisms

$$\begin{aligned}
d_0^{\mathrm{wss}}\colon \mathbb{F}_2[h_{2,1}^{\pm},v_1,v_2^8]/v_1^2 &\otimes x_{1,1}t_{1,1}N_{\{(1,2),\dots,(1,k-1)\}} \\
&\xrightarrow{\cong} \mathbb{F}_2[h_{2,1}^{\pm},v_1,v_2^8]/v_1^2 \otimes Q_1(x_{1,1}x_{1,2})N_{\{(1,3),\dots,(1,k)\}},\\
d_0^{\mathrm{wss}}\colon \mathbb{F}_2[h_{2,1}^{\pm},v_1,v_2^8]/v_1^2 &\otimes x_{1,2}t_{1,2}N_{\{(1,1),(1,3),\dots,(1,k-1)\}} \\
&\xrightarrow{\cong} \mathbb{F}_2[h_{2,1}^{\pm},v_1,v_2^8]/v_1^2 \otimes Q_1(x_{1,2}x_{1,3})N_{\{(1,1),(1,4),\dots,(1,k)\}},
\end{aligned}$$

and, for $k>2$ odd, the WSS d_0 gives isomorphisms

$$\begin{aligned}
d_0^{\mathrm{wss}}\colon \mathbb{F}_2[h_{2,1}^{\pm},v_2^8] &\otimes x_{1,1}t_{1,1}N_{\{(1,2),\dots,(1,k-1)\}} \\
&\xrightarrow{\cong} \mathbb{F}_2[h_{2,1}^{\pm},v_2^8] \otimes Q_1(x_{1,1}x_{1,2})N_{\{(1,3),\dots,(1,k)\}},\\
d_0^{\mathrm{wss}}\colon \mathbb{F}_2[h_{2,1}^{\pm},v_2^8] &\otimes x_{1,2}t_{1,2}N_{\{(1,1),(1,3),\dots,(1,k-1)\}} \\
&\xrightarrow{\cong} \mathbb{F}_2[h_{2,1}^{\pm},v_2^8] \otimes Q_1(x_{1,2}x_{1,3})N_{\{(1,1),(1,4),\dots,(1,k)\}}. \qquad \square
\end{aligned}$$

We shall denote the elements of the Mahowald–Tangora wedge [32] in $\mathrm{Ext}_{A_*}(\mathbb{F}_2)$ by[8]

$$v_1^i h_{2,1}^j g^2, \quad i \geq 0,\ j \geq 0.$$

Recall that the Mahowald operator

$$M = \langle g_2, h_0^3, - \rangle$$

leads to an infinite collection of wedges

$$M^k(v_1^i h_{2,1}^j g^2) \in \mathrm{Ext}_{A_*}(\mathbb{F}_2)$$

with nonzero image in

$$\mathrm{Ext}_{B_*}(\mathbb{F}_2) = \mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)[v_3],$$

where B_* is the quotient algebra

(3.16) $$B_* := \mathbb{F}_2[\zeta_1, \zeta_2, \zeta_3, \zeta_4]/(\zeta_1^8, \zeta_2^4, \zeta_3^2, \zeta_4^2)$$

of A_* [33; 23]. The existence of the element $\Delta^2 g^2 \in \mathrm{Ext}_{A_*}(\mathbb{F}_2)$ gives elements

$$\Delta^{2m} M^k(v_1^i h_{2,1}^{j+8m} g^2) \in \mathrm{Ext}_{A_*}(\mathbb{F}_2).$$

These elements are all linearly independent, since they project to linearly independent elements of $\mathrm{Ext}_{B_*}(\mathbb{F}_2)$.

The following proposition gives the elements of $\mathrm{Ext}_{A(2)_*}$ that some of the remaining $h_{2,1}$–towers in $\mathrm{Ext}_{A(2)_*}$ detect in the algebraic tmf resolution:

Proposition 3.17 *The following table lists, for $i \geq 0$, $m \geq 0$ and $j \geq 4$, an $A(2)_*$–comodule M, an $h_{2,1}$–tower in $g^{-1}\,\mathrm{Ext}_{A(2)_*}(M)$, the corresponding $h_{2,1}$–tower in $\mathrm{Ext}_{A(2)_*}(M)$, and an $h_{2,1}$–tower in $\mathrm{Ext}_{A_*}(\mathbb{F}_2)$ that it detects in the algebraic tmf resolution (assuming the latter is nonzero):*

M	$g^{-1}\,\mathrm{Ext}_{A(2)_*}(M)$	$\mathrm{Ext}_{A(2)_*}(M)$	$\mathrm{Ext}_{A_*}(\mathbb{F}_2)$
$\mathbb{F}_2$	$\Delta^{2m} v_1^i h_{2,1}^{j+8m+8}$	$\Delta^{2m} v_1^i h_{2,1}^{j+8m} g^2$	$\Delta^{2m} v_1^i h_{2,1}^{j+8m} g^2$
$\underline{\mathrm{bo}}_1$	$\Delta^{2m} h_{2,1}^{j+8m+4} Q_1(x_{1,1})$	$\Delta^{2m} h_{2,1}^{j+8m+4} \zeta_2^4$	$\Delta^{2m} h_{2,1}^{j+8m} n$
$\underline{\mathrm{bo}}_2$	$\Delta^{2m} h_{2,1}^{j+8m+6} Q_1(x_{2,1})$	$\Delta^{2m} h_{2,1}^{j+8m+1} g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})$	$\Delta^{2m} h_{2,1}^{j+8m} Q_2$
	$\Delta^{2m} v_1^{i+2} h_{2,1}^{j+8m+11} x_{1,1} t_{1,1}$	$\Delta^{2m} v_1^{i+2} h_{2,1}^{j+8m+2} g^2 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)$	$\Delta^{2m} v_1^i h_{2,1}^{j+8m} M g^2$

[8] This notation is slightly misleading, as there are a few wedge elements for which the P operator does not take the element we are denoting by $v_1^i x$ to the element we are denoting by $v_1^{i+4} x$, but we justify this notation by the fact that the wedge elements map to elements with such names in $\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$.

(Note that the notation Q_2 in the above table refers to the name of the generator of $\mathrm{Ext}_{A_*}^{7,57+7}(\mathbb{F}_2)$, and *not* the Milnor generator $Q_2 \in A$.)

Proof The classes corresponding to $\Delta^{2m} v_1^i h_{2,1}^k$ are clear, because they are in the image of the map

$$\mathrm{Ext}_{A_*}(\mathbb{F}_2) \to \mathrm{Ext}_{A(2)_*}(\mathbb{F}_2).$$

In the case of the classes corresponding to $\Delta^{2m} h_{2,1}^k n$ and $\Delta^{2m} h_{2,1}^k Q_2$, we consider the $h_{2,1}^j$–multiples of n and $Q_2 \in \mathrm{Ext}_{A_*}(\mathbb{F}_2)$ for $j \geq 4$:

$$gn, \quad gt, \quad rn, \quad mn, \quad g^2 n, \quad \dots, \quad gQ_2, \quad gC_0, \quad rQ_2, \quad mQ_2, \quad g^2 Q_2, \quad \dots.$$

It suffices to show that

$$n, \quad t, \quad Q_2, \quad C_0$$

are detected in the algebraic tmf resolution by

$$h_{2,1}^4 \zeta_2^4 + \alpha_1, \quad h_{2,1}^5 \zeta_2^4 + \alpha_2, \quad h_{2,1}^6 v_0^{-2} v_2^2 \zeta_1^{16} + \alpha_3, \quad h_{2,1}^7 v_0^{-2} v_2^2 \zeta_1^{16} + \alpha_4, \tag{3.18}$$

where $g\alpha_i = r\alpha_i = m\alpha_i = 0$.

Examination of a computer calculation of $\mathrm{Ext}_{A_*}(\overline{A /\!/ A(2)}_*^{\otimes 2})$ reveals that none of the elements n, t, Q_2 and C_0 are in the image of the map

$$\mathrm{Ext}_{A_*}^{*,*}(\overline{A /\!/ A(2)}_*^{\otimes 2}) \to \mathrm{Ext}_{A_*}^{*+2,*}(\mathbb{F}_2). \tag{3.19}$$

Since the elements n, t, Q_2 and C_0 map to zero in $\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$, they must therefore be detected on the 1–line of the algebraic tmf resolution. Examination of the relevant Ext charts reveals the only possibility is for the elements to be detected by classes of the form (3.18).

If we consider the class $Mg \in \mathrm{Ext}_{A_*}(\mathbb{F}_2)$, one can check both that it is not in the image of (3.19), and that the only class in $\mathrm{Ext}_{A(2)_*}(\overline{A /\!/ A(2)_*})$ which can detect it is the class

$$e_0^2 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4) \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2).$$

It follows from the multiplicative structure of the wedge and the fact that

$$g e_0^2 = v_1^2 h_{2,1}^2 g^2,$$

that the elements $v_1^i h_{2,1}^j M g^2 \in \mathrm{Ext}_{A_*}(\mathbb{F}_2)$ are detected by

$$v_1^{i+2} h_{2,1}^{j+2} g^2 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4) \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2)$$

for $i \geq 0$ and $j \geq 4$. □

4 The MASS for $M(8, v_1^8)$

In this and the following sections, we shall use the notation

$$x[k]$$

to denote an element of $\mathrm{Ext}_{A(2)_*}(M \otimes H(8, v_1^8))$ detected by an element

$$x \in \mathrm{Ext}_{A(2)_*}(M)$$

on the k–cell of $H(8, v_1^8)$ for $k \in \{0, 1, 17, 18\}$.

The MASS for $\mathrm{tmf}_* M(8, v_1^8)$

The computation of $\mathrm{Ext}_{A(2)_*}(H(8, v_1^8))$ is depicted in Figure 5. In this figure, solid dots correspond to classes carried by the "0–cell" of $H(8, v_1^8)$, and open circles correspond to classes carried by the "1–cell" of $H(8, v_1^8)$. The large solid circles correspond to h_0–torsion-free classes of $\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$ on the 0–cell of $H(8, v_1^8)$. The classes with solid boxes around them support $h_{2,1}$–towers. Everything is v_2^8–periodic.

Figure 6 depicts the differentials in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$ through the same range; the complete computation of this MASS can be similarly accomplished. An explanation of how to determine these differentials can be found in [7].

The algebraic tmf resolution for $H(8, v_1^8)$

The following lemma explains that, in our $H(8, v_1^8)$ computations, we may disregard terms coming from $\mathrm{Ext}_{A(1)_*}$ in the sequence of spectral sequences (2.9):

Lemma 4.1 [7, Lemma 8.8] *In the algebraic* tmf *resolution for $M = H(8, v_1^8)$, the terms*

$$\mathrm{Ext}_{A(1)_*}(\text{something})$$

in (2.9) *do not contribute to* $\mathrm{Ext}^{s,t}_{A_*}(H(8, v_1^8))$ *if*

$$s > \tfrac{1}{7}(t-s) + \tfrac{51}{7}.$$

For $n > 0$ and $i_1, \dots, i_n > 0$, the terms

$$\mathrm{Ext}^{s,t}_{A(2)_*}(\underline{\mathrm{bo}}_{i_1} \otimes \cdots \otimes \underline{\mathrm{bo}}_{i_n} \otimes H(8, v_1^8))$$

that are the terms in the algebraic tmf resolution for $H(8, v_1^8)$ are in some sense less complicated than $\mathrm{Ext}_{A(2)_*}(H(8, v_1^8))$.

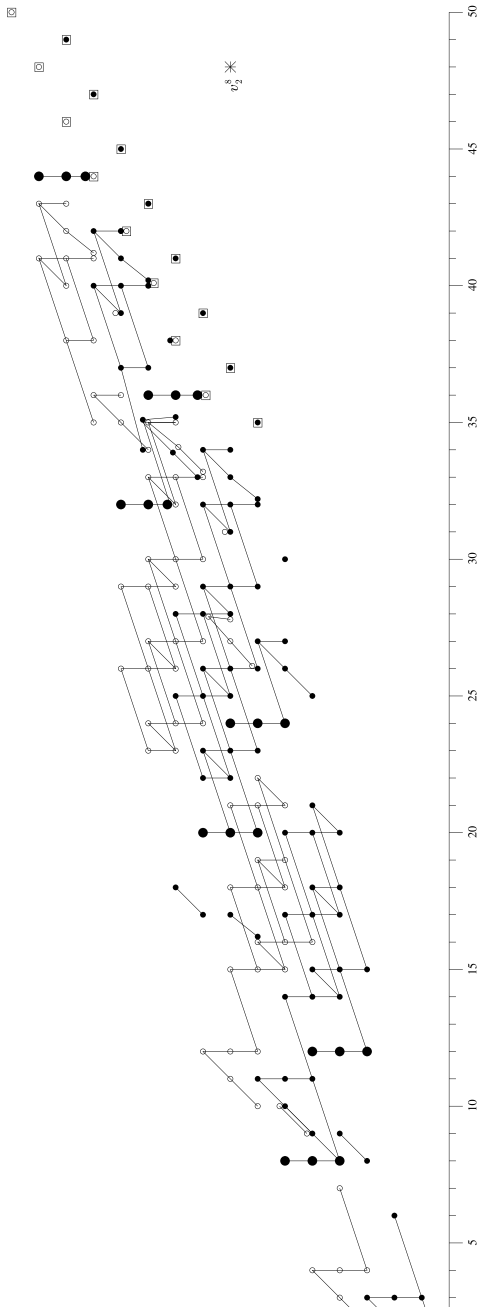


Figure 5: The groups $\mathrm{Ext}_{A(2)_*}(H(8, v_1^8))$.

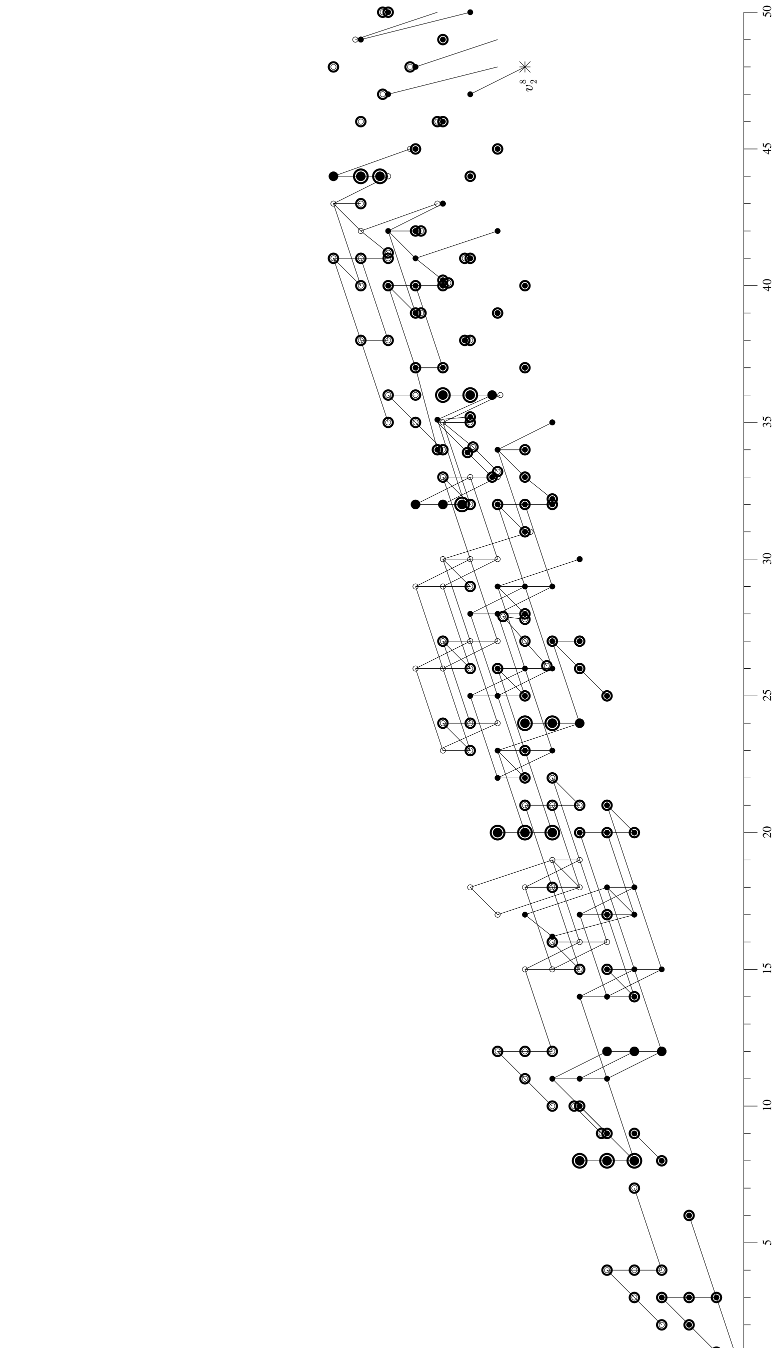


Figure 6: The MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$.

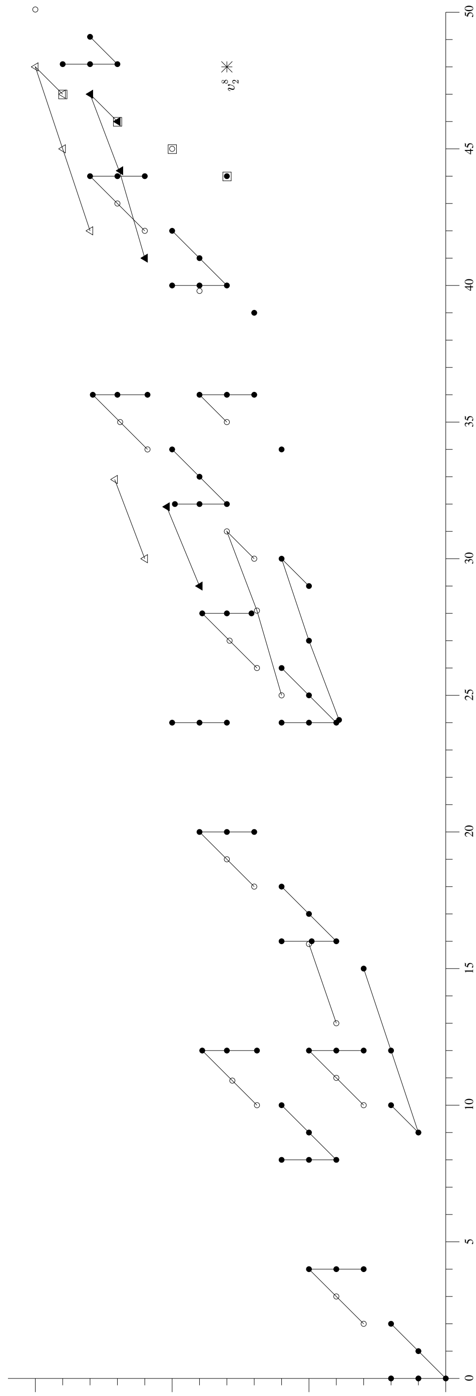


Figure 7: $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes H(8, v_1^8))$.

Most of the features of these computations can already be seen in the computation of $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes H(8, v_1^8))$, which is displayed in Figure 7. This computation was performed by taking the computation of $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1)$ (see for example [6]) and running the long exact sequences in Ext associated to the cofiber sequences

$$\Sigma^3 \underline{\mathrm{bo}}_1[-3] \xrightarrow{h_0^3} \underline{\mathrm{bo}}_1 \to \underline{\mathrm{bo}}_1 \otimes H(8),$$
$$\Sigma^{24} \underline{\mathrm{bo}}_1 \otimes H(8)[-8] \xrightarrow{v_1^8} \underline{\mathrm{bo}}_1 \otimes H(8) \to \underline{\mathrm{bo}}_1 \otimes H(8, v_1^8).$$

In Figure 7, as before, solid dots represent generators carried by the 0–cell of $H(8, v_1^8)$ and open circles are carried by the 1–cell. Unlike the case of $\mathrm{Ext}_{A(2)_*}(H(8))$, there is v_1^8–torsion in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes H(8))$. This results in classes in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes H(8, v_1^8))$ carried by the 17–cell and the 18–cell of $H(8, v_1^8)$, which are represented by solid triangles and open triangles, respectively. A box around a generator indicates that it actually carries a copy of $\mathbb{F}_2[h_{2,1}]$. As before, everything is v_2^8–periodic.

One can similarly compute

$$\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes k} \otimes H(8, v_1^8))$$

for larger values of k by applying the same method to the corresponding computations of

$$\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes k})$$

in [6]. We do not bother to record the complete results of these computations for small values of k, but will freely use them in what follows. The spectral sequences (2.9) imply these computations control $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_I)$.

$h_{2,1}$–towers in the algebraic tmf resolution for $H(8, v_1^8)$

Theorem 3.12 has the following implication for the g–local algebraic tmf resolution of $H(8, v_1^8)$:

$$\begin{aligned} &h_{2,1}^{-1} \mathrm{Ext}^{*,*}_{E[Q_1, P_2^1]}(A /\!/ A(2)_*^{\otimes n} \otimes H(8, v_1^8)) \\ &= \mathbb{F}_2[h_{2,1}^{\pm}] \otimes \Bigg(\mathbb{F}_2[v_1]/v_1^8 \otimes H(8)\{x_{J'} t_{J'}\}_{J'} \oplus \bigoplus_{|J| \text{ odd}} N_J \otimes H(8, v_1^8)\{x_{J'} t_{J'}\}_{J \cap J' = \varnothing} \\ &\qquad \oplus \bigoplus_{|J| \neq 0 \text{ even}} \mathbb{F}_2[v_1]/v_1^2 \otimes N_J \otimes H(8, v_1^8)\{x_{J'} t_{J'}\}_{J \cap J' = \varnothing} \Bigg), \end{aligned}$$

where J and J' range over the subsets of

$$\{(i, j) : 1 \le i,\ 1 \le j \le n\}.$$

This leads to the following twist in the analog of Proposition 3.15:

Proposition 4.2 *In $g^{-1\,\mathrm{wss}}E_0(H(8,v_1^8))$, all of the $h_{2,1}$–towers coming from*

$$\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes k} \otimes H(8,v_1^8))$$

for $k \geq 3$ are either the source of a nontrivial d_0–differential or the target of a d_0–differential. For $k=2$, the $h_{2,1}$–towers

$$v_1^{\epsilon} h_{2,1}^{j} Q_1(x_{1,1}x_{1,2})[n]$$

are killed for $\epsilon \in \{0,1\}$ and $n \in \{0,1\}$ (but the corresponding towers with $n \in \{17,18\}$ are *not* killed).

Proof Everything is identical to the proof of Proposition 3.15, except that the differentials

$$d_0^{\mathrm{wss}}\colon \mathbb{F}_2[v_1,h_{2,1}^{\pm}]/v_1^8\{x_{1,1}t_{1,1}\}\otimes H(8) \to \mathbb{F}_2[v_1,h_{2,1}^{\pm}]/v_1^2\{Q_1(x_{1,1}x_{1,2})\}\otimes H(8,v_1^8)$$

now have nontrivial kernel and cokernel. □

We now give elements of $\mathrm{Ext}_{A_*}(H(8,v_1^8))$ which these remaining $h_{2,1}$–towers detect in the algebraic tmf resolution. Note that, as pointed out in [33], the Mahowald operator satisfies

$$h_0^3 M(x) = 0,$$

which implies that, for any $x \in \mathrm{Ext}_{A_*}(\mathbb{F}_2)$, there exists a lift

$$M(x)[1] \in \mathrm{Ext}_{A_*}(H(8))$$

and thus an element $M(x)[1] \in \mathrm{Ext}_{A_*}(H(8,v_1^8))$. Furthermore, the element $\Delta^2 = v_2^8$ exists in $\mathrm{Ext}_{A_*}(H(8,v_1^8))$ (see Lemma 5.1 below). We conclude that, for $0 \leq i \leq 7$, $j,k,l \geq 0$ and $\epsilon \in \{0,1\}$, the wedge elements

$$v_1^i h_{2,1}^j \Delta^{2k} M^l g^2[\epsilon] \in \mathrm{Ext}_{A_*}(H(8,v_1^8))$$

exist, and we see they are linearly independent by mapping to $\mathrm{Ext}_{B_*}(H(8,v_1^8))$ (where B_* is as defined in (3.16)).

Proposition 4.3 *The following table lists, for $m \geq 0$, $0 \leq i \leq 7$, $0 \leq i' \leq 5$, $j \geq 4$, $k \in \{0,1,17,18\}$ and $\epsilon,\epsilon' \in \{0,1\}$, an $A(2)_*$–comodule M, an $h_{2,1}$–tower in $g^{-1}\,\mathrm{Ext}_{A(2)_*}(M \otimes H(8,v_1^8))$, the corresponding $h_{2,1}$–tower in $\mathrm{Ext}_{A(2)_*}(M \otimes H(8,v_1^8))$ and an $h_{2,1}$–tower in $\mathrm{Ext}_{A_*}(H(8,v_1^8))$ that it detects in the algebraic tmf resolution*:

M	$g^{-1}\,\mathrm{Ext}_{A(2)_*}(M\otimes H(8,v_1^8))$	$\mathrm{Ext}_{A(2)_*}(M\otimes H(8,v_1^8))$	$\mathrm{Ext}_{A_*}(H(8,v_1^8))$
$\mathbb{F}_2$	$\Delta^{2m}v_1^i h_{2,1}^{j+8}[\epsilon]$	$\Delta^{2m}v_1^i h_{2,1}^{j}g^2[\epsilon]$	$\Delta^{2m}v_1^i h_{2,1}^{j}g^2[\epsilon]$
$\underline{\mathrm{bo}}_1$	$\Delta^{2m}h_{2,1}^{j+4}Q_1(x_{1,1})[k]$	$\Delta^{2m}h_{2,1}^{j+4}\zeta_2^4[k]$	$\Delta^{2m}h_{2,1}^{j}n[k]$
$\underline{\mathrm{bo}}_2$	$\Delta^{2m}h_{2,1}^{j+6}Q_1(x_{2,1})[k]$	$\Delta^{2m}h_{2,1}^{j+1}g(h_{2,1}v_0^{-2}v_2^2\zeta_1^{16})[k]$	$\Delta^{2m}h_{2,1}^{j}Q_2[k]$
	$\Delta^{2m}v_1^{i'+2}h_{2,1}^{j+11}x_{1,1}t_{1,1}[\epsilon]$	$\Delta^{2m}v_1^{i'+2}h_{2,1}^{j+2}\cdot g^2(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)[\epsilon]$	$\Delta^{2m}v_1^{i'}h_{2,1}^{j}Mg^2[\epsilon]$
$\underline{\mathrm{bo}}_1^{\otimes 2}$	$v_1^{\epsilon'}\Delta^{2m}h_{2,1}^{j+11}\cdot Q_1(x_{1,1}x_{1,2})[17+\epsilon]$	$\Delta^{2m}v_1^{\epsilon'}h_{2,1}^{j+2}\cdot g^2(v_0^{-1}v_2^2[\zeta_1^8,\zeta_2^4])[17+\epsilon]$	$\Delta^{2m}v_1^{6+\epsilon'}h_{2,1}^{j}Mg^2[\epsilon]$

Proof The cases of

$$\Delta^{2m}v_1^i h_{2,1}^j g^2[\epsilon],\quad \Delta^{2m}h_{2,1}^j n[\epsilon],\quad \Delta^{2m}h_{2,1}^j Q_2[\epsilon],\quad \Delta^{2m}v_1^{i'}h_{2,1}^j Mg^2[\epsilon]$$

follow immediately from Proposition 3.17 since all of these elements are annihilated by v_0^3.

The elements

$$h_{2,1}^{j+4}\zeta_2^4\in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1),\quad h_{2,1}^{j+6}\zeta_1^{16}\in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2)$$

lift to elements

$$\begin{aligned} &h_{2,1}^{j+4}\zeta_2^4[17+\epsilon]\in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1\otimes H(8,v_1^8)),\\ &h_{2,1}^{j+6}\zeta_1^{16}[17+\epsilon]\in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2\otimes H(8,v_1^8)).\end{aligned}\tag{4.4}$$

One can explicitly check that the lifts (4.4) are permanent cycles in the algebraic tmf resolution. Therefore they detect the desired elements

$$h_{2,1}^j n[17+\epsilon], h_{2,1}^j Q_2[17+\epsilon]\in \mathrm{Ext}_{A_*}(H(8,v_1^8)).$$

Applying case (5) of the geometric boundary theorem [5, Lemma A.4.1] to the triangle

$$H(8,v_1^8)[-1]\to \Sigma^{24}H(8)[-8]\xrightarrow{v_1^8} H(8)\to H(8,v_1^8)$$

and the differential

$$d_1(v_1^{\epsilon'}h_{2,1}^{j+2}g^2(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)) = v_1^{\epsilon'}h_{2,1}^{j+2}g^2(v_0^{-1}v_2^2[\zeta_1^8,\zeta_2^4])$$

in the algebraic tmf resolution for $\Sigma^{24}H(8)[-8]$ (3.14), we find that the images of the elements

$$v_1^{8+\epsilon'}h_{2,1}^j M(g^2)[\epsilon]\in \mathrm{Ext}_{A_*}(H(8))$$

under the map

$$\mathrm{Ext}_{A_*}(H(8)) \to \mathrm{Ext}_{A_*}(H(8, v_1^8))$$

are detected by the elements

$$v_1^{\epsilon'} h_{2,1}^{j+2} g^2 (v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[17+\epsilon]$$

in the algebraic tmf resolution for $H(8, v_1^8)$. □

5 The v_2^{32}–self-map on $M(8, v_1^8)$

We now endeavor to prove Theorem 1.8. We first recall the following lemma:

Lemma 5.1 [7, Lemma 7.6] *The element*

$$v_2^8 \in \mathrm{Ext}_{A(2)_*}^{8,48+8}(H(8, v_1^8))$$

is a permanent cycle in the algebraic tmf *resolution, and gives rise to an element*

$$v_2^8 \in \mathrm{Ext}_{A_*}^{8,48+8}(H(8, v_1^8)).$$

It follows from the Leibniz rule that v_2^{32} persists to the E_4–page of the MASS for $M(8, v_1^8)$. Our task will then be reduced to showing that $d_r(v_2^{32}) = 0$ for $r \geq 4$. We will do this by identifying the potential targets of such a differential, and show that they are either the source or target of shorter differentials. This will necessitate lifting certain differentials from the MASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^n \wedge M(8, v_1^8)$ to the MASS for $M(8, v_1^8)$.

As explained in [8, Section 7.4], work of the second author, Davis and Rezk [31; 15] implies that the algebraic map

$$\mathrm{Ext}_{A(2)}(\Sigma^8 \underline{\mathrm{bo}}_1 \oplus \Sigma^{16} \underline{\mathrm{bo}}_2) \to \mathrm{Ext}_{A(2)_*}(\overline{A /\!/ A(2)}_*)$$

realizes to a map

$$\mathrm{tmf} \wedge \overline{\mathrm{tmf}}_2 \to \mathrm{tmf} \wedge \overline{\mathrm{tmf}}, \tag{5.2}$$

where $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}_2$ is a spectrum built out of $\mathrm{tmf} \wedge \Sigma^8 \mathrm{bo}_1$ and $\mathrm{tmf} \wedge \Sigma^{16} \mathrm{bo}_2$. They furthermore show that there is a map

$$\Sigma^{32} \mathrm{tmf} \to \mathrm{tmf} \wedge \overline{\mathrm{tmf}}_2, \tag{5.3}$$

which geometrically realizes the inclusion of the direct summand (2.9),

$$\mathrm{Ext}_{A(2)_*}(\Sigma^{33} \mathbb{F}_2[-1]) \hookrightarrow \mathrm{Ext}_{A(2)_*}(\Sigma^{16} \underline{\mathrm{bo}}_2) \subset \mathrm{Ext}_{A(2)_*}(\Sigma^8 \underline{\mathrm{bo}}_1 \oplus \Sigma^{16} \underline{\mathrm{bo}}_2).$$

The attaching map from $\mathrm{tmf} \wedge \mathrm{bo}_2$ to $\mathrm{tmf} \wedge \mathrm{bo}_1$ in the spectrum $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}_2$ induces d_3–differentials from the $h_{2,1}$–towers in bo_2 to the $h_{2,1}$–towers in bo_1 in the ASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}$ under the map (5.2). Furthermore, there are differentials in the ASSs for $\mathrm{tmf} \wedge \mathrm{bo}_1$, $\mathrm{tmf} \wedge \mathrm{bo}_2$ and tmf, which induce differentials in the ASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}$ under the maps (5.2) and (5.3). We wish to study when these differentials (and more generally differentials in the ASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^n$) lift via the tmf resolution to differentials in the ASS for the sphere.

To this end we consider the partial totalizations

$$T^n := \mathrm{Tot}^n(\mathrm{tmf}^{\bullet+1})$$

of the cosimplicial tmf resolution of the sphere, so that we have

$$S \simeq \varprojlim_n T^n$$

and fiber sequences

$$\Sigma^{-n}\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^n \to T^n \to T^{n-1}.$$

The spectrum T^n is a ring spectrum, and in particular has a unit

$$S \to T^n.$$

We let

$$\underline{T}^n = \mathrm{Tot}^n(A/\!/A(2)_*^{\otimes \bullet+1}) \tag{5.4}$$

denote the corresponding construction in the stable homotopy category of A_*–comodules. There is a MASS

$$\mathrm{Ext}^{*,*}_{A_*}(\underline{T}^n \otimes H(8, v_1^8)) \Rightarrow T^n_* M(8, v_1^8)$$

and the algebraic tmf resolution for $H(8, v_1^8)$ truncates to give an algebraic tmf resolution

$$\bigoplus_{i=0}^{n} \mathrm{Ext}^{*,*}_{A(2)_*}(\overline{A/\!/A(2)}_*^{\otimes i} \otimes H(8, v_1^8)) \Rightarrow \mathrm{Ext}_{A_*}(\underline{T}^n \otimes H(8, v_1^8)).$$

The following lemma will be our key to lifting the desired differentials:

Lemma 5.5 *Suppose x is an element of $\mathrm{Ext}_{A_*}(H(8, v_1^8))$ which is detected in the n–line of the algebraic tmf resolution for $H(8, v_1^8)$ by an element*

$$x' \in \mathrm{Ext}_{A(2)_*}(\overline{A/\!/A(2)}_*^{\otimes n} \otimes H(8, v_1^8)).$$

Furthermore, suppose that, in the MASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^n \wedge M(8, v_1^8)$, there is a differential

$$d_r^{\mathrm{mass}}(x') = y'$$

and that, for $2 \le r' < r$, we have

$$d_{r'}^{\mathrm{mass}}(x) = 0$$

in the MASS for the $M(8, v_1^8)$. Then either

(1) *the differential*
$$d_r^{\mathrm{mass}}(x)$$
in the ASS for $M(8, v_1^8)$ is detected by y' in the algebraic tmf resolution; *or*

(2) *the element y' is the target of a differential in the algebraic tmf resolution for $H(8, v_1^8)$*, *or*, *in the algebraic tmf resolution for $\underline{T}^n \otimes H(8, v_1^8)$*, *$y'$ detects an element of $\mathrm{Ext}_{A_*}(\underline{T}^n \otimes H(8, v_1^8))$ which is zero in ${}^{\mathrm{mass}}E_r(T^n \wedge M(8, v_1^8))$.*

Proof Consider the maps of algebraic tmf resolutions and MASSs induced from the zigzag

$$M(8, v_1^8) \xrightarrow{\alpha} T^n \wedge M(8, v_1^8) \xleftarrow{\beta} \Sigma^{-n}\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^n \wedge M(8, v_1^8).$$

Define

$$\bar{x} := \alpha_*(x) \in \mathrm{Ext}_{A_*}(\underline{T}^n \otimes H(8, v_1^8))$$

Then $\bar{x}$ is detected by x', regarded as an element of the algebraic tmf resolution for $T^n \wedge M(8, v_1^8)$. In particular, this means that

$$\bar{x} = \beta_*(x')$$

Therefore, the differential

$$d_r^{\mathrm{mass}}(x') = y'$$

in the MASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^n \wedge M(8, v_1^8)$ maps to a differential

$$d_r^{\mathrm{mass}}(\bar{x}) = \bar{y} := \beta_*(y')$$

in the MASS for $T^n \wedge M(8, v_1^8)$. In particular, either

(1) $\bar{y}$ is nonzero in ${}^{\mathrm{mass}}E_r(T^n \wedge M(8, v_1^8))$ and is detected by y' in the algebraic tmf resolution for $\underline{T}^n \otimes H(8, v_1^8)$, or

(2) either $\bar{y} = 0$ in ${}^{\mathrm{mass}}E_r(T^n \wedge M(8, v_1^8))$ or y' is killed in the algebraic tmf resolution for $\underline{T}^n \otimes H(8, v_1^8)$.

If the latter is true, then y' is killed in the algebraic tmf resolution for $H(8, v_1^8)$, since the algebraic tmf resolution for $\underline{T}^n \otimes H(8, v_1^8)$ is a truncation of the algebraic tmf resolution for $H(8, v_1^8)$.

If we are in case (2), we are done. If we are in case (1), consider the differential

$$y := d_r^{\mathrm{mass}}(x)$$

in the MASS for $M(8, v_1^8)$ (which is defined by hypothesis). We must have

$$\alpha_*(y) = \bar{y}.$$

Therefore, $d_r^{\mathrm{mass}}(x)$ is detected by y' in the algebraic tmf resolution. □

Remark 5.6 We will primarily be applying Lemma 5.5 to the following two cases:

Case 1 $(x = \Delta^{2m} h_{2,1}^j Q_2[k])$ Suppose that we can prove

$$d_2^{\mathrm{ass}}(\Delta^{2m} h_{2,1}^j Q_2[k]) = 0$$

in the MASS for $M(8, v_1^8)$. The element $\Delta^{2m} h_{2,1}^j Q_2[k]$ is detected by

$$\Delta^{2m} h_{2,1}^{j+1} g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})[k] \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2 \otimes H(8, v_1^8))$$

in the algebraic tmf resolution, and it is proven in [8] that, in the ASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}$, there is a differential

$$\begin{aligned} d_3^{\mathrm{ass}}&(\Delta^{2m} h_{2,1}^{j+1} g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})) \\ &= \Delta^{2m} h_{2,1}^{j+4} g(h_{2,1} \zeta_2^4) + \epsilon(m) \Delta^{2m-4} h_{2,1}^{j+20} g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16}), \end{aligned}$$

where

$$\epsilon(m) = \begin{cases} 1 & \text{if } m \equiv 2 \bmod 4, \\ 0 & \text{otherwise.} \end{cases}$$

Lifting this differential to $\mathrm{tmf} \wedge \overline{\mathrm{tmf}} \wedge M(8, v_1^8)$, Lemma 5.5 implies that either the target of the differential $d_3^{\mathrm{ass}}(\Delta^{2m} h_{2,1}^j Q_2[k])$ in the MASS for $M(8, v_1^8)$ is detected by

$$\Delta^{2m} h_{2,1}^{j+4} g(h_{2,1} \zeta_2^4)[k] + \epsilon(m) \Delta^{2m-4} h_{2,1}^{j+20} g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})[k]$$

in the algebraic tmf resolution, or

$$\Delta^{2m} h_{2,1}^{j+4} g(h_{2,1} \zeta_2^4)[k] + \epsilon(m) \Delta^{2m-4} h_{2,1}^{j+20} g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})[k]$$

is the target of a differential in the algebraic tmf resolution or detects an element of $\mathrm{Ext}_{A_*}(\underline{T}^1 \otimes H(8, v_1^8))$ which is zero on the E_3–page of the MASS for $T^1 \wedge M(8, v_1^8)$.

Case 2 $(x = M \Delta^2 v_1^i h_{2,1}^{j+8}[\epsilon]$ for $\epsilon \in \{0, 1\}$ and $0 \le i \le 4)$ The element $M \Delta^2 v_1^i h_{2,1}^{j+8}[\epsilon]$ is detected by

$$\Delta^2 v_1^{i+2} h_{2,1}^{j+10} (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[\epsilon]$$

in the algebraic tmf resolution for $H(8,v_1^8)$, and the map (5.3) implies there is a differential

$$d_2^{\mathrm{mass}}(\Delta^2 v_1^{i+2} h_{2,1}^{j+10}(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)[\epsilon]) = v_1^{i+3}h_{2,1}^{j+19}(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)[\epsilon]$$

in the MASS for $\mathrm{tmf}\wedge\overline{\mathrm{tmf}}\wedge M(8,v_1^8)$.

Then Lemma 5.5 implies that either $d_2^{\mathrm{mass}}(M\Delta^2 v_1^i h_{2,1}^{j+8}[\epsilon])$ is detected by

$$v_1^{i+3}h_{2,1}^{j+19}(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)[\epsilon]$$

in the algebraic tmf resolution, or $v_1^{i+3}h_{2,1}^{j+19}(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)[\epsilon]$ is killed in the tmf resolution for $H(8,v_1^8)$ or it detects an element which is zero in the E_2–term of the MASS for $T^1\wedge M(8,v_1^8)$. However, the element

$$Mv_1^{i+1}h_{2,1}^{j+17}[\epsilon]\in \mathrm{Ext}_{A_*}(H(8,v_1^8))$$

is nonzero, and is detected by $v_1^{i+3}h_{2,1}^{j+19}(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)[\epsilon]$ in the algebraic tmf resolution for $H(8,v_1^8)$. We conclude that $v_1^{i+3}h_{2,1}^{j+19}(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)[\epsilon]$ is not killed in the algebraic tmf resolution for $H(8,v_1^8)$. Since the algebraic tmf resolution for $\underline{T}^1\otimes H(8,v_1^8)$ is a truncation of the algebraic tmf resolution for $H(8,v_1^8)$, we conclude that $v_1^{i+3}h_{2,1}^{j+19}(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)[\epsilon]$ detects a nontrivial element of the E_2–page of the MASS for $T^1\wedge M(8,v_1^8)$. We conclude that

$$d_2^{\mathrm{mass}}(M\Delta^2 v_1^i h_{2,1}^{j+8}[\epsilon])$$

is nontrivial in the MASS for $M(8,v_1^8)$, and is detected in the algebraic tmf resolution by $v_1^{i+3}h_{2,1}^{j+19}(v_0^{-1}v_2^2\zeta_1^8\zeta_2^4)[\epsilon]$.

Proof of Theorem 1.8 By Proposition 2.3, it suffices to prove that

$$v_2^{32}\in \mathrm{Ext}_{A_*}(H(8,v_1^8))$$

is a permanent cycle in the MASS. Furthermore, since $v_2^8\in {}^{\mathrm{mass}}E_2(M(8,v_1^8))$, the Leibniz rule implies that $v_2^{32}\in {}^{\mathrm{mass}}E_4(M(8,v_1^8))$. We therefore are left with eliminating possible targets of $d_r^{\mathrm{mass}}(v_2^{32})$ for $r\geq 4$.

Suppose that $d_r(v_2^{32})$ is nontrivial for $r\geq 4$. We successively consider terms in the algebraic tmf resolution which could detect $d_r(v_2^{32})$, and then eliminate these possibilities one by one.

The only terms in the algebraic tmf resolution E_1–page which can contribute to $\mathrm{Ext}_{A_*}^{s,191+s}(H(8, v_1^8))$ for $s \geq 36$ are

- $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes s})$ for $0 \leq s \leq 6$, and
- $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes s} \otimes \underline{\mathrm{bo}}_2)$ for $0 \leq s \leq 2$.

Furthermore, $\underline{\mathrm{bo}}_1^{\otimes s}$ only contributes $h_{2,1}$–towers in this range for $s = 5, 6$. We list these contributions below, except we do not list elements in $h_{2,1}$–towers coming from $\underline{\mathrm{bo}}_1^{\otimes s}$ for $s \geq 2$ which are zero in the WSS E_1–term (see Proposition 4.2). Also, since v_2^{32} is a permanent cycle in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$, we can disregard any terms coming from $\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$ (the 0–line of the algebraic tmf resolution). Finally, we do not include any terms which can be eliminated through the application of Case 2 of Remark 5.6.

We now eliminate these possibilities one by one. We will consider the terms in order of *reverse* algebraic tmf filtration.

$\underline{\mathbf{bo}}_{\mathbf{1}}^{\otimes \mathbf{4}}$ In the modified May spectral sequence (3.3), there is a differential

$$d_8^{\mathrm{mmss}}(b_{2,2}h_3^2) = h_3^5$$

which lifts under the map Φ_* of (3.4) to a nontrivial differential

$$d_1^{\mathrm{wss}}([\zeta_1^8 \mid \zeta_1^8 \mid \zeta_2^4 \mid \zeta_2^4]) = [\zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8]$$

in the WSS for $\mathbb{F}_2$, and this implies a nontrivial differential

$$d_1^{\mathrm{wss}}(v_1^4 \Delta^6 h_1^2 [\zeta_1^8 \mid \zeta_1^8 \mid \zeta_2^4 \mid \zeta_2^4][1]) = v_1^4 \Delta^6 h_1^2 [\zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8][1]$$

in the WSS for $H(8, v_1^8)$.

$\underline{\mathbf{bo}}_{\mathbf{1}}^{\otimes \mathbf{2}} \otimes \underline{\mathbf{bo}}_{\mathbf{2}}$ In the cobar complex for $\mathbb{F}_2[\zeta_1^8, \zeta_2^4]$, we find

$$d([\zeta_1^8, \zeta_2^4] \mid \zeta_1^8 \zeta_2^4) \quad \text{and} \quad \mathrm{d}(\imath_1^8 \mid \imath_1^8 \imath_2^4 \mid \imath_2^4 + \imath_2^4 \mid \imath_1^8 \imath_2^4 \mid \imath_1^8)$$

are linearly independent, and

$$d([\zeta_1^8, \zeta_2^4] \mid \zeta_1^8 \zeta_2^4 + \zeta_1^8 \zeta_2^4 \mid [\zeta_1^8, \zeta_2^4]) = 0.$$

However,

$$d(\zeta_1^8 \zeta_2^4 \mid \zeta_1^8 \zeta_2^4) = [\zeta_1^8, \zeta_2^4] \mid \zeta_1^8 \zeta_2^4 + \zeta_1^8 \zeta_2^4 \mid [\zeta_1^8, \zeta_2^4].$$

The elements are thus eliminated by multiplying the computations above with $v_1^{-2} v_2^4 h_{2,1}^{22}$ and lifting them to the top cell of $H(8, v_1^8)$.

$\underline{\mathbf{bo}}_{\mathbf{1}}^{\otimes \mathbf{3}}$ Note that

$$\mathrm{Ext}_{A_*}^{10,10+48}(\mathbb{F}_2) = 0.$$

We conclude that the class

$$v_1^4 c_0 h_1 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4) \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_2)$$

must either support or be the target of a differential in the algebraic tmf resolution, for otherwise it would give a nonzero element of $\mathrm{Ext}_{A_*}^{10,10+48}(\mathbb{F}_2)$. However, by examination, there are no classes in $\mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$ which can kill $v_1^4 c_0 h_1 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)$ in the algebraic tmf resolution, so there must be a nontrivial differential

$$d_r(v_1^4 c_0 h_1 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4))$$

in the algebraic tmf resolution for $\mathbb{F}_2$. Since the target of this differential must be h_1–torsion, there is only one possibility:

$$d_2(v_1^4 c_0 h_1 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)) = v_1^4 h_1^2 v_2^2 \zeta_1^8 \mid \zeta_1^8 \mid \zeta_2^4.$$

It follows that we have

$$d_2(v_1^4 c_0 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)) = v_1^4 h_1 v_2^2 \zeta_1^8 \mid \zeta_1^8 \mid \zeta_2^4.$$

This differential lifts to a differential

$$d_2(v_1^4 c_0 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]) = v_1^4 h_1 v_2^2 \zeta_1^8 \mid \zeta_1^8 \mid \zeta_2^4[1]$$

in the algebraic tmf resolution for $H(8, v_1^8)$. Multiplying by Δ^6, we have

$$d_2(\Delta^6 v_1^4 c_0 (v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]) = \Delta^6 v_1^4 h_1 v_2^2 \zeta_1^8 \mid \zeta_1^8 \mid \zeta_2^4[1].$$

$\underline{\mathbf{bo}}_{\mathbf{1}} \otimes \underline{\mathbf{bo}}_{\mathbf{2}}$ There is a differential

$$d_0^{\mathrm{wss}}(\zeta_2^{12}) = [\zeta_2^4, \zeta_2^8]$$

in the WSS for $\mathbb{F}_2$ which lifts to a differential

$$d_0^{\mathrm{wss}}(v_1 h_{2,1}^{21} g (v_0^{-1} v_2^2 \zeta_2^{12})) = v_1 h_{2,1}^{21} g (v_0^{-1} v_2^2 [\zeta_2^4, \zeta_2^8]).$$

We therefore only have to consider one of the two potential elements. In the modified May spectral sequence (3.3), there is a differential

$$d_8^{\mathrm{mmss}}(h_{2,3}) = h_{1,3} h_{1,4}$$

which lifts to a differential

$$d_1^{\mathrm{wss}}(\zeta_2^8) = \zeta_1^8 \mid \zeta_1^{16}.$$

using the map Φ_* of (3.3), and gives a differential

$$d_1^{\mathrm{wss}}(\zeta_2^4 \mid \zeta_2^8) = \zeta_2^4 \mid \zeta_1^8 \mid \zeta_1^{16}.$$

The elements

$$v_1 g(v_0^{-1} v_2^2 \zeta_2^4 \mid \zeta_2^8) \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes \underline{\mathrm{bo}}_2)$$

and

$$v_1 g(v_0^{-1} v_2^2 \zeta_2^4 \mid \zeta_1^8 \mid \zeta_1^{16}) \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2} \otimes \underline{\mathrm{bo}}_2)$$

support $h_{2,1}$–towers which are nontrivial in ${}^{\mathrm{wss}}E_1$. Therefore, we have a nontrivial differential

$$d_1^{\mathrm{wss}}(v_1 h_{2,1}^{21} g(v_0^{-1} v_2^2 \zeta_2^4 \mid \zeta_2^8)) = v_1 h_{2,1}^{21} g(v_0^{-1} v_2^2 \zeta_2^4 \mid \zeta_1^8 \mid \zeta_1^{16}).$$

This differential lifts to the top cell of $H(8, v_1^8)$ to give

$$d_1^{\mathrm{wss}}(v_1 h_{2,1}^{21} g(v_0^{-1} v_2^2 \zeta_2^4 \mid \zeta_2^8)[18]) = v_1 h_{2,1}^{21} g(v_0^{-1} v_2^2 \zeta_2^4 \mid \zeta_1^8 \mid \zeta_1^{16})[18]$$

in the WSS for $H(8, v_1^8)$.

$\underline{\mathbf{bo}}_{\mathbf{1}}^{\otimes \mathbf{2}}$ The element

$$h_{2,1}^5 \Delta^4 v_1 g(v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[18]$$

detects the element

$$\Delta^4 \cdot MP\Delta h_0^2 e_0[18]$$

in the algebraic tmf resolution for $H(8, v_1^8)$. Regarding this element as an element in the MASS for $\mathrm{tmf} \wedge \mathrm{bo}_1^2$, there is a nontrivial differential

$$d_3^{\mathrm{mass}}(h_{2,1}^5 \Delta^4 v_1 g(v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[18]) = h_{2,1}^{24} v_1 g(v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[18].$$

By applying $(-)^{\wedge_{\mathrm{tmf}} 2}$ to the map of tmf–modules (5.2), we may consider the composite

(5.7) $$\mathrm{tmf} \wedge \mathrm{bo}_1^2 \hookrightarrow (\mathrm{tmf} \wedge \overline{\mathrm{tmf}}_2)^{\wedge_{\mathrm{tmf}} 2} \to \mathrm{tmf} \wedge \overline{\mathrm{tmf}}^2.$$

The differential above maps to a nontrivial differential between elements of the same name in the MASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^2$. We wish to apply Lemma 5.5. We must have

$$d_2^{\mathrm{mass}}(\Delta^4 \cdot MP\Delta h_0^2 e_0[18]) = 0$$

in the MASS for $M(8, v_1^8)$, since there are no elements in the algebraic tmf resolution for $H(8, v_1^8)$ which could detect a target for this differential. Thus Lemma 5.5 implies that either

$$d_3^{\mathrm{mass}}(\Delta^4 \cdot MP\Delta h_0^2 e_0[18])$$

is nontrivial and detected by $h_{2,1}^{24} v_1 g(v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[18]$, or

$$h_{2,1}^{24} v_1 g(v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[18]$$

is killed in the algebraic tmf resolution for $H(8, v_1^8)$, or detects an element which is killed in the MASS for $T^2 \wedge M(8, v_1^8)$. The only such possibility is for

$$\Delta^2 h_{2,1}^{23} \zeta_2^4[17]$$

to detect the source of a d_2–differential in the MASS for $T^2 \wedge M(8, v_1^8)$ to do such a killing. Projecting onto the top Moore space of $M(8, v_1^8)$, this would imply

$$\Delta^2 h_{2,1}^{23} \zeta_2^4$$

detects an element in the algebraic tmf resolution for the sphere which supports a nontrivial d_2–differential in the ASS for the sphere. However, $\Delta^2 h_{2,1}^{23} \zeta_2^4$ detects

$$\Delta^2 g^5 \cdot \Delta h_2 c_1$$

in the ASS for the sphere, and there is a differential

$$d_2^{\mathrm{ass}}(\Delta^2 g^5 \cdot \Delta h_2 c_1) = d_2^{\mathrm{ass}}(\Delta^2 g^2) \cdot g^3 \cdot \Delta h_2 c_1 = \Delta^2 h_2^2 g^2 e_0 \cdot g^3 \cdot \Delta h_2 c_1.$$

However, $\Delta^2 h_2^2 e_0 \cdot \Delta h_2 c_1 = 0$ in $\mathrm{Ext}_{A_*}(\mathbb{F}_2)$ [13], so this d_2^{ass} is zero.

We now turn our attention to the other potential target coming from $\underline{\mathrm{bo}}_1^{\otimes 2}$,

$$h_{2,1}^{15} \Delta^2 g (v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[18].$$

This element detects

$$\Delta^2 g^2 v_1^6 h_{2,1} M g^3[0]$$

in the algebraic tmf resolution for $M(8, v_1^8)$. However, in the ASS for the sphere, $v_1^6 h_{2,1} g^3$ is a d_2–cycle, and so there is a differential

$$d_2^{\mathrm{ass}}(\Delta^2 g^2 \cdot v_1^6 h_{2,1} g^3) = d_2^{\mathrm{ass}}(\Delta^2 g^2) \cdot v_1^6 h_{2,1} g^3 = \Delta^2 h_2^2 g^2 e_0 \cdot v_1^6 h_{2,1} g^3 = v_1^7 h_{2,1}^{22} g^2.$$

Applying $M(-) = \langle -, h_0^3, g_2 \rangle$ and mapping under the inclusion of the bottom cell of $M(8, v_1^8)$, we get a nontrivial differential

$$d_2^{\mathrm{mass}}(\Delta^2 g^2 \cdot v_1^6 h_{2,1} M g^3[0]) = v_1^7 h_{2,1}^{22} M g^2[0].$$

$\underline{\mathbf{bo}}_{\mathbf{1}}$ The element

$$h_{2,1}^{31} g (h_{2,1} \zeta_2^4)$$

detects

$$g^8 n \in \mathrm{Ext}_{A_*}(\mathbb{F}_2)$$

in the algebraic tmf resolution for $\mathbb{F}_2$ (Proposition 3.17). This element can be eliminated by Case 1 of Remark 5.6, but we can also handle it manually using low-dimensional calculations in the ASS for the sphere. There is a differential

$$d_3(mQ_2) = g^3 n$$

in the ASS for the sphere [24], from which it follows that $g^8 n$ is zero on the E_4–page of the ASS of the sphere, and hence $g^8 n[0]$ is zero on the E_4–page of the MASS for $M(8, v_1^8)$.

For the element

$$h_{2,1}^{18} \Delta^2 g(h_{2,1} \zeta_2^4)[17],$$

we wish to employ Case 1 of Remark 5.6, using the differential

$$d_3^{\mathrm{mass}}(h_{2,1}^{15} \Delta^2 g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})[17]) = h_{2,1}^{18} \Delta^2 g(h_{2,1} \zeta_2^4)[17]$$

in the MASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}} \wedge M(8, v_1^8)$. Note that

$$h_{2,1}^{15} \Delta^2 g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})[17]$$

detects the element

$$C'' \cdot \Delta^2 g^2 [17]$$

in the algebraic tmf resolution. Observe [25; 13] that we have

$$d_2(C'' \cdot \Delta^2 g^2) = C'' \cdot d_2(\Delta^2 g^2) = g^2 \cdot C'' \Delta h_2^2 e_0 = g^2 \cdot 0 = 0.$$

It follows that $d_2(C'' \cdot \Delta^2 g^2[17])$ is in the image of the map

$$\mathrm{Ext}_{A_*}(H(8)) \to \mathrm{Ext}_{A_*}(H(8, v_1^8)),$$

but a check of the algebraic tmf resolution for $H(8, v_1^8)$ reveals there are no possible targets in this bidegree. We therefore have

$$d_2(C'' \cdot \Delta^2 g^2[17]) = 0.$$

Therefore, the hypotheses of Lemma 5.5 are satisfied. It follows that

$$h_{2,1}^{18} \Delta^2 g(h_{2,1} \zeta_2^4)[17]$$

either is killed in the algebraic tmf resolution for $H(8, v_1^8)$, or detects an element in the MASS which is killed by $d_3(C'' \cdot \Delta^2 g^2[17])$, or detects an element which killed by a d_2–differential in the MASS for $T^1 \wedge M(8, v_1^8)$. We just need to eliminate this last possibility.

Any possible source for such a d_2–differential would necessarily be detected on the 0–line of the algebraic tmf resolution and would not support a nontrivial d_2 in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$. The only such possibility is

$$\Delta^4 h_{21}^{19}[1].$$

However, we can express this element as the Hurewicz image of the element
$$gm \cdot \Delta^4 \cdot g^2[1]$$
in the MASS for $M(8, v_1^8)$. This element is therefore necessarily a d_2–cycle, since it is a product of d_2–cycles.

$\underline{\mathrm{bo}}_2$ We begin with the element
$$h_{2,1}^5 \Delta^4 g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})[18]$$
which detects the element
$$\Delta^4 g Q_2[18]$$
in the MASS for $M(8, v_1^8)$. We are in Case 1 of Remark 5.6. An elementary check using the charts of [25] reveals that the element gQ_2 in the ASS for the sphere lifts to a d_2–cycle
$$gQ_2[18]$$
supported by the top cell of $H(8, v_1^8)$. Since Δ^4 is a d_2–cycle in the MASS for $M(8, v_1^8)$, we deduce that
$$\Delta^4 g Q_2[18]$$
is a d_2–cycle. We therefore deduce that
$$d_3^{\mathrm{mass}}(\Delta^4 g Q_2[18])$$
either is detected by
$$\Delta^4 h_{2,1}^8 g(h_{2,1}\zeta_2^4)[18] + h_{2,1}^{24} g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})$$
in the algebraic tmf resolution for $H(8, v_1^8)$, or
$$\Delta^4 h_{2,1}^8 g(h_{2,1}\zeta_2^4)[18] + h_{2,1}^{24} g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})$$
is killed in the algebraic tmf resolution for $H(8, v_1^8)$ or detects an element which is killed in the MASS for $T^1 \wedge M(8, v_1^8)$. The only possible sources of such algebraic tmf resolution differentials are wedge elements coming from $\mathrm{Ext}_{A(2)_*}(H(8, v_1^8))$, and we know these all must be permanent cycles in the algebraic tmf resolution because they detect the corresponding wedge elements of $\mathrm{Ext}_{A_*}(H(8, v_1^8))$. The only elements of the algebraic tmf resolution which can detect an element which could support a d_2–differential killing
$$\Delta^4 h_{2,1}^8 g(h_{2,1}\zeta_2^4)[18] + h_{2,1}^{24} g(h_{2,1} v_0^{-2} v_2^2 \zeta_1^{16})$$
in the MASS for $T^1 \wedge M(8, v_1^8)$ are the elements

(5.8) $$\Delta^2 v_1^6 h_{2,1}^{23}[0] \quad \text{and} \quad \Delta^2 v_1^3 h_{2,1}^{24}[1].$$

However, using the map of spectral sequences

$$ {}^{\mathrm{mass}}E_2^{*,*}(T^1 \wedge M(8, v_1^8)) \to {}^{\mathrm{mass}}E_2^{*,*}(\mathrm{tmf} \wedge M(8, v_1^8)), $$

we can eliminate these possibilities on the basis that the elements (5.8) support nontrivial d_2–differentials in the MASS for $M(8, v_1^8)$.

We are left with eliminating

$$ v_1^2 h_{2,1}^{31}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1] $$

as possibly detecting $d_5^{\mathrm{mass}}(v_2^{32})$ in the MASS for $M(8, v_1^8)$. This is the trickiest obstruction to eliminate. In the MASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}} \wedge M(8, v_1^8)$, there is a differential

$$ d_2^{\mathrm{mass}}(\Delta^2 v_1 h_{2,1}^{22}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]) = v_1^2 h_{2,1}^{31}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]. $$

The *problem* is that, in the WSS for $H(8, v_1^8)$, there is a nontrivial differential

$$ d_0^{\mathrm{wss}}(\Delta^2 v_1 h_{2,1}^{22}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]) = \Delta^2 v_1 h_{2,1}^{22}(v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[1]. $$

Sublemma 5.9 *The element v_2^{32} is a permanent cycle in the MASS for $T^1 \wedge M(8, v_1^8)$.*

$\underline{\mathrm{bo}}_1$	$h_{2,1}^{31} g(h_{2,1}\zeta_2^4)[0]$ $h_{2,1}^{18} \Delta^2 g(h_{2,1}\zeta_2^4)[17]$
$\underline{\mathrm{bo}}_2$	$h_{2,1}^{5} \Delta^4 g(h_{2,1}\zeta_1^{16})[18]$ $v_1^2 h_{2,1}^{31}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]$
$\underline{\mathrm{bo}}_1^{\otimes 2}$	$h_{2,1}^{5} \Delta^4 v_1 g(v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[18]$ $h_{2,1}^{15} \Delta^2 g(v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4])[18]$
$\underline{\mathrm{bo}}_1 \otimes \underline{\mathrm{bo}}_2$	$v_1 h_{2,1}^{21} g(v_0^{-1} v_2^2 [\zeta_2^4, \zeta_2^8])[18]$ $v_1 h_{2,1}^{21} g(v_0^{-1} v_2^2 [\zeta_2^8, \zeta_2^4])[18]$
$\underline{\mathrm{bo}}_1^3$	$v_1^4 \Delta^6 h_1(v_2^2 \zeta_1^8 \mid \zeta_1^8 \mid \zeta_2^4)[1]$
$\underline{\mathrm{bo}}_1^{\otimes 2} \otimes \underline{\mathrm{bo}}_2$	$v_1 h_{2,1}^{18} g(v_0^{-2} v_2^4 [\zeta_1^8, \zeta_2^4] \mid \zeta_1^8 \zeta_2^4)[18]$ $v_1 h_{2,1}^{18} g(v_0^{-2} v_2^4 (\zeta_1^8 \mid \zeta_1^8 \zeta_2^4 \mid \zeta_2^4 + \zeta_2^4 \mid \zeta_1^8 \zeta_2^4 \mid \zeta_1^8))[18]$ $v_1 h_{2,1}^{18} g(v_0^{-2} v_2^4 \zeta_1^8 \zeta_2^4 \mid [\zeta_1^8, \zeta_2^4])[18]$
$\underline{\mathrm{bo}}_1^4$	$v_1^4 \Delta^6 h_1^2 \zeta_1^8 \mid \zeta_1^8 \mid \zeta_2^4 \mid \zeta_2^4[1]$

Table 1: List of potential targets of $d_r^{\mathrm{mass}}(v_2^{32})$ for $r \geq 4$.

Proof The elements of the algebraic tmf resolution which could possibly detect the target of a differential

$$d_r^{\mathrm{mass}}(v_2^{32}), \quad r \geq 4,$$

in the MASS for $T^1 \wedge M(8, v_1^8)$ consist of those terms in Table 1 coming from $\underline{\mathrm{bo}}_1$ and $\underline{\mathrm{bo}}_2$.

Using (5.3), there is a map

$$\Sigma^{31}\mathrm{tmf} \wedge M(8, v_1^8) \to \Sigma^{-1}\mathrm{tmf} \wedge \overline{\mathrm{tmf}} \to T^1$$

and we therefore have a differential

$$d_2^{\mathrm{mass}}(\Delta^2 v_1 h_{2,1}^{22}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]) = v_1^2 h_{2,1}^{31}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]$$

in the MASS for $T^1 \wedge M(8, v_1^8)$. Therefore, $v_1^2 h_{2,1}^{31}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]$ cannot be the target of a differential $d_5^{\mathrm{mass}}(v_2^{32})$ in the MASS for $T^1 \wedge M(8, v_1^8)$.

Our previous arguments eliminate all the other possibilities. □

Suppose now for the purpose of generating a contradiction that the differential

$$d_5^{\mathrm{mass}}(v_2^{32})$$

in the MASS for $M(8, v_1^8)$ is nontrivial and detected by $v_1^2 h_{2,1}^{31}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]$ in the algebraic tmf resolution for $H(8, v_1^8)$. Consider the fiber sequence

$$\Sigma^{-2}\overline{\mathrm{tmf}}^2 \wedge M(8, v_1^8) \to M(8, v_1^8) \to T^1 \wedge M(8, v_1^8) \xrightarrow{\partial} \Sigma^{-1}\overline{\mathrm{tmf}}^2.$$

We have proven that v_2^{32} exists in $\pi_{192} T^1 \wedge M(8, v_1^8)$, and, because our assumption implies that v_2^{32} does not lift to $\pi_{192} M(8, v_1^8)$, we must have

$$0 \neq \partial(v_2^{32}) \in \pi_{191} \Sigma^{-2}\overline{\mathrm{tmf}}^2 \wedge M(8, v_1^8).$$

Sublemma 5.10 *There exists a choice of $v_2^{32} \in \pi_{192} T^1 \wedge M(8, v_1^8)$ such that $\partial(v_2^{32})$ has modified Adams filtration* 34.

Proof Let $X^{\langle k \rangle}$ denote the k^{th} modified Adams cover of X — so that the MASS for $X^{\langle k \rangle}$ is the truncation of the MASS for X obtained by only considering terms in ${}^{\mathrm{mass}}E_2^{s,t}(X)$ for $s \geq k$ — and let $X_{\langle k \rangle}$ denote the cofiber

$$X^{\langle k+1 \rangle} \to X \to X_{\langle k \rangle}$$

Then we have fiber sequences

$$M(8, v_1^8)_{\langle k\rangle} \to (T^1 \wedge M(8, v_1^8))_{\langle k\rangle} \to (\Sigma^{-1}\overline{\mathrm{tmf}}^2 \wedge M(8, v_1^8))_{\langle k-2\rangle}.$$

Define $\widetilde{M}_{\langle k\rangle}$ to be the homotopy pullback

$$\begin{tikzcd} \widetilde{M}_{\langle k\rangle} \arrow[r] \arrow[d] & T^1 \wedge M(8, v_1^8) \arrow[d] \\ M(8, v_1^8)_{\langle k\rangle} \arrow[r] & (T^1 \wedge M(8, v_1^8))_{\langle k\rangle} \end{tikzcd}$$

Then the algebraic tmf resolution for $\widetilde{M}_{\langle k\rangle}$ is the truncation of the algebraic tmf resolution for $M(8, v_1^8)$ obtained by omitting, for $n \geq 2$, all terms of

$$\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_{i_1} \otimes \cdots \otimes \underline{\mathrm{bo}}_{i_n} \otimes H(8, v_1^8))$$

of cohomological degree greater than $k-n$. It follows from the map of algebraic tmf resolutions and MASSs associated to the map

$$M(8, v_1^8) \to \widetilde{M}_{\langle k\rangle}$$

that there is a differential

$$d_5^{\mathrm{mass}}(v_2^{32}) = v_1^2 h_{2,1}^{31}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]$$

in the MASS for $\widetilde{M}_{\langle k\rangle}$. This differential is nontrivial in the MASS for $\widetilde{M}_{\langle 36\rangle}$, because it is nontrivial in the MASS for $M(8, v_1^8)$, and any intervening differentials killing the target in the algebraic tmf resolution or MASS for $\widetilde{M}_{\langle 36\rangle}$ would lift to $M(8, v_1^8)$ because the spectral sequences are isomorphic in the relevant range. The same is not true in the case of $\widetilde{M}_{\langle 35\rangle}$, where

$$d_0^{\mathrm{wss}}(\Delta^2 v_1 h_{2,1}^{22}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]) = 0$$

and therefore $\Delta^2 v_1 h_{2,1}^{22}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]$ persists to the E_2–term of the MASS

$$d_2^{\mathrm{mass}}(\Delta^2 v_1 h_{2,1}^{22}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]) = v_1^2 h_{2,1}^{31}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1].$$

Therefore, the proof of Sublemma 5.9 goes through with $T^1 \wedge M(8, v_1^8)$ replaced by $\widetilde{M}_{\langle 35\rangle}$ to show that there exists an element

$$\widetilde{v_2^{32}} \in \pi_{192} \widetilde{M}_{\langle 35\rangle}$$

which is detected by v_2^{32} in the MASS. Consider the diagram

$$\begin{tikzcd} & & \Sigma^{-1}\overline{\mathrm{tmf}}^2 \arrow[d] \\ \widetilde{M}_{\langle 36\rangle} \arrow[r] \arrow[d] & T^1 \wedge M(8,v_1^8) \arrow[r, "\partial'"'] \arrow[ur, "\partial"] \arrow[d, equal] & (\Sigma^{-1}\overline{\mathrm{tmf}}^2 \wedge M(8,v_1^8))_{\langle 34\rangle} \arrow[d] \\ \widetilde{M}_{\langle 35\rangle} \arrow[r] & T^1 \wedge M(8,v_1^8) \arrow[r, "\partial''"'] & (\Sigma^{-1}\overline{\mathrm{tmf}}^2 \wedge M(8,v_1^8))_{\langle 33\rangle} \end{tikzcd}$$

where the rows are cofiber sequences. The element $\widetilde{v_2^{32}} \in \pi_{192}\widetilde{M}_{\langle 35\rangle}$ maps to an element $v_2^{32} \in T^1 \wedge M(8,v_1^8)$ with

$$\partial''(v_2^{32}) = 0.$$

However, since $d_5^{\mathrm{mass}}(v_2^{32})$ is nontrivial in the MASS for $\widetilde{M}_{\langle 36\rangle}$, the element $v_2^{32} \in \pi_{192}T^1 \wedge M(8,v_1^8)$ cannot lift to $\widetilde{M}_{\langle 36\rangle}$, and therefore

$$\partial'(v_2^{32}) \neq 0.$$

It follows that $\partial(v_2^{32})$ has modified Adams filtration 34. □

However, we have:

Sublemma 5.11 *There are no elements of $\pi_{191}\Sigma^{-2}\overline{\mathrm{tmf}}^2 \wedge M(8,v_1^8)$ of modified Adams filtration* 34.

Proof The only possible elements in the algebraic tmf resolution for $\overline{\mathrm{tmf}}^2 \wedge M(8,v_1^8)$ which could contribute to modified Adams filtration 34 in this degree are

(5.12) $$\Delta^2 v_1 h_{2,1}^{22}(v_0^{-1}v_2^2[\zeta_1^8,\zeta_2^4])[1] \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2} \otimes H(8,v_1^8))$$

and the elements of Table 1 of algebraic tmf filtration greater than 1 in the appropriate modified Adams filtration. However, the previous arguments eliminate all of the candidates coming from Table 1, so we are left with eliminating (5.12). We wish to lift the differential

$$d_3^{\mathrm{mass}}(\Delta^6 v_1 h_{2,1}^3(v_0^{-1}v_2^2[\zeta_1^8,\zeta_2^4])[1]) = \Delta^2 v_1 h_{2,1}^{22}(v_0^{-1}v_2^2[\zeta_1^8,\zeta_2^4])[1]$$

in the MASS for $\mathrm{tmf}\wedge\overline{\mathrm{tmf}}^2\wedge M(8,v_1^8)$ to a differential in the MASS for $\overline{\mathrm{tmf}}^2\wedge M(8,v_1^8)$. We therefore must argue that

$$d_2^{\mathrm{mass}}(\Delta^6 v_1 h_{2,1}^3(v_0^{-1}v_2^2[\zeta_1^8,\zeta_2^4])[1]) = 0$$

in the MASS for $\overline{\mathrm{tmf}}^2 \wedge M(8, v_1^8)$. We will therefore argue there are no elements in the algebraic tmf resolution for $\overline{\mathrm{tmf}}^2 \wedge M(8, v_1^8)$ which could detect the target of such a d_2. Ignoring any possibilities which are eliminated by Proposition 4.2, the only possibilities are

$$\begin{aligned} &\Delta^6 v_1^4 h_1 v_0^{-1} v_2^2 \zeta_1^8 \mid [\zeta_1^8, \zeta_2^4][1], && \Delta^6 v_1^4 h_0^2 \zeta_1^8 \mid [\zeta_1^8, \zeta_2^4] \mid \zeta_1^8 \mid \zeta_1^8[0], \\ &\Delta^6 v_1^4 h_1 v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4] \mid \zeta_1^8[1], && \Delta^6 v_1^4 h_0^2 \zeta_1^8 \mid \zeta_1^8 \mid [\zeta_1^8, \zeta_2^4] \mid \zeta_1^8[0], \\ &\Delta^6 v_1^4 h_0^2 [\zeta_1^8, \zeta_2^4] \mid \zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8[0], && \Delta^6 v_1^4 h_0^2 \zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8 \mid [\zeta_1^8, \zeta_2^4][0]. \end{aligned}$$

However, these are killed by the respective WSS differentials

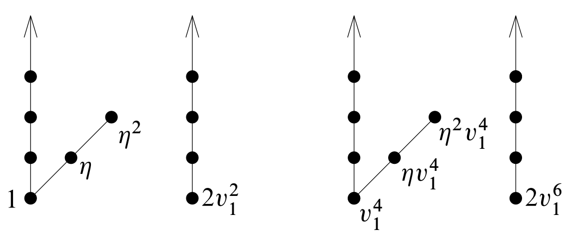
$$\begin{aligned} &d_0^{\mathrm{wss}} \Delta^6 v_1^4 h_1 v_0^{-1} v_2^2 \zeta_1^8 \mid \zeta_1^8 \zeta_2^4[1], && d_0^{\mathrm{wss}} \Delta^6 v_1^4 h_0^2 \zeta_1^8 \mid \zeta_1^8 \zeta_2^4 \mid \zeta_1^8 \mid \zeta_1^8[0], \\ &d_0^{\mathrm{wss}} \Delta^6 v_1^4 h_1 v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4 \mid \zeta_1^8[1], && d_0^{\mathrm{wss}} \Delta^6 v_1^4 h_0^2 \zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8 \zeta_2^4 \mid \zeta_1^8[0], \\ &d_0^{\mathrm{wss}} \Delta^6 v_1^4 h_0^2 \zeta_1^8 \zeta_2^4 \mid \zeta_1^8 \mid \zeta_1^8 \mid |\zeta_1^8[0], && d_0^{\mathrm{wss}} \Delta^6 v_1^4 h_0^2 \zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8 \mid \zeta_1^8 \zeta_2^4[0]. \end{aligned} \qquad \square$$

Thus we have arrived at a contradiction, as we have produced an element of modified Adams filtration 34, and subsequently showed no such elements exist. We conclude that our supposition, that the differential $d_5^{\mathrm{mass}}(v_2^{32})$ in the MASS for $M(8, v_1^8)$ is nontrivial and detected by $v_1^2 h_{2,1}^{31}(v_0^{-1} v_2^2 \zeta_1^8 \zeta_2^4)[1]$ in the algebraic tmf resolution, is false. $\square$

6 Determination of elements not in the tmf Hurewicz image

Theorem 6.1 *The elements of* tmf_* *not in the subgroup described in Theorem 1.2 are not in the Hurewicz image.*

We first recall some well-known K–theory computations. Recall that $\pi_*\mathrm{KO}$ is given by the v_1^4–periodic pattern



Let

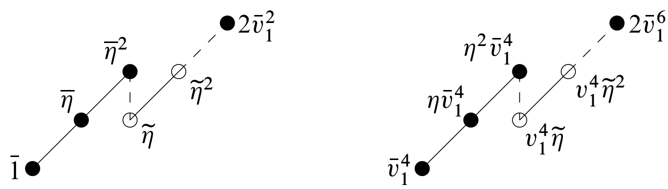
$$M(2^\infty) := \varinjlim_i M(2^i)$$

denote the Moore spectrum for $\mathbb{Z}/2^\infty$.

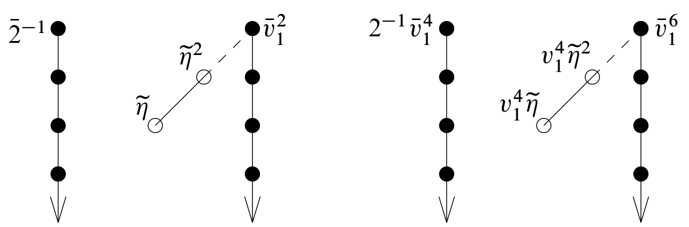
Consider the diagram of cofiber sequences

(6.2)
$$\begin{tikzcd}
\Sigma^{-1}\mathrm{KO}\wedge M(2) \arrow[r, "p"] \arrow[d, "\cdot 2^{-1}"'] & \mathrm{KO} \arrow[r, "\cdot 2"] \arrow[d, equal] & \mathrm{KO} \arrow[r, "\overline{(\cdot)}"] \arrow[d, "\cdot 2^{-1}"] & \mathrm{KO}\wedge M(2) \arrow[d, "\cdot 2^{-1}"] \\
\Sigma^{-1}\mathrm{KO}\wedge M(2^\infty) \arrow[r, "p"'] & \mathrm{KO} \arrow[r] & \mathrm{KO}_{\mathbb{Q}} \arrow[r, "\overline{(\cdot)}"'] & \mathrm{KO}\wedge M(2^\infty)
\end{tikzcd}$$

The groups $\mathrm{KO}_* M(2)$ are well known to be given by the v_1^4–periodic pattern



where we denote lifts of elements of KO_* along the map p of diagram (6.2) with a tilde, and the images of the map $\overline{(\cdot)}$ with a bar. It then follows easily from the map of long exact sequences coming from the above diagram that $\mathrm{KO}_* M(2^\infty)$ is given by the v_1^4–periodic pattern



where again we denote lifts over the map p with a tilde, and images under the map $\overline{(\cdot)}$ with a bar. The infinite sequences of dots going down represent the elements 2^{-i} in $\mathbb{Z}/2^\infty = \mathbb{Q}/\mathbb{Z}_{(2)}$.

Proof of Theorem 6.1 Recall [27, Corollary 3] that we have an equivalence

$$c_4^{-1}\mathrm{tmf} \simeq \mathrm{KO}[j^{-1}],$$

where $j^{-1} = \Delta/c_4^3$. Applying π_0 to this equivalence, we have a commutative diagram

$$\begin{tikzcd}
S \arrow[r] \arrow[d] & \mathrm{KO} \arrow[dr, hook] & \\
\mathrm{tmf} \arrow[r] & c_4^{-1}\mathrm{tmf} \arrow[r, "\simeq"'] & \mathrm{KO}[j^{-1}]
\end{tikzcd}$$

Consider the diagram

$$\begin{tikzcd}
\pi_* S \arrow[d, "h"'] & \pi_{*+1} M(2^\infty) \arrow[l, "p"'] \arrow[r] \arrow[d, "h"'] & \mathrm{KO}_{*+1} M(2^\infty) \arrow[dr, hook, "i"] & \\
\mathrm{tmf}_* \arrow[d, "L"'] & \mathrm{tmf}_{*+1} M(2^\infty) \arrow[l] \arrow[r, "L"] & c_4^{-1}\mathrm{tmf}_{*+1} M(2^\infty) \arrow[r, equal] \arrow[dll, "p'"] & \mathrm{KO}_{*+1} M(2^\infty)[j^{-1}] \\
c_4^{-1}\mathrm{tmf}_* & & &
\end{tikzcd}$$

Suppose that $x \in \mathrm{tmf}_{>0}$ has nontrivial image in $L(x) \in c_4^{-1}\mathrm{tmf}_*$ and that $x = h(y)$. Since y is torsion, it lifts over p to an element

$$\tilde{y} \in \pi_{*+1} M(2^\infty).$$

The commutativity of the diagram implies that

$$0 \neq L(x) \in \mathrm{Im}(p' \circ i)$$

and this implies that

$$L(x) \in \bigl\{c_4^k \eta^l : k \geq 0,\, l \in \{1, 2\}\bigr\}.$$

Now consider elements of the form

$$x = \alpha \Delta^k \nu \in \mathrm{tmf}_*$$

with $\alpha \not\equiv 0 \bmod 8$. Suppose that $x = h(y)$. Lift y to an element

$$\tilde{y} \in \pi_{*+1} M(2^\infty).$$

Then we have

$$Lh(\tilde{y}) = \tfrac{1}{8}\,\overline{\alpha \Delta^k v_1^2} = \tfrac{1}{4}\,\overline{\alpha v_1^{12k+2}}\, j^{-k} \neq 0.$$

But the commutativity of the diagram implies that $Lh(\tilde{y})$ is in the image of i, which implies that $k = 0$. □

7 Lifting the remaining elements of tmf_* to π_*^s

Multiplicative generators of the Hurewicz image below the 192–stem

In this section, we determine a set of elements which multiplicatively generate the tmf Hurewicz image below the 192–stem. The results in this subsection drastically reduce the number of classes which we must lift in the sequel.

Lemma 7.1 *The Hurewicz map $S \to \mathrm{tmf}$ is a map of ring spectra. In particular, it preserves multiplication.*

Corollary 7.2 *Suppose $\alpha = \beta\gamma$ is a product of elements $\beta, \gamma \in \pi_*(\mathrm{tmf})$ with lifts $\widetilde{\beta}, \widetilde{\gamma} \in \pi_*(S)$. Then $\widetilde{\beta}\widetilde{\gamma} \in \pi_*(S)$ must be a lift of α.*

With this in mind, it suffices to find a subset of the Hurewicz image which generates the entire Hurewicz image up to the 192–stem under products. Our desired generating subset is given in Corollary 7.16. We will obtain our generating set by listing generators in lemmas and then recording their products in corollaries, until we have exhausted the tmf Hurewicz image up to stem 192.

Lemma 7.3 *The classes $\eta \in \pi_1(\mathrm{tmf})$, $\nu \in \pi_3(\mathrm{tmf})$, $\epsilon \in \pi_8(\mathrm{tmf})$, $\kappa \in \pi_{14}(\mathrm{tmf})$, $\overline{\kappa} \in \pi_{20}(\mathrm{tmf})$, $u \in \pi_{39}(\mathrm{tmf})$ and $w \in \pi_{45}(\mathrm{tmf})$ are in the Hurewicz image.*

Proof The elements η, ν, ϵ, κ, $\overline{\kappa}$, u and w are all well-known elements of π_*^s, detected in the Adams spectral sequence by h_1, h_2, c_0, d_0, g, $\Delta h_1 d_0$ and $\Delta h_1 g$ [22, Table 8]. These elements have nontrivial images under the map of Adams spectral sequences induced by the unit map $S \to \mathrm{tmf}$. The lemma is therefore somewhat tautological, as the corresponding elements in tmf were *defined* in Section 1 to be the Hurewicz images of these elements. □

Lemma 7.4 *The class $q \in \pi_{32}(\mathrm{tmf})$ is in the Hurewicz image.*

Proof See the proof of Lemma 7.18(1). □

Corollary 7.5 *The classes $\eta^2 \in \pi_2(\mathrm{tmf})$, $\nu^2 \in \pi_6(\mathrm{tmf})$, $\nu^3 = \epsilon\eta \in \pi_9(\mathrm{tmf})$, $\kappa\eta \in \pi_{15}(\mathrm{tmf})$, $\kappa\nu \in \pi_{17}(\mathrm{tmf})$, $\overline{\kappa}\eta \in \pi_{21}(\mathrm{tmf})$, $\overline{\kappa}\eta^2 = \kappa\epsilon \in \pi_{22}(\mathrm{tmf})$, $\overline{\kappa}\epsilon = \kappa^2 \in \pi_{28}(\mathrm{tmf})$, $q\eta \in \pi_{33}(\mathrm{tmf})$, $\overline{\kappa}\kappa \in \pi_{34}(\mathrm{tmf})$, $\overline{\kappa}\kappa\eta \in \pi_{35}(\mathrm{tmf})$, $\overline{\kappa}^2 \in \pi_{40}(\mathrm{tmf})$, $\overline{\kappa}^2\eta \in \pi_{41}(\mathrm{tmf})$, $\overline{\kappa}^2\eta^2 = \kappa^3 \in \pi_{42}(\mathrm{tmf})$, $w\eta \in \pi_{46}(\mathrm{tmf})$, $\overline{\kappa}q \in \pi_{52}(\mathrm{tmf})$, $\overline{\kappa}q\eta \in \pi_{53}(\mathrm{tmf})$, $\overline{\kappa}^2\kappa \in \pi_{54}(\mathrm{tmf})$, $\overline{\kappa}u \in \pi_{59}(\mathrm{tmf})$, $\overline{\kappa}^3 \in \pi_{60}(\mathrm{tmf})$, $\overline{\kappa}w \in \pi_{65}(\mathrm{tmf})$, $\overline{\kappa}w\eta \in \pi_{66}(\mathrm{tmf})$, $\overline{\kappa}^4 \in \pi_{80}(\mathrm{tmf})$, $\overline{\kappa}^2 w \in \pi_{85}(\mathrm{tmf})$, $w^2 \in \pi_{90}(\mathrm{tmf})$, $\overline{\kappa}^5 \in \pi_{100}(\mathrm{tmf})$, $\overline{\kappa}^3 w \in \pi_{105}(\mathrm{tmf})$, $\overline{\kappa}w^2 \in \pi_{110}(\mathrm{tmf})$, $\overline{\kappa}^4 w \in \pi_{125}(\mathrm{tmf})$ and $\overline{\kappa}^2 w^2 \in \pi_{130}(\mathrm{tmf})$ are in the Hurewicz image.*

Lemma 7.6 *The classes $\{\nu\Delta^2\}\nu \in \pi_{54}(\mathrm{tmf})$, $\{\nu\Delta^2\}\kappa \in \pi_{65}(\mathrm{tmf})$ and $\{\eta^2\Delta^2\}\overline{\kappa} \in \pi_{70}(\mathrm{tmf})$ are in the Hurewicz image.*

Proof See Lemma 7.21. □

Corollary 7.7 *The classes $\{\nu\Delta^2\}\nu^2 \in \pi_{57}(\mathrm{tmf})$ and $\{\nu\Delta^2\}\kappa\nu \in \pi_{68}(\mathrm{tmf})$ are in the Hurewicz image.*

Lemma 7.8 *The classes* $\{\nu\Delta^4\}\nu \in \pi_{102}(\mathrm{tmf})$, $\{\epsilon\Delta^4\} \in \pi_{104}(\mathrm{tmf})$, $\{\kappa\Delta^4\} \in \pi_{110}(\mathrm{tmf})$, $2\Delta^4\bar{\kappa} \in \pi_{116}(\mathrm{tmf})$ *and* $\{\eta\Delta^4\}\bar{\kappa} \in \pi_{117}(\mathrm{tmf})$ *are in the Hurewicz image.*

Proof See Lemmas 7.22 and 7.23. □

Corollary 7.9 *The classes* $\{\epsilon\Delta^4\}\eta \in \pi_{105}(\mathrm{tmf})$, $\{\kappa\Delta^4\}\eta \in \pi_{111}(\mathrm{tmf})$, $\{\kappa\Delta^4\}\nu \in \pi_{113}(\mathrm{tmf})$, $\{\kappa\Delta^4\}\nu^2 \in \pi_{116}(\mathrm{tmf})$, $\{\eta\Delta^4\}\bar{\kappa}\eta \in \pi_{118}(\mathrm{tmf})$, $\{\kappa\Delta^4\}\kappa \in \pi_{124}(\mathrm{tmf})$, $\{\kappa\Delta^4\}\bar{\kappa} \in \pi_{130}(\mathrm{tmf})$, $\{\kappa\Delta^4\}\bar{\kappa}\eta \in \pi_{131}(\mathrm{tmf})$, $\{\eta\Delta^4\}\bar{\kappa}^2 \in \pi_{137}(\mathrm{tmf})$ *and* $\{\eta\Delta^4\}\bar{\kappa}^2\eta \in \pi_{138}(\mathrm{tmf})$ *are in the Hurewicz image.*

Lemma 7.10 *The class* $\{q\Delta^4\} \in \pi_{128}(\mathrm{tmf})$ *is in the Hurewicz image.*

Proof See Lemma 7.24. □

Corollary 7.11 *The classes* $\{q\Delta^4\}\eta \in \pi_{129}(\mathrm{tmf})$, $\{q\Delta^4\}\kappa = w\eta\Delta^4 \in \pi_{142}(\mathrm{tmf})$, $\{q\Delta^4\}\bar{\kappa} \in \pi_{148}(\mathrm{tmf})$, $\{q\Delta^4\}\bar{\kappa}\eta \in \pi_{149}(\mathrm{tmf})$ *and* $\{q\Delta^4\}\bar{\kappa}\eta^2 \in \pi_{150}(\mathrm{tmf})$ *are in the Hurewicz image.*

Lemma 7.12 *The class* $\Delta^4 u \in \pi_{135}(\mathrm{tmf})$ *is in the Hurewicz image.*

Proof See Lemma 7.25. □

Corollary 7.13 *The classes* $\Delta^4 u\eta \in \pi_{136}(\mathrm{tmf})$ *and* $\Delta^4 u\bar{\kappa} \in \pi_{155}(\mathrm{tmf})$ *are in the Hurewicz image.*

Lemma 7.14 *The classes* $\{\nu\Delta^6\}\nu \in \pi_{150}(\mathrm{tmf})$ *and* $\{\nu\Delta^6\}\kappa \in \pi_{161}(\mathrm{tmf})$ *are in the Hurewicz image.*

Proof See Lemma 7.26. □

Corollary 7.15 *The classes* $\{\nu\Delta^6\}\nu^2 \in \pi_{153}$, $\{\nu\Delta^6\}\nu^3 \in \pi_{156}$, $\{\nu\Delta^6\}\kappa\eta \in \pi_{162}(\mathrm{tmf})$ *and* $\{\nu\Delta^6\}\kappa\nu \in \pi_{164}(\mathrm{tmf})$ *are in the Hurewicz image.*

Thus our calculation of the Hurewicz image up to dimension 192 has been reduced to showing that the following list of elements is in the Hurewicz image:

Corollary 7.16 *Up to dimension* 192, *the Hurewicz image is generated under multiplication by*

$$\big\{\eta, \nu, \epsilon, \kappa, \bar{\kappa}, q, u, w, \{\nu\Delta^2\}\nu, \{\nu\Delta^2\}\kappa, \{\eta^2\Delta^2\}\bar{\kappa}, \{\nu\Delta^4\}\nu, \{\epsilon\Delta^4\}, \{\kappa\Delta^4\}, 2\Delta^4\bar{\kappa}, \{\eta\Delta^4\}\bar{\kappa}, \{q\Delta^4\}, \Delta^4 u, \{\nu\Delta^6\}\nu, \{\nu\Delta^6\}\kappa\big\}.$$

Lifting generators

We will now describe our method for lifting generators. Given an element $x \in \mathrm{tmf}_*$, we want to lift it to an element $y \in \pi_*^s$. To this end, we consider the diagram of (M)ASSs

$$\begin{tikzcd}
& \mathrm{Ext}_{A(2)_*}(H(8,v_1^8)) \arrow[rr, Rightarrow] \arrow[dd] & & \mathrm{tmf}_{*+18}M(8,v_1^8) \arrow[dd] \\
\mathrm{Ext}_{A_*}(H(8,v_1^8)) \arrow[ur] \arrow[rr, Rightarrow] \arrow[dd] & & \pi_{*+18}M(8,v_1^8) \arrow[ur] \arrow[dd] & \\
& \mathrm{Ext}_{A(2)_*}(\mathbb{F}_2) \arrow[rr, Rightarrow] & & \mathrm{tmf}_* \\
\mathrm{Ext}_{A_*}(\mathbb{F}_2) \arrow[ur] \arrow[rr, Rightarrow] & & \pi_*^s \arrow[ur] &
\end{tikzcd}$$

First, we identify an element

$$x' \in \mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$$

which detects the element x in the ASS for tmf_*, and then we identify an element

$$\tilde{x}' \in \mathrm{Ext}_{A(2)_*}(H(8,v_1^8))$$

which maps to it. This element $\tilde{x}'$ can be regarded as an element of the 0–line of the algebraic tmf resolution for $\mathrm{Ext}_{A_*}(H(8,v_1^8))$. We will show that the element $\tilde{x}'$ is a permanent cycle in the algebraic tmf resolution, and thus lifts to an element

$$\tilde{y}' \in \mathrm{Ext}_{A_*}(H(8,v_1^8)).$$

We will then show that the element $\tilde{y}'$ is a permanent cycle in the MASS for $M(8,v_1^8)$, and hence detects an element

$$\tilde{y} \in \pi_* M(8,v_1^8).$$

Let $y \in \pi_*^s$ be the projection of $\tilde{y}$ to the top cell. It then follows that the image of y in tmf_* equals x, modulo terms of higher Adams filtration (AF). Furthermore, using the v_2^{32}–self-map on $M(8,v_1^8)$, we deduce that the element

$$v_2^{32k}\tilde{y} \in \pi_* M(8,v_1^8)$$

projects on the top cell to an element $v_2^{32k}y \in \pi_*^s$ whose image in tmf_* is $\Delta^{8k}x$ modulo terms of higher Adams filtration. Finally, Theorem 6.1 eliminates the potential ambiguity caused by elements of higher Adams filtration, since the elements of higher Adams filtration are v_1^4–periodic.

We will show all of the generators of Corollary 7.16 except for η, ν and ϵ actually come from the top cell of $M(8, v_1^8)$, and thus v_2^{32}–periodicity extends our work below dimension 192 to all dimensions. It turns out that ν^2 and ϵ do not come from the top cell of $M(8, v_1^8)$. In order to show that the elements

$$\Delta^{8k}\nu^2, \Delta^{8k}\epsilon \in \pi_*\mathrm{tmf}$$

are in the Hurewicz image for $k > 0$, we will instead show that $\Delta^8\nu^2$ and $\Delta^8\epsilon$ come from the top cell of $M(8, v_1^8)$ (Lemma 7.27).

Lemma 7.17 *The following classes lift to the top cell of* $M(8, v_1^8)$:

(1) $\kappa \in \pi_{14}(\mathrm{tmf})$.

(2) $\bar{\kappa} \in \pi_{20}(\mathrm{tmf})$.

Proof We will check that each element lifts using the AHSS:

(1) Since κ is 2–torsion (and thus 8–torsion), it lifts to $\kappa[1] \in \pi_{15}(M(8))$. Inspection of [25, page 3] in stems 31 and 32 and AF $\geq$ 12 reveals that there are no classes which could detect $v_1^8\kappa[1]$. Therefore $\kappa[1]$ lifts to $\kappa[18] \in \pi_{32}(M(8, v_1^8))$.

(2) Since $\bar{\kappa}$ is 8–torsion, it lifts to $\bar{\kappa}[1] \in \pi_{21}(M(8))$. Inspection of [25, page 3] in stems 36 and 37 and AF $\geq$ 12 reveals that there are no classes which could detect $v_1^8\bar{\kappa}[1]$. Therefore $\bar{\kappa}[1]$ lifts to $\bar{\kappa}[18] \in \pi_{38}(M(8, v_1^8))$. □

Lemma 7.18 *The following classes lift to the top cell of* $M(8, v_1^8)$:

(1) $q \in \pi_{32}(\mathrm{tmf})$.

(2) $u \in \pi_{39}(\mathrm{tmf})$.

(3) $w \in \pi_{45}(\mathrm{tmf})$.

Proof We will check that each element lifts using the Atiyah–Hirzebruch spectral sequence (AHSS).

(1) We begin with $q \in \pi_{32}(\mathrm{tmf})$, which we will define to be the unique nontrivial c_4–torsion class detected by the element

$$v_2^4 c_0 \in \mathrm{Ext}^{7,7+32}_{A(2)_*}(\mathbb{F}_2)$$

in the ASS for tmf. The element $v_2^4c_0$ does *not* lift to Ext_{A_*}. Nevertheless, we claim that there is an element[9] $\tilde{q} \in \pi_{32}^s$ detected by the element

$$\Delta h_1 h_3 \in \mathrm{Ext}_{A_*}^{6,6+32}(\mathbb{F}_2)$$

in the ASS for the sphere, which maps to q under the tmf Hurewicz homomorphism. Our strategy will be to argue that $\tilde{q}$ and q lift to

$$\tilde{q}[18] \in \pi_{50} M(8, v_1^8) \quad \text{and} \quad q[18] \in \mathrm{tmf}_{50} M(8, v_1^8),$$

respectively, and that the element which detects $\tilde{q}[18]$ in the MASS for $M(8, v_1^8)$ maps to the element which detects $q[18]$ in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$ under the map

$$\mathrm{Ext}_{A_*}(H(8, v_1^8)) \to \mathrm{Ext}_{A(2)_*}(H(8, v_1^8)). \tag{7.19}$$

Inspection of [25, page 3] in stem 32 and AF ≥ 7 reveals that $\tilde{q}$ is 2–torsion (and thus 8–torsion), so $\tilde{q}$ lifts to $\tilde{q}[1] \in \pi_{33}(M(8))$. Inspection of [25, page 3] in stems 48 and 49 and AF ≥ 14 reveals that there are no classes which could detect $v_1^8\tilde{q}[1]$. Therefore $\tilde{q}[1]$ lifts to $\tilde{q}[18] \in \pi_{50}(M(8, v_1^8))$. A similar but easier analysis reveals that the lift $q[18]$ exists.

The elements $\Delta h_1 h_3 \in \mathrm{Ext}_{A_*}(\mathbb{F}_2)$ and $v_2^4 c_0 \in \mathrm{Ext}_{A(2)_*}(\mathbb{F}_2)$ are h_0–torsion, and hence lift to elements

$$\Delta h_1 h_3[1] \in \mathrm{Ext}_{A_*}(H(8)), \quad v_2^4 c_0[1] \in \mathrm{Ext}_{A(2)_*}(H(8))$$

which detect $\tilde{q}[1] \in \pi_{33} M(8)$ and $q[1] \in \mathrm{tmf}_{33} M(8)$, respectively, in the MASS. To identify the elements which detect $\tilde{q}[18]$ and $q[18]$ in the MASS, we make use of the geometric boundary theorem [5, Appendix A].[10] The differentials

$$d_3(v_1^2 h_{2,1} g^2[1]) = v_1^8 \Delta h_3 h_1[1], \quad d_4(v_1^2 h_{2,1} g^2[1]) = v_1^8 v_2^4 c_0[1]$$

in the MASSs for $M(8)$ and $\mathrm{tmf} \wedge M(8)$, respectively, imply that $\tilde{q}[18] \in \pi_{50} M(8, v_1^8)$ and $q[18] \in \mathrm{tmf}_{50} M(8, v_1^8)$ are detected by

$$v_1^2 h_{2,1} g^2[1] \in \mathrm{Ext}_{A_*}(H(8, v_1^8)), \quad v_1^2 h_{2,1} g^2[1] \in \mathrm{Ext}_{A(2)_*}(H(8, v_1^8))$$

in the MASSs for $M(8, v_1^8)$ and $\mathrm{tmf} \wedge M(8, v_1^8)$, respectively, and the former maps to the latter under the map (7.19).

[9]The element we are calling $\tilde{q} \in \pi_{32}^s$ is traditionally called q, but we add the tilde to distinguish it from the element we are calling q in $\pi_{32}\mathrm{tmf}$.

[10]We are specifically using case (5) of the geometric boundary theorem since the relevant class (denoted by $p_*(y)$ in the theorem statement) is a permanent cycle. We will be using this argument repeatedly in subsequent proofs in this section, and for brevity will simply say "by the geometric boundary theorem …" in these subsequent instances.

(2) Since $u \in \pi_{39}\mathrm{tmf}$ is detected by an element of $\mathrm{Ext}_{A(2)_*}$ in the image of the map

$$\mathrm{Ext}_{A_*}(\mathbb{F}_2) \to \mathrm{Ext}_{A(2)_*}(\mathbb{F}_2), \tag{7.20}$$

we immediately see that the element $u \in \pi_{39}(S)$ maps to it. We are left with lifting $u \in \pi^s_{39}$ to the top cell of $M(8, v_1^8)$. Inspection of [25, page 3] in stem 39 and $\mathrm{AF} \geq 10$ reveals that u is 2–torsion (and thus 8–torsion), so u lifts to $u[1] \in \pi_{40}(M(8))$. Inspection of [25, page 3] in stems 55 and 56 and $\mathrm{AF} \geq 17$ reveals that there are no classes which could detect $v_1^8 u[1]$. Therefore $u[1]$ lifts to $u[18] \in \pi_{57}(M(8, v_1^8))$.

(3) The element $w \in \pi_{45}\mathrm{tmf}$ is detected by an element which is in the image of the map (7.20), and thus we deduce that $w \in \pi_{45}(S)$ maps to it. A similar argument to the case above shows that w lifts to $w[18] \in \pi_{63}(M(8, v_1^8))$. □

Lemma 7.21 *The following classes lift to the top cell of $M(8, v_1^8)$:*

1. $\Delta^2 \nu^2 \in \pi_{54}(\mathrm{tmf})$.
2. $\Delta^2 \kappa \nu \in \pi_{65}(\mathrm{tmf})$.
3. $\Delta^2 \eta^2 \overline{\kappa} \in \pi_{70}(\mathrm{tmf})$.

Proof We follow the proof of [7, Theorem 11.1] (which builds on [7, Example 9.5 and Proposition 10.1]).

(1) We begin with $\Delta^2 \nu^2 \in \pi_{54}(\mathrm{tmf})$. This class lifts to an element

$$\Delta^2 \nu^2[1] \in \mathrm{tmf}_{55}(M(8))$$

which is detected by

$$v_2^8 h_2^2[1] \in \mathrm{Ext}^{12,55+12}_{A(2)_*}(H(8))$$

in the MASS for $\mathrm{tmf} \wedge M(8)$. Let

$$\Delta^2 \nu^2[18] \in \mathrm{tmf}_{72}(M(8, v_1^8))$$

be a lift of $\Delta^2 \nu^2[1]$. In the MASS for $\mathrm{tmf} \wedge M(8)$, there is a differential

$$d_2(v_2^{10} v_1^4 h_2 h_0[1]) = v_2^8 v_1^8 h_2^2[1].$$

Since $v_2^{10} v_1^4 h_2 h_0[1]$ is a permanent cycle in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$, it follows from the geometric boundary theorem that $\Delta^2 \nu^2[18]$ is detected by $v_2^{10} v_1^4 h_2 h_0[1]$ in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$. In particular, we see that $\Delta^2 \nu^2[18]$ has modified Adams filtration (MAF) 18 and stem 72.

We now check that $v_2^{10}v_1^4h_2h_0[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8,v_1^8)$. Its relative position[11] is $t-s=65$ and $\mathrm{AF}=17$, its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2}\otimes H(8,v_1^8))$ is $t-s=58$ and $\mathrm{AF}=16$, and its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 3}\otimes H(8,v_1^8))$ is $t-s=51$ and $\mathrm{AF}=15$, the last of which lies above the vanishing line. Inspection of the relevant charts shows that $v_2^{10}v_1^4h_2h_0[1]$ cannot support a nontrivial d_1–differential since the target bidegrees are zero. Therefore $v_2^{10}v_1^4h_2h_0[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8,v_1^8)$ and therefore it detects an element $\{v_2^{10}v_1^4h_2h_0[1]\}$ in $\mathrm{Ext}_{A_*}(H(8,v_1^8))$.

Finally, inspection of the same algebraic tmf resolution charts reveals that there are no possible targets for a nontrivial differential supported by $\{v_2^{10}v_1^4h_2h_0[1]\}$ in the MASS for $M(8,v_1^8)$. Therefore $\{v_2^{10}v_1^4h_2h_0[1]\}$ is a permanent cycle which detects a lift of $\Delta^2\nu^2$.

(2) The class $\Delta^2\kappa\nu\in\pi_{65}(\mathrm{tmf})$ lifts to an element

$$\Delta^2\kappa\nu[1]\in\mathrm{tmf}_{66}(M(8))$$

which is detected by

$$v_2^8h_2d_0[1]\in\mathrm{Ext}_{A(2)_*}^{15,66+15}(H(8))$$

in the MASS for $\mathrm{tmf}\wedge M(8)$. Lift $\Delta^2\kappa\nu[1]$ to an element

$$\Delta^2\kappa\nu[18]\in\mathrm{tmf}_{83}(M(8,v_1^8)).$$

In the MASS for $\mathrm{tmf}\wedge M(8)$, there is a differential

$$d_2(v_2^{10}v_1^4d_0h_0[1])=v_2^8v_1^8h_2d_0[1].$$

By the geometric boundary theorem, $v_2^8\kappa\nu[18]$ is detected by $v_2^{10}v_1^4d_0h_0[1]$ in the MASS for $\mathrm{tmf}\wedge M(8,v_1^8)$. In particular, we see that $\Delta^2\kappa\nu[18]$ has MAF 21 and stem 83.

We now check that $v_2^{10}v_1^4d_0h_0[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8,v_1^8)$. Its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1\otimes H(8,v_1^8))$ is $t-s=76$ and $\mathrm{AF}=20$, its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2}\otimes H(8,v_1^8))$ is $t-s=69$ and $\mathrm{AF}=19$, and its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 3}\otimes H(8,v_1^8))$ is $t-s=62$ and $\mathrm{AF}=18$, the last of which has targets only above the vanishing line. Inspection of the relevant charts shows

[11] We will say that $x\in\mathrm{Ext}_{A(2)_*}(H(8,v_1^8))$ has *relative position* $(t-s,s)$ in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_I\otimes H(8,v_1^8))$ if the image of a differential supported by x in the algebraic tmf resolution lies in $\mathrm{Ext}_{A(2)_*}^{s+1,t}(\underline{\mathrm{bo}}_I\otimes H(8,v_1^8))$, and the image of a differential supported by x in the MASS could be detected in the algebraic tmf resolution by an element in $\mathrm{Ext}_{A(2)_*}^{s+r,t-r+1}(\underline{\mathrm{bo}}_1\otimes H(8,v_1^8))$. In other words, if you were to pretend x is an element in $\mathrm{Ext}_{A(2)_*}^{s,t}(\underline{\mathrm{bo}}_I\otimes H(8,v_1^8))$, then d_r–differentials in the algebraic tmf resolution "look" like Adams d_1's, and d_r–differentials in the MASS "look" like Adams d_r's.

that $v_2^{10}v_1^4d_0h_0[1]$ cannot support a nontrivial d_1–differential since the target bidegrees are zero. Therefore $v_2^{10}v_1^4d_0h_0[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$ and detects an element $\{v_2^{10}v_1^4d_0h_0[1]\}$ in $\mathrm{Ext}_{A_*}(H(8, v_1^8))$.

Finally, inspection of the same charts reveals that there are no possible targets for a nontrivial differential supported by $\{v_2^{10}v_1^4d_0h_0[1]\}$ in the MASS for $M(8, v_1^8)$. Therefore $\{v_2^{10}v_1^4d_0h_0[1]\}$ is a permanent cycle.

(3) The class $\Delta^2\eta^2\overline{\kappa} \in \pi_{70}(\mathrm{tmf})$ lifts to an element

$$\Delta^2\eta^2\overline{\kappa}[1] \in \mathrm{tmf}_{71}(M(8))$$

which is detected by

$$g^2h_{2,1}^6[1] \in \mathrm{Ext}_{A(2)_*}^{16,71+16}(H(8))$$

in the MASS for $\mathrm{tmf} \wedge M(8)$. Lift $\Delta^2\eta^2\overline{\kappa}[1]$ to an element

$$\Delta^2\eta^2\overline{\kappa}[18] \in \mathrm{tmf}_{88}(M(8, v_1^8)).$$

In the MASS for $\mathrm{tmf} \wedge M(8)$, there is a differential

$$d_2(v_2^8v_1^4d_0e_0[1]) = g^2v_1^8h_{2,1}^6[1].$$

By the geometric boundary theorem, $\Delta^2\eta^2\overline{\kappa}[18]$ is detected by $v_2^8v_1^4d_0e_0[1]$ in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$. In particular, we see that $\Delta^2\eta^2\overline{\kappa}[18]$ has MAF 24 and stem 88.

We now check that $v_2^8v_1^4d_0e_0[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$. Its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes H(8, v_1^8))$ is $t-s = 81$ and $\mathrm{AF} = 23$ and its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2} \otimes H(8, v_1^8))$ is $t-s = 74$ and $\mathrm{AF} = 22$, the latter of which lies above the vanishing line. Inspection of the relevant charts shows that $v_2^8v_1^4d_0e_0[1]$ cannot support a nontrivial differential in the algebraic tmf resolution for $H(8, v_1^8)$ since the target bidegrees are zero. Therefore $v_2^8v_1^4d_0e_0[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$ and therefore lifts to an element $\{v_2^8v_1^4d_0e_0[1]\}$ in $\mathrm{Ext}_{A_*}(H(8, v_1^8))$.

Finally, inspection of the same charts reveals that there are no possible targets for a nontrivial differential supported by $\{v_2^8v_1^4d_0e_0[1]\}$ in the MASS for $M(8, v_1^8)$. Therefore $\{v_2^8v_1^4d_0e_0[1]\}$ is a permanent cycle in the MASS for $M(8, v_1^8)$. $\square$

Lemma 7.22 *The following classes lift to the top cell of $M(8, v_1^8)$:*

(1) $\Delta^4\nu^2 \in \pi_{102}(\mathrm{tmf})$, $\Delta^4\epsilon \in \pi_{104}(\mathrm{tmf})$, $\Delta^4\kappa \in \pi_{110}(\mathrm{tmf})$.

(2) $\Delta^4 2\overline{\kappa} \in \pi_{116}(\mathrm{tmf})$.

Proof (1) These classes were lifted in [7, Theorem 11.1].

(2) The class $\Delta^4 2\bar{\kappa} \in \pi_{116}(\mathrm{tmf})$ lifts to an element

$$\Delta^4 2\bar{\kappa}[1] \in \mathrm{tmf}_{117}(M(8))$$

which is detected by

$$v_2^{16} h_0 g[1] \in \mathrm{Ext}_{A(2)_*}^{23,117+23}(H(8))$$

in the MASS for $\mathrm{tmf} \wedge M(8)$. Lift $\Delta^4 2\bar{\kappa}[1]$ to an element

$$\Delta^4 2\bar{\kappa}[18] \in \mathrm{tmf}_{134}(M(8, v_1^8)).$$

In the MASS for $\mathrm{tmf} \wedge M(8)$, there is a differential

$$d_2(v_2^{18} v_1^4 d_0 h_2[1]) = v_2^{16} v_1^8 h_0 g[1].$$

By the geometric boundary theorem, $\Delta^4 2\bar{\kappa}[18]$ is detected by $v_2^{18} v_1^4 d_0 h_2[1]$ in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$. In particular, we see that $\Delta^4 2\bar{\kappa}[18]$ has MAF 29 and stem 134.

We now check that $v_2^{18} v_1^4 d_0 h_2[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$. Its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes H(8, v_1^8))$ is $t-s = 127$ and $\mathrm{AF} = 28$, its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2} \otimes H(8, v_1^8))$ is $t-s = 120$ and $\mathrm{AF} = 27$, and its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 3} \otimes H(8, v_1^8))$ is $t-s = 113$ and $\mathrm{AF} = 26$, the last of which lies above the vanishing line. Inspection of the relevant charts shows that $v_2^{16} 2\bar{\kappa}[18]$ cannot support a nontrivial d_1–differential since the target bidegrees are zero. Therefore $v_2^{16} 2\bar{\kappa}[18]$ is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$ and lifts to an element $v_2^{16} 2\bar{\kappa}[18]$ in $\mathrm{Ext}_{A_*}(H(8, v_1^8))$.

Finally, inspection of the same charts reveals that there are no possible targets for a nontrivial differential supported by $v_2^{16} 2\bar{\kappa}[18]$ in the MASS for $M(8, v_1^8)$. Therefore $v_2^{16} 2\bar{\kappa}[18]$ is a permanent cycle. □

Contrary to the previous cases, there are several potential obstructions to lifting $\Delta^4 \bar{\kappa}\eta \in \pi_{117}(\mathrm{tmf})$ to the top cell of $M(8, v_1^8)$ which are tricky to resolve. However, since this element is 2–torsion and v_1^4–torsion, we may instead attempt to lift it to the top cell of the generalized Moore spectrum $M(2, v_1^4)$ of [6], where the potential obstructions are much simpler to analyze. It then follows from the fact that the composite

$$\Sigma^8 M(2, v_1^4) \xrightarrow{\cdot 4v_1^4} M(8, v_1^8) \to S^{18}$$

is projection onto the top cell of $M(2, v_1^4)$ that $\Delta^4 \bar{\kappa}\eta$ does lift to the top cell of $M(8, v_1^8)$.

Lemma 7.23 *The class $\Delta^4\overline{\kappa}\eta \in \pi_{117}(\mathrm{tmf})$ lifts to the top cell of $M(2, v_1^4)$.*

Proof The class $\Delta^4\eta\overline{\kappa} \in \pi_{117}(\mathrm{tmf})$ lifts to an element

$$\Delta^4\eta\overline{\kappa}[1] \in \mathrm{tmf}_{118}(M(2))$$

which is detected by

$$v_2^{16}h_1g[1] \in \mathrm{Ext}^{21,118+21}(H(2))$$

in the MASS for $\mathrm{tmf} \wedge M(2)$. Lift $\Delta^4\eta\overline{\kappa}[1]$ to an element

$$\Delta^4\eta\overline{\kappa}[10] \in \mathrm{tmf}_{127}(M(2, v_1^4)).$$

In the MASS for $\mathrm{tmf} \wedge M(2)$, there is a differential

$$d_3(v_2^{20}h_2^2[1]) = v_2^{16}v_1^4h_1g[1].$$

It follows from the geometric boundary theorem that $\Delta^4\eta\overline{\kappa}[10]$ is detected by $v_2^{20}h_2^2[1]$ in the MASS for $\mathrm{tmf} \wedge M(2, v_1^4)$. In particular, we see that $\Delta^4\eta\overline{\kappa}[10]$ has MAF 24 and stem 127.

We now check that $v_2^{20}h_2^2[1]$ is a permanent cycle in the algebraic tmf resolution for $H(2, v_1^4)$. Its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes H(2, v_1^4))$ is $t-s = 120$ and $\mathrm{AF} = 23$, its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2} \otimes H(2, v_1^4))$ is $t-s = 113$ and $\mathrm{AF} = 22$, and its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 3} \otimes H(2, v_1^4))$ is $t-s = 106$ and $\mathrm{AF} = 21$. Inspection of the relevant charts [6, Figures 6.4–6.5] shows that there is potentially a nontrivial differential

$$d_1(v_2^{20}h_2^2[1]) = x_{119,24},$$

in the algebraic tmf resolution, where

$$x_{119,24} \in \mathrm{Ext}_{A(2)_*}^{24,119+24}(\underline{\mathrm{bo}}_1 \otimes H(2, v_1^4)),$$

but, since $v_2^{20}h_2^2[1]$ is v_2^{16}–divisible and $x_{119,24}$ is not, this differential cannot occur (compare with the proof of [7, Proposition 10.1]). Therefore $v_2^{20}h_2^2[1]$ is a permanent cycle in the algebraic tmf resolution for $H(2, v_1^4)$ and therefore lifts to an element $\{v_2^{20}h_2^2[1]\}$ in $\mathrm{Ext}_{A_*}(H(2, v_1^4))$.

Finally, inspection of the same charts reveals that there are no possible nontrivial differentials supported by $\{v_2^{20}h_2^2[1]\}$ in the MASS for $M(2, v_1^4)$. Therefore $\{v_2^{20}h_2^2[1]\}$ is a permanent cycle in the MASS for $M(2, v_1^4)$. □

Lemma 7.24 *The class $\Delta^4 q \in \pi_{128}(\mathrm{tmf})$ lifts to the top cell of $M(8, v_1^8)$.*

Proof The class $\Delta^4 q \in \pi_{128}(\mathrm{tmf})$ lifts to an element

$$\Delta^4 q[1] \in \mathrm{tmf}_{129}(M(8))$$

which is detected by

$$v_2^{20} c_0[1] \in \mathrm{Ext}_{A(2)_*}^{23,129+23}(H(8))$$

in the MASS for $\mathrm{tmf} \wedge M(8)$. Lift $\Delta^4 q[1]$ to an element

$$\Delta^4 q[18] \in \mathrm{tmf}_{146}(M(8, v_1^8)).$$

In the MASS for $\mathrm{tmf} \wedge M(8)$, there is a differential

$$d_4(v_2^{16} g^2 h_{2,1} v_1^2[1]) = v_2^{20} v_1^8 c_0[1].$$

By the geometric boundary theorem, $\Delta^4 q[18]$ is detected by $v_2^{16} g^2 h_{2,1} v_1^2[1]$ in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$. In particular, we see that $\Delta^4 q[18]$ has MAF 29 and stem 146.

We now check that $v_2^{16} g^2 h_{2,1} v_1^2[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$. Its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes H(8, v_1^8))$ is $t-s = 139$ and $\mathrm{AF} = 28$, its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2} \otimes H(8, v_1^8))$ is $t-s = 132$ and $\mathrm{AF} = 27$, and its relative position in $\mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 3} \otimes H(8, v_1^8))$ is $t-s = 125$ and $\mathrm{AF} = 26$.

The proof of Lemma 7.18(1) implies that the element

$$g^2 h_{2,1} v_1^2[1] \in \mathrm{Ext}_{A(2)_*}(H(8, v_1^8))$$

is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$. It follows from Lemma 5.1 that

$$v_2^{16} g^2 h_{2,1} v_1^2[1]$$

is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$, and detects an element

$$v_2^{16} \cdot \{g^2 h_{2,1} v_1^2[1]\} \in \mathrm{Ext}_{A_*}(H(8, v_1^8))$$

which persists to the E_3–page of the MASS for $M(8, v_1^8)$.

The only possibility for this element to support a nontrivial MASS differential is for it to support a d_3–differential whose target to by detected by the element

$$v_1 h_{2,1}^{19}(v_0^{-1} v_2^2[\zeta_1^8, \zeta_2^4])[18] \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2} \otimes H(8, v_1^8))$$

in the algebraic tmf resolution for $H(8, v_1^8)$.

We wish to use Lemma 5.5 to argue that the element $v_1 h_{2,1}^{19}(v_0^{-1} v_2^2[\zeta_1^8, \zeta_2^4])[18]$ detects an element in $\mathrm{Ext}_{A_*}(H(8, v_1^8))$ which is zero in the E_3–page of the MASS. In the MASS for $\mathrm{bo}_1^2 \wedge M(8, v_1^8)$, there is a differential

$$d_2(v_2^8 h_{2,1}^{10}(v_0^{-1} v_2^2[\zeta_1^8, \zeta_2^4])[18]) = v_1 h_{2,1}^{19}(v_0^{-1} v_2^2[\zeta_1^8, \zeta_2^4])[18].$$

Using the map

$$\Sigma^{16}\mathrm{tmf} \wedge \mathrm{bo}_1^2 \wedge M(8, v_1^8) \hookrightarrow \mathrm{tmf} \wedge \overline{\mathrm{tmf}}^2 \wedge M(8, v_1^8)$$

we get the same differential in the MASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^2 \wedge M(8, v_1^8)$. By Proposition 4.3, the element $v_2^8 h_{2,1}^{10}(v_0^{-1} v_2^2[\zeta_1^8, \zeta_2^4])[18]$ is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$, detecting the element

$$\Delta^2 v_1^6 M(g^2)[1] \in \mathrm{Ext}_{A_*}(H(8, v_1^8)).$$

Therefore the hypotheses of Lemma 5.5 are satisfied, and we deduce that

$$v_1 h_{2,1}^{19}(v_0^{-1} v_2^2[\zeta_1^8, \zeta_2^4])[18]$$

detects an element which is zero in the E_3–page of the MASS, and hence cannot be the target of a nontrivial d_3–differential in the MASS. □

Lemma 7.25 *The class $\Delta^4 u \in \pi_{135}(\mathrm{tmf})$ lifts to the top cell of $M(8, v_1^8)$.*

Proof The class $\Delta^4 u \in \pi_{135}(\mathrm{tmf})$ lifts to an element

$$\Delta^4 u[1] \in \mathrm{tmf}_{136}(M(8))$$

which is detected by

$$v_2^{16} v_1^2 x_{35}[1] \in \mathrm{Ext}_{A(2)_*}^{25,136+25}(H(8))$$

in the MASS for $\mathrm{tmf} \wedge M(8)$. Lift $\Delta^4 u[1]$ to an element

$$\Delta^4 u[18] \in \mathrm{tmf}_{153}(M(8, v_1^8)).$$

There is a differential in the MASS for $\mathrm{tmf} \wedge M(8)$,

$$d_4(v_2^{16} v_1^3 h_{2,1}^2 g^2[1]) = v_2^{16} v_1^{10} x_{35}[1],$$

so, by the geometric boundary theorem, $\Delta^4 u[18]$ is detected by $v_2^{16} v_1^3 h_{2,1}^2 g^2[1]$ in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$. In particular, $\Delta^4 u[18]$ has MAF 31 and stem 153.

We now check that $v_2^{16} v_1^3 h_{2,1}^2 g^2[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8, v_1^8)$. Note that $v_1^3 h_{2,1}^2 g^2[1]$ detects $u[18]$ in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$. In Lemma 7.18, we established that $u[18]$ lifts to $M(8, v_1^8)$, and therefore $v_1^3 h_{2,1}^2 g^2[1]$ is a permanent cycle in the algebraic tmf resolution and it detects a permanent cycle in the MASS for $M(8, v_1^8)$. It follows from Lemma 5.1 that

$$v_2^{16} v_1^3 h_{2,1}^2 g^2[1]$$

is a permanent cycle in the algebraic tmf resolution and detects an element

$$v_2^{16} \cdot \{v_1^3 h_{2,1}^2 g^2[1]\} \in \mathrm{Ext}_{A_*}(H(8, v_1^8)).$$

Inspection of the relevant charts shows that the only possible nontrivial MASS differentials supported by this element would be

$$d_2(v_2^{16} \cdot \{v_1^2 h_{2,1}^2 g^2[1]\}) = \{v_2^8 h_{2,1}^{15} \zeta_2^4[18]\}.$$

However, we have

$$d_2(v_2^{16} \cdot \{v_1^3 h_{2,1}^2 g^2[1]\}) = 0,$$

since it is a product of d_2–cycles. □

Lemma 7.26 *The following classes lift to the top cell of* $M(8, v_1^8)$:

(1) $\Delta^6 \nu^2 \in \pi_{150}(\mathrm{tmf})$.

(2) $\Delta^6 \kappa \nu \in \pi_{161}(\mathrm{tmf})$.

Proof (1) The class $\Delta^6 \nu^2 \in \pi_{150}(\mathrm{tmf})$ lifts to an element

$$\Delta^6 \nu^2[1] \in \mathrm{tmf}_{151}(M(8))$$

which is detected by

$$v_2^{24} h_2^2[1] \in \mathrm{Ext}_{A(2)_*}^{28,151+28}(H(8))$$

in the MASS for $\mathrm{tmf} \wedge M(8)$. Lift $\Delta^6 \nu^2[1]$ to an element

$$\Delta^6 \nu^2[18] \in \mathrm{tmf}_{168}(M(8, v_1^8)).$$

In the MASS for $\mathrm{tmf} \wedge M(8)$, there is a differential

$$d_2(v_2^{26} v_1^4 h_2 h_0[1]) = v_2^{24} v_1^8 h_2^2[1].$$

By the geometric boundary theorem, $\Delta^6 \nu^2[18]$ is detected by $v_2^{26} v_1^4 h_2 h_0[1]$ in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$. In particular, we see that $\Delta^6 \nu^2[18]$ has MAF 34 and stem 168.

In Lemma 7.21(1), we showed that $v_2^{10} v_1^4 h_2 h_0[1]$ is a permanent cycle in the algebraic tmf resolution, detecting an element

$$\{v_2^{10} v_1^4 h_2 h_0[1]\} \in \mathrm{Ext}_{A_*}(H(8, v_1^8))$$

in the algebraic tmf resolution for $H(8, v_1^8)$. By Lemma 5.1, this is also true of $v_2^{26} v_1^4 h_2 h_0[1]$.

Lemma 5.1 implies that $d_2(v_2^{16}) = 0$ in the MASS for $M(8, v_1^8)$. By Lemma 7.21(1), it follows that

$$d_2(v_2^{16} \cdot \{v_2^{10} v_1^4 h_2 h_0[1]\}) = 0.$$

Inspection of the algebraic tmf resolution charts reveals that there are no possible targets of a longer MASS differential supported by $v_2^{16} \cdot \{v_2^{10} v_1^4 h_2 h_0[1]\}$.

(2) The class $\Delta^6\kappa\nu \in \pi_{161}(\mathrm{tmf})$ lifts to an element

$$\Delta^6\kappa\nu[1] \in \mathrm{tmf}_{162}(M(8))$$

which is detected by

$$v_2^{24}d_0h_2[1] \in \mathrm{Ext}_{A(2)_*}^{31,161+31}(H(8))$$

in the MASS for $\mathrm{tmf} \wedge M(8)$. Lift $\Delta^6\kappa\nu[1]$ to an element

$$\Delta^6\kappa\nu[18] \in \mathrm{tmf}_{179}(M(8,v_1^8)).$$

In the MASS for $\mathrm{tmf} \wedge M(8)$, there is a differential

$$d_2(v_2^{26}v_1^4h_0d_0[1]) = v_2^{24}v_1^8h_2d_0[1].$$

By the geometric boundary theorem, $\Delta^6\kappa\nu[18]$ is detected by $v_2^{26}v_1^4h_0d_0[1]$ in the MASS for $\mathrm{tmf} \wedge M(8,v_1^8)$. In particular, we see that $\Delta^6\kappa\nu[18]$ has MAF 37 and stem 179.

We showed in Lemma 7.21 that $v_2^{10}v_1^4h_0d_0[1]$ is a permanent cycle in the algebraic tmf resolution. By Lemma 5.1, it follows that $v_2^{26}v_1^4h_0d_0[1]$ is a permanent cycle in the algebraic tmf resolution for $H(8,v_1^8)$ and lifts to an element $\{v_2^{26}v_1^4h_0d_0[1]\}$ in $\mathrm{Ext}_{A_*}(H(8,v_1^8))$.

Finally, inspection of the algebraic tmf resolution charts reveals that there are no possible nontrivial differentials on $\{v_2^{26}v_1^4h_0d_0[1]\}$ in the MASS for $M(8,v_1^8)$. Therefore $\{v_2^{26}v_1^4h_0d_0[1]\}$ is a permanent cycle. □

Lemma 7.27 *The classes $\Delta^8\nu^2 \in \pi_{198}\mathrm{tmf}$ and $\Delta^8\epsilon \in \pi_{200}\mathrm{tmf}$ lift to the top cell of $M(8,v_1^8)$.*

Proof The classes $\Delta^8\nu^2 \in \pi_{198}(\mathrm{tmf})$ and $\Delta^8\epsilon \in \pi_{200}\mathrm{tmf}$ lift to elements

$$\Delta^8\nu^2[1] \in \mathrm{tmf}_{199}(M(8)), \quad \Delta^8\epsilon[1] \in \mathrm{tmf}_{201}(M(8))$$

which are detected by

$$v_2^{32}h_2^2[1] \in \mathrm{Ext}_{A(2)_*}^{36,199+36}(H(8)), \quad v_2^{32}c_0[1] \in \mathrm{Ext}_{A(2)_*}^{37,201+37}(H(8))$$

in the MASS for $\mathrm{tmf} \wedge M(8)$. Lift $\Delta^8\nu^2[1]$ and $\Delta^8\epsilon[1]$ to elements

$$\Delta^8\nu^2[18] \in \mathrm{tmf}_{210}(M(8,v_1^8)), \quad \Delta^8\epsilon[18] \in \mathrm{tmf}_{212}(M(8,v_1^8)).$$

In the MASS for $\mathrm{tmf} \wedge M(8)$, there are differentials

$$d_2(v_2^{32} v_1^4 h_0 h_2 v_2^2[1]) = v_2^{32} v_1^8 h_2^2[1], \quad d_3(v_2^{32} v_1^4 e_0[1]) = v_2^{32} v_1^8 c_0[1].$$

By the geometric boundary theorem, $\Delta^8 \nu^2[18]$ is detected by $v_2^{34} v_1^4 h_0 h_2[1]$ and $\Delta^8 \epsilon[18]$ is detected by $v_2^{32} v_1^4 e_0[1]$ in the MASS for $\mathrm{tmf} \wedge M(8, v_1^8)$.

In [7, Theorem 11.1] the classes $\Delta^4 \nu^2[18] \in \pi_{120} M(8, v_1^8)$ and $\Delta^4 \epsilon[18] \in \pi_{122} M(8, v_1^8)$ were produced by showing that the elements

$$v_2^{18} v_1^4 h_0 h_2[1] \in \mathrm{Ext}_{A(2)_*}^{26,120+26}(H(8, v_1^8)), \quad v_2^{16} v_1^4 e_0[1] \in \mathrm{Ext}_{A(2)_*}^{26,122+26}(H(8, v_1^8))$$

detect via the algebraic tmf resolution elements

$$\{v_2^{18} v_1^4 h_0 h_2[1]\} \in \mathrm{Ext}_{A_*}^{26,120+26}(H(8, v_1^8)), \quad \{v_2^{16} v_1^4 e_0[1]\} \in \mathrm{Ext}_{A_*}^{26,122+26}(H(8, v_1^8)),$$

which are permanent cycles in the MASS for $M(8, v_1^8)$.

Since the element $v_2^{16} \in \mathrm{Ext}_{A_*}(H(8, v_1^8))$ is the square of the element v_2^8, we have $d_2(v_2^{16}) = 0$. We deduce that the elements

$$v_2^{16} \cdot \{v_2^{18} v_1^4 h_0 h_2[1]\} \in \mathrm{Ext}_{A_*}^{26,120+26}(H(8, v_1^8)),$$
$$v_2^{16} \cdot \{v_2^{16} v_1^4 e_0[1]\} \in \mathrm{Ext}_{A_*}^{26,122+26}(H(8, v_1^8))$$

persist to the E_3–page of the MASS for $M(8, v_1^8)$. If we can show they are permanent cycles, we are done.

We begin with $\{v_2^{34} v_1^4 h_0 h_2[1]\}$. Examination of the algebraic tmf resolution for $M(8, v_1^8)$ reveals that the only possibility of a nontrivial differential in the MASS supported by this element would be a $d_4(\{v_2^{34} v_1^4 h_0 h_2[1]\})$, which would be detected by

$$h_{2,1}^{33} v_1 v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4][18] \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1^{\otimes 2} \otimes H(8, v_1^8)).$$

In the MASS for $\mathrm{tmf} \wedge \mathrm{bo}_1^{\wedge 2}$ there is a differential

$$d_2^{\mathrm{mass}}(\Delta^2 h_{2,1}^{24} v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4][18]) = h_{2,1}^{33} v_1 v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4][18].$$

Using the map (5.7) we deduce that there is a corresponding differential in the MASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}}^{\wedge 2}$. The elements

$$h_{2,1}^{33} v_1 v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4][18], \quad \Delta^2 h_{2,1}^{24} v_1 v_0^{-1} v_2^2 [\zeta_1^8, \zeta_2^4][18]$$

respectively detect

$$v_1^7 h_{2,1}^{23} M g^2[1] \in \mathrm{Ext}_{A_*}(H(8, v_1^8)), \quad \Delta^2 v_1^6 h_{2,1}^{14} M g^2[1] \in \mathrm{Ext}_{A_*}(H(8, v_1^8))$$

in the algebraic tmf resolution for $M(8, v_1^8)$. We therefore deduce from Lemma 5.5 that $v_1^7 h_{2,1}^{23} M g^2[1]$ is killed by

$$d_2^{\mathrm{mass}}(\Delta^2 v_1^6 h_{2,1}^{14} M g^2[1])$$

in the MASS for $M(8, v_1^8)$. Therefore it cannot be the target of a nontrivial d_4^{mass}.

We now consider $\{v_2^{32} v_1^4 e_0[1]\}$. Examination of the algebraic tmf resolution for $M(8, v_1^8)$ reveals that the only possibility of a nontrivial differential in the MASS supported by this element would be a $d_4(\{v_2^{32} v_1^4 e_0[1]\})$, which would be detected by

$$\Delta^2 h_{2,1}^{28} \zeta_2^4[18] \in \mathrm{Ext}_{A(2)_*}(\underline{\mathrm{bo}}_1 \otimes H(8, v_1^8))$$

in the algebraic tmf resolution for $M(8, v_1^8)$. To eliminate this possibility we wish to employ Case 1 of Remark 5.6, using the differential

$$d_3^{\mathrm{mass}}(\Delta^2 h_{2,1}^{25} v_0^{-2} v_2^2 \zeta_1^{16}[18]) = \Delta^2 h_{2,1}^{28} \zeta_2^4[18]$$

in the MASS for $\mathrm{tmf} \wedge \overline{\mathrm{tmf}} \wedge M(8, v_1^8)$. The element $\Delta^2 h_{2,1}^{25} v_0^{-2} v_2^2 \zeta_1^{16}[18]$ detects the element

$$\Delta^2 h_{2,1}^{19} Q_2[18] \in \mathrm{Ext}_{A_*}(H(8, v_1^8))$$

in the algebraic tmf resolution for $M(8, v_1^8)$. We just need to check that there is no possibility for $\Delta^2 h_{2,1}^{19} Q_2[18]$ to support a nontrivial d_2^{mass} in the MASS for $M(8, v_1^8)$. However, examination of the algebraic tmf resolution for $M(8, v_1^8)$ reveals there are no classes which could detect the target of such a nontrivial d_2^{mass}. □

References

[1] **J F Adams**, *On the groups $J(X)$, IV*, Topology 5 (1966) 21–71 MR Zbl

[2] **M G Barratt**, **J D S Jones**, **M E Mahowald**, *Relations amongst Toda brackets and the Kervaire invariant in dimension* 62, J. London Math. Soc. 30 (1984) 533–550 MR Zbl

[3] **T Bauer**, *Computation of the homotopy of the spectrum* `tmf`, from "Groups, homotopy and configuration spaces" (N Iwase, T Kohno, R Levi, D Tamaki, J Wu, editors), Geom. Topol. Monogr. 13, Geom. Topol. Publ., Coventry (2008) 11–40 MR Zbl

[4] **A Beaudry**, **M Behrens**, **P Bhattacharya**, **D Culver**, **Z Xu**, *On the E_2–term of the bo–Adams spectral sequence*, J. Topol. 13 (2020) 356–415 MR Zbl

[5] **M Behrens**, *The Goodwillie tower and the EHP sequence*, Mem. Amer. Math. Soc. 1026, Amer. Math. Soc., Providence, RI (2012) MR Zbl

[6] **M Behrens**, **M Hill**, **M J Hopkins**, **M Mahowald**, *On the existence of a v_2^{32}–self map on $M(1,4)$ at the prime* 2, Homology Homotopy Appl. 10 (2008) 45–84 MR Zbl

[7] **M Behrens**, **M Hill**, **M J Hopkins**, **M Mahowald**, *Detecting exotic spheres in low dimensions using* coker J, J. Lond. Math. Soc. 101 (2020) 1173–1218 MR Zbl

[8] **M Behrens**, **K Ormsby**, **N Stapleton**, **V Stojanoska**, *On the ring of cooperations for 2–primary connective topological modular forms*, J. Topol. 12 (2019) 577–657 MR Zbl

[9] **E Belmont**, **K Shimomura**, *Beta families arising from a v_2^9 self map on $S/(3, v_1^8)$*, preprint (2021) arXiv 2109.01059

[10] **P Bhattacharya**, **I Bobkova**, **B Thomas**, *The P_2^1 Margolis homology of connective topological modular forms*, Homology Homotopy Appl. 23 (2021) 379–402 MR Zbl

[11] **R Bott**, *The stable homotopy of the classical groups*, Ann. of Math. 70 (1959) 313–337 MR Zbl

[12] **W Browder**, *The Kervaire invariant of framed manifolds and its generalization*, Ann. of Math. 90 (1969) 157–186 MR Zbl

[13] **R R Bruner**, *The cohomology of the mod* 2 *Steenrod algebra*: *a computer computation*, preprint (2019) Available at `http://www.rrb.wayne.edu/papers/cohom.pdf`

[14] **R R Bruner**, **J Rognes**, *The Adams spectral sequence for topological modular forms*, Mathematical Surveys and Monographs 253, Amer. Math. Soc., Providence, RI (2021) MR Zbl

[15] **D M Davis**, **M Mahowald**, *Connective versions of* TMF(3), Int. J. Mod. Math. 5 (2010) 223–252 MR Zbl

[16] **C L Douglas**, **J Francis**, **A G Henriques**, **M A Hill** (editors), *Topological modular forms*, Mathematical Surveys and Monographs 201, Amer. Math. Soc., Providence, RI (2014) MR Zbl

[17] **P G Goerss**, **J D S Jones**, **M E Mahowald**, *Some generalized Brown–Gitler spectra*, Trans. Amer. Math. Soc. 294 (1986) 113–132 MR Zbl

[18] **M A Hill**, **M J Hopkins**, **D C Ravenel**, *On the nonexistence of elements of Kervaire invariant one*, Ann. of Math. 184 (2016) 1–262 MR Zbl

[19] **M J Hopkins**, **M Mahowald**, *From elliptic curves to homotopy theory*, from "Topological modular forms" (C L Douglas, J Francis, A G Henriques, M A Hill, editors), Mathematical Surveys and Monographs 201, Amer. Math. Soc., Providence, RI (2014) 261–285 MR Zbl

[20] **M J Hopkins**, **J H Smith**, *Nilpotence and stable homotopy theory, II*, Ann. of Math. 148 (1998) 1–49 MR Zbl

[21] **M Hovey**, *Homotopy theory of comodules over a Hopf algebroid*, from "Homotopy theory: relations with algebraic geometry, group cohomology, and algebraic K–theory" (P Goerss, S Priddy, editors), Contemp. Math. 346, Amer. Math. Soc., Providence, RI (2004) 261–304 MR Zbl

[22] **D C Isaksen**, *Stable stems*, Mem. Amer. Math. Soc. 1269, Amer. Math. Soc., Providence, RI (2019) MR Zbl

[23] **D C Isaksen**, *The Mahowald operator in the cohomology of the Steenrod algebra*, preprint (2020) arXiv 2001.01758

[24] **D C Isaksen**, **G Wang**, **Z Xu**, *Stable homotopy groups of spheres: from dimension* 0 *to* 90, preprint (2020) arXiv 2001.04511

[25] **D C Isaksen**, **G Wang**, **Z Xu**, *Classical and $\mathbb{C}$–motivic Adams charts*, electronic resource (2022) Available at `https://dx.doi.org/10.5281/zenodo.6987157`

[26] **M A Kervaire**, **J W Milnor**, *Groups of homotopy spheres, I*, Ann. of Math. 77 (1963) 504–537 MR Zbl

[27] **G Laures**, *$K(1)$–local topological modular forms*, Invent. Math. 157 (2004) 371–403 MR Zbl

[28] **W Lellmann**, **M Mahowald**, *The bo–Adams spectral sequence*, Trans. Amer. Math. Soc. 300 (1987) 593–623 MR Zbl

[29] **G Li**, **X D Shi**, **G Wang**, **Z Xu**, *Hurewicz images of real bordism theory and real Johnson–Wilson theories*, Adv. Math. 342 (2019) 67–115 MR Zbl

[30] **M Mahowald**, *bo–resolutions*, Pacific J. Math. 92 (1981) 365–383 MR Zbl

[31] **M Mahowald**, **C Rezk**, *Topological modular forms of level* 3, Pure Appl. Math. Q. 5 (2009) 853–872 MR Zbl

[32] **M Mahowald**, **M Tangora**, *An infinite subalgebra of* $\mathrm{Ext}_A(Z_2, Z_2)$, Trans. Amer. Math. Soc. 132 (1968) 263–274 MR Zbl

[33] **H Margolis**, **S Priddy**, **M Tangora**, *Another systematic phenomenon in the cohomology of the Steenrod algebra*, Topology 10 (1970) 43–46 MR Zbl

[34] **A Mathew**, *The homology of tmf*, Homology Homotopy Appl. 18 (2016) 1–29 MR Zbl

[35] **E E Moise**, *Affine structures in* 3*–manifolds, V: The triangulation theorem and Hauptvermutung*, Ann. of Math. 56 (1952) 96–114 MR Zbl

[36] **G Wang**, **Z Xu**, *The triviality of the* 61*–stem in the stable homotopy groups of spheres*, Ann. of Math. 186 (2017) 501–580 MR Zbl

Department of Mathematics, University of Notre Dame
Notre Dame, IN, United States

Department of Mathematics, Northwestern University
Evanston, IL, United States

Department of Mathematics, University of Virginia
Charlottesville, VA, United States

`mbehren1@nd.edu`, `mbp6pj@virginia.edu`

Proposed: Haynes R Miller Received: 17 February 2021
Seconded: Stefan Schwede, Jesper Grodal Revised: 18 November 2021

Geometry & Topology **27**:7 (2023) 2833–2897
DOI: 10.2140/gt.2023.27.2833
Published: 19 September 2023

Hamiltonian no-torsion

MARCELO S ATALLAH
EGOR SHELUKHIN

In 2002, Polterovich established that on closed aspherical symplectic manifolds, Hamiltonian diffeomorphisms of finite order, also called Hamiltonian torsion, must be trivial. We prove the first higher-dimensional Hamiltonian no-torsion theorems beyond that of Polterovich, by considering the dynamical aspects of the problem. Our results are threefold.

First, we show that closed symplectic Calabi–Yau and negative monotone symplectic manifolds admit no Hamiltonian torsion. A key role is played by a new notion of a Hamiltonian diffeomorphism with nonisolated fixed points.

Second, going beyond topological constraints by means of Smith theory in filtered Floer homology, barcodes and quantum Steenrod powers, we prove that every closed positive monotone symplectic manifold admitting Hamiltonian torsion is geometrically uniruled by pseudoholomorphic spheres. In fact, we produce nontrivial homological counts of such curves, answering a close variant of Problem 24 from the introductory monograph of McDuff and Salamon. This provides additional no-torsion results and obstructions to Hamiltonian actions of compact Lie groups, related to a celebrated result of McDuff from 2009, and lattices such as $\mathrm{SL}(k,\mathbb{Z})$ for $k \geq 2$. We also prove that there is no Hamiltonian torsion diffeomorphism with noncontractible orbits.

Third, by defining a new invariant of a Hamiltonian diffeomorphism, we prove a first nontrivial symplectic analogue of Newman's 1931 theorem on finite groups of transformations. Namely, for each monotone symplectic manifold there exists a neighborhood of the identity in the Hamiltonian group endowed with Hofer's metric or Viterbo's spectral metric that contains no finite subgroups.

37J11, 53D22, 53D40; 57R17, 57S17

1 Introduction and main results

1.1 Introduction

The question of the existence of finite group actions on manifolds has been of interest in topology for a long time. It was in order to study this question that P A Smith [99] developed in the 1930s what is now called Smith theory for cohomology with $\mathbb{F}_p$ coefficients in the context of continuous actions of finite p-groups. We refer the reader to Borel [4], Bredon [5], Floyd [25] and Hsiang [45] for references on Smith theory.

Quite a lot of progress regarding this question has been obtained in low-dimensional topology (see for example Morgan [62]) and in smooth topology in arbitrary dimension (see for example Mundet i Riera [78]). As a first easy example, we remark that it is not hard to classify finite group actions on closed surfaces. Further progress was made in low-dimensional symplectic topology (Chen and Kwasik [10]) ruling out symplectic finite group actions acting trivially on homology on certain symplectic Calabi–Yau 4–manifolds (see also Wu and Liu [109]) by means of tools such as Seiberg–Witten theory, which are available only in dimension four.

In higher-dimensional symplectic topology,[1] while the existence of general symplectic finite group actions has to the best of our knowledge not been ruled out in any given setting,[2] it was shown by Polterovich [72] that nontrivial *Hamiltonian* finite group actions, which we refer to as *Hamiltonian torsion*, on symplectically aspherical manifolds do not exist. Essentially, the only other constraints on symplectic and Hamiltonian finite group actions in higher dimensions were obtained by Mundet i Riera [77], showing, roughly speaking, that finite groups acting in a Hamiltonian way (or symplectically in the simply connected case) must be approximately abelian: specifically, they satisfy the Jordan property. In turn, abelian Hamiltonian finite group actions do exist on closed symplectic manifolds such as toric varieties, which tend to have a lot of pseudoholomorphic curves. These developments, as well as further results that we describe below, have motivated Problem 24 from the list of problems that are "appealing in their own right and central to symplectic topology" in the monograph [59] of McDuff and Salamon. This problem seeks obstructions to the existence of Hamiltonian torsion related to the scarcity of pseudoholomorphic curves in the manifold. One of the goals of this paper is to produce a solution to a close version of Problem 24, proving a result which is, in a way, stronger.

[1] That is, in dimension $2n \geq 6$.

[2] See Section 1.2.4, however.

Another goal is to study the metric rigidity properties of Hamiltonian torsion, also alluded to in the presentation of this problem. Finally, we prove a topological rigidity result: all periodic orbits of a Hamiltonian isotopy whose time-one map is torsion must be contractible.

To motivate Problem 24 further, and to introduce a few important notions, we add that Hamiltonian actions of cyclic groups on rational ruled symplectic 4–manifolds — that is, symplectic S^2–bundles over S^2 — were recently shown to be induced by S^1–actions; see Chen [9] and Chiang and Kessler [11]. However, this is false for general symplectic 4–manifolds; see Remark 7. The strongest restriction to date on manifolds admitting nontrivial Hamiltonian S^1–actions was obtained by McDuff [57], who showed that all such manifolds must be *uniruled*, in the sense that at least one genus-zero k–point Gromov–Witten invariant for $k \geq 3$ involving the point class must not vanish. Of course, rational ruled symplectic 4–manifolds satisfy this condition, with $k = 3$: they are *strongly uniruled.* Either condition implies that these manifolds are *geometrically uniruled*: for each ω–compatible almost complex structure J and each point $p \in M$, there is a J–holomorphic sphere[3] passing through p. Finally, in Shelukhin [93] a new notion of uniruledness, $\mathbb{F}_p$–Steenrod uniruledness, was introduced for $p = 2$, and was generalized to odd primes $p > 2$ by work in progress of Shelukhin and Wilkins [97]; the quantum Steenrod p^{th} power of the cohomology class Poincaré dual to the point class is defined and deformed in the sense of not coinciding with the classical Steenrod p^{th} power. This notion similarly implies geometric uniruledness. It is currently not known whether it implies uniruledness in the sense of McDuff, but it is expected to do so; see Seidel [91] and Seidel and Wilkins [92] for first steps in this direction.

This paper proves the first higher-dimensional Hamiltonian no-torsion results since that of Polterovich, which hold beyond the symplectically aspherical case. Firstly, we prove that, in addition to symplectically aspherical manifolds, symplectically Calabi–Yau and negative monotone symplectic manifolds do not admit Hamiltonian torsion. An elementary argument then shows that if a closed symplectic manifold M admits Hamiltonian torsion, then it has a spherical homology class A such that $\langle c_1(TM), A\rangle > 0$ and $\langle [\omega], A\rangle > 0$; see Corollary 2. Our results have a similar flavor to the result of McDuff for S^1–actions: indeed, negative monotone and Calabi–Yau manifolds are not geometrically uniruled, and neither are the symplectically aspherical ones.

[3]This is a smooth map $u\colon \mathbb{CP}^1 \to M$ satisfying $Du \circ j = J \circ Du$ for the standard complex structure j on $\mathbb{CP}^1$. Such spheres and their significance in symplectic topology were discovered by Gromov [38]. We refer to McDuff and Salamon [58] for a detailed modern description of this notion.

Going far beyond topological restrictions, we further study restrictions on Hamiltonian torsion in the (positive) monotone case. Using recently discovered techniques, we show that in this case the existence of nontrivial Hamiltonian torsion implies $\mathbb{F}_p$–Steenrod uniruledness for certain primes p, and hence geometric uniruledness. This again fits well with the result of McDuff and in fact provides a partial solution to Problem 24 from the monograph [59] of McDuff and Salamon. Studying the properties of the quantum Steenrod operations and their relation to Gromov–Witten invariants further — see Seidel and Wilkins [92] and Wilkins [106; 107] for first inroads in this direction — might show that our solution is in fact quite complete. Furthermore, we are tempted to conjecture the following analogue of the result of McDuff.

Conjecture 1 Each closed symplectic manifold with nontrivial Hamiltonian torsion must be uniruled.

Before addressing further results on the metric properties of Hamiltonian torsion diffeomorphisms when they exist in the monotone case, we comment on our methods of proof. The main general idea of the paper is to treat such a diffeomorphism as a Hamiltonian dynamical system, despite the fact that it exhibits very simple periodic dynamics. Indeed, quite paradoxically, studying its asymptotic behavior for large iterations is effective, as it yields new topological and Floer-theoretical properties of such diffeomorphisms.

Curiously enough, on a more technical level, our arguments involve a recently discovered analogue of Smith theory in filtered Hamiltonian Floer homology (see Seidel [90], Shelukhin [95] and Shelukhin and Zhao [98]), and related notions of quantum Steenrod powers (see Shelukhin and Wilkins [97] and Wilkins [106; 107]). Previously these methods were applied to questions of existence of infinitely many periodic points (see again Shelukhin and Wilkins [97] and Shelukhin [95]) and, more restrictively, of obstructions on manifolds to admit Hamiltonian pseudorotations (see Shelukhin [93; 94] and Çineli, Ginzburg and Gürel [7]). In fact, a more precise general theme of this paper is that a Hamiltonian diffeomorphism of finite order behaves in many senses like a counterexample to the Conley conjecture. For instance, the statement of Corollary 2 is analogous to that of [36, Theorem 1.1] that provides the most general setting wherein the Conley conjecture is known to hold.

Our third and last series of results studies the metric rigidity of Hamiltonian torsion and related maps. We start by proving that the spectral norm (see Oh [65], Schwarz [86] and Viterbo [104]) of a Hamiltonian torsion element ϕ of order k on a closed rational symplectic manifold (ie a manifold for which $\langle[\omega], \pi_2(M)\rangle = \rho\cdot\mathbb{Z}$ with $\rho > 0$) satisfies $\gamma(\phi) \geq \rho/k$, and as an immediate consequence, the same estimate applies for the Hofer norm (see Hofer [40] and Lalonde and McDuff [51]).

More importantly, in our final main result, we prove that in the monotone case, given $\phi \in \mathrm{Ham}(M,\omega) \setminus \{\mathrm{id}\}$ of order k, ie with $\phi^k = \mathrm{id}$, there exists $m \in \mathbb{Z}/k\mathbb{Z}$ such that

$$\gamma(\phi^m) \geq \frac{\rho}{3}. \tag{1}$$

This last result should be considered a Hamiltonian analogue of the celebrated result of Newman [63] (see also Dress [16] and Smith [100]), the C^0–distance having been replaced by the spectral distance. Moreover we prove the stronger statement that if k is prime, then $\gamma(\phi^m) \geq \rho\lfloor k/2\rfloor/k$ for a certain $m \in \mathbb{Z}/k\mathbb{Z}$, and provide a similar statement in the context of Hamiltonian pseudorotations.

The bound (1) can further be seen to imply Newman's result in a special case, as follows. By Shelukhin [96, Theorem C] (see also Kawamoto [47]), when $M = \mathbb{CP}^n$ is the complex projective space with the standard symplectic form normalized so that $\mathbb{CP}^1$ has area 1, there is a constant c_n, depending only on the dimension, such that for all $\phi \in \mathrm{Ham}(M,\omega)$, the usual C^0–distance of ϕ to the identity satisfies

$$d_{C^0}(\phi, \mathrm{id}) \geq c_n \gamma(\phi).$$

Hence, if ϕ is of finite order, then by (1) there exists $m \in \mathbb{Z}$ such that

$$d_{C^0}(\phi^m, \mathrm{id}) \geq \frac{c_n}{3}.$$

It would be very interesting to see if the results of this paper can be extended to the case of Hamiltonian homeomorphisms, as defined in Buhovsky, Humilière and Seyfaddini [6]. This generalization does not seem to be straightforward because we use the properties of the linearization of the Hamiltonian diffeomorphism at its fixed points, as well as Smith theory in filtered Floer homology, which is not in general stable in the C^0–topology.

We close the introduction by noting that we expect that our results in the monotone case should extend to the semipositive case, once the relevant results of [95] and [97] have been generalized to the requisite setting. Since these generalizations would not considerably differ, in a conceptual way, from the arguments presented in this paper, but would necessitate more lengthy technical proofs, we defer their investigation to further publications.

1.2 Main results

We start with the following theorem of Polterovich [72], originally stated in the case where $\pi_2(M) = 0$. For the reader's convenience we include its proof in Section 5.4.

Theorem A (Polterovich) *Let (M,ω) be a closed symplectically aspherical symplectic manifold. If G is a finite group, then each homomorphism $G \to \mathrm{Ham}(M,\omega)$ is trivial.*

In this paper we prove a number of additional "no-torsion" theorems of this kind, going beyond the symplectically aspherical case, and study the metric properties of Hamiltonian diffeomorphisms of finite order when such obstructions do not hold. Our conditions on the manifold that imply the absence of Hamiltonian torsion are of two kinds: the first is purely topological, and the second, perhaps more surprisingly, is in terms of pseudoholomorphic curves.

1.2.1 Topological conditions The first set of results of this paper is as follows.

Theorem B *Let (M,ω) be a closed negative monotone or closed symplectically Calabi–Yau symplectic manifold. If G is a finite group, then each homomorphism $G \to \mathrm{Ham}(M,\omega)$ is trivial.*

A simple exercise in linear algebra shows that the class of manifolds, which we call *symplectically nonpositive*, covered by Theorems A and B can be described concisely as those closed symplectic manifolds (M,ω) for which

$$\langle [\omega], A\rangle \cdot \langle c_1(TM), A\rangle \leq 0 \quad \text{for all } A \in \pi_2(M).$$

In other words, the following holds.

Corollary 2 *If a closed symplectic manifold (M,ω) admits a nontrivial homomorphism $G \to \mathrm{Ham}(M,\omega)$ from a finite group, then there exists an $A \in \pi_2(M)$ such that $\langle [\omega], A\rangle > 0$ and $\langle c_1(TM), A\rangle > 0$.*

For details of this implication see [36, Proof of Theorem 4.1].

Theorem B follows directly from Theorems C and D below. These two steps essentially generalize the notion of a perfect Hamiltonian diffeomorphism, ie one that has a finite number of contractible periodic points of all periods, to the case of compact path-connected isolated sets of fixed points. We call such an isolated set of fixed points of $\phi \in \mathrm{Ham}(M,\omega)$ a *generalized fixed point* of ϕ. Recall that a fixed point x of a Hamiltonian diffeomorphism $\phi = \phi_H^1$ is called contractible whenever the homotopy class $\alpha(x,\phi)$ of the path $\alpha(x,H) = \{\phi_H^t(x)\}$ for a Hamiltonian H generating ϕ is trivial. This class does not depend on the choice of Hamiltonian, by a classical argument in Floer theory. We call a generalized fixed point $\mathcal{F}$ of ϕ contractible if all fixed points $x \in \mathcal{F}$ are contractible. We denote by $\mathscr{F}$ the generalized periodic orbit, consisting of all $\alpha(x,H)$ for $x \in \mathcal{F}$, corresponding to the generalized fixed point $\mathcal{F}$. This is a subset of the free loop space $\mathcal{L}M$ of M. If $\mathcal{F}$ is contractible, we show that there exists a *capping* $\bar{\mathscr{F}}$ of $\mathscr{F}$, which is a lift of $\mathscr{F}$ to a suitable cover of the connected component $\mathcal{L}_{\mathrm{pt}}M$ of the loop space consisting of contractible loops. Finally, and crucially, we introduce the following notion: we call a generalized fixed point $\mathscr{F}$ *index-constant* if the mean-index $\Delta(H,\bar{x})$ for $\bar{x} \in \bar{\mathscr{F}}$ is constant as a function of $\bar{x} \in \bar{\mathscr{F}}$ (which is in turn determined by $x \in \mathcal{F}$ and the capping $\bar{\mathscr{F}}$). We refer to Section 2.1.3 for the definition of the mean-index.

We call $\phi \in \mathrm{Ham}(M,\omega)$ a *generalized perfect* Hamiltonian diffeomorphism if there exists a sequence $k_j \to \infty$ of iterations satisfying the following two properties: first, it contains a subsequence $l_i = k_{j_i}$ with $l_i \mid l_{i+1}$ for all i; second, for all $j \in \mathbb{Z}_{>0}$ the diffeomorphism ϕ^{k_j} has only a finite set, which does not depend on j, of contractible generalized fixed points, *which are all index-constant*.

Finally, we call a diffeomorphism ϕ with a finite number of (contractible) generalized fixed points *weakly nondegenerate* if for each (contractible) fixed point x of ϕ, the spectrum of the differential $D(\phi)_x$ at x contains points different from $1 \in \mathbb{C}$. Using the existence of ω–compatible almost complex structures invariant under a Hamiltonian diffeomorphism of finite order, we prove the following structural result.

Theorem C *Let (M,ω) be a closed symplectic manifold. Then a torsion Hamiltonian diffeomorphism $\phi \in \mathrm{Ham}(M,\omega)$ is a weakly nondegenerate generalized perfect Hamiltonian diffeomorphism. In fact, it is Floer–Morse–Bott and its generalized fixed points are symplectic submanifolds.*

While we do not require this for Theorem C, for most of our applications it is sufficient to assume that ϕ is p–torsion for a prime p, that is, $\phi^p = \mathrm{id}$.

Following the index arguments of Salamon and Zehnder [84], and their generalization due to Ginzburg and Gürel [34], we prove the following obstruction to the existence of weakly nondegenerate generalized perfect Hamiltonian diffeomorphisms.

Theorem D *Let a closed symplectic manifold (M,ω) be negative monotone or symplectically Calabi–Yau. Then (M,ω) does not admit weakly nondegenerate generalized perfect Hamiltonian diffeomorphisms.*

Together with Theorem C, Theorem D immediately implies Theorem B. In fact, in view of Cauchy's theorem for finite groups, to rule out all Hamiltonian finite group actions it is sufficient to rule out all Hamiltonian torsion of prime order. One can say that, almost paradoxically, we use the *large-time asymptotic behavior* of our Hamiltonian system to study its *periodic* dynamics! This is the main general idea of this paper.

As easy examples show, generalized perfect Hamiltonian diffeomorphisms do indeed exist on the manifolds of Theorem D if one drops the weak nondegeneracy assumption. For example, one can take $T^2 = S^1 \times S^1$, where $S^1 = \mathbb{R}/\mathbb{Z}$, to be the standard torus with (x,y) denoting a general point, and $\omega_{\mathrm{st}} = dx \wedge dy$ the standard symplectic form, and pick $\phi \in \mathrm{Ham}(T^2,\omega_{\mathrm{st}})$ given by $\phi = \phi_H^t$ for $t > 0$, with $H \in C^\infty(T^2,\mathbb{R})$ given by $H(x,y) = \cos(2\pi y)$. It is easy to see that the set of contractible periodic points of ϕ consists precisely of the two isolated sets $\{y = 0\}$ and $\{y = \frac{1}{2}\}$.

1.2.2 Conditions in terms of pseudoholomorphic curves Our second set of results deals with the class of monotone symplectic manifolds. It is evident that far more than topological conditions is necessary to rule out Hamiltonian torsion in this case, since each Hamiltonian S^1–manifold, such as $\mathbb{CP}^n$ for example, admits Hamiltonian torsion. We formulate our restriction on the existence of Hamiltonian torsion geometrically as follows. For an ω–compatible almost complex structure J on a closed symplectic manifold (M,ω), we say that the manifold is *geometrically uniruled* if for each point $p \in M$, there exists a nonconstant J–holomorphic sphere $u\colon \mathbb{CP}^1 \to M$ such that $p \in \mathrm{im}(u)$.

Theorem E *Let (M,ω) be a closed monotone symplectic manifold that is not geometrically uniruled for some ω–compatible almost complex structure J. Then each homomorphism $G \to \mathrm{Ham}(M,\omega)$, where G is a finite group, is trivial.*

This is a corollary of the following more precise result involving the quantum Steenrod power operations.

Theorem F *Let (M,ω) be a closed monotone symplectic manifold that admits a Hamiltonian diffeomorphism of order $d > 1$. Then the p^{th} quantum Steenrod power of the cohomology class $\mu \in H^{2n}(M;\mathbb{F}_p)$ Poincaré dual to the point class is deformed for all primes p coprime to d.*

Theorem F follows directly from Theorems G and I below.

Theorems E and F provide an obstruction to the existence of Hamiltonian diffeomorphisms of finite order in terms of pseudoholomorphic curves. The existence of an obstruction of this type was conjectured by McDuff and Salamon, and publicized as Problem 24 in their introductory monograph [59]. Therefore we provide a solution to a reasonable variant of this problem. Indeed, further investigations into the enumerative nature of quantum Steenrod operations might prove that our solution is in fact complete in the framework of monotone symplectic manifolds. Such investigations were initiated in Seidel and Wilkins [92] and Wilkins [106; 107].

In particular, in the special case where (M,ω) has minimal Chern number $N=n+1$, we deduce from Theorem F and the work of Seidel and Wilkins [92], as in Shelukhin [93], that nontrivial Hamiltonian torsion implies that the quantum product $[\mathrm{pt}]*[\mathrm{pt}]$ does not vanish. This means that the manifold is strongly rationally connected: it implies strong uniruledness, and moreover that for each pair of distinct points p_1, p_2 in M, and each ω–compatible almost complex structure J, there exists a J–holomorphic sphere in M passing through p_1 and p_2.

As mentioned above, the proof of Theorem F relies on two steps: Theorems G and I. These steps are aimed at showing that torsion Hamiltonian diffeomorphisms of closed monotone symplectic manifolds, which by Theorem C are generalized perfect and weakly nondegenerate, are moreover homologically minimal in the following sense. To formulate it precisely, we first discuss a useful technical notion.

Let $\mathbb{K}$ be a coefficient field. For a generalized fixed point $\mathcal{F}$ of a Hamiltonian diffeomorphism ψ, we define a generalized version $HF^{\mathrm{loc}}(\psi,\mathcal{F})$ of local Floer homology. Such notions date back to the original work of Floer [24; 23] and have been revisited a number of times: for example by Pozniak in [76]. It is naturally $\mathbb{Z}/2\mathbb{Z}$–graded.[4]

We call a Hamiltonian diffeomorphism a *generalized* $\mathbb{K}$ *pseudorotation* with the sequence k_j if it is generalized perfect with the sequence k_j and, further, $HF^{\mathrm{loc}}(\psi,\mathcal{F})\neq 0$ for all $\mathcal{F}\in\pi_0(\mathrm{Fix}(\psi))$ and the homological count

$$N(\psi,\mathbb{K}) := \sum_{\mathcal{F}\in\pi_0(\mathrm{Fix}(\psi))} \dim_{\mathbb{K}} HF^{\mathrm{loc}}(\psi,\mathcal{F})$$

of generalized fixed points of $\psi=\phi^{k_j}$ satisfies

$$N(\psi,\mathbb{K}) = \dim_{\mathbb{K}} H_*(M;\mathbb{K}) \quad \text{for all } j\in\mathbb{Z}_{>0}.$$

[4]We also define a $\mathbb{Z}$–graded version for a capped generalized 1–periodic point $\bar{\mathscr{F}}$ lifting $\mathcal{F}$.

We recall that usually an $\mathbb{F}_p$ pseudorotation is defined analogously, with the sequence $k_j = p^{j-1}$, and with the additional hypothesis that each $\mathcal{F} \in \pi_0(\mathrm{Fix}(\psi))$ for $\psi = \phi^{k_j}$ consists of a single point. Unless otherwise stated, a *generalized* $\mathbb{F}_p$ *pseudorotation* will be considered with the sequence $k_j = p^{j-1}$.

In view of the discussion in Shelukhin [95; 96], this homological minimality for a Hamiltonian diffeomorphism ψ with a finite number of generalized fixed points is equivalent to the absence of finite bars in the barcode $\mathcal{B}(\psi)$ of ψ, a notion of recent interest in symplectic topology; see eg Kislev and Shelukhin [48], Polterovich and Shelukhin [74], Polterovich, Shelukhin and Stojisavljević [75] and Shelukhin [95; 96]. It also implies the equality

$$\mathrm{Spec}^{\mathrm{ess}}(F;\mathbb{K}) = \mathrm{Spec}^{\mathrm{vis}}(F;\mathbb{K})$$

between two homologically defined subsets of the spectrum associated to a Hamiltonian $F \in \mathcal{H}$ generating ψ. Recall that the spectrum $\mathrm{Spec}(F)$ of F is the set of critical values of the action functional of F. For a coefficient field $\mathbb{K}$, there is a nested sequence of subsets

$$\mathrm{Spec}^{\mathrm{ess}}(F;\mathbb{K}) \subset \mathrm{Spec}^{\mathrm{vis}}(F;\mathbb{K}) \subset \mathrm{Spec}(F).$$

Here the *essential spectrum* $\mathrm{Spec}^{\mathrm{ess}}(F;\mathbb{K})$ is the set of values of all spectral invariants associated to F, in other words the set of starting points of infinite bars in the barcode of F. The *visible spectrum* $\mathrm{Spec}^{\mathrm{vis}}(F;\mathbb{K})$ is the set of action values of capped (generalized) periodic orbits of F that have nonzero local Floer homology, in other words the set of endpoints of all bars in the barcode. It is not hard to modify the definitions of the two homological spectra to include multiplicities, in which case their equality would be equivalent to homological minimality.

The first step in the proof of Theorem F, which is nontrivial and uses Smith theory in filtered Floer homology (cf [95; 98]), is the following reduction.

Theorem G *Let (M,ω) be a closed monotone symplectic manifold. Suppose that $\phi \in \mathrm{Ham}(M,\omega)$ is a Hamiltonian diffeomorphism of prime order $q \geq 2$. Then*:

(i) *For each prime p different from q, the q–torsion diffeomorphism ϕ is a weakly nondegenerate generalized pseudorotation over $\mathbb{F}_p$, with the sequence k_j given by the monotone increasing ordering of the set*

$$\{k \in \mathbb{Z}_{>0} \mid k \neq 0 \ (\mathrm{mod}\ q)\}.$$

(ii) *Moreover, for each Hamiltonian H generating ϕ, and each coefficient field $\mathbb{K}$ of characteristic p coprime to q, we have*

$$\mathrm{Spec}^{\mathrm{ess}}(H;\mathbb{K}) = \mathrm{Spec}^{\mathrm{vis}}(H;\mathbb{K}),$$

and for all k coprime to q, we have

$$\mathrm{Spec}^{\mathrm{ess}}(H^{(k)};\mathbb{Q}) = k\cdot\mathrm{Spec}^{\mathrm{ess}}(H;\mathbb{Q}) + \rho\cdot\mathbb{Z}.$$

(iii) *Finally, part* (i) *holds also for $p = q$, and in part* (ii), *equalities*

$$\mathrm{Spec}^{\mathrm{ess}}(H;\mathbb{K}) = \mathrm{Spec}^{\mathrm{vis}}(H;\mathbb{K}),$$

$$\mathrm{Spec}^{\mathrm{ess}}(H^{(k)};\mathbb{K}) = k\cdot\mathrm{Spec}^{\mathrm{ess}}(H;\mathbb{K}) + \rho\cdot\mathbb{Z}$$

hold with arbitrary coefficient field $\mathbb{K}$, and moreover,

$$\mathrm{Spec}^{\mathrm{vis}}(H;\mathbb{K}) = \mathrm{Spec}(H).$$

The proof of Theorem G appears in Section 5.9. For the moment, we briefly explain the approach used to prove Theorem G(i). Following the main theme of the proof of Theorem C, we use information about large iterations of H to study the periodic Hamiltonian diffeomorphism $\phi = \phi_H^1$ that it generates. More precisely, let $\psi = \phi^k$ with k coprime to q. Combining the theory of barcodes of Hamiltonian diffeomorphisms (see Proposition 23), and Smith-type inequalities in filtered Floer homology (see Theorem N), we observe that for the *bar-lengths*

$$\beta_1(\psi,\mathbb{F}_p) \le \cdots \le \beta_{K(\psi,\mathbb{F}_p)}(\psi,\mathbb{F}_p)$$

of ψ, we have the following inequality. Set

$$\beta_{\mathrm{tot}}(\psi,\mathbb{F}_p) = \beta_1(\psi,\mathbb{F}_p) + \cdots + \beta_{K(\psi,\mathbb{F}_p)}(\psi,\mathbb{F}_p)$$

to be the *total bar-length* of ψ. Then

$$\beta_{\mathrm{tot}}(\psi^{p^m},\mathbb{F}_p) \ge p^m\cdot\beta_{\mathrm{tot}}(\psi,\mathbb{F}_p).$$

However, $\beta_{\mathrm{tot}}(\psi^{p^m},\mathbb{F}_p)$ is bounded, since it can take at most $q-1$ values. This implies

$$\beta_{\mathrm{tot}}(\psi,\mathbb{F}_p) = 0,$$

which in turn implies part (i), by the theory of barcodes; see Proposition 23.

Remark 3 We separate part (iii) of Theorem G because it requires a different proof, relying on Proposition 5 below. The first statement of part (iii) is obtained via Proposition 5 by classical Smith theory combined with the classical homological Arnol'd conjecture, outlined in Chiang and Kessler [11, Remark 7.1] with details for $p = 2$. One could also obtain this statement by a suitable generalization of Theorem N on Smith theory in filtered Floer homology, which is, however, out of the scope of this paper.

Remark 4 When the order q is not prime, a version of Theorem G still holds. We leave its somewhat lengthier formulation to the interested reader, since we do not require it for our arguments, only observing that part (i) holds under the assumption that p does not divide q, and the sequence of iterations is given by $\{k \in \mathbb{Z}_{>0} \mid \gcd(k,q) = 1\}$ and part (ii) holds as stated.

The following statement is a key component of the proof of Theorem G(iii). It relies on the generalization of the Morse–Bott theory of Pozniak [76, Theorem 3.4.11] to the situation with signs and orientations, as in for example Schmaschke [85, Chapter 9], Fukaya, Oh, Ohta and Ono [28, Chapter 8], or Wehrheim and Woodward [105]. However, it is not entirely straightforward, because as classical examples show, it is false in the general Floer–Morse–Bott situation, while in our case it holds because of the existence of special ω–compatible almost complex structures adapted to the situation.

Proposition 5 *Let (M,ω) be a closed symplectic manifold, and $\phi \in \mathrm{Ham}(M,\omega)$ a Hamiltonian diffeomorphism of finite order $d \geq 2$. Let $\mathcal{F}$ be a path-connected component of the fixed-point set of ϕ. Finally, let R be a commutative unital ring. Then the local Floer homology of ϕ at $\mathcal{F}$ with coefficients in R satisfies*

$$HF^{\mathrm{loc}}(\phi,\mathcal{F}) \cong H(\mathcal{F};R).$$

The proof of Theorem G has the following by-product, which is a new analogue, for Hamiltonian torsion, of the classical consequence of Floer theory, whereby the map $\pi_1(\mathrm{Ham}(M,\omega)) \to \pi_1(M)$ is trivial.

Theorem H *Let (M,ω) be a closed monotone symplectic manifold, and let ϕ in $\mathrm{Ham}(M,\omega)$ be a Hamiltonian diffeomorphism of finite order. Then **all** the fixed points of ϕ are contractible.*

The second step in the argument proving Theorem F is the following statement. It essentially follows the arguments of Shelukhin [94] and Shelukhin and Wilkins [97].

Theorem I *Let (M,ω) be a closed monotone symplectic manifold that admits a weakly nondegenerate generalized $\mathbb{F}_p$ pseudorotation for a prime $p \geq 2$. Then the p^{th} quantum Steenrod power of the cohomology class $\mu \in H^{2n}(M;\mathbb{F}_p)$ Poincaré dual to the point class is deformed.*

Theorems G and I immediately imply Theorem F and therefore, by a Gromov compactness argument, Theorem E.

1.2.3 Applications to actions of Lie groups and lattices To conclude the discussion of our first two sets of results, we discuss their implications to the question of existence of Hamiltonian actions of possibly disconnected Lie groups, and lattices in Lie groups, on closed symplectic manifolds.

A well-known result of Delzant [15] (see [73] for an alternative argument) implies that a simple Lie group can only act nontrivially on a closed symplectic manifold if it is compact. A compact zero-dimensional Lie group is finite, whence Theorems B and E provide topological and geometrical obstructions to their actions. The identity component K_0 of a compact Lie group K of positive dimension is a compact connected Lie group of positive dimension, and as such admits a maximal torus $T \cong (S^1)^k$ of positive dimension, whose conjugates cover the whole group K_0. Therefore, the absence of Hamiltonian torsion, as in Theorems A, B, E and F, implies that a nontrivial K–action yields a nontrivial K_0–action, since otherwise it would factor through K/K_0, which is finite. This in turn yields a nontrivial T–action and a fortiori a nontrivial S^1–action. A celebrated result of McDuff [57] then shows that nontrivial S^1–actions imply uniruledness in the sense of k–point genus-zero Gromov–Witten invariants, and hence geometric uniruledness. We therefore obtain the following result.

Corollary 6 *Let (M, ω) be a closed positive monotone symplectic manifold that is not geometrically uniruled, or a negative monotone or symplectically Calabi–Yau symplectic manifold. Then each homomorphism $K \to \mathrm{Ham}(M, \omega)$ for a compact Lie group K must be trivial.*

Moreover, by a simple continuity argument, a nontrivial continuous S^1–action implies a nontrivial $\mathbb{Z}/p\mathbb{Z}$–action for each prime p. Therefore Theorems B and E imply the above corollary for symplectically aspherical, symplectically Calabi–Yau, negative monotone, or monotone symplectic manifolds directly, without relying on the result of McDuff. Moreover, Theorem F also implies that if a positive monotone symplectic manifold admits a nontrivial Hamiltonian S^1–action, it must be $\mathbb{F}_p$–Steenrod uniruled for all primes p. It is seen from examples due to Seidel and Wilkins [92] that there exist closed monotone symplectic manifolds that are uniruled in the sense of Gromov–Witten invariants, and yet not $\mathbb{F}_p$–Steenrod uniruled for certain primes p. More precisely, the monotone blowup M of $\mathbb{CP}^2$ at 6 points is not $\mathbb{F}_2$–Steenrod uniruled, but is evidently uniruled in the Gromov–Witten sense.

The following discussion shows that for a certain nonmonotone 6–point blowup of $\mathbb{CP}^2$ there exists a Hamiltonian involution that cannot be inscribed into an S^1–action.

Note that [92, Example 1.7] and Theorem F imply that the *monotone* blowup M admits no Hamiltonian torsion of order other than 2. It would be interesting to construct a nontrivial Hamiltonian involution of M or prove that it does not exist.

Remark 7 In [12], Chiang and Kessler gave an example of a symplectic involution, ie $\phi \in \mathrm{Symp}(M_0)$ such that $\phi^2 = \mathrm{id}$, of a certain nonmonotone 6–point blowup M_0 of the standard $\mathbb{CP}^2$, with blowup sizes $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{3}{16}, \frac{1}{8}$. This involution belongs to the symplectic Torelli group $\mathrm{Symp}^h(M_0)$ of symplectomorphisms acting trivially on homology, and has the property that it does not belong to any S^1–subgroup of $\mathrm{Ham}(M_0)$. Li, Li and Wu [53] showed in particular that the mapping-class group $\pi_0 \,\mathrm{Symp}^h(M_0)$ is isomorphic to the quotient $G_6 = P_6(S^2)/Z$ of the spherical pure braid group $P_6(S^2)$ on 6 strands by its center $Z \cong \mathbb{Z}/2\mathbb{Z}$. It is well known that G_6 has no torsion; see González-Meneses [37] for a beautiful account of related subjects. This implies that $\phi \in \mathrm{Symp}_0(M_0) = \mathrm{Ham}(M_0)$, showing that ϕ is a Hamiltonian involution that does not belong to any S^1–subgroup.

We note that McDuff's theorem was proven by showing that certain loops of Hamiltonian diffeomorphisms in a blow-up of the manifold are nontrivial, and detectable by Seidel's representation [87]. It would be interesting to investigate the existence of nontrivial Hamiltonian loops associated to Hamiltonian diffeomorphisms of finite order. For a Hamiltonian H generating $\phi \in \mathrm{Ham}(M,\omega)$ of order d, the Hamiltonian $H^{(d)}$ generates a loop homotopic to $\{(\phi_H^t)^d\}$. The noncontractibility of this loop is not obvious since for a rotation $\phi_{2\pi/3}$ of S^2 by angle $2\pi/3$ about the z–axis, the loop $\{\phi^3_{t\cdot 2\pi/3}\}$ is not contractible in $\mathrm{Ham}(S^2,\omega_{\mathrm{st}})$, while the loop $\{\phi^3_{-t\cdot 4\pi/3}\}$ is contractible therein, yet $\phi_{-4\pi/3} = \phi_{2\pi/3}$.

Finally we can argue, following the work of Polterovich [72] on the Hamiltonian Zimmer conjecture, that $\mathrm{SL}(k,\mathbb{Z})$ for $k \geq 2$ has no nontrivial Hamiltonian actions on symplectically aspherical, symplectically Calabi–Yau, negative monotone, or monotone and not geometrically uniruled closed symplectic manifolds. Indeed, it is well known that $\mathrm{SL}(k,\mathbb{Z})$ for $k \geq 2$ is generated by elements of finite order. We remark, however, that the case of finite-index subgroups of $\mathrm{SL}(k,\mathbb{Z})$ with $k \geq 3$ is much more difficult and seems to be currently out of reach of our methods.

1.2.4 Symplectic actions It makes sense to study finite group actions by more general symplectic diffeomorphisms than Hamiltonian ones. In particular, a classical statement in the topology of hyperbolic surfaces is that diffeomorphisms of finite order cannot be isotopic to the identity. Further progress in this direction was made in low-dimensional symplectic topology; see Chen [9], Chen and Kwasik [10] and Wu and Liu [109]. In this section we collect remarks and results in the higher-dimensional setting.

Let us denote by $\mathrm{Symp}(M,\omega)$ the group of diffeomorphisms preserving the symplectic form, and by $\mathrm{Symp}_0(M,\omega)$ its identity component. Of course $\mathrm{Ham}(M,\omega)$ is a subgroup of $\mathrm{Symp}_0(M,\omega)$.

We first make the observation that if $\mathrm{Ham}(M,\omega)$ and $\mathrm{Symp}_0(M,\omega)$ coincide, Hamiltonian no-torsion theorems yield no-torsion theorems for elements of $\mathrm{Symp}_0(M,\omega)$. Let $\Gamma_\omega \subset H^1(M,\mathbb{R})$ be the well-known *flux group*, defined as the image of the map $\mathrm{Flux}\colon \pi_1(\mathrm{Symp}(M,\omega)) \to H^1(M,\mathbb{R})$ given by integrating ω over the two-cycle traced by a loop of symplectomorphisms applied to one-cycles. It is a finitely generated abelian group. The exact sequence

$$1 \to \mathrm{Ham}(M,\omega) \to \mathrm{Symp}_0(M,\omega) \to H^1(M,\mathbb{R})/\,\Gamma_\omega \to 1$$

therefore implies that $\mathrm{Ham}(M,\omega) = \mathrm{Symp}_0(M,\omega)$ if and only if $H^1(M,\mathbb{R}) = 0$. Second, following Polterovich [72, Example 1.3.C], by the same exact sequence we note that whenever $\Gamma_\omega = 0$, all torsion elements in $\mathrm{Symp}_0(M,\omega)$ must in fact be Hamiltonian. By a result of McDuff [56, Theorem 1], this happens for homologically monotone and negative monotone symplectic manifolds, ie when $[\omega] = \kappa \cdot c_1(TM)$ for some $\kappa \neq 0$ as elements of $H^2(M,\mathbb{R})$. By a result of Kędra [49], this also holds for closed symplectically aspherical manifolds (M,ω), ie when $[\omega] = 0$ on $\pi_2(M)$, of nonvanishing Euler characteristic or when the center of $\pi_1(M)$ is trivial; see also Kędra, Kotschick and Morita [50].

We expect that the methods developed in this paper will yield new results on torsion in symplectomorphism groups and plan to investigate this in a further publication.

1.2.5 Metric properties Our third and final set of results studies the metric properties of Hamiltonian torsion diffeomorphisms, in cases that are not ruled out by our previous arguments, for example on $\mathbb{CP}^n$.

Recall that the *spectral pseudonorm* of a Hamiltonian $H \in C^\infty(S^1 \times M,\mathbb{R})$ on a closed symplectic manifold (M,ω) is defined in terms of Hamiltonian spectral invariants as

$$\gamma(H) = c([M],H) + c([M],\overline{H}),$$

and the *spectral norm* of $\phi \in \mathrm{Ham}(M,\omega)$ is set as

$$\gamma(\phi) = \inf_{\phi_H^1=\phi} \gamma(H).$$

We refer to Section 2 for a more in-depth discussion of this interesting notion, remarking for now that this is a conjugation-invariant and nondegenerate norm on $\mathrm{Ham}(M,\omega)$, yielding a bi-invariant metric

$$d_\gamma(f,g)=\gamma(gf^{-1}).$$

This was shown in large generality in Oh [65], Schwarz [86] and Viterbo [104].

Furthermore, whenever defined, $\gamma(\phi)$ provides a lower bound on the celebrated Hofer distance $d_{\mathrm{Hofer}}(\phi,\mathrm{id})$, defined as

$$d_{\mathrm{Hofer}}(\phi,\mathrm{id})=\inf_{\phi_H^1=\phi}\int_0^1 \max_M H(t,-)-\min_M H(t,-)\,dt;$$

see Hofer [40] and Lalonde and McDuff [51]. Finally in Buhovsky, Humilière and Seyfaddini [6], Kawamoto [47] and Shelukhin [96] it was shown, in various degrees of generality, that $\gamma(\phi)$ is bounded by the C^0–distance $d_{C^0}(\phi,\mathrm{id})$ of ϕ to the identity, at least in a small d_{C^0}–neighborhood of the identity.

These and numerous other recent results show that the spectral norm γ is an important measure of a Hamiltonian diffeomorphism. Here, we provide lower bounds on $\gamma(\phi)$, under the assumption that ϕ is of finite order. Our first result is relatively general and quite straightforward, and follows essentially from the homogeneity of the action functional under iteration. However, it underlines the fact that the finite order condition implies certain metric rigidity.

Theorem J *Let (M,ω) be a closed rational symplectic manifold, with rationality constant $\rho>0$, ie $\langle[\omega],\pi_2(M)\rangle=\rho\cdot\mathbb{Z}$. Suppose that $\phi\in\mathrm{Ham}(M,\omega)$ is a nontrivial Hamiltonian diffeomorphism of order d, ie $\phi^d=\mathrm{id}$. Then $\gamma(\phi)\geq\rho/d$.*

As a further consequence of Theorem G, which requires considerably more complex methods, we obtain the following analogue of Newman's theorem for the spectral norm of Hamiltonian torsion elements. This result is the first nontrivial result of its kind in symplectic topology, and is implicitly conjectured in the formulation of [59, Problem 24].

Theorem K *Let (M,ω) be a closed monotone symplectic manifold of rationality constant $\rho>0$. Consider a Hamiltonian diffeomorphism $\phi\in\mathrm{Ham}(M,\omega)$ of order $d>1$. Then there exists $m\in\mathbb{Z}/d\mathbb{Z}$ such that*

$$\gamma(\phi^m)\geq\frac{\rho}{3}.$$

Here the coefficients are in an arbitrary field $\mathbb{K}$. *In fact, if* $d = p$ *is prime, we prove the stronger statement that there exists* $m \in \mathbb{Z}/p\mathbb{Z}$ *such that*

$$\gamma(\phi^m) \geq \frac{\rho \cdot \lfloor p/2 \rfloor}{p}.$$

The key notion in the proof of this result is a new invariant of a Hamiltonian diffeomorphism $\phi \in \mathrm{Ham}(M, \omega)$, which we call the *spectral length* $l(\phi, \mathbb{K})$ of ϕ with coefficients in a field $\mathbb{K}$. It is defined as the minimal diameter of $\mathrm{Spec}^{\mathrm{ess}}(H; \mathbb{K}) \cap I$ over intervals $I = (a - \rho, a] \subset \mathbb{R}$ of length ρ, where H is a Hamiltonian with $\phi_H^1 = \phi$. In particular, we show that this minimum does not depend on the choice of the Hamiltonian H. We show the key property that $l(\phi, \mathbb{K}) \leq \gamma(\phi, \mathbb{K})$ and that, in our case, the spectral length behaves in a controlled way with respect to iterations. By a combinatorial analysis of our situation we consequently deduce Theorem K. We expect $l(\phi, \mathbb{K})$ to have additional applications in quantitative symplectic topology, which we plan to investigate.

Theorem K is generally speaking sharp, as can be seen from the rotation ϕ of S^2 by $2\pi/3$ about the z–axis. In this case $\phi^3 = \mathrm{id}$ and $\gamma(\phi) = \gamma(\phi^2) = \gamma(\phi^{-1}) = \rho/3$, where ρ is the area of the sphere. Observe moreover that the lower bound in Theorem K does not depend on the order of ϕ. In particular if $d = 2$, then Theorem J gives the stronger lower bound $\gamma(\phi) \geq \rho/2$, which is again sharp for the π–rotation of S^2 about the z–axis. We recall that Newman's theorem is a directly analogous assertion, but for the C^0–distance to the identity, in the setting of homeomorphisms of smooth manifolds. In contrast to our result, the constant in Newman's theorem is not explicit.

Finally, we remark that analogous statements hold for generalized $\mathbb{F}_p$ pseudorotations ϕ with sufficiently large admissible sequences. For example, for the sequence $k_j = p^{j-1}$, we get the lower bound $\gamma(\phi^{k_j}) \geq \rho/(p+1)$ for some $j \in \mathbb{Z}_{>0}$, which is saturated by the rotation of S^2 by $2\pi/(p+1)$ about the z–axis. For the sequence $k_j = j$, we obtain the following lower bound, which is saturated by any $2\pi\theta$–rotation on S^2 about the z–axis, where $\theta \notin \mathbb{Q}$.

Theorem L *Let* $\phi \in \mathrm{Ham}(M, \omega)$ *be a generalized* $\mathbb{K}$ *pseudorotation with sequence* $k_j = j$ *on a closed monotone symplectic manifold* (M, ω) *with rationality constant* ρ. *Then*

$$\sup_{j \in \mathbb{Z}_{>0}} \gamma(\phi^{k_j}) \geq \frac{\rho}{2},$$

the coefficients being taken in $\mathbb{K}$.

This result is new in this generality even for strongly nondegenerate pseudorotations. Moreover, Theorem L applies to irrational elements of effective Hamiltonian S^1–actions, and Theorem K applies to rational elements. In particular, by considering the element $\left[\frac{1}{2}\right] \in S^1 = \mathbb{R}/\mathbb{Z}$, we obtain that the Hofer length of such a Hamiltonian S^1–action is at least ρ. In the case of semifree S^1–actions, this lower bound can be deduced from McDuff and Slimowitz [60], where it is also proven that the S^1–action is Hofer length-minimizing among Hamiltonian loops in the same free homotopy class. Our results do not prove such homotopical minimality. However, they do apply in the case where the action is not semifree, where no such results are known. In fact such Hamiltonian loops may well be nullhomotopic; see also Karshon and Pearl [46] for more general shortening results in this case. Finally, we observe that in the special case where (M,ω) is a complex projective space, a similar result to Theorem L can be obtained in a different way by following the methods of Ginzburg and Gürel [35].

2 Preliminary material

2.1 Basic setup

In this section, we recall established aspects of the theory of Hamiltonian diffeomorphisms on symplectic manifolds. Throughout the article, (M,ω) denotes a $2n$–dimensional closed symplectic manifold.

Definition 8 (monotone, negative monotone and symplectically Calabi–Yau) Suppose that the cohomology class of the symplectic form ω is proportional to the first Chern class, ie

$$[\omega] = \kappa \cdot c_1(TM)$$

for some $\kappa \neq 0$, on the image $H_2^S(M;\mathbb{Z})$ of the Hurewicz map $\pi_2(M) \to H_2(M;\mathbb{Z})$. If $\kappa < 0$ we call (M,ω) *negative monotone*, and if $\kappa > 0$ we call it (*positive*) *monotone*. If the first Chern class $c_1(TM)$ vanishes on the image of the Hurewicz map, we say that (M,ω) is *symplectically Calabi–Yau*.

The symplectic manifold (M,ω) is called *rational* whenever $\mathcal{P}_\omega = \langle [\omega], H_2^S(M;\mathbb{Z})\rangle$ is a discrete subgroup of $\mathbb{R}$. If $\mathcal{P}_\omega \neq 0$, then $\mathcal{P}_\omega = \rho\cdot\mathbb{Z}$ for $\rho > 0$, which we call the *rationality constant* of (M,ω). If $\mathcal{P}_\omega = 0$ we call (M,ω) *symplectically aspherical*.[5]

[5] In the literature the additional condition $\langle c_1(TM), H_2^S(M;\mathbb{Z})\rangle = 0$, which we do not require, is often imposed. This condition allows one to introduce a $\mathbb{Z}$–grading on the Floer complex, which we do not require once $\mathcal{P}_\omega = 0$.

Finally we recall that the minimal Chern number of (M,ω) is the index

$$N = N_M = [\mathbb{Z} : I]$$

in $\mathbb{Z}$ of the subgroup $I = \operatorname{im}\bigl(c_1(TM)\colon \pi_2(M) \to \mathbb{Z}\bigr)$. Namely, $[\mathbb{Z} : I] = |\mathbb{Z}/I|$ is the cardinality of the quotient group $\mathbb{Z}/I$.

2.1.1 Hamiltonian isotopies and diffeomorphisms We next consider normalized 1–periodic Hamiltonian functions $H \in \mathcal{H} \subset C^\infty(\mathrm{S}^1 \times M, \mathbb{R})$, where $\mathcal{H}$ is the space of Hamiltonians normalized so that $H(t,-)$ has zero ω^n–mean for all $t \in [0,1]$. For each $H \in \mathcal{H}$ we have the corresponding time-dependent vector field X_H^t defined by the relation $\omega(X_H^t, \cdot) = -dH_t$. In particular, to each Hamiltonian function we can associate a Hamiltonian isotopy $\{\phi_H^t\}$ induced by X_H^t and its time-one map $\phi_H = \phi_H^1$. We omit the H from this notation whenever it is clear from context. Such maps ϕ_H are called Hamiltonian diffeomorphisms and they form a group denoted by $\operatorname{Ham}(M,\omega)$.

For a Hamiltonian diffeomorphism $\phi \in \operatorname{Ham}(M,\omega)$, we denote the set of its *contractible* fixed points by $\operatorname{Fix}(\phi)$. Contractible means the homotopy class $\alpha(x,\phi)$ of the path $\alpha(x,H) = \{\phi_H^t(x)\}$ for a Hamiltonian $H \in \mathcal{H}$ generating ϕ is trivial. This class does not depend on the choice of Hamiltonian, by a classical argument in Floer theory. We write $x^{(k)}$ for the image of $x \in \operatorname{Fix}(\phi)$ under the inclusion $\operatorname{Fix}(\phi) \subset \operatorname{Fix}(\phi^k)$.

We denote by $H^{(k)} \in C^\infty(\mathrm{S}^1 \times M, \mathbb{R})$ the k^{th} iteration of a Hamiltonian function H, given by $H^{(k)}(t,x) = kH(kt,x)$. Note that $\phi_{H^{(k)}} = \phi_H^k$. There is a bijective correspondence between $\operatorname{Fix}(\phi_H)$ and contractible 1–periodic orbits of the isotopy $\{\phi_H^t\}$, thus for $x \in \operatorname{Fix}(\phi_H)$, we denote by $x(t)$ the 1–periodic orbit given by $x(t) = \phi_H^t(x)$ and, similarly, by $x^{(k)}(t)$ the 1–periodic orbit given by $x^{(k)}(t) = \phi_{H^{(k)}}^t(x^{(k)})$.

2.1.2 The Hamiltonian action functional Let $x\colon \mathrm{S}^1 \to M$ be a contractible loop. It is then possible to extend this map to a *capping* of x, namely, a map $\bar{x}\colon \mathrm{D}^2 \to M$ such that $\bar{x}|_{\mathrm{S}^1} = x$. Let $\mathcal{L}_{\mathrm{pt}}M$ denote the space of contractible loops in M and consider the equivalence relation on capped loops given by

$$(x,\bar{x}) \sim (y,\bar{y}) \qquad \Longleftrightarrow \qquad x = y \quad \text{and} \quad \bar{x} \,\#\, (-\bar{y}) \in \ker([\omega]) \cap \ker(c_1),$$

where $\bar{x} \,\#\, (-\bar{y})$ stands for gluing the disks along their boundaries with the orientation of $\bar{y}$ reversed. Here $\ker([\omega])$ and $\ker(c_1)$ denote the kernels of the homomorphisms $H_2^S(M;\mathbb{Z}) \to \mathbb{R}$ induced by $[\omega]$ and $c_1(TM)$. The quotient space $\widetilde{\mathcal{L}}_{\mathrm{pt}}M$ of capped loops by the above equivalence relation is a covering of $\mathcal{L}_{\mathrm{pt}}M$ with the group of deck transformations isomorphic to $\Gamma = H_2^S(M;\mathbb{Z})/\bigl(\ker([\omega]) \cap \ker(c_1)\bigr)$. Note that if (M,ω) is positive or negative monotone or symplectically Calabi–Yau, then $\ker([\omega]) \cap \ker(c_1) = \ker([\omega])$, whence $\Gamma = H_2^S(M;\mathbb{Z})/\ker([\omega])$. Note also that $\Gamma \cong \pi_2(M)/\bigl(\ker([\omega]') \cap \ker(c_1')\bigr)$, where the maps $[\omega]', c_1' \colon \pi_2(M) \to \mathbb{R}$ are the compositions of $[\omega]$ and c_1 with the Hurewicz homomorphism $\pi_2(M) \to H_2^S(M;\mathbb{Z})$.

We write $(x,\bar{x})$, or simply $\bar{x}$, for the equivalence class of the capped loop. With this notation, to each $A \in \Gamma$ we associate the deck transformation sending a capped loop $\bar{x}$ to $\bar{x} \# A$. We define the Hamiltonian action functional $\mathcal{A}_H : \widetilde{\mathcal{L}}_{\mathrm{pt}} M \to \mathbb{R}$ of a 1–periodic Hamiltonian H by

$$\mathcal{A}_H(\bar{x}) = \int_0^1 H(t,x(t))\,dt - \int_{\bar{x}} \omega.$$

Observe that the critical points of the Hamiltonian action functional are exactly $(x,\bar{x})$ for x a contractible 1–periodic orbit satisfying the equation $x'(t) = X_H^t(x(t))$. We denote by $\mathcal{O}(H)$ the set of such orbits, and by $\widetilde{\mathcal{O}}(H)$ the set of critical points of $\mathcal{A}_H$. The *action spectrum* of H is defined as $\mathrm{Spec}(H) = \mathcal{A}_H(\widetilde{\mathcal{O}}(H))$. We remark, following [86], that in the rational case the action spectrum is a closed nowhere-dense subset of $\mathbb{R}$. In addition, if $A \in \Gamma$ then

$$\mathcal{A}_H(\bar{x} \# A) = \mathcal{A}_H(\bar{x}) - \int_A \omega,$$

and for $\bar{x}^{(k)}$, the k^{th} iteration of $\bar{x}$ with the naturally inherited capping, we have

$$\mathcal{A}_{H^{(k)}}(\bar{x}^{(k)}) = k\mathcal{A}_H(\bar{x}).$$

Definition 9 (nondegenerate and weakly nondegenerate orbits) A 1–periodic orbit x of H is called *nondegenerate* if 1 is not an eigenvalue of the linearized time-one map $D(\phi_H^1)_{x(0)}$ at $x(0)$. We call x *weakly nondegenerate* if there exists at least one eigenvalue of $D(\phi_H^1)_{x(0)}$ different from 1. We say that a Hamiltonian H is *nondegenerate* (resp. *weakly nondegenerate*) if all its 1–periodic orbits are nondegenerate (resp. weakly nondegenerate).

The nondegeneracy of an orbit x of H is equivalent to

$$\mathrm{graph}(\phi_H) = \{(x,\phi_H(x)) \mid x \in M\}$$

intersecting the diagonal $\Delta_M \subset M \times M$ transversely at $(x(0),x(0))$. Following [84], for any Hamiltonian H and $\epsilon > 0$, there exists a nondegenerate Hamiltonian H' satisfying $\|H - H'\|_{C^2} < \epsilon$. This fact is key in the definition of filtered Floer homology of degenerate Hamiltonians and for local Floer homology.

2.1.3 Mean-index and the Conley–Zehnder index Following [84; 34], the *mean-index* $\Delta(H,\bar{x})$ of a capped orbit $\bar{x}$ of a possibly degenerate Hamiltonian H is a real number measuring the sum of the angles swept by certain eigenvalues of $\{D(\phi_H^t)_{x(t)}\}$ lying on the unit circle. Here a trivialization induced by the capping is used in order to view $\{D(\phi_H^t)_{x(t)}\}$ as a path in $\mathrm{Sp}(2n,\mathbb{R})$. One can show that for the time-one map $\phi=\phi_H$ generated by the Hamiltonian H, the mean-index depends only on the class $\tilde{\phi}$ of $\{\phi_H^t\}_{t\in[0,1]}$ in the universal cover $\widetilde{\mathrm{Ham}}(M,\omega)$, making the notation $\Delta(\tilde{\phi},\bar{x})$ suitable. In addition, the mean-index depends continuously on $\tilde{\phi}$ in the C^1–topology and on the capped orbit, and it behaves well with iterations,

$$\Delta(\tilde{\phi}^k,\bar{x}^{(k)})=k\cdot\Delta(\tilde{\phi},\bar{x}).$$

Meanwhile, the *Conley–Zehnder index* $\mathrm{CZ}(H,\bar{x})$ of a nondegenerate capped 1–periodic orbit $\bar{x}$ is integer-valued, and roughly measures the winding number of the abovementioned eigenvalues. Once again, the index only depends on $\tilde{\phi}$, so we can also write $\mathrm{CZ}(H,\bar{x})=\mathrm{CZ}(\tilde{\phi},\bar{x})$. We shall use the same normalization as in [34], namely, $\mathrm{CZ}(H,\bar{x})=n$ if x is a nondegenerate maximum of an autonomous Hamiltonian H with small Hessian and $\bar{x}$ is the constant capping. We shall omit the H or $\tilde{\phi}$ in the notation when it is clear from the context. We remark that for an element $A\in\Gamma$,

$$\Delta(\bar{x}\,\#\,A)=\Delta(\bar{x})-2\langle c_1(TM),A\rangle \quad\text{and}\quad \mathrm{CZ}(\bar{x}\,\#\,A)=\mathrm{CZ}(\bar{x})-2\langle c_1(TM),A\rangle.$$

Also, in the case that $\bar{x}$ is nondegenerate, we have

$$|\Delta(\bar{x})-\mathrm{CZ}(\bar{x})|<n. \tag{2}$$

Following [79; 73; 18], we observe that a version of the Conley–Zehnder index can be defined even in the case where the capped orbit is degenerate. It is called the Robbin–Salamon index, and it coincides with the usual Conley–Zehnder index in the nondegenerate case. Furthermore, we note that the mean-index can be equivalently defined by

$$\Delta(\tilde{\phi}_H,\bar{x})=\lim_{k\to\infty}\frac{1}{k}\,\mathrm{CZ}(\tilde{\phi}_H^k,\bar{x}^{(k)}), \tag{3}$$

where we are slightly abusing notation in the sense that CZ here means the Robbin–Salamon index so as to include the degenerate case. The limit in (3) exists, as the Robbin–Salamon index is a quasimorphism $\mathrm{CZ}\colon\widetilde{\mathrm{Sp}}(2n,\mathbb{R})\to\mathbb{R}$; see eg [14] and [18, Section 3.3.4]. In particular, as can be seen directly from its definition in [84], the mean-index is induced by a homogeneous quasimorphism $\Delta\colon\widetilde{\mathrm{Sp}}(2n,\mathbb{R})\to\mathbb{R}$.

Moreover, this map is continuous, and satisfies the additivity property

$$\Delta(\Phi\Psi) = \Delta(\Phi) + \Delta(\Psi)$$

for all $\Phi \in \pi_1(\mathrm{Sp}(2n,\mathbb{R})) \subset \widetilde{\mathrm{Sp}}(2n,\mathbb{R})$ and all $\Psi \in \widetilde{\mathrm{Sp}}(2n,\mathbb{R})$.

2.2 Floer theory

Floer theory was first introduced by A Floer [21; 22; 23] as a generalization of the Morse–Novikov homology for the Hamiltonian action functional defined above. We refer to [67] for details on the constructions described in this subsection, and to [1; 88; 110] for a discussion of canonical orientations.

2.2.1 Filtered and total Floer homology We review the construction of *filtered Floer homology* in order to recall some basic properties and set notation.

Let H be a nondegenerate 1–periodic Hamiltonian on a rational symplectic manifold (M,ω) and $\mathbb{K}$ a fixed base field. For $a \in \mathbb{R}\backslash\mathrm{Spec}(H)$ and $\{J_t \in \mathcal{J}(M,\omega)\}_{t\in \mathrm{S}^1}$ a generic loop of ω–compatible almost complex structures, set

$$CF_k(H;J)^{<a} = \Big\{\sum \lambda_{\bar{x}} \cdot \bar{x} \;\Big|\; \bar{x} \in \widetilde{\mathcal{O}}(H),\, \mathrm{CZ}(\bar{x}) = k,\, \lambda_{\bar{x}} \in \mathbb{K},\, \mathcal{A}_H(\bar{x}) < a\Big\},$$

where $\#\{\lambda_{\bar{x}} \neq 0 \mid \mathcal{A}_H(\bar{x}) > c\} < \infty$ for every $c \in \mathbb{R}$. Intuitively, it is the vector space over $\mathbb{K}$ generated by the critical points of the Hamiltonian action functional of filtration level $< a$. The graded $\mathbb{K}$–vector space $CF_*(H,J)^{<a}$ is endowed with the Floer differential $d_{H;J}$, which is defined as the signed count of isolated solutions (quotiented out by the $\mathbb{R}$–action) of the asymptotic boundary value problem on maps $u\colon \mathbb{R}\times \mathrm{S}^1 \to M$ defined by the negative gradient of $\mathcal{A}_H$; see [83; 84]. In other words, the boundary operator counts the finite-energy solutions to the Floer equation,

$$\frac{\partial u}{\partial s} + J_t(u)\Big(\frac{\partial u}{\partial t} - X_H^t(u)\Big) = 0 \quad \text{such that } E(u) = \int_{\mathbb{R}}\int_{\mathrm{S}^1}\Big\|\frac{\partial u}{\partial s}\Big\|^2 dt\, ds < \infty,$$

which converge as s tends to $\pm\infty$ to periodic orbits x_- and x_+ such that the capping $\bar{x}_- \# u$ is equivalent to $\bar{x}_+$ and $\mathrm{CZ}(\bar{x}_-) - \mathrm{CZ}(\bar{x}_+) = 1$. In this case the Floer trajectory u satisfies $E(u) = \mathcal{A}_H(\bar{x}_-) - \mathcal{A}_H(\bar{x}_+)$. We thus obtain the *filtered Floer chain complex* $(CF_*(H;J)^{<a}, d_{H;J})$, which is a subcomplex of the *total Floer chain complex* (corresponding to $a = +\infty$). Furthermore, for an interval $I = (a,b)$ with $a < b$ and $a,b \in (\mathbb{R}\setminus \mathrm{Spec}(H)) \cup \{\pm\infty\}$, we define the Floer complex in the action window I as the quotient complex

$$CF_*(H;J)^I = CF_*(H;J)^{<b}/CF_*(H;J)^{<a},$$

where $CF_*(H;J)^{<-\infty}=0$. The resulting homology of this complex $HF_*(H)^I$ is the *Floer homology* of H *in the action window* I and it is independent of the generic choice of almost complex structure J. So the (total) Floer homology of H can be obtained by setting $a=-\infty$ and $b=+\infty$. We note that in the positive and negative monotone case $CF_*(H;J)$ is naturally a module over the Novikov field $\Lambda_{\mathbb{K}}=\mathbb{K}[\![q^{-1},q]$ with q a variable of degree $2N$. Indeed we define $q^{-1}\cdot\bar{x}=\bar{x}\,\#\,A_0$ for A_0 the generator of Γ with $\langle c_1(TM),A_0\rangle=N$, and extend it to a module structure in the natural way. In the Calabi–Yau case, $CF_*(H;J)$ is a module over the Novikov field

$$\Lambda_{\mathbb{K},\omega}=\Big\{\sum a_iT^{\lambda_i}\;\Big|\;a_i\in\mathbb{K},\,\lambda_i\in\mathcal{P}_\omega,\,\lambda_i\to\infty\Big\}.$$

While we shall not use it in the paper, we remark that in the general case, it is a module over the Novikov field

$$\Lambda_{\mathbb{K},\omega,c_1}=\Big\{\sum a_iT^{A_i}\;\Big|\;a_i\in\mathbb{K},\,A_i\in\Gamma,\,\omega(A_i)\to\infty\Big\}.$$

Observe that by interpolating between distinct Hamiltonians through generic families and writing the Floer continuation map, where the Hamiltonian perturbation term and the almost complex structure depend on the $\mathbb{R}$–coordinate, one can show that $HF_*(H)$ does not depend on the Hamiltonian, and $HF_*(H)^I$ depends only on the homotopy class of $\{\phi_H^t\}_{t\in[0,1]}$ in the universal cover $\widetilde{\mathrm{Ham}}(M,\omega)$ of the Hamiltonian group $\mathrm{Ham}(M,\omega)$. Also, when M is rational the above construction extends by a standard continuity argument to degenerate Hamiltonians.

Remark 10 Theorems B, D and J partially deal with negative monotone or general spherically rational symplectic manifolds. It is important to emphasize that for our arguments to apply to this case in full generality, we must make use of the machinery of virtual cycles to guarantee that the Floer differential is well defined. In this case, the ground field $\mathbb{K}$ should be of characteristic zero. Our arguments are not sensitive to the specific approach to questions of transversality. We refer to [33; 55; 28; 80] for early works on the subject, subsequently augmented in [32, Chapters 15–20; 31, Section 9; 30, Section 8; 29, Section 19]. We refer to [32, Chapter 1.4] for an overview of other approaches to virtual fundamental cycles. This includes the theory of polyfolds initiated in [42; 43; 44]; see [20] for a recent survey. We also note that [69] provides foundations of Hamiltonian Floer theory in full generality. Furthermore, we mention the following cases where classical transversality techniques are applicable. First, if (M,ω) is a semipositive[6] symplectic manifold — that is, if (M,ω) is symplectically

[6]The terminology "weakly monotone" also appears in the literature for the same notion.

aspherical, symplectically Calabi–Yau, positive monotone, or if the minimal Chern number of (M,ω) is $N \geq n-2$ — then classical transversality applies by [41]. Second, if the manifold is homologically rational, ie the symplectic form can be scaled so that all of its periods are integers, then classical transversality applies by [8] following [13].

2.2.2 The irrational case In this paper we also consider the case in which the manifold M is symplectically Calabi–Yau, which includes the possibility of it being irrational. In this case we have to work a little harder if H is degenerate since the continuation argument above does not work as before, as nonspectral a,b for H do not necessarily remain nonspectral even for arbitrarily small perturbations H_1 of H. Moreover, the resulting homology groups depend on the choice of nondegenerate perturbation H_1. We shall follow [39] to work around this issue.

For a fixed Hamiltonian H and action window $I=(a,b)$ with $a<b$, where $a,b \in (\mathbb{R} \setminus \mathrm{Spec}(H)) \cup \{\pm\infty\}$, consider the set of nondegenerate perturbations $\widetilde{H}$ whose action spectra do not include a and b and $H \leq \widetilde{H}$, ie $H(t,x) \leq \widetilde{H}(t,x)$ for all $x \in M$ and $t \in \mathrm{S}^1$. Note that such perturbations $\widetilde{H}$ of a mean-normalized H will in general not be mean-normalized. However, this does not present an issue for our purposes. Observe that $\leq$ induces a partial order in the set of perturbations. In addition, by considering a monotone decreasing homotopy $\widetilde{H}^s$ from $\widetilde{H}^0$ to $\widetilde{H}^1$, one obtains an induced homomorphism between the Floer homology groups. These give rise to continuation maps $HF_*(H')^I \to HF_*(H'')^I$ whenever $H'' \leq H'$. Therefore, we can define the filtered Floer homology of H by taking the direct limit

$$HF_*(H)^I = \varinjlim HF_*(H')^I$$

over the homology groups of the perturbations satisfying the aforementioned conditions. We remark that in the case where H is nondegenerate or M is rational, this definition coincides with the usual filtered Floer homology groups.

2.2.3 Local Floer homology In this section we shall follow [34] in order to briefly review the construction of the *local Floer homology* of a Hamiltonian H at a capping $\bar{x}$ of an isolated 1–periodic orbit x.

Since $\bar{x}$ is isolated we can find an isolating neighborhood U of x in the extended phase-space $\mathrm{S}^1 \times M$ whose closure does not intersect the image $\{(t,y(t))\}_{t\in[0,1]}$ of any other orbit y of H. For a sufficiently C^2–small nondegenerate perturbation H' of H, the orbit x splits into finitely many 1–periodic orbits $\mathcal{O}(H',x)$ of H', which are contained in U and whose cappings are inherited from $\bar{x}$. We denote by $\mathcal{O}(H',\bar{x})$

the capped 1–periodic orbits $\overline{x}$ splits into. Moreover, we can also guarantee that any Floer trajectory and any broken trajectory between capped orbits in $\mathcal{O}(H', \overline{x})$ are contained in U. For a base field $\mathbb{K}$ we consider the vector space $CF_*(H, \overline{x})$ generated by $\mathcal{O}(H', \overline{x})$, which by the above observation naturally inherits a Floer differential and a grading by the Conley–Zehnder index. The homology of this chain complex is independent of the choice of the perturbation H' once it is close enough to H, and it is called the local Floer homology of H at $\overline{x}$; it is denoted by $HF_*^{\mathrm{loc}}(H, \overline{x})$. This group depends only on the class $\widetilde{\phi}$ of $\{\phi_H^t\}$ in the universal cover $\widetilde{\mathrm{Ham}}(M, \omega)$, and the capped orbit $\overline{x}$. Namely, homotopic paths have choices of cappings of orbits corresponding to a fixed point $x \in \mathrm{Fix}(\phi)$ in bijection, and the corresponding groups are canonically isomorphic. Hence we write $HF_*^{\mathrm{loc}}(H, \overline{x}) = HF_*^{\mathrm{loc}}(\widetilde{\phi}, \overline{x})$. If we ignore the $\mathbb{Z}$–grading, then the group depends only on $\phi = \phi_H^1$ and $x \in \mathrm{Fix}(\phi)$. In this case, we write $HF^{\mathrm{loc}}(\phi, x)$ for the corresponding local homology group, which is naturally only $\mathbb{Z}/(2)$–graded.

Let $\overline{x}$ be a capped 1–periodic orbit of a Hamiltonian H. We define the *support* of $HF_*^{\mathrm{loc}}(H, \overline{x})$ to be the collection of integers k such that $HF_k^{\mathrm{loc}}(H, \overline{x}) \neq 0$. By the continuity of the mean-index and by equation (2), it follows that $HF_*^{\mathrm{loc}}(H, \overline{x})$ is supported in the interval $[\Delta(\overline{x}) - n, \Delta(\overline{x}) + n]$. One can show that if x is weakly nondegenerate then $HF_*^{\mathrm{loc}}(H, \overline{x})$ is actually supported in $(\Delta(\overline{x}) - n, \Delta(\overline{x}) + n)$. We shall explore the idea behind the proof of this second fact later as we use the same argument to prove a similar claim in slightly greater generality. Namely, we extend it to an isolated compact path-connected family of contractible fixed points.

2.3 Quantum homology and PSS isomorphism

In the present section we describe the quantum homology of a symplectic manifold. It might be helpful to think of it as the Hamiltonian Floer homology in the case where the Hamiltonian is a C^2–small time-independent Morse function. Alternatively, one may consider it as the cascade approach [26] to Morse homology for the unperturbed symplectic area functional on the space $\widetilde{\mathcal{L}}_{\mathrm{pt}} M$. For a more detailed exposition of these subjects we refer to [52; 67; 89].

2.3.1 Quantum homology Fix a ground field $\mathbb{K}$. Consider the Novikov field $\Lambda = \Lambda_{\mathbb{K}} = \mathbb{K}[\![q^{-1}, q]$ of (M, ω) in the positive and negative monotone cases, where $\deg(q) = 2N$ and $\Lambda = \Lambda_{\mathbb{K},\omega}$ in the Calabi–Yau case. We set

$$QH(M) = QH(M, \mathbb{K}) = H_*(M; \Lambda)$$

as a Λ–module. This module has the structure of a graded-commutative unital algebra over $\Lambda_{\mathbb{K}}$ whose product, denoted by $*$, is defined in terms of 3–point genus-zero Gromov–Witten invariants [54; 58; 81; 82; 108]. It can be thought of as a deformation of the usual intersection product on homology. As in the classical homology algebra, the unit for this quantum product is the fundamental class $[M]$ of M.

2.3.2 Piunikhin–Salamon–Schwarz isomorphism Under our conventions for the Conley–Zehnder index, one obtains a map

$$\mathrm{PSS}\colon QH_*(M) \to HF_{*-n}(H)$$

by counting (for generic auxiliary data) certain isolated configurations. More precisely, the configurations considered consist of negative gradient trajectories $\gamma\colon(-\infty,0]\to M$ of a generic Morse–Smale pair[7] incident at $\gamma(0)$ with the asymptotic $\lim_{s\to-\infty} u(s,\cdot)$ of a map $u\colon \mathbb{R}\times \mathrm{S}^1 \to M$ of finite energy, satisfying the Floer equation

$$\frac{\partial u}{\partial s} + J_t(u)\left(\frac{\partial u}{\partial t} - X_K^t(u)\right) = 0$$

for $(s,t)\in\mathbb{R}\times\mathrm{S}^1$ and $K(s,t)\in C^\infty(M,\mathbb{R})$ a small perturbation of $\beta(s)H_t$ such that $K(s,t)=\beta(s)H_t$ for $s\ll -1$ and for $s\gg +1$. Here $\beta\colon\mathbb{R}\to[0,1]$ is a smooth function satisfying $\beta(s)=0$ for $s\ll -1$ and $\beta(s)=1$ for $s\gg +1$. This map produces an isomorphism of $\Lambda_{\mathbb{K}}$–modules, which intertwines the quantum product on $QH(M)$ with the pair-of-pants product on Hamiltonian Floer homology. This map is called the *Piunikhin–Salamon–Schwarz isomorphism.*

2.4 Spectral invariants in Floer theory

We review the theory of spectral invariants following the works of [73; 34; 67], which contain a more exhaustive list of properties and finer details of the construction.

Let (M,ω) be a closed symplectic manifold, H a generic Hamiltonian and $\{J_t\}_{t\in\mathrm{S}^1}$ a loop of ω–compatible almost complex structures. For $a\in\mathbb{R}\setminus\mathrm{Spec}(H)$, the inclusion of the filtered Floer complex into the total complex induces a homomorphism

$$i_a\colon HF(H)^{<a}\to HF(H).$$

[7]That is, a Morse function $f\in C^\infty(M,\mathbb{R})$ and Riemannian metric g on M, satisfying the Morse–Smale condition.

For each $\alpha_M \in QH_*(M) \setminus \{0\}$, using the PSS isomorphism $QH_*(M) \cong HF_{*-n}(H)$ we then define

$$c(\alpha_M, H) = \inf\{a \in \mathbb{R} \mid \mathrm{PSS}(\alpha_M) \in \mathrm{im}(i_a)\}.$$

It is not hard to see that the spectral invariants do not depend on the choice of an almost complex structure. In addition, for $H \in \mathcal{H}$ the spectral invariant $c(\alpha_M, H)$ depends only on the class $\widetilde{\phi}_H$ of $\{\phi_H^t\}$ in the universal cover $\widetilde{\mathrm{Ham}}(M,\omega)$; consequently, we also denote it by $c(\alpha_M, \widetilde{\phi}_H) = c(\alpha_M, H)$.

Definition 11 (non-Archimedean valuation) Let Λ be a field. A non-Archimedean valuation on Λ is a function $\nu\colon \Lambda \to \mathbb{R} \cup \{+\infty\}$ such that

(i) $\nu(x) = +\infty$ if and only if $x = 0$,

(ii) $\nu(xy) = \nu(x) + \nu(y)$ for all $x, y \in \Lambda$,

(iii) $\nu(x+y) \geq \min\{\nu(x), \nu(y)\}$ for all $x, y \in \Lambda$.

The Novikov field $\Lambda_{\mathbb{K}} = \mathbb{K}[\![q^{-1}, q]$ possesses a non-Archimedean valuation

$$\nu\colon \Lambda_{\mathbb{K}} \to \mathbb{R} \cup \{+\infty\}$$

given by setting $\nu(0) = +\infty$ and

$$\nu\Big(\sum a_j q^j\Big) = -\max\{j \mid a_j \neq 0\}. \tag{4}$$

Spectral invariants enjoy a wealth of useful properties, established by Schwarz [86], Viterbo [104], Oh [58; 64; 66] and generalized by Usher [101; 102], all of which hold for closed rational symplectic manifolds, using the machinery of virtual cycles as discussed in Remark 10 if necessary. We summarize some of the relevant properties for our purposes:

(i) **Spectrality** For each $\alpha_M \in QH(M) \setminus \{0\}$ and $H \in \mathcal{H}$,

$$c(\alpha_M, \widetilde{\phi}_H) \in \mathrm{Spec}(H).$$

(ii) **Stability** For any $H, G \in \mathcal{H}$ and $\alpha_M \in QH(M) \setminus \{0\}$,

$$\int_0^1 \min_M (H_t - G_t)\, dt \leq c(\alpha_M, \widetilde{\phi}_H) - c(\alpha_M, \widetilde{\phi}_G) \leq \int_0^1 \max_M (H_t - G_t)\, dt.$$

(iii) **Triangle inequality** For any $H, G \in \mathcal{H}$ and $\alpha_M, \alpha'_M \in QH(M) \setminus \{0\}$,

$$c(\alpha_M * \alpha'_M, \widetilde{\phi}_H \widetilde{\phi}_G) \leq c(\alpha_M, \widetilde{\phi}_H) + c(\alpha'_M, \widetilde{\phi}_G).$$

(iv) **Value at identity** For every $\alpha_M \in QH(M) \setminus \{0\}$,

$$c(\alpha_M, \widetilde{\mathrm{id}}) = -\rho \cdot \nu(\alpha_M),$$

where ρ is the rationality constant of (M, ω) and ν is as in (4).

(v) **Novikov action** For all $H \in \mathcal{H}$, $\alpha_M \in QH(M)$ and $\lambda \in \Lambda_{\mathbb{K}}$,

$$c(\lambda\alpha_M, H) = c(\alpha_M, H) - \rho \cdot \nu(\lambda).$$

(vi) **Non-Archimedean property** For all $\alpha_M, \alpha'_M \in QH(M)$,

$$c(\alpha_M + \alpha'_M, H) \leq \max\{c(\alpha_M, H), c(\alpha'_M, H)\}.$$

By the continuity property, the spectral invariants are defined for all $H \in \mathcal{H}$, and all the above listed properties apply in this generality. Further, we observe that for $\alpha_M \in QH(M)$ satisfying $\alpha_M * \alpha_M = \alpha_M$, the triangle inequality for the spectral invariants implies

$$c(\alpha_M, \widetilde{\phi}_{H^{(k)}}) = c(\alpha_M, \widetilde{\phi}_H^k) \leq k \cdot c(\alpha_M, \widetilde{\phi}_H).$$

2.4.1 Spectral norm For a Hamiltonian $H \in \mathcal{H}$, we define its spectral pseudonorm by

$$\gamma(H) = c([M], \widetilde{\phi}_H) + c([M], \widetilde{\phi}_{\overline{H}}), \tag{5}$$

where $\overline{H}$ is the Hamiltonian function $\overline{H}(t, x) = -H(1-t, x)$. A result of [65; 86; 104] shows that

$$\gamma(\phi) = \inf_{\phi_H^1 = \phi} \gamma(H)$$

defines a nondegenerate norm $\gamma : \mathrm{Ham}(M, \omega) \to \mathbb{R}_{\geq 0}$ and yields a bi-invariant distance $\gamma(\phi, \phi') = \gamma(\phi'\phi^{-1})$. We call $\gamma(\phi)$ the *spectral norm* of ϕ and $\gamma(\phi, \phi')$ the *spectral distance* between ϕ and ϕ'.

2.4.2 Carrier of the spectral invariant In this section we review the definition of *carriers of the spectral invariant*, mainly following [34]. We observe that while we are going to introduce the notion of carriers specifically for the fundamental class $[M] \in QH(M)$, it can be done so for any nontrivial quantum homology class μ.

First, we fix $\alpha_M = [M]$ and write $c(H) = c(\widetilde{\phi}_H) = c([M], \widetilde{\phi}_H)$. Observe that in the case of a nondegenerate Hamiltonian H, we have

$$c(\widetilde{\phi}_H) = \inf\{\mathcal{A}_H(\sigma) \mid \sigma \in CF_n(H),\ \mathrm{PSS}([M]) = [\sigma]\},$$

where $\mathcal{A}_H(\sigma) = \max\{\mathcal{A}_H(\overline{x}) \mid \lambda_{\overline{x}} \neq 0\}$ for $\sigma = \sum \lambda_{\overline{x}} \overline{x}$. That is, it is the maximum action of a capped orbit $\overline{x}$ entering $\sigma \in CF_n(H)$. By the spectrality property of spectral invariants, the infimum is obtained. Consequently, there exists a cycle σ satisfying $[\sigma] = \mathrm{PSS}([M])$ such that $\mathcal{A}_H(\bar{x}) = c(\widetilde{\phi}_H)$ for an orbit $\bar{x}$ entering σ. We call $\bar{x}$ the *carrier of the spectral invariant* and observe that in order to guarantee its uniqueness, all the 1–periodic orbits of H need to have distinct action values. In order to generalize the notion of carriers to the case where H is degenerate and has isolated orbits, we first recall that for each C^2–small nondegenerate perturbation H', every capped 1–periodic orbit $\bar{x}$ splits into several nondegenerate 1–periodic orbits $\mathcal{O}(H', \bar{x})$, with their capping inherited from $\bar{x}$.

Definition 12 (carrier of degenerate H with isolated orbits) A capped 1–periodic orbit $\bar{x}$ is said to be a carrier of the spectral invariant if there exists a sequence $\{H'_k\}$ of nondegenerate perturbations C^2–converging to H such that for each k, one of the orbits in $\mathcal{O}(H'_k, \bar{x})$ is a carrier of the spectral invariant for H'_k. A uncapped orbit is said to be a carrier if it becomes one for a suitable capping.

As in the nondegenerate case, the uniqueness of the carrier follows from all the 1–periodic orbits having distinct action values. In this case, the carrier becomes independent of the choice of sequence $\{H'_k\}$. In addition, the definition of a carrier and the continuity of the action functional and of the mean-index readily yield

$$c(\widetilde{\phi}_H) = \mathcal{A}_H(\bar{x}) \quad \text{and} \quad 0 \le \Delta(\widetilde{\phi}_H, \bar{x}) \le 2n,$$

where the inequalities can be made strict in the case where the orbit x is weakly nondegenerate. In [34] the following result was obtained.

Lemma 13 *Suppose H only has isolated 1–periodic orbits, and let $\bar{x}$ be a carrier of the spectral invariant of the fundamental class. Then $HF_n^{\mathrm{loc}}(H, \bar{x}) \neq 0$.*

In Section 3.1.2 below, we generalize this statement to the case of isolated path-connected sets of periodic orbits, and also to arbitrary quantum homology classes.

3 Isolated connected sets of periodic points

3.1 Generalized perfect Hamiltonians

Recall that a Hamiltonian H is called perfect if it has a finite number of contractible periodic points of all periods. We consider the more general condition where H has finitely many isolated path-connected families of periodic orbits, which in turn implies that $\mathrm{Fix}(\phi_H)$ is composed of finitely many isolated path-connected sets.

Definition 14 A Hamiltonian diffeomorphism $\phi \in \mathrm{Ham}(M,\omega)$ is *generalized perfect* whenever the following conditions are met:

(i) $\mathrm{Fix}(\phi)$ has finitely many isolated compact path-connected components.

(ii) There exists a sequence of integers $k_i \to \infty$ which contains a subsequence $l_i = k_{j_i}$ with $l_i \mid l_{i+1}$ for all i, and for which $\mathrm{Fix}(\phi^{k_i}) = \mathrm{Fix}(\phi)$ for all i.

(iii) For each isolated path-connected component $\mathcal{F}$ of $\mathrm{Fix}(\phi)$, and for all i, the mean-index $\Delta(H^{(k_i)}, \bar{x}^{(k_i)})$, where $x \in \mathcal{F}$ and $\bar{x} \in \bar{\mathscr{F}}$, is a constant function of $x \in \mathcal{F}$. We denote this constant by $\Delta(H^{(k_i)}, \bar{\mathscr{F}}^{(k_i)})$.

An isolated path-connected component $\mathcal{F} \subset \mathrm{Fix}(\phi)$ can be thought of as, and is indeed called in this paper, a *generalized fixed point*. In this section we slightly generalize some of the theory discussed in Section 2, allowing us to treat generalized perfect Hamiltonians. We observe that the third condition in Definition 14 is not vacuous: indeed, one can construct an example of a generalized fixed point $\mathcal{F}$ where the mean-index is not a constant function of $x \in \mathcal{F}$, by means of the Hamiltonian suspension construction [71, Section 3.1] applied to an appropriate contractible Hamiltonian loop of S^2. However, as stated in Theorem C, a p–torsion Hamiltonian diffeomorphism is weakly nondegenerate generalized perfect: in particular, the mean-index is constant on each generalized fixed point.

3.1.1 Lifts of generalized 1–periodic orbits Let (M,ω) be a closed symplectic manifold and H a Hamiltonian function generating a Hamiltonian diffeomorphism ϕ_H on M whose set of contractible fixed points consists of a finite number of path-connected components. Denote the path-connected components of $\mathrm{Fix}(\phi_H)$ by $\mathcal{F}_1, \dots, \mathcal{F}_m$. For each j and each $x \in \mathcal{F}_j$, there is a corresponding contractible loop $x(t) = \phi_H^t(x)$, thus to each isolated fixed-point set $\mathcal{F}_j$ we can associate a subset $\mathscr{F}_j$ of the space $\mathcal{L}_{\mathrm{pt}}M$ of all contractible loops in M. It is natural to ask whether the generalized orbits $\mathscr{F}_j$ lift to the Γ–cover $\widetilde{\mathcal{L}}_{\mathrm{pt}}M$ in a suitable manner, namely, if the preimage under the projection $\mathrm{Pr}\colon \widetilde{\mathcal{L}}_{\mathrm{pt}}M \to \mathcal{L}_{\mathrm{pt}}M$ is composed of isolated path-connected "copies" of $\mathscr{F}_j$. We show that the lift exists, and we denote by $\bar{\mathscr{F}}_j$ a particular lift of $\mathscr{F}_j$. This is analogous to a capping of an orbit in the case of a usual Hamiltonian.

Consider the set $\mathscr{F} \subset \mathcal{L}_{\mathrm{pt}}M$ associated to $\mathcal{F} \in \pi_0(\mathrm{Fix}(\phi_H))$ and let $i : \mathscr{F} \to \mathcal{L}_{\mathrm{pt}}M$ be the natural inclusion map. Formally, we are asking when, given a loop $x_0 \in \mathscr{F}$ and $\bar{x}_0 \in \mathrm{Pr}^{-1}(\{x_0\})$, a lift of i exists: namely, a unique map $f : \mathscr{F} \to \widetilde{\mathcal{L}}_{\mathrm{pt}}M$ such that $f(x_0) = \bar{x}_0$ and $\mathrm{Pr} \circ f = i$. From the theory of covering spaces, the existence of the lift is equivalent to $i_*(\pi_1(\mathscr{F}, x_0)) \subset \mathrm{Pr}_*(\pi_1(\widetilde{\mathcal{L}}_{\mathrm{pt}}M, \bar{x}_0))$.

Proposition 15 *Let (M,ω) be a symplectic manifold in one of the three classes considered in this paper, and ϕ_H a generalized perfect Hamiltonian diffeomorphism. Then each generalized orbit $\mathscr{F}$ can be lifted to $\overline{\mathscr{F}}$ in a unique manner specified by a loop $x_0 \in \mathscr{F}$ and an element in its fiber $\bar{x}_0 \in \mathrm{Pr}^{-1}(x_0)$.*

Proof Let γ be a loop in $\mathscr{F}$ such that $\gamma_0 = x_0$. We show that we can find a loop $\tilde{\gamma}$ in $\widetilde{\mathcal{L}}_{\mathrm{pt}}M$ such that $i \circ \gamma = \mathrm{Pr} \circ \tilde{\gamma}$, which implies the claim of the theorem. We build $\tilde{\gamma}$ in a natural way by defining the capping at γ_s to be given by gluing the "cylinder" given by traversing the loop γ from 0 to s to the capping $\bar{x}_0$. To see that the capped orbits $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ are equivalent in $\widetilde{\mathcal{L}}_{\mathrm{pt}}M$, we show that

$$\int_{\mathrm{T}^2} \gamma^*\omega = 0 \tag{6}$$

for every loop γ in $\mathscr{F}$. We can then guarantee the existence of a lift. Equation (6) follows from the continuity of $\mathcal{A}_H$ and the fact that $\mathrm{Spec}(H)$ has zero measure in $\mathbb{R}$. Indeed, $\mathcal{A}_H(\tilde{\gamma}_s) = \mathcal{A}_H(\tilde{\gamma}_0)$ for every s, otherwise, the fact that $\tilde{\gamma}_s$ is a critical point for each s would imply that $\mathcal{A}_H\big(\bigcup_{0\le t\le s}\tilde{\gamma}_t\big)$ is a positive measure subset of $\mathrm{Spec}(H)$. Finally, $\mathcal{A}_H(\tilde{\gamma}_1) = \mathcal{A}_H(\tilde{\gamma}_0)$ amounts to fulfilling the sufficient condition given by equation (6). Therefore for the three classes we consider, the proof is complete since in this case $\Gamma \cong \pi_2(M)/\ker([\omega])$ and hence it is only necessary to verify (6). Alternatively, one can prove that $\langle [T^2], \gamma^*(c_1)\rangle = 0$ directly, by replacing γ with a map $\gamma_1 : S^2 \to M$ with $\langle [T^2], \gamma^*(c_1)\rangle = \langle [S^2], \gamma_1^*(c_1)\rangle$, which vanishes by our assumption on the manifold and (6). $\square$

3.1.2 Generalized local Floer homology In this section, we define a version of local Floer homology for a generalized capped orbit $\overline{\mathscr{F}} \subset \widetilde{\mathcal{L}}_{\mathrm{pt}}M$ of a 1–periodic Hamiltonian H in a way that is closely related to what was done in [61; 34]. The only differences are that we are beyond the symplectically aspherical case and we are dealing with path-connected components of $\mathrm{Fix}(\phi_H)$ instead of isolated points. The proofs of [61] used to define the local Floer homology are valid in this case with nearly no modifications. The notion of local Floer homology in a more general setting goes back to the original work of Floer [24; 23] and has been revisited a number of times, for example in the work of Pozniak [76]. The main ingredients of the construction are as follows.

For each $\mathcal{F} \in \pi_0(\mathrm{Fix}(\phi_H))$, we can find an isolating neighborhood $U_{\mathcal{F}}$ of the corresponding generalized 1–periodic orbit $\mathscr{F}$ in the extended phase-space $\mathrm{S}^1 \times M$, ie

$$\{(t, \phi_H^t(x)) \mid t \in \mathrm{S}^1,\ x \in \mathcal{F}\} \subset U_{\mathcal{F}}.$$

Moreover, we can choose this collection of neighborhoods to be pairwise disjoint: $U_{\mathcal{F}}$ is disjoint from $U_{\mathcal{F}'}$ for each pair of distinct generalized fixed points $\mathcal{F}$ and $\mathcal{F}'$. Such an open set $U_{\mathcal{F}}$ in the extended phase-space can be constructed, using the isotopy ϕ_H^t, from an open neighborhood of $\mathcal{F}$ in M. Hence by a slight abuse of notation we think of $U_{\mathcal{F}}$ as a neighborhood of $\mathcal{F}$ in M.

Now there exists an $\epsilon > 0$ small enough that for any nondegenerate Hamiltonian perturbation H' satisfying $\|H - H'\|_{C^2} < \epsilon$, the set of orbits $\mathcal{O}(H', \mathscr{F})$ which $\mathscr{F}$ splits into is contained in $U_{\mathcal{F}}$, and so is every (broken) Floer trajectory connecting any such two orbits; see Lemma 21. We can now consider the complex $CF_*(H', \bar{\mathscr{F}})$ over a ground field $\mathbb{K}$ generated by the capped 1–periodic orbits $\mathcal{O}(H', \bar{\mathscr{F}})$ which $\bar{\mathscr{F}}$ splits into, where the cappings are naturally produced from the specific lift $\bar{\mathscr{F}}$. One can see that this complex is graded by the Conley–Zehnder index and has a well-defined differential. By a standard continuation argument, one can show that the homology of this complex is independent of the nondegenerate perturbation (once it is small enough) and of the choice of almost complex structure. We refer to the resulting homology as the local Floer homology of H at $\bar{\mathscr{F}}$, and denote it by $HF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}})$. Write

$$\Delta^{\min}(H, \mathscr{F}) = \min_{\bar{x} \in \bar{\mathscr{F}}} \Delta(H, \bar{x}) \quad \text{and} \quad \Delta^{\max}(H, \mathscr{F}) = \max_{\bar{x} \in \bar{\mathscr{F}}} \Delta(H, \bar{x})$$

for the minimum and maximum of the mean-indices $\Delta(H, \bar{x})$ for $\bar{x} \in \bar{\mathscr{F}}$.

We claim that if $\mathscr{F}$ is a family of weakly nondegenerate orbits, then the support of $HF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}})$ satisfies

$$\operatorname{Supp}(HF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}})) \subset (\Delta^{\min}(H, \bar{\mathscr{F}}) - n, \Delta^{\max}(H, \bar{\mathscr{F}}) + n). \tag{7}$$

In fact, by a simple argument following from the continuity of the mean-index and inequality (2), one obtains that $\operatorname{Supp}(HF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}}))$ satisfies the nonstrict version of (7). In order to obtain the strict inequalities, we use the assumption that $\mathcal{F}$ is weakly nondegenerate, and its compactness, to argue as in [84]. In the situation where the Hamiltonian is generalized perfect, we obtain the following.

Lemma 16 *Suppose H is a weakly nondegenerate generalized perfect Hamiltonian and let $\bar{\mathscr{F}}$ be a generalized capped orbit of H. Then $HF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}})$ is supported in the open interval $(\Delta(H, \bar{\mathscr{F}}) - n, \Delta(H, \bar{\mathscr{F}}) + n)$.*

Furthermore, the notion of action carriers discussed in Section 2.4 remains valid in this generalized setting by altering isolated fixed points to generalized fixed points in

Definition 12. Thus, the spectral invariant $c([M], H)$ is carried by a capped generalized periodic orbit $\bar{\mathscr{F}}$ of H. In this case, we have the following generalization of Lemma 13, whose proof, once Lemma 21 below is taken into account, follows just as in [34].

Lemma 17 *Suppose H has only a finite number of generalized fixed points, and let $\bar{\mathscr{F}}$ be a carrier of the spectral invariant of the fundamental class. In this case $HF_n^{\mathrm{loc}}(H, \bar{\mathscr{F}}) \neq 0$.*

Remark 18 Consider $\mathcal{F} \in \pi_0(\mathrm{Fix}(\phi))$ and $\mathscr{F} \subset \mathcal{L}_{\mathrm{pt}} M$ the associated generalized 1–periodic orbit. We remark that different choices of lifts $\bar{\mathscr{F}}$ result in isomorphic local Floer homology groups, with a shift in index given by an integer multiple of $2N$. In particular, if $A \in \Gamma$, then

$$HF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}} \# A) \cong HF_{*+2\langle c_1(TM), A\rangle}^{\mathrm{loc}}(H, \bar{\mathscr{F}}),$$

where $\bar{\mathscr{F}} \# A$ denotes the unique choice of lift containing the capped orbit $\bar{x} \# A$ for $x \in \mathcal{F}$ and $\bar{x} \in \bar{\mathscr{F}}$. From this discussion, we conclude that $\dim_{\mathbb{K}} HF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}})$ does not depend on the capping of $\mathscr{F}$. Hence, the notation $\dim_{\mathbb{K}} HF_*^{\mathrm{loc}}(H, \mathcal{F})$ is justified in this case. Furthermore, when (M, ω) is symplectically Calabi–Yau the local Floer homology does not depend on the choice of lift, thus we denote it by $HF_*^{\mathrm{loc}}(H, \mathcal{F})$. This is analogous to the effect of recapping on local Floer homology in the case of isolated fixed points.

We shall require a slightly more general statement about carriers of quantum homology classes. The definition of a carrier $\bar{\mathscr{F}}$ of a quantum homology class $\alpha_M \in QH(M)$ is the same as for the fundamental class, with $[M]$ replaced by α_M. We then have the following result.

Lemma 19 *Let $\alpha_M \in QH_k(M) \setminus \{0\}$ be a homogeneous element of degree k. Suppose H has only finitely many (contractible) generalized fixed points and let $\bar{\mathscr{F}}$ be a carrier of the spectral invariant of α_M. Then $HF_{k-n}^{\mathrm{loc}}(H, \bar{\mathscr{F}}) \neq 0$.*

In fact a stronger result is true, of which this statement is a direct consequence. It was proven as [93, Theorem D] in the context of ϕ_H^1 with isolated fixed points, but its proof adapts essentially immediately to the context of a finite number of (contractible) generalized fixed points. Indeed, our case differs from the one in [93] by replacing fixed points by generalized fixed points everywhere, hence the only technical difference consists in establishing Lemma 21. We recall that the proof relies on homological perturbation techniques, starting from the decomposition of Section 3.1.4. It constitutes a Novikov-field version of the canonical Λ^0–complexes from [95]. The goal of these arguments is to introduce a new complex which calculates the same total homology but replaces each local Floer complex $CF^{\mathrm{loc}}_*(H,\bar{\mathscr{F}})$, which depends on a sufficiently C^2–small perturbation H_1 of H, by its homology $HF^{\mathrm{loc}}_*(H,\bar{\mathscr{F}})$. This is the local Floer homology of H at $\bar{\mathscr{F}}$, which does not depend on H_1. Note that since we work over a field, $CF^{\mathrm{loc}}_*(H,\bar{\mathscr{F}})$ is chain-homotopy equivalent to $HF^{\mathrm{loc}}_*(H,\bar{\mathscr{F}})$ with the zero differential. The complex obtained from the Floer complex of H_1 in this way computes the same total homology, as desired, but is also strict in the sense of strictly decreasing a natural filtration. Furthermore, it allows us to compute directly the filtered Floer homology of H.

Theorem M *Let (M,ω) be a closed symplectic manifold which is positive or negative monotone. Consider the class $\tilde{\phi}\in\widetilde{\mathrm{Ham}}(M,\omega)$ of the Hamiltonian flow $\{\phi^t_H\}_{t\in[0,1]}$ of $H\in\mathcal{H}$, with $\mathrm{Fix}(\phi^1_H)$ consisting of a finite number of generalized fixed points. Let $\mathbb{K}$ be a ground field which is arbitrary in the positive monotone case and of characteristic zero in the negative monotone case. Then there is a filtered homotopy-canonical complex $(C(H),d_H)$ over the Novikov field $\Lambda_{\mathbb{K}}$ on the action-completion of*

$$\bigoplus HF^{\mathrm{loc}}_*(\tilde{\phi},\bar{\mathscr{F}}),$$

the sum running over all capped generalized 1–periodic orbits $\bar{\mathscr{F}}\in\tilde{\mathcal{O}}(H)$. Specifically, $C(H)$ consists of infinite sums $x=\sum y_i$ where $y_i\in HF^{\mathrm{loc}}_*(\tilde{\phi},\bar{\mathscr{F}}_i)$ with $\mathcal{A}_H(\bar{\mathscr{F}}_i)\xrightarrow{i\to\infty}-\infty$. The complex $(C(H),d_H)$ is free and graded over $\Lambda_{\mathbb{K}}$, and is strict, ie $\mathcal{A}_H(d_H(y))<\mathcal{A}_H(y)$ for all $y\in C(H)$, with respect to the non-Archimedean action-filtration $\mathcal{A}_H$ on $C(H)$, defined by

$$\mathcal{A}_H\Big(\sum\lambda_j y_j\Big)=\max\{-\nu(\lambda_j)+\mathcal{A}_H(y_j)\},\quad \mathcal{A}_H(y_j)=\mathcal{A}_H(\bar{\mathscr{F}}_{i(j)}). \tag{8}$$

Here $\{y_j\}$ is a $\Lambda_{\mathbb{K}}$–basis of $C(H)$ that is determined by $\{y_j\mid i(j)=i\}$ being a basis of $HF^{\mathrm{loc}}_*(\tilde{\phi},\bar{\mathscr{F}}_i)$, where $\mathrm{Fix}(\phi)=\{\mathcal{F}_i\}$ and for each i, $\bar{\mathscr{F}}_i$ is a choice of a lift of the generalized 1–periodic orbit $\mathscr{F}_i$ corresponding to $\mathcal{F}_i$ to a capped generalized periodic orbit in $\tilde{\mathcal{O}}(H)$. Furthermore, for all $a\in\mathbb{R}\setminus\mathrm{Spec}(H)$, the filtered homology $HF(H)^{<a}$ is given by $HF(C(H)^{<a})$, where $C(H)^{<a}=(\mathcal{A}_H)^{-1}(-\infty,a)$. In particular,

$$HF(H)=H(C(H),d_H)\cong QH(M;\Lambda_{\mathbb{K}}).$$

Moreover, for all $a\le b$ with $a,b\in(\mathbb{R}\setminus\mathrm{Spec}(H))\cup\{\infty\}$, the comparison map $HF(H)^{<a}\to HF(H)^{<b}$ is induced by the inclusion $C(H)^{<a}\to C(H)^{<b}$.

Definition 20 (visible spectrum) We define the visible spectrum of a Hamiltonian function H as
$$\mathrm{Spec}^{\mathrm{vis}}(H) = \{\mathcal{A}_H(\bar{\mathscr{F}}) \mid HF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}}) \neq 0\},$$
where $\mathcal{A}_H(\bar{\mathscr{F}})$ denotes the action of any capped orbit $\bar{x} \in \bar{\mathscr{F}}$ for a lift $\bar{\mathscr{F}}$ associated to a generalized fixed point $\mathcal{F} \subset \mathrm{Fix}(\phi_H)$. Indeed, an argument similar to the proof of Proposition 15 shows that the restriction $\mathcal{A}_H|_{\bar{\mathscr{F}}}$ is constant. It is clear that $\mathrm{Spec}^{\mathrm{vis}}(H) \subset \mathrm{Spec}(H)$. In the context of barcodes (see Section 3.1.5), the visible spectrum corresponds to the endpoints of all bars of the barcode $\mathcal{B}(H)$ associated to the filtered Floer homology of H.

3.1.3 Crossing energy We show that for a C^2–small perturbation H' of a generalized perfect Hamiltonian H on a closed symplectic manifold (M, ω), every Floer trajectory u connecting orbits of H' contained in distinct isolating neighborhoods has energy bounded below by a constant independent of the perturbation. This is an important technical step.

Lemma 21 *There exist $\delta > 0$ and $\epsilon > 0$ such that for every nondegenerate perturbation H' of H satisfying $\|H - H'\|_{C^2} < \epsilon$, every orbit in $\mathcal{O}(H', \mathscr{F}_j)$ is contained in $U_{\mathcal{F}_j}$ for $j = 1, \dots, m$, every Floer trajectory u connecting capped orbits in distinct isolating neighborhoods satisfies $E(u) > \delta$, and every Floer trajectory connecting capped orbits in the same $\mathcal{O}(H', \bar{\mathscr{F}}_j)$ is contained in $U_{\mathcal{F}_j}$. Finally, if (M, ω) is rational, every Floer trajectory u connecting capped orbits in $\mathcal{O}(H', \bar{\mathscr{F}}_j)$, $\mathcal{O}(H', \bar{\mathscr{F}}'_j)$, for different cappings $\bar{\mathscr{F}}_j, \bar{\mathscr{F}}'_j$ of the same $\mathscr{F}_j$, has energy $E(u) \geq \rho/2$.*

Proof Suppose there exists a sequence of nondegenerate Hamiltonians $\{H'_k\}$ that C^2–converges to H, and a sequence of Floer trajectories u_k of H'_k connecting orbits in distinct isolating neighborhoods such that $E(u_k) \to 0$. Since H has finitely many generalized fixed points, we may suppose without loss of generality that all the Floer trajectories u_k connect orbits in $U_{\mathcal{F}}$ to orbits in $U_{\mathcal{F}'}$, where $\mathcal{F}, \mathcal{F}' \in \pi_0(\mathrm{Fix}(\phi_H))$ are distinct.

By a compactness result of [19], and arguing as in [61], we obtain the existence of a Floer trajectory u of H connecting an orbit in $U_{\mathcal{F}}$ to an orbit in $U_{\mathcal{F}'}$ such that $E(u) = 0$. Thus,
$$\frac{\partial u}{\partial s} = 0 \quad \text{and} \quad \frac{\partial u}{\partial t} = X_H^t,$$
which, in turn, implies that for each s, the loop $u_s = u(s, \cdot)$ is a 1–periodic orbit of H. This contradicts the fact that the generalized fixed points of H are isolated.

Note that in the above argument, if for each k, u_k connects orbits in the same $U_{\mathcal{F}_j}$ but is not contained in $U_{\mathcal{F}_j}$, then for k sufficiently large, $E(u_k) > \delta$ again. Indeed, otherwise we would again reach a contradiction by a compactness argument. However, if now u_k connects orbits in the same $\mathcal{O}(H'_k, \bar{\mathscr{F}}_j)$, then its energy, given by the difference of actions of its two asymptotic capped orbits, tends to zero as $k \to \infty$. We conclude that u_k must be contained in $U_{\mathcal{F}_j}$ for all k sufficiently large.

Finally, we remark that if u_k connects orbits in $\mathcal{O}(H'_k, \bar{\mathscr{F}}_j)$ and in $\mathcal{O}(H'_k, \bar{\mathscr{F}}'_j)$, where $\bar{\mathscr{F}}_j$ and $\bar{\mathscr{F}}'_j$ are different cappings of $\mathscr{F}_j$, then, if the symplectic manifold is rational, $E(u_k) \geq \rho - o(1)$ as $k \to \infty$. □

3.1.4 Decomposition of Floer differential An important feature related to local Floer homology concerns the decomposition of the full differential defined on the complex $CF_*(H')$ into the sum of local differentials of complexes $CF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}})$ — for all the different lifts of the finitely many generalized fixed points — and into an additional component we shall call D. Note that here, H' is a nondegenerate Hamiltonian C^2–close enough to H in the aforementioned sense. Namely, for each chain $\sigma \in CF_*(H')$, we have

$$\partial\sigma = \sum \tilde{\partial}_{\bar{\mathscr{F}}}\sigma + D\sigma, \tag{9}$$

where $\tilde{\partial}_{\bar{\mathscr{F}}}$ represents an extension of the local differential of the complex $CF_*^{\mathrm{loc}}(H, \bar{\mathscr{F}})$ obtained by setting $\tilde{\partial}_{\bar{\mathscr{F}}}\bar{x} = 0$ for every capped orbit which does not belong $\mathcal{O}(H', \bar{\mathscr{F}})$. Loosely speaking, D only "counts" Floer trajectories connecting orbits contained in disjoint isolating open sets $U_{\mathcal{F}}$.

Remark 22 Suppose that σ is a chain in the complex $CF_*(H')$ and $\bar{z}$ is an orbit entering $D\sigma$. Naturally, there exists $0 \leq k \leq m$ such that $\bar{z} \in CF_*(H, \bar{\mathscr{F}}_k)$ for a particular lift of $\mathscr{F}_k$, and a Floer trajectory u connecting an orbit $\bar{y} \in CF_*(H, \bar{\mathscr{F}}_l)$ to $\bar{z}$ for $l \neq k$ (and a particular lift of $\mathscr{F}_l$). We then obtain

$$\mathcal{A}_{H'}(\bar{z}) = \mathcal{A}_{H'}(\bar{y}) - E(u) < \mathcal{A}'_H(\bar{y}) - \delta, \tag{10}$$

where the first equality comes from the fact that the energy of a Floer trajectory connecting two capped orbits is equal to their action difference, and the δ comes from the uniform lower bound for the crossing energy from Lemma 21. In other words,

$$\mathcal{A}_{H'}(Dx) < \mathcal{A}_{H'}(x) - \delta$$

for all $x \neq 0$ in $CF_*(H')$.

3.1.5 Barcodes of Hamiltonian diffeomorphisms The proof of Theorem G uses notions and results regarding barcodes of Hamiltonian diffeomorphisms, in the case where they have a finite number of contractible generalized fixed points. Hitherto, this theory was developed mostly for the case where the generalized fixed points are in fact points, yet given Lemma 21, all relevant results generalize to our situation. In the next section we describe the main Smith-type inequality regarding the behavior of barcodes under iteration.

We will summarize the properties necessary for us, and refer to [74; 75; 103; 48; 96; 95] for further discussion of this notion, in the context of continuity in the Hofer distance and the spectral distance in particular. For convenience, we work in the setting of monotone symplectic manifolds, yet natural analogues of various statements exist in the semipositive, rational and general settings.

Proposition 23 *Let (M,ω) be a monotone symplectic manifold with $\mathcal{P}_\omega = \rho\cdot\mathbb{Z}$, and consider $\phi \in \mathrm{Ham}(M,\omega)$ with $\mathrm{Fix}(\phi)$ consisting of a finite number of generalized fixed points. Let $\mathbb{K}$ be a coefficient field. Let H be a Hamiltonian generating ϕ. Then $\mathrm{Spec}(H) \subset \mathbb{R}$ is a discrete subset, and there exists a countable collection*

$$\mathcal{B}(H) = \mathcal{B}(H;\mathbb{K}) = \{(I_i, m_i)\}_{i\in\mathcal{I}},$$

*called the **barcode** of H with coefficients in $\mathbb{K}$, of intervals I_i in $\mathbb{R}$ of the form $I_i = (a_i, b_i]$ or $I_i = (a_i,\infty)$, called **bars** with multiplicities $m_i \in \mathbb{Z}_{>0}$, such that the following properties hold:*

(i) *The group $\rho\cdot\mathbb{Z}$ acts on $\mathcal{B}(H)$ in the sense that for all $k\in\mathbb{Z}$ and all $(I,m)\in\mathcal{B}$, we have $(I+\rho k, m)\in\mathcal{B}$.*

(ii) *For each window $J=(a,b)$ in $\mathbb{R}$ with $a,b\notin \mathrm{Spec}(H)$, only a finite number of intervals I with $(I,m)\in\mathcal{B}$ have endpoints in J. Furthermore,*
$$\dim_{\mathbb{K}} HF(H)^J = \sum_{\substack{(I,m)\in\mathcal{B}(H)\\ \#\partial I\cap J=1}} m,$$
where for an interval $I=(a,b]$, we set $\partial I = \{a,b\}$, and for $I=(a,\infty)$, we set $\partial I=\{a\}$.

(iii) *For $a\in\mathrm{Spec}(H)$ and $\epsilon>0$ sufficiently small that $(a-\epsilon,a+\epsilon)\cap\mathrm{Spec}(H)=\{a\}$, we have*
$$\dim_{\mathbb{K}} HF(H)^{(a-\epsilon,a+\epsilon)} = \sum_{\substack{(I,m)\in\mathcal{B}(H)\\ a\in\partial I}} m,$$
$$\dim_{\mathbb{K}} HF(H)^{(a-\epsilon,a+\epsilon)} = \sum_{\mathcal{A}(\bar{\mathscr{F}})=a} \dim_{\mathbb{K}} HF^{\mathrm{loc}}(H,\bar{\mathscr{F}}).$$

(iv) *There are $K(\phi, \mathbb{K})$ orbits of finite bars counted with multiplicity, and $B(\mathbb{K})$ orbits of infinite bars counted with multiplicity, under the $\rho \cdot \mathbb{Z}$ action on $\mathcal{B}(H)$. These numbers satisfy*

$$B(\mathbb{K}) = \dim_{\mathbb{K}} H_*(M;\mathbb{K}) \quad \text{and} \quad N(\phi, \mathbb{K}) = 2K(\phi, \mathbb{K}) + B(\mathbb{K}),$$

where

$$N(\phi, \mathbb{K}) = \sum \dim_{\mathbb{K}} HF^{\text{loc}}(\phi, \mathcal{F})$$

is the ***homological count of the fixed points*** *of ϕ, the sum running over all the set $\pi_0(\mathrm{Fix}(\phi))$ of its generalized fixed points.*

(v) *To each orbit $[(I, m)]$, with $I = (a, b]$, of finite bars, there corresponds a bar-length $|I| = b - a$, counted with multiplicity m. There are hence $K(\phi, \mathbb{K})$* ***bar-lengths*** *corresponding to the orbits of finite bars,*

$$0 < \beta_1(\phi, \mathbb{K}) \le \cdots \le \beta_{K(\phi,\mathbb{K})}(\phi, \mathbb{K}),$$

which depend only on ϕ. We call

$$\beta(\phi, \mathbb{K}) = \beta_{K(\phi,\mathbb{K})}(\phi, \mathbb{K})$$

the ***boundary-depth*** *of ϕ, and*

$$\beta_{\text{tot}}(\phi, \mathbb{K}) = \sum_{1 \le j \le K(\phi,\mathbb{K})} \beta_j(\phi, \mathbb{K})$$

its ***total bar-length****.*

(vi) *Each spectral invariant $c(\alpha, H) \in \mathrm{Spec}(H)$ for $\alpha \in QH_*(M) \setminus \{0\}$ is a starting point of an infinite bar in $\mathcal{B}(H)$, and each such starting point is given by a spectral invariant.*[8]

(vii) *If H' is another Hamiltonian generating ϕ, then $\mathcal{B}(H') = \mathcal{B}(H)[c]$ for a certain constant $c \in \mathbb{R}$, where $\mathcal{B}(H)[c] = \{(I_i - c, m_i)\}_{i \in \mathcal{I}}$.*

(viii) *If $\mathbb{K}$ is a field extension of $\mathbb{F}$ and H is a Hamiltonian, then $\mathcal{B}(H;\mathbb{K}) = \mathcal{B}(H;\mathbb{F})$. In particular, $\mathcal{B}(H;\mathbb{K}) = \mathcal{B}(H;\mathbb{F}_p)$ if $\mathrm{char}(\mathbb{K}) = p$, and $\mathcal{B}(H;\mathbb{K}) = \mathcal{B}(H;\mathbb{Q})$ if $\mathrm{char}(\mathbb{K}) = 0$.*

[8] In fact, representatives for the set of orbits of infinite bars counting with multiplicity, can be obtained as spectral invariants of an orthogonal basis of $QH_*(M)$ over the Novikov field $\Lambda_{\mathbb{K}}$, with respect to the non-Archimedean filtration $l_H(-) = c(-, H)$. As we shall not require this stronger statement, we refer to [95; 96] for a discussion of the relevant notions.

3.1.6 Smith theory in filtered Floer homology One of the fundamental results of [95] is the following Smith-type inequality, that readily adapts to our setting by Lemma 21 and its generalization to the situation of branched covers of the cylinder as in [98, Proposition 9]. We refer to [95, Theorem D] for a detailed argument in the case of isolated fixed points, and observe that our generalization below is formulated in such a way that the same proof applies verbatim, by replacing fixed points by generalized fixed points everywhere.

Theorem N *Let (M,ω) be a monotone symplectic manifold, p a prime number, and $\phi \in \mathrm{Ham}(M,\omega)$ with $\mathrm{Fix}(\phi)$ and $\mathrm{Fix}(\phi^p)$ each consisting of a finite number of generalized fixed points, and such that the natural inclusion $\mathrm{Fix}(\phi) \to \mathrm{Fix}(\phi^p)$ restricts to a homeomorphism from each generalized fixed point $\mathcal{F}$ of ϕ to a generalized fixed point of ϕ^p, which we denote by $\mathcal{F}^{(p)}$. Then*

$$\beta_{\mathrm{tot}}(\phi^p,\mathbb{F}_p) \geq p\cdot\beta_{\mathrm{tot}}(\phi,\mathbb{F}_p).$$

This inequality will be the key component in the proof of Theorem G.

A somewhat simpler statement than Theorem N is the Smith inequality in generalized local Floer homology, whose proof is precisely as in [98] together with the crossing energy argument of Lemma 21.

Proposition 24 *Let (M,ω) be a closed symplectic manifold, p a prime number and $\phi \in \mathrm{Ham}(M,\omega)$. Suppose that $\mathrm{Fix}(\phi)$ and $\mathrm{Fix}(\phi^p)$ each consist of a finite number of generalized fixed points. Let $\mathcal{F}$ be a generalized fixed point of ϕ, such that the natural inclusion $\mathrm{Fix}(\phi)\to\mathrm{Fix}(\phi^p)$ restricted to $\mathcal{F}$ is a homeomorphism onto $\mathcal{F}^{(p)}$. Then*

$$\dim_{\mathbb{F}_p} HF^{\mathrm{loc}}(\phi,\mathcal{F}) \leq \dim_{\mathbb{F}_p} HF^{\mathrm{loc}}(\phi^p,\mathcal{F}^{(p)}).$$

3.1.7 Quantum Steenrod operations Quantum Steenrod operations are remarkable algebraic maps

$$Q\mathrm{St}_p\colon QH^*(M;\mathbb{F}_p) \to QH^*(M;\mathbb{F}_p)[\![u]\!]\langle\theta\rangle$$

for p a prime number, u a formal variable of degree 2, and θ a formal variable of degree 1. As is the usual quantum product, $Q\mathrm{St}_p$ is essentially defined by certain counts of configurations consisting of holomorphic curves in M incident with negative gradient trajectories of Morse functions in M. The main difference is that $Q\mathrm{St}_p$ uses p input and 1 output trajectories, and the counts are carried out in families parametrized by the classifying space $B(\mathbb{Z}/p\mathbb{Z})$ of $\mathbb{Z}/p\mathbb{Z}$. The investigation of the enumerative significance of these counts, in terms of various Gromov–Witten invariants, and its implications for mirror symmetry was started in [91; 92].

These operations were first proposed by Fukaya [27], and were formally introduced for $p = 2$ by Wilkins in [106]. They were then studied in [107] in relation to the equivariant pair-of-pants product of Seidel [90]. For a definition for $p > 2$ odd, see [91; 97]. The significance of quantum Steenrod operations in Hamiltonian dynamics was first observed in [93], and was further investigated in [7; 94; 97]. While for the moment these operations are defined in the setting of monotone symplectic manifolds, it is expected that they will be generalized to the semipositive (also called weakly monotone) setting.

One particular property of quantum Steenrod operations that we use in this paper, which was first observed in [93] for $p = 2$, and proved in [97] for $p > 2$, is that whenever

$$QSt_p(\mu) \neq u^{(p-1)n}\mu, \tag{11}$$

where $\mu \in H^{2n}(M;\mathbb{F}_p)$ is the cohomology class Poincaré dual to the point class, the symplectic manifold (M,ω) is geometrically uniruled: for each ω–compatible almost complex structure J on M, and each point $x \in M$, there exists a J–holomorphic sphere $u\colon \mathbb{CP}^1 \to M$ such that $x \in \mathrm{im}(u)$. Hence, we call a (monotone) symplectic manifold $\mathbb{F}_p$*–Steenrod uniruled* if condition (11) holds. The algebraic significance of this condition is that $u^{(p-1)n}\mu = \mathrm{St}_p(\mu)$, where St_p is the (slightly reformulated) total Steenrod p^{th} power of the class μ, and in general,

$$QSt_p = \mathrm{St}_p + O(q),$$

where $O(q)$ is a collection of terms involving the quantum variable q to power at least 1. These terms correspond to configurations involving J–holomorphic spheres of positive symplectic area, hence condition (11) means that the quantum Steenrod power of the point cohomology class is *deformed* by holomorphic spheres.

3.2 Floer cohomology

At times it shall be convenient to work with Floer cohomology and quantum cohomology of closed symplectic manifolds, instead of homology. All the preliminary results above adapt naturally to this setting. In fact, we may define

$$CF^*(H,J) = CF_{n-*}(\overline{H},\bar{J}),$$

where $\overline{H}(t,x) = -H(1-t,x)$, $\bar{J}_t(x) = J_{1-t}(x)$. The usual action functional $\mathcal{A}_H$ on the left-hand side takes the form $(-\mathcal{A}_{\overline{H}})$ on the right-hand side. Note that hereby the cohomological differential increases the filtration, the triangle inequality for spectral invariants has the opposite inequality, and infinite bars in the barcode are of the form $(-\infty, b)$. Local Floer cohomology is defined in the same way as for homology. Action carriers, and contribution to local Floer cohomology hold similarly: $c(\mu, H)$ for $\mu \in QH^{2n}(M)$ is carried by a capped generalized periodic orbit $\bar{\mathscr{F}}$ of H if, in the same sense as for homology, $\bar{\mathscr{F}}$ is a *lowest* action term in a *highest* minimal action representative of the image $\mathrm{PSS}_H(\mu)$ of μ under the PSS isomorphism [70] from the quantum cohomology $QH^*(M) \to HF^{*-n}(H)$ to the filtered Floer cohomology of the Hamiltonian H. For (M,ω) rational, in particular monotone, for each nonzero class $\mu \in QH^*(M)$, and for $H \in \mathcal{H}$ with $\#\pi_0(\mathrm{Fix}(\phi_H^1)) < \infty$, we have that $c(\mu, H)$ is carried by at least one generalized capped 1–periodic orbit $\bar{\mathscr{F}}$ of H. Furthermore, if μ is a homogeneous class of degree k, and $\bar{\mathscr{F}}$ carries $c(\mu, H)$, then $HF_{\mathrm{loc}}^{k-n}(H, \bar{\mathscr{F}}) \neq 0$. We refer to [52] for further discussion of the comparison between Floer homology and Floer cohomology.

4 Cluster structure of the essential spectrum

Definition 25 (essential spectrum) We define the essential spectrum of a Hamiltonian function H as

$$\mathrm{Spec}^{\mathrm{ess}}(H) = \{c(\alpha, H) \mid \alpha \in QH_*(M) \setminus \{0\}\}.$$

Observe that the spectrality property of the spectral invariants is equivalent to the inclusion $\mathrm{Spec}^{\mathrm{ess}}(H) \subset \mathrm{Spec}(H)$. In fact, Lemma 19 implies that $\mathrm{Spec}^{\mathrm{ess}}(H) \subset \mathrm{Spec}^{\mathrm{vis}}(H)$ for Hamiltonian diffeomorphisms with a finite number of (contractible) generalized fixed points. In the context of barcodes (see Section 3.1.5), the essential spectrum corresponds to the endpoints of infinite bars of the barcode $\mathcal{B}(H)$ associated to the filtered Floer homology of H.

We next show that whenever $\gamma(H) < \rho$, the essential spectrum has a cluster structure determined by the subset produced by quantum homology classes of valuation 0.

Proposition 26 *Suppose M is a monotone symplectic manifold and H a Hamiltonian function on M. Then*

$$0 \leq c([M], \widetilde{\phi}_H) - c(\alpha, \widetilde{\phi}_H) \leq \gamma(H)$$

for all $\alpha \in QH(M)$ such that $\nu(\alpha) = 0$, including all $\alpha \in H_*(M) \subset QH(M)$.

Proof By the triangle inequality and the value at identity properties of the spectral invariant,

$$c(\alpha,\widetilde{\phi}_H)=c(\alpha*[M],\widetilde{\mathrm{id}}\,\widetilde{\phi}_H)\le c(\alpha,\widetilde{\mathrm{id}})+c([M],\widetilde{\phi}_H)=c([M],\widetilde{\phi}_H)$$

for all $\alpha\in QH(M)$ such that $\nu(\alpha)=0$. In addition,

$$0=c(\alpha,\widetilde{\mathrm{id}})=c(\alpha*[M],\widetilde{\phi}_H\widetilde{\phi}_{\overline{H}})\le c(\alpha,\widetilde{\phi}_H)+c([M],\widetilde{\phi}_{\overline{H}}).$$

Combining both inequalities we obtain

$$0\le c([M],\widetilde{\phi}_H)-c(\alpha,\widetilde{\phi}_H)\le c([M],\widetilde{\phi}_H)+c([M],\widetilde{\phi}_{\overline{H}})=\gamma(H),$$

which concludes the proof of the proposition. □

Proposition 27 *Let M be a monotone symplectic manifold with rationality constant $\rho>0$, let H be a Hamiltonian function on M with $\gamma(H)<\rho$ and let $\alpha\in QH(M)$. Then*

$$c([M],\widetilde{\phi}_H)-\rho<c(\alpha,\widetilde{\phi}_H)\le c([M],\widetilde{\phi}_H)$$

if, and only if, $\nu(\alpha)=0$.

Proof If $\nu(\alpha)=0$, then Proposition 26 and the hypothesis that $\gamma(H)<\rho$ imply that

$$c([M],\widetilde{\phi}_H)-\rho<c(\alpha,\widetilde{\phi}_H)\le c([M],\widetilde{\phi}_H).$$

Conversely, let $x_1,\dots,x_B$ be a homogeneous basis of $H_*(M)\subset QH(M)$ and write $c=c([M],\widetilde{\phi}_H)$. Then, by Proposition 26, we have $c(x_k,\widetilde{\phi}_H)\in(c-\rho,c]$ for all $1\le k\le B$. Also, for $q^j\in\Lambda_{\mathbb{K}}$, the equality $c(q^jx_k,\widetilde{\phi}_H)=c(x_k,\widetilde{\phi}_H)+j\rho$ implies that $c(q^jx_k,\widetilde{\phi}_H)\notin(c-\rho,c]$ for all $j\neq 0$. Thus, if $c(\alpha,\widetilde{\phi}_H)\in(c-\rho,c]$ for

$$\alpha=\lambda x_k=\sum a_jq^jx_k,$$

where $\lambda\in\Lambda$, the non-Archimedean property of the spectral invariant imposes that

$$\alpha=a_0x_k+\sum_{j<0}a_jq^jx_k,$$

which in turn implies $\nu(\alpha)=0$. In general, $\alpha\in QH(M)$ is of the form $\sum\lambda_kx_k$, where $\lambda_k=\sum a_j^{(k)}q^j$. Consequently, if $c(\alpha,\widetilde{\phi}_H)\in(c-\rho,c]$, we may argue as before to conclude

$$\alpha=\sum_k\Bigl(a_0^{(k)}x_k+\sum_{j<0}a_j^{(k)}q^jx_k\Bigr). \tag{12}$$

Thus, $\nu(\alpha)=0$, which concludes the proof of the claim. □

Remark 28 The above propositions are valid, albeit with minor modifications to the proofs, in the more general case where M is only assumed to be rational. If M is negative monotone, then the base field $\mathbb{K}$ is required to be of characteristic zero; see Remark 10.

Let $\phi \in \mathrm{Ham}(M,\omega)$, and suppose $\gamma(\phi) < \rho$. We can, therefore, find a Hamiltonian function H generating ϕ such that $\gamma(H) < \rho$. Our goal is to extract information from the cluster structure of H in order to bound $\gamma(\phi)$ from below. First we set notation. Put $\mathrm{S}^1_\rho = \mathbb{R}/\rho\cdot\mathbb{Z}$ and, for $a \in \mathbb{R}$, let $[a] \in \mathrm{S}^1_\rho$ be its equivalence class. For $\theta \in \mathrm{S}^1_\rho$, define

$$\Gamma_\theta = \{(a-\rho, a] \mid a \in \mathbb{R},\ [a] = \theta\}.$$

Note that the intervals in Γ_θ are disjoint and their union covers the real line. In addition, observe that, modulo $\rho\cdot\mathbb{Z}$, the set $\mathrm{Spec}^{\mathrm{ess}}(H) \cap \mathrm{I}$ does not depend on the interval $\mathrm{I} \in \Gamma_\theta$.

Definition 29 (spectral length) We define the *θ–parsed spectral length of H* as

$$l(H,\Gamma_\theta) = \mathrm{diam}(\mathrm{Spec}^{\mathrm{ess}}(H) \cap \mathrm{I}) = \sup\{|a-b| \mid a, b \in \mathrm{Spec}^{\mathrm{ess}}(H) \cap \mathrm{I}\},$$

where $\mathrm{I} \in \Gamma_\theta$ is arbitrary. For $\Gamma_H = \Gamma_{[c([M],H)]}$ we call $l(H,\Gamma_H)$ the *fundamental length of H*. Finally, we define the *spectral length* of $\phi \in \mathrm{Ham}(M,\omega)$ as

$$l(\phi) = \inf\{l(H,\Gamma_\theta) \mid \theta \in \mathrm{S}^1_\rho\}, \tag{13}$$

where H is any Hamiltonian function generating ϕ. The right-hand side of (13) does not depend on the choice of Hamiltonian: indeed, if H' is another Hamiltonian generating ϕ, then $\mathrm{Spec}^{\mathrm{ess}}(H') = \mathrm{Spec}^{\mathrm{ess}}(H) + c$ for a certain $c \in \mathbb{R}$ by Proposition 23(vii). (Another proof using Seidel elements is also possible.)

Remark 30 The following alternative definition of $l(\phi)$ helps calculate it in examples. Set $\pi : \mathbb{R} \to \mathrm{S}^1_\rho$ for the natural projection: $\pi(a) = [a]$. The image $\pi(\mathrm{Spec}^{\mathrm{ess}}(H)) \subset \mathrm{S}^1_\rho$ is then a finite set. Hence its complement consists of a finite number of open intervals $\{K_j\}_{j=1}^m$. In terms of these intervals,

$$l(\phi) = \rho - \max_j |K_j|,$$

where for an interval K in S^1_ρ we denote by $|K|$ the length of K with respect to the standard metric. Yet again, we may reformulate $l(\phi)$ more intuitively as the smallest length of an interval containing $\pi(\mathrm{Spec}^{\mathrm{ess}}(H))$, that is,

$$l(\phi) = \inf\{|L| \mid L \supset \pi(\mathrm{Spec}^{\mathrm{ess}}(H))\},$$

the infimum running over intervals L in S^1_ρ.

Remark 31 We can also define an a priori larger invariant $l'(\phi) \geq l(\phi)$ of ϕ by

$$l'(\phi) = \inf_{\phi_H^1 = \phi} l(H, \Gamma_H).$$

However, we find $l(\phi)$ more convenient for this paper.

Lemma 32 *The fundamental length of a Hamiltonian H satisfies*

$$l(H, \Gamma_H) \leq \gamma(H).$$

If, in addition, $\gamma(H) < \rho$, then we have equality, ie

$$l(H, \Gamma_H) = \gamma(H).$$

Proof By definition the θ–parsed spectral length of H is bounded above by ρ for any choice of $\theta \in \mathrm{S}^1$; in particular, $l(H, \Gamma_H) \leq \rho$. Thus, we need only to consider the case where $\gamma(H) < \rho$. Equation (12) in the proof of Proposition 27, or alternatively Proposition 23(vi), implies that $\#\{\mathrm{Spec}^{\mathrm{ess}}(H) \cap \mathrm{I}\} < \infty$ for $\mathrm{I} \in \Gamma_H$ and hence for $\mathrm{I} \in \Gamma_\theta$ for any $\theta \in \mathrm{S}^1_\rho$. Thus by Proposition 27 the fundamental length of H is given by

$$l(H, \Gamma_H) = c([M], H) - c(\alpha_{\min,H}, H),$$

where $\alpha_{\min,H} \in QH(M)$ has zero valuation. Consequently, Proposition 26 implies that $l(H, \Gamma_H) \leq \gamma(H)$. To prove equality, we observe that by the Poincaré duality property of spectral invariants (see [68; 17]) and the fact that the set $\mathrm{Spec}^{\mathrm{ess}}(H) \cap \mathrm{I}$ is finite, there exists $\beta \in QH(M) \setminus \{0\}$ such that $c(\beta, H) = -c([M], \overline{H})$. By adding $\gamma(H)$ to both sides of the equality we obtain $c(\beta, H) + \gamma(H) = c([M], H)$, which implies

$$c([M], H) - \rho < c(\beta, H) \leq c([M], H).$$

Therefore, $\gamma(H) \leq l(H, \Gamma_H)$, which gives us the claimed equality. □

Lemma 33 *Let ϕ be a Hamiltonian diffeomorphism. Then, $l(\phi) \leq \gamma(\phi)$.*

Proof Let H be any Hamiltonian function that generates ϕ. By definition $l(H, \Gamma_\theta) < \rho$ for every $\theta \in \mathrm{S}^1_\rho$; in particular, we have $l(\phi) < \rho$. Hence, if $\gamma(\phi) \geq \rho$, the desired inequality holds trivially. Therefore, we may suppose that $\gamma(\phi) < \rho$, in which case we may take H such that $\gamma(H) < \rho$. Consequently, Lemma 32 implies $l(H, \Gamma_H) = \gamma(H)$; in particular, we have that $l(\phi) \leq \gamma(H)$. If H' is any other Hamiltonian function generating ϕ, with $\gamma(H') \leq \gamma(H)$, the same argument implies $l(\phi) \leq \gamma(H')$. Thus, we conclude that $l(\phi) \leq \gamma(\phi)$. □

Remark 34 Lemma 32 immediately implies that if $\gamma(\phi) < \rho$, then $l'(\phi) = \gamma(\phi)$. It is not clear that the same holds for $l(\phi)$. However, we can prove that if $\gamma(\phi) < \rho/2$, then $l(\phi) = \gamma(\phi)$. Indeed, if $\gamma(H) < \gamma(\phi) + \epsilon < \rho/2$, by Lemma 32 we have $l(H, \Gamma_H) = \gamma(H) < \rho/2$. However, this implies that for arbitrary $\theta \in \mathrm{S}^1_\rho$, either $l(H, \Gamma_\theta) = l(H, \Gamma_H)$, if the partitions of $\mathrm{Spec}^{\mathrm{ess}}(H)$ into clusters corresponding to Γ_H and Γ_θ coincide, or $l(H, \Gamma_\theta) \geq \rho - l(H, \Gamma_H) > \rho/2 > l(H, \Gamma_H)$, if they do not. Hence, by taking the infima, $l(\phi) = \gamma(\phi)$.

Lemma 35 *Let ϕ be a generalized Hamiltonian $\mathbb{K}$ pseudorotation with sequence $k_j = j$ and take a Hamiltonian H generating ϕ. Suppose that all the distances between pairs of points in $\mathrm{Spec}^{\mathrm{ess}}(H)$ are rational multiples of ρ. Then there exists a positive integer m such that $\gamma(\phi^m) \geq \rho$.*

Proof Fix the base coefficient field $\mathbb{K}$ for all homological notions in the proof. We can suppose $\gamma(\phi) < \rho$, otherwise the implication of the theorem would be true for $m = 1$. Furthermore, we note that the hypothesis of the theorem is independent of the choice of Hamiltonian function; thus, we may suppose that $\gamma(H) < \rho$, which by Lemma 32, implies $l(H, \Gamma_H) = \gamma(H)$. Hence, we have a cluster structure determined by finitely many values of the essential spectrum of H belonging to the interval

$$\mathrm{I}_H = (c([M], H) - \rho, c([M], H)].$$

Thus, setting

$$\mathrm{Spec}^{\mathrm{ess}}(H) \cap \mathrm{I}_H = \{c_1, \dots, c_B\},$$

by the hypothesis of the proposition we have

$$c_i - c_j = \frac{a_{ij}}{b_{ij}} \rho \in \rho \cdot \mathbb{Q} \cap (-\rho, \rho)$$

for all $i \neq j$. Note that any pair of points $\alpha, \beta \in \mathrm{Spec}^{\mathrm{ess}}(H)$ are of the form $\alpha = c_i + k\rho$ and $\beta = c_j + l\rho$ for integers l and k. Thus their difference is of the form

$$\alpha - \beta = \left(\frac{a_{ij}}{b_{ij}} + (k - l)\right)\rho. \tag{14}$$

Now, let m be the integer given by $\prod_{i<j} b_{ij}$. The facts that

$$\mathrm{Fix}(\phi^m) = \mathrm{Fix}(\phi), \quad \mathrm{Spec}^{\mathrm{ess}}(H) = \mathrm{Spec}^{\mathrm{vis}}(H) \quad \text{and} \quad HF^{\mathrm{loc}}(\phi^{k_j}, \mathcal{F}^{(k_j)}) \neq 0$$

for all generalized fixed points $\mathcal{F}$ of ϕ imply

$$\mathrm{Spec}^{\mathrm{ess}}(H^{(m)}) = m \cdot \mathrm{Spec}^{\mathrm{ess}}(H) + \rho \cdot \mathbb{Z}. \tag{15}$$

As a consequence of equations (14) and (15) and the definition of m, we have that $\mathrm{Spec}^{\mathrm{ess}}(H^{(m)}) = \rho \cdot \mathbb{Z} + c$ for a suitable constant $c \in \mathbb{R}$. Hence, $l(F, \Gamma_\theta) = 0$ for any Hamiltonian F generating ϕ^m and $\theta \in \mathrm{S}^1_\rho$. If $\gamma(F) < \rho$, then by Lemma 32 $\gamma(F) = l(F, \Gamma_F) = 0$, which is absurd since this would imply $\phi^m = \mathrm{id}$. Hence $\gamma(\phi^m) \geq \rho$. □

5 Proofs

5.1 Proof of Theorem C

Let (M, ω) be a closed symplectic manifold and consider a nontrivial $\phi \in \mathrm{Ham}(M, \omega)$ such that $\phi^p = \mathrm{id}$ for an integer[9] p. We can construct a Riemannian metric $\langle \cdot, \cdot \rangle$ which is invariant under the action of the group

$$G = \{\mathrm{id}, \phi, \ldots, \phi^{p-1}\},$$

a fact that is true for any compact Lie group G. In other words, ϕ is an isometry with respect to this metric. We first show that $\mathrm{Fix}(\phi)$ is composed of finitely many isolated path-connected components.

Let $x \in \mathcal{F} \subset \mathrm{Fix}(\phi)$, where $\mathcal{F}$ is the path-connected component of x. We claim that there exists a neighborhood of x which does not intersect any other connected component of $\mathrm{Fix}(\phi)$. Suppose the contrary. Then x would be a limit point of $\mathrm{Fix}(\phi) \setminus \mathcal{F}$. In particular, if $B_\epsilon(x)$ is a normal ball of radius ϵ around x, then there exists a point $y \in B_\epsilon(x) \cap (\mathrm{Fix}(\phi) \setminus \mathcal{F})$ and we can consider the unique minimizing geodesic γ given by the exponential map, satisfying $\gamma(0) = x$ and $\gamma(1) = y$. However, ϕ is an isometry so we have that $\tilde{\gamma} = \phi \circ \gamma$ is also a minimizing geodesic satisfying $\tilde{\gamma}(0) = x$ and $\tilde{\gamma}(1) = y$, hence by uniqueness we must have $\mathrm{Image}(\gamma) \subset \mathrm{Fix}(\mathcal{F})$, contradicting the fact that y was in a distinct path-connected component. Since $\mathcal{F}$ is compact we can choose the radius ϵ of the normal ball uniformly so that $\mathcal{F}$ is in fact isolated, which by the compactness of M implies that there are only finitely many path-connected components.

Furthermore, if k is coprime to p then we have $\mathrm{Fix}(\phi^k) = \mathrm{Fix}(\phi)$. In fact, since p and k are relatively prime there exist integers a_k and b_k such that $a_k k + b_k p = 1$. Thus,

$$\phi = \phi^{a_k k + b_k p} = \phi^{a_k k} \phi^{b_k p} = \phi^{a_k k}.$$

[9]While we do not use it in this proof, it might help the reader to first assume that p is a prime.

So if x is a fixed point of ϕ^k then the above equality shows that x is also a fixed point of ϕ. Conversely, if x is a fixed point of ϕ it is clearly a fixed point for any of its iterations. Finally, the same argument shows that if x is contractible as a fixed point of ϕ^k it is also contractible as a fixed point of ϕ, and vice versa. Therefore $\mathrm{Fix}(\phi^k) = \mathrm{Fix}(\phi)$.

To show that ϕ is weakly nondegenerate we utilize the fact if M is connected and $f \in \mathrm{Iso}(M, \langle \cdot, \cdot \rangle)$ is such that $f(x) = x$ and $D(f)_x = \mathrm{id}_{T_x M}$ for a point $x \in M$, then $f = \mathrm{id}_M$. This can be proven by considering the nonempty closed set

$$S = \{y \in M \mid f(y) = y, D(f)_y = \mathrm{id}_{T_y M}\},$$

and noting that the existence of normal balls implies that S is also open. Applied to our context, we must then show that for every $x \in \mathrm{Fix}(\phi)$, $D(\phi)_x$ must have at least one eigenvalue different from 1, otherwise ϕ would have to be trivial. One way to see this is by noting that as $D(\phi)_x \in \mathrm{Sp}_{2n}(T_x M)$ is an element of finite order, its Jordan form is diagonal, hence it is trivial if and only if all its eigenvalues are equal to 1.

A slight modification of the above arguments, which amounts to the slice theorem [2, Theorem I.2.1], shows first that each connected component $\mathcal{F}$ of the fixed-point set of ϕ is a closed connected submanifold of M (and hence is path-connected). Moreover, for each $\mathcal{F}$ and $x \in \mathcal{F}$, $\ker(D(\phi)_x - \mathrm{id}_{T_x M}) = T_x \mathcal{F}$, which is to say that the graph of ϕ intersects the diagonal $\Delta \subset M \times M^-$ *cleanly*. In other words, ϕ is a Floer–Morse–Bott Hamiltonian diffeomorphism.

Finally, to prove that for a generalized fixed point $\mathcal{F}$ of ϕ, and capping $\overline{\mathscr{F}}$ of its corresponding generalized periodic orbit $\mathscr{F}$, the mean-index $\Delta(H, \bar{x})$ is constant as a function of $x \in \mathcal{F}$, we argue as follows. We shall prove that for a fixed $x_0 \in \mathcal{F}$, the function $f : \mathcal{F} \to \mathbb{R}$, given by $f(x) = \Delta(H, \bar{x}) - \Delta(H, \bar{x}_0)$, has integer values. By continuity of the mean-index this implies that f is identically constant, and as $f(x_0) = 0$, it is identically zero. This shows the required statement.

First we prove that f has integer values. Similarly to the case of a Riemannian metric, by [59, Proposition 2.5.6] we can find an ω–compatible almost complex structure J on M that is preserved by ϕ. This allows us to consider $D(\phi)_x \in \mathrm{Sp}_{2n}(T_x M)$ for all $x \in \mathcal{F}$ a unitary matrix, which has diagonal Jordan form, and is determined up to conjugation by its spectrum with geometric multiplicities. Furthermore its spectrum lies in the finite set $\mu_p \subset \mathbb{C}$ of p^{th} roots of unity. Therefore by continuity of the spectrum in the operator norm, which holds for normal and hence for unitary matrices in particular, the spectrum of $D(\phi)_x$ does not depend on $x \in \mathcal{F}$, and all $D(\phi)_x$ for $x \in \mathcal{F}$ are conjugate by appropriate unitary isomorphisms. Therefore $D(\phi)_x$ and $D(\phi)_{x_0}$ can be connected to the identity by conjugate paths, which therefore have equal mean-indices. Now, as the paths obtained from $D(\phi_H^t)_x$ and $D(\phi_H^t)_{x_0}$ by means of the cappings differ from these conjugate paths by suitable loops Φ and Φ_0 in the symplectic group, we obtain that $f(x) = \Delta(H, \overline{x}) - \Delta(H, \overline{x}_0) = \Delta(\Phi) - \Delta(\Phi_0) \in \mathbb{Z}$.

Finally, observe that with $D(\phi)_x$ being (ω_x, J_x)–unitary, $T_x\mathcal{F}$ is J_x–invariant, and the tangent space T_xM splits as a symplectic direct sum $T_x\mathcal{F} \oplus \mathcal{N}_x$, where $\mathcal{N}_x$ is the normal bundle to $\mathcal{F}$ at x (in fact this splitting can be obtained by taking $\mathcal{N}_x$ to be the Hermitian orthogonal complement of $T_x\mathcal{F}$). In particular, $\mathcal{F}$ is a symplectic submanifold of (M, ω).

Hence, the above discussion shows that ϕ is generalized perfect with sequence k_j being the monotone-increasing ordering of the set

$$\{k \in \mathbb{Z}_{>0} \mid \gcd(k, p) = 1\}. \qquad \square$$

Remark 36 We have just seen that a p–torsion Hamiltonian diffeomorphism ϕ is weakly nondegenerate generalized perfect. In our setting it is enough to consider the case where ϕ has prime order. In fact, if ϕ has order $d \geq 2$ and l is a prime that divides d, ie there is an integer m such that $d = lm$, we consider the Hamiltonian diffeomorphism $\psi = \phi^m$, which, in turn, has prime order. Equivalently, by Cauchy's theorem, if G is a finite group then for every prime p dividing its order there exists an element of order p.

5.2 Proof of Proposition 5

We first observe that by the universal coefficient formula, it is sufficient to prove the statement for $R = \mathbb{Z}$.

Now from [85, Chapter 9 and the proof of Theorem 2.3.2] as well as the translation of [76, Theorem 3.4.11] from the setting of Lagrangian clean intersections to the Floer–Morse–Bott Hamiltonian setting [3, Theorem 5.2.2] by means of the graph construction, it is direct to see that there is an isomorphism

$$HF^{\mathrm{loc}}(\phi, \mathcal{F}) \cong H(\mathcal{F}; \mathcal{O} \times_{\pm 1} \mathbb{Z})$$

of the local Floer homology and the homology of $\mathcal{F}$ with coefficients in a $\mathbb{Z}$–local system $\mathcal{O} \times_{\pm 1} \mathbb{Z}$, with structure group $\{\pm 1\} \cong \mathbb{Z}/2\mathbb{Z}$, associated to a double cover $\mathcal{O}$

of $\mathcal{F}$ that we describe below. It is the goal of the proof to show that *in our case* this local system is trivial.

The local system $\mathcal{O}$ is defined as follows. For $x, y \in \mathcal{F}$, consider the space $P_{x,y}(\mathcal{F})$ of smooth maps $\gamma\colon \mathbb{R} \to \mathcal{F}$ such that

$$\lim_{s\to-\infty} \gamma(s) = x \quad \text{and} \quad \lim_{s\to\infty} \gamma(s) = y$$

for which the convergence is exponential with derivatives. Let $u_\gamma\colon \mathbb{R} \times \mathrm{S}^1 \to M$ denote the cylinder $u_\gamma(s,t) = \phi_H^t(\gamma(s))$. Look at the bundle $E_\gamma \to \mathbb{R} \times \mathrm{S}^1$ given by $E_\gamma = (u_\gamma)^* TM$. Now for each sufficiently small positive number $\epsilon \in (0, \epsilon_0)$, where ϵ_0 depends only on H and $\mathcal{F}$, consider real Cauchy–Riemann differential operators

$$D_\gamma\colon W^{1,p,\epsilon}(E_\gamma) \to L^{p,\epsilon}(E_\gamma)$$

between Sobolev spaces of sections of E_γ with ϵ–exponential decay as $|s| \to \infty$, that over $(-\infty, -C)$ and (C, ∞), for a large $C > 0$, coincide with real Cauchy–Riemann operators determined by a choice of an ω–compatible almost complex structure $\{J_t\} \in \mathcal{J}_M$ and connections whose parallel transport over the curve $\{(s,t)\}_{t\in[0,1]}$ (with s fixed) is determined by the linearization of ϕ_H^t at $\gamma(s)$. For $\epsilon > 0$ sufficiently small, all these operators are Fredholm. Moreover, with the auxiliary data of connections and complex structures forming a contractible space, all these operators are furthermore homotopic to each other in the space of Fredholm operators. It is shown in [85] and [28, Chapter 8] that for $\gamma, \gamma' \in P_{x,y}(\mathcal{F})$, the orientation torsors $|D_\gamma|$ and $|D_{\gamma'}|$ of the determinant spaces $\det(D_\gamma)$ and $\det(D_{\gamma'})$ are canonically isomorphic.[10] We can therefore fix $x \in \mathcal{F}$, and set our local system $\mathcal{O} \to \mathcal{F}$ to be induced from the sets $|D_\gamma|$ for $\gamma \in P_{x,y}(\mathcal{F})$ with $y \in \mathcal{F}$, with the natural identifications provided by this isomorphism.

Now we prove that $\mathcal{O}$ is trivial in our case. Suppose $\gamma \in P_{x,y}(\mathcal{F})$. It is sufficient to show that $\det(D_\gamma)$ is canonically oriented. Now, as in the proof of Theorem C, in our case there exists an ω–compatible almost complex structure J on M which is invariant under ϕ. In particular, $D(\phi)_x\colon T_xM \to T_xM$ is (J_x, ω_x)–unitary for all $x \in \mathcal{F}$. This, together with the fact that the universal cover $\widetilde{\mathrm{Sp}}(2n, \mathbb{R})$ deformation-retracts to the universal cover $\widetilde{U}(n)$ of its unitary subgroup, implies that D_γ is homotopic in the space of Fredholm operators, canonically up to a contractible choice of auxiliary

[10] Recall that the determinant line of a Fredholm operator D is the real vector space of dimension one defined as $\det(D) = \det(\mathrm{coker}(D))^\vee \otimes \det(\ker(D))$, where for a real finite-dimensional vector space V of dimension d, $\det(V) = \Lambda^d(V)$, and for a real vector space l of dimension one, its orientation torsor over the group ± 1 is $|l| = (l \setminus \{0\})/(\mathbb{R}_{>0})$.

data, to a real Cauchy–Riemann operator $D\colon W^{1,p,\epsilon}(E_\gamma) \to L^{p,\epsilon}(E_\gamma)$ corresponding to a J–unitary connection. Call the homotopy[11] $\{D^r\}_{r\in[0,1]}$, where $D^0 = D_\gamma$ and $D^1 = D$. But such operators D are in fact complex Cauchy–Riemann operators, their kernels and cokernels are complex vector spaces, and hence their determinants are canonically oriented. Hence $|D|$ and $|D_\gamma|$ admit canonical elements o and o_γ. By a similar argument, following the definition of the isomorphisms $\psi_{\gamma,\gamma'}\colon |D_\gamma| \xrightarrow{\sim} |D_{\gamma'}|$ from [85; 28, Chapter 8], with the key point being that orientation gluing is natural with respect to homotopies [85, Lemma 9.4.1], we see that $\psi_{\gamma,\gamma'}(o_\gamma) = o_{\gamma'}$. Therefore $\mathcal{O}$ admits a continuous section, and hence is trivial. This finishes the proof. □

5.3 Proof of Theorem J

Consider $\phi \in \mathrm{Ham}(M,\omega)\setminus\{\mathrm{id}\}$ such that $\phi^d = \mathrm{id}$, and let H be a Hamiltonian function generating ϕ. Then $\gamma(H) > 0$ by the nondegeneracy of the spectral norm. Since ϕ has finite order d we have that $\{\phi^t_{H^{(d)}}\}_{t\in[0,1]}$ is a Hamiltonian loop, which, in addition to the fact that M has rationality constant $\rho > 0$, implies that $\mathrm{Spec}(H^{(d)}) = a + \rho\cdot\mathbb{Z}$ for a real constant a. One can show by a quick calculation that $\mathrm{Spec}(\overline{H}^{(d)}) = -a + \rho\cdot\mathbb{Z}$.

Furthermore, observe that $\mathrm{Spec}(H) \subset \mathrm{Spec}(H^{(d)})/d$. In fact, if $c \in \mathrm{Spec}(H)$ then there exists a 1–periodic capped orbit $\overline{x} \in \widetilde{\mathcal{O}}(H)$ such that $\mathcal{A}_H(\overline{x}) = c$. Consequently, $\mathcal{A}_{H^{(d)}}(\overline{x}^{(d)}) = d\cdot\mathcal{A}_H(\overline{x}) = d\cdot c$, which implies the claim when added to the fact that $\overline{x}^d$ is a critical point of $\mathcal{A}_{H^{(d)}}$.

Finally, the above observations imply that $\gamma(H) \in (\rho/d)\cdot\mathbb{Z}$. In particular, the fact $\gamma(H) \geq 0$ implies $\gamma(H) \geq \rho/d$. Since H was an arbitrary Hamiltonian generating ϕ, it is clear that $\gamma(\phi) \geq \rho/d$. □

5.4 Proof of Theorem A

Similarly to the case of Theorem J, $\phi^d = \mathrm{id}$ implies, in the symplectically aspherical setting, that for a Hamiltonian H generating ϕ, we have $\mathrm{Spec}(H^{(d)}) = \{a\}$ and $\mathrm{Spec}(\overline{H}^{(d)}) = \{-a\}$ for a constant $a \in \mathbb{R}$, so $\mathrm{Spec}(H) \subset \mathrm{Spec}(H^{(d)})/d = \{a/d\}$ consists of at most one point. Since $\mathrm{Spec}(H)$ contains $c([M],H)$, we obtain that

[11] In fact we apply a homotopy depending smoothly on $x_0 \in \mathcal{F}$ from the symplectic connections on $x^*(TM) \to S^1$ for $x(t) = \phi^t_H(x_0)$ given by the linearized flow of ϕ^t_H to unitary connections, while at all times preserving their monodromies $D(\phi^1_H)_{x_0}$ over S^1 for all $x_0 \in \mathcal{F}$. This means in particular that the kernels of the asymptotic operators for fixed x_0 do not depend on the homotopy parameter r, up to natural identification. This and the compactness of $\mathcal{F}$ imply that the $\epsilon > 0$ above can be chosen sufficiently small that all the operators D^r along the homotopy are indeed Fredholm as operators $W^{1,p,\epsilon}(E_\gamma) \to L^{p,\epsilon}(E_\gamma)$.

$c([M], H) = a/d$. Similarly, $c([M], \overline{H}) = -a/d$. This means that $\gamma(H) = 0$ and hence $\gamma(\phi) = 0$, which implies by nondegeneracy of γ that $\phi = \mathrm{id}$. This finishes the proof. □

5.5 Proof of Theorem K

Consider $\phi \in \mathrm{Ham}(M, \omega) \setminus \{\mathrm{id}\}$ such that $\phi^p = \mathrm{id}$ for a prime number p. Fix a coefficient field $\mathbb{K}$. We show that there exists a positive integer m such that

$$\gamma(\phi^m) \geq \frac{\lfloor p/2 \rfloor}{p} \cdot \rho, \tag{16}$$

where $\lfloor p/2 \rfloor$ denotes the floor of $p/2$. We may suppose $p \geq 3$, since the case $p = 2$ is settled by Theorem J for $m = 1$. In this case, note that $\lfloor p/2 \rfloor = (p-1)/2$. Supposing that $\gamma(\phi) < \rho(p-1)/2p$, we can find a Hamiltonian H generating ϕ satisfying $\gamma(H) < \rho(p-1)/2p$. In the proof of Theorem J we saw that $\gamma(H)$ must be a positive integer multiple of ρ/p. Therefore, by Lemma 32 we can find a positive integer $r \leq (p-3)/2$ such that

$$l(H, \Gamma_H) = \gamma(H) = \frac{r\rho}{p}. \tag{17}$$

In particular, we have that $2r < p$, which combined with the fact that $p > 2$ implies that there exist integers a, b such that $a(2r) + bp = 1$. Observe that b must be an odd integer, since $a(2r)$ is even while p and 1 are odd. Let k be the integer such that $-b = 2k + 1$. Furthermore, note that $a \neq 0$, and set $m = |a|$. There are two cases to be considered, depending on the sign of the integer a:

- If $a > 0$, we have that $m(2r) - (2k+1)p = 1$, which implies

$$\frac{mr}{p} - \frac{p+1}{2p} = k, \tag{18}$$

where $(p+1)/2 = \lceil p/2 \rceil$. Furthermore, since m and p are coprime, Theorem G implies that

$$\mathrm{Spec}^{\mathrm{ess}}(H^{(m)}) = m \cdot \mathrm{Spec}^{\mathrm{ess}}(H) + \rho\mathbb{Z}. \tag{19}$$

Combining (17), (18) and (19) we obtain that there exist $c_0, c_1 \in \mathrm{Spec}^{\mathrm{ess}}(H)$ such that

$$mc_1 - mc_0 = \frac{mr\rho}{p} = \frac{(p+1)\rho}{2p} + k\rho.$$

In addition, $mc_1 + j\rho$ and $mc_0 + j\rho$ belong to the essential spectrum of $H^{(m)}$ for every integer j. We conclude that for each $\theta \in \mathrm{S}^1_\rho$ and $\mathrm{I} \in \Gamma_\theta$, there exists an integer l such that either

$$mc_1 + l\rho, mc_0 + (k+l)\rho \in \mathrm{I} \quad \text{or} \quad mc_1 + l\rho, mc_0 + (k+l+1)\rho \in \mathrm{I}.$$

Consequently,

$$\begin{aligned} l(H^{(m)}, \Gamma_\theta) &\geq \min\{mc_1 - mc_0 - k\rho, mc_0 - mc_1 + (k+1)\rho\} \\ &= \min\left\{\frac{(p+1)\rho}{2p}, \frac{(p-1)\rho}{2p}\right\} = \frac{(p-1)\rho}{2p}. \end{aligned}$$

Since θ was arbitrary, we conclude

$$l(\phi^m) \geq \frac{\lfloor p/2 \rfloor}{p} \cdot \rho. \tag{20}$$

- If $a < 0$, an analogous argument can be made to show that once again (20) is valid.

Hence, by Lemma 33 we obtain the inequality (16).

5.6 Proof of Theorem L

Consider a generalized pseudorotation ϕ as in Lemma 35. As a consequence of this lemma, we may suppose that there exist $c_1, c_2 \in \mathrm{Spec}^{\mathrm{ess}}(H)$ such that $c_1 - c_2 \in \rho\cdot(\mathbb{R}\setminus\mathbb{Q})$, otherwise $\gamma(\phi^m) \geq \rho$ for some positive integer m. Since the orbit of any irrational rotation in S^1 is dense, for every $\epsilon > 0$ there exists an integer m_ϵ such that

$$\frac{\rho}{2} - \epsilon < d_{S^1_\rho}([c_1], [m_\epsilon \cdot c_2]) \leq \frac{\rho}{2},$$

where for $x \in \mathbb{R}$ we denote by $[x] \in S^1_\rho = \mathbb{R}/\rho\mathbb{Z}$ its equivalence class, and $d_{S^1_\rho}$ is the distance function on S^1_ρ coming from the standard flat metric on $\mathbb{R}$. Therefore, arguing as in the proof of Theorem K we conclude

$$\sup_{k \in \mathbb{Z}_{>0}} \gamma(\phi^k) \geq \frac{\rho}{2}. \qquad \square$$

The proofs of Theorems D and I rely on the following observations regarding the mean-index. First, let $\tilde{\phi}$ be a lift of ϕ to the universal cover $\widetilde{\mathrm{Ham}}(M,\omega)$ of $\mathrm{Ham}(M,\omega)$. As our path-connected isolated fixed-point sets are weakly nondegenerate, if the capping $\overline{\mathscr{F}}$ of the generalized 1–periodic orbit $\mathscr{F}$ corresponding to an isolated fixed-point set

$\mathcal{F} \subset \mathrm{Fix}(\phi)$ carries a cohomology class μ of Conley–Zehnder index n in $HF^n(\tilde{\phi}) \cong QH^{2n}(M, \Lambda_{\mathbb{K}})$, for a coefficient field $\mathbb{K}$, then its mean-index $\Delta = \Delta(\tilde{\phi}, \bar{\mathscr{F}})$ satisfies $\Delta - n < n < \Delta + n$. Hence,

$$\Delta(\tilde{\phi}, \bar{\mathscr{F}}) \in (0, 2n). \tag{21}$$

Similarly, if $\bar{\mathscr{F}}$ carries a homology class $u \in HF_n(\tilde{\phi}) \cong QH_{2n}(M, \Lambda_{\mathbb{K}})$, then (21) holds. Both of these implications follow from Lemma 19, equation (7) and Section 3.2. We will specifically use the case $u = [M]$, which follows from Lemma 17.

5.7 Proof of Theorem D

We first treat the negative monotone case. Choose $H \in \mathcal{H}$ so that the path $\{\phi_H^t\}_{t \in [0,1]}$ represents the class $\tilde{\phi}$ lifting ϕ. Let k_i be the sequence associated to ϕ as a generalized perfect Hamiltonian diffeomorphism. By the pigeonhole principle applied to the subsequence l_i with $l_i \,|\, l_{i+1}$ for all i, there exists an isolated fixed-point set $\mathcal{F} \subset \mathrm{Fix}(\phi)$, and an increasing subsequence of k_i, which we renumber and denote by r_i, such that $c([M], H^{(r_i)})$ is carried by a capping $\bar{\mathscr{G}}_i$ of the isolated set of 1–periodic orbits of the r_i–iterated Hamiltonian $H^{(r_i)}$ corresponding to $\mathcal{F}^{(r_i)}$. Set $\bar{\mathscr{G}} = \bar{\mathscr{G}}_1$. Since r_1 divides all r_i, by taking a power of ϕ we can assume that $r_1 = 1$.

Write $\bar{\mathscr{G}}_i$ as a recapped iteration of $\bar{\mathscr{G}}$, ie

$$\bar{\mathscr{G}}_i = \bar{\mathscr{G}}^{(r_i)} \# A_i. \tag{22}$$

We claim that for r_i large, $\omega(A_i) \geq 0$ and $c_1(A_i) > 0$, contradicting negative monotonicity. Indeed, write $\mathcal{A}_i$ for the action functional of $H^{(r_i)}$, and $\mathcal{A} := \mathcal{A}_1$. Then by (22) and the triangle inequality for spectral invariants,

$$r_i \mathcal{A}(\bar{\mathscr{G}}) - \omega(A_i) = \mathcal{A}_i(\bar{\mathscr{G}}_i) = c([M], H^{(r_i)}) \leq r_i c([M], H) = r_i \mathcal{A}(\bar{\mathscr{G}}). \tag{23}$$

Hence,

$$\omega(A_i) \geq 0.$$

However, as $\bar{\mathscr{G}}_i$ carries $c([M], H^{(r_i)})$, by (21) we have $\Delta(H^{(r_i)}, \bar{\mathscr{G}}_i) \in (0, 2n)$ and also $\Delta(H, \bar{\mathscr{G}}) \in (0, 2n)$. Hence $r_i \Delta(H, \bar{\mathscr{G}}) > 2n$ for r_i large enough, and

$$2n > \Delta(H^{(r_i)}, \bar{\mathscr{G}}_i) = r_i \Delta(H, \bar{\mathscr{G}}) - 2c_1(A_i). \tag{24}$$

Therefore

$$c_1(A_i) > 0,$$

which finishes the proof.

We now prove the symplectic Calabi–Yau case of the theorem. In this case, the mean-index of each capped fixed-point set $\bar{\mathscr{F}}$ does not depend on the capping. Hence we write $\Delta(H, \mathcal{F})$ for each generalized fixed point $\mathcal{F}$ for this mean-index. Then for each positive sequence $k_i \to \infty$ of iterations with ϕ^{k_i} having a fixed finite number of weakly nondegenerate generalized fixed points, we argue as follows. For each $\mathcal{F} \in \pi_0(\mathrm{Fix}(\phi))$,

$$\Delta(H^{(k_i)}, \mathcal{F}^{(k_i)}) = (k_i/k_1)\Delta(H^{(k_1)}, \mathcal{F}^{(k_1)}).$$

Hence, if $\Delta(H^{(k_1)}, \mathcal{F}^{(k_1)}) > 0$ then $\Delta(H^{(k_i)}, \mathcal{F}^{(k_i)}) > 2n$ for all k_i sufficiently large, and if $\Delta(H^{(k_1)}, \mathcal{F}^{(k_1)}) \leq 0$ then $\Delta(H^{(k_i)}, \mathcal{F}^{(k_i)}) \leq 0$ for all k_i. Now, as each $\mathcal{F}$ is weakly nondegenerate, we obtain by the same argument as for the proof of the support property of local Floer homology, Lemma 16, that for all k_i sufficiently large, $H^{(k_i)}$ admits a C^2–small nondegenerate Hamiltonian perturbation H_i without capped periodic orbits of Conley–Zehnder index n. However, this is in contradiction to the existence of the PSS isomorphism. Specifically, in this case $HF_n(H_i) = 0$ by definition of Floer homology, and by the PSS isomorphism $HF_n(H_i) \cong QH_{2n}(M) \neq 0$. Indeed $[M] \in QH_{2n}(M)$ is nonzero. □

The following result was first proven in [93] in the setting of a *pseudorotation* assuming that the quantum Steenrod square of the point cohomology class is undeformed, or in other words that (M, ω) is not $\mathbb{F}_2$–Steenrod uniruled. We observe that the same statement holds for generalized pseudorotations, with essentially the same proof, and with a small modification following [97], for all primes p. Here $\mu \in QH^{2n}(M, \Lambda_{\mathbb{F}_p})$ denotes the cohomology class Poincaré dual to the point.

Theorem O *Let ψ be a generalized $\mathbb{F}_p$ pseudorotation with sequence $k_j = p^{j-1}$ of a closed monotone symplectic manifold (M, ω) that is not $\mathbb{F}_p$–Steenrod uniruled. Then*

$$c(\mu, \tilde{\psi}^p) \geq p \cdot c(\mu, \tilde{\psi}) \tag{25}$$

for each $\tilde{\psi} \in \widetilde{\mathrm{Ham}}(M, \omega)$ covering ψ.

We proceed to the proof of Theorem I.

5.8 Proof of Theorem I

Choose $H \in \mathcal{H}$ so that the path $\{\phi_H^t\}_{t \in [0,1]}$ represents the class $\tilde{\phi}$ lifting ϕ. By the pigeonhole principle, there exists an isolated fixed-point set $\mathcal{F} \subset \mathrm{Fix}(\phi)$, and an increasing sequence k_i such that $c(\mu, H^{(r_i)})$ for $r_i = p^{k_i}$ is carried by a capping $\bar{\mathscr{G}}_i$ of the isolated set of 1–periodic orbits of the r_i–iterated Hamiltonian $H^{(r_i)}$ corresponding to $\mathcal{F}^{(r_i)}$. By taking a power of ϕ, we can assume that $r_1 = 1$, and set $\overline{\mathscr{G}} = \overline{\mathscr{G}}_1$. Write $\overline{\mathscr{G}}_i$ as a recapped iteration of $\overline{\mathscr{G}}$, ie

$$\overline{\mathscr{G}}_i = \overline{\mathscr{G}}^{(r_i)} \# A_i. \tag{26}$$

We claim that for r_i large, we get $\omega(A_i) \le 0$ and $c_1(A_i) > 0$, contradicting monotonicity. Indeed, write $\mathcal{A}_i$ for the action functional of $H^{(r_i)}$, and set $\mathcal{A} := \mathcal{A}_1$. Then by (26) and Theorem O,

$$r_i \mathcal{A}(\overline{\mathscr{G}}) - \omega(A_i) = \mathcal{A}_i(\overline{\mathscr{G}}_i) = c(\mu, H^{(r_i)}) \ge r_i c(\mu, H) = r_i \mathcal{A}(\overline{\mathscr{G}}).$$

Hence,

$$\omega(A_i) \le 0.$$

However, as $\overline{\mathscr{G}}_i$ carries $c(\mu, H^{(r_i)})$, by (21) we have $\Delta(H^{(r_i)}, \overline{\mathscr{G}}_i) \in (0, 2n)$ and also $\Delta(H, \overline{\mathscr{G}}) \in (0, 2n)$. Hence $r_i \Delta(H, \overline{\mathscr{G}}) > 2n$ for r_i large enough, and

$$2n > \Delta(H^{(r_i)}, \overline{\mathscr{G}}_i) = r_i \Delta(H, \overline{\mathscr{G}}) - 2c_1(A_i).$$

Therefore,

$$c_1(A_i) > 0.$$

5.9 Proof of Theorem G

Suppose that $\phi \in \mathrm{Ham}(M, \omega) \setminus \{\mathrm{id}\}$ is of prime order $q \ge 2$. Let $p \ge 2$ be a prime different from q. In particular, $\phi^{j \cdot p^k} \ne \mathrm{id}$ for all $k \in \mathbb{Z}$ and $1 \le j \le q-1$.

Write

$$B(\phi, \mathbb{F}_p) = \max_{1 \le j \le q-1} \beta_{\mathrm{tot}}(\phi^j, \mathbb{F}_p).$$

By Theorem N we obtain for $1 \le j \le q-1$ that

$$B(\phi, \mathbb{F}_p) \ge \beta_{\mathrm{tot}}(\phi^{j \cdot p^k}, \mathbb{F}_p) \ge p^k \beta_{\mathrm{tot}}(\phi^j, \mathbb{F}_p).$$

Choosing a sufficiently large positive k, this implies that for all $1 \le j \le q-1$,

$$\beta_{\mathrm{tot}}(\phi^j, \mathbb{F}_p) = 0,$$

whence by Proposition 23 all such ϕ^j are generalized $\mathbb{F}_p$ pseudorotations. They are weakly nondegenerate by Theorem C. In other words, the equality

$$\mathrm{Spec}^{\mathrm{vis}}(H; \mathbb{F}_p) = \mathrm{Spec}^{\mathrm{ess}}(H; \mathbb{F}_p)$$

follows directly from the fact that $\beta_{\mathrm{tot}}(\phi, \mathbb{F}_p) = 0$. This finishes the proof of part (i).

Let us prove that $\mathrm{Spec}^{\mathrm{vis}}(H^{(k)};\mathbb{Q}) = k\cdot\mathrm{Spec}^{\mathrm{vis}}(H;\mathbb{Q})+\rho\cdot\mathbb{Z}$ for all $k\in\mathbb{Z}$ coprime with q. By the universal coefficient formula in local Floer homology, it is sufficient to prove the identity $\mathrm{Spec}^{\mathrm{vis}}(H^{(k)};\mathbb{F}_p) = k\cdot\mathrm{Spec}^{\mathrm{vis}}(H;\mathbb{F}_p)+\rho\cdot\mathbb{Z}$ for coefficients in $\mathbb{F}_p$ for an infinite sequence of primes p. Consider the primes p for which $p=a \pmod q$, where $a\in(\mathbb{F}_q)^*$ is a cyclic generator of the multiplicative group $(\mathbb{F}_q)^*=\mathrm{GL}(1,\mathbb{F}_q)$ of $\mathbb{F}_q$. In this case the set $\{\phi^{p^j}\mid j\in\mathbb{Z}_{\geq 0}\}$ coincides with

$$\{\phi^k\mid 1\leq k\leq q-1\}=\{\phi^k\mid k\neq 0 \pmod q\}.$$

Let $\bar{\mathscr{F}}$ be a capped generalized periodic orbit of H. It is enough to prove that

$$\dim_{\mathbb{F}_p} HF^{\mathrm{loc}}(H^{(p^j)},\bar{\mathscr{F}}^{(p^j)})=\dim_{\mathbb{F}_p} HF^{\mathrm{loc}}(H,\bar{\mathscr{F}})$$

for all $j\in\mathbb{Z}_{\geq 0}$. Indeed, as explained above, each capped generalized fixed point of ϕ^k is a recapping of a p^j–iterated capped generalized fixed point of ϕ.

We know by the Smith inequality in generalized local Floer homology, Proposition 24, that $\dim_{\mathbb{F}_p} HF^{\mathrm{loc}}(H^{(p^j)},\bar{\mathscr{F}}^{(p^j)})$ is an increasing function of j. However, by the finite-order condition it takes only a finite number of values. Therefore it must be identically constant. This finishes the proof of part (ii).

Now we prove part (iii) relying on Proposition 5. First let $p=q$. Then for $\psi=\phi^k$, with k coprime to p,

$$N(\psi,\mathbb{F}_p)=\sum\dim_{\mathbb{F}_p} HF^{\mathrm{loc}}(\psi,\mathcal{F})=\sum\dim_{\mathbb{F}_p} H(\mathcal{F};\mathbb{F}_p), \tag{27}$$

the sum running over all contractible generalized fixed points $\mathcal{F}$ of ψ, since by Proposition 5,

$$HF^{\mathrm{loc}}(\psi,\mathcal{F})\cong H(\mathcal{F};\mathbb{F}_p)$$

for all generalized fixed points $\mathcal{F}$. We remark that $H(\mathcal{F};\mathbb{F}_p)\neq 0$. By Proposition 23, we know that

$$N(\psi,\mathbb{F}_p)\geq\dim_{\mathbb{F}_p} H(M;\mathbb{F}_p).$$

On the other hand, by the classical Smith inequality [99; 25; 4], we have

$$\sum\dim_{\mathbb{F}_p} H(\mathcal{F};\mathbb{F}_p)\leq\dim_{\mathbb{F}_p} H(M;\mathbb{F}_p), \tag{28}$$

the sum running over all the generalized fixed points of ψ. This yields

$$N(\psi,\mathbb{F}_p)=\dim_{\mathbb{F}_p} H(M;\mathbb{F}_p).$$

This finishes the proof of the first statement of part (iii).

To prove the second statement of part (iii), we first note that for $\mathrm{char}(\mathbb{K}) = p$,

$$\mathrm{Spec}^{\mathrm{ess}}(H,\mathbb{K}) = \mathrm{Spec}^{\mathrm{ess}}(H,\mathbb{F}_p) \quad \text{and} \quad \mathrm{Spec}^{\mathrm{vis}}(H,\mathbb{K}) = \mathrm{Spec}^{\mathrm{vis}}(H,\mathbb{F}_p)$$

by Proposition 23, and the equality

$$\mathrm{Spec}^{\mathrm{ess}}(H,\mathbb{F}_p) = \mathrm{Spec}^{\mathrm{vis}}(H,\mathbb{F}_p)$$

follows by the first statement of part (iii) and Proposition 23. For the last part, we note that by Proposition 5, $\mathrm{Spec}^{\mathrm{vis}}(H,\mathbb{K}) = \mathrm{Spec}(H)$ because

$$\dim HF^{\mathrm{loc}}(H,\bar{\mathscr{F}}) = \dim HF^{\mathrm{loc}}(\phi,\mathcal{F}) = \dim H(\mathcal{F};\mathbb{K}) > 0$$

for all capped contractible generalized 1–periodic orbits $\bar{\mathscr{F}}$ of H. Now for k coprime to q, $\mathrm{Spec}(H^{(k)}) = \{\mathcal{A}_{H^{(k)}}(\bar{\mathscr{F}}^{(k)} \# A)\}$, where the set runs over all $A \in \Gamma$, and $\bar{\mathscr{F}}$ runs over all capped contractible generalized 1–periodic orbits $\bar{\mathscr{F}}$ of H. Indeed, all the contractible generalized fixed points of ϕ^k are of the form $\mathcal{F}^{(k)}$ for $\mathcal{F}$ a contractible generalized fixed point of ϕ, and the identity quickly follows. Now using the homogeneity and the recapping properties of the action functional, we obtain

$$\mathrm{Spec}(H^{(k)}) = k\cdot\mathrm{Spec}(H) + \rho\cdot\mathbb{Z}.$$

Combined with the identities $\mathrm{Spec}^{\mathrm{ess}}(H^{(k)};\mathbb{K}) = \mathrm{Spec}^{\mathrm{vis}}(H^{(k)};\mathbb{K}) = \mathrm{Spec}(H^{(k)})$ and $\mathrm{Spec}^{\mathrm{ess}}(H;\mathbb{K}) = \mathrm{Spec}^{\mathrm{vis}}(H;\mathbb{K}) = \mathrm{Spec}(H)$, this finishes the proof.

5.10 Proof of Theorem H

First assume that ψ is of prime order p. Then the proof follows from equations (27) and (28). Indeed, the upper bound holds for *all* the generalized fixed points of ψ, and the lower bound $N(\psi,\mathbb{F}_p) \geq \dim_{\mathbb{F}_p} H(M;\mathbb{F}_p)$ takes into account *only contractible* generalized fixed points. If ψ had a noncontractible generalized fixed point, it would contribute $\dim_{\mathbb{F}_p} H(\mathcal{F};\mathbb{F}_p) > 0$ to the sum, making the equality impossible. Alternatively, one can argue by means of a suitable generalization of Theorem N with $p \neq q$.

Now suppose that ψ is of order p^k, with $k \geq 1$. As in Section 5.9, by Proposition 5

$$HF^{\mathrm{loc}}(\psi,\mathcal{F}) \cong H(\mathcal{F};\mathbb{F}_p)$$

for all generalized fixed points $\mathcal{F}$; and

$$N(\psi,\mathbb{F}_p) = \sum \dim_{\mathbb{F}_p} HF^{\mathrm{loc}}(\psi,\mathcal{F}) = \sum \dim_{\mathbb{F}_p} H(\mathcal{F};\mathbb{F}_p), \tag{29}$$

the sum running over all contractible generalized fixed points $\mathcal{F}$ of ψ. Moreover, by Proposition 23, we have

$$N(\psi, \mathbb{F}_p) \geq \dim_{\mathbb{F}_p} H(M; \mathbb{F}_p).$$

Finally, by the Smith inequality for finite p–groups [99; 25; 4] we again have

$$\sum \dim_{\mathbb{F}_p} H(\mathcal{F}; \mathbb{F}_p) \leq \dim_{\mathbb{F}_p} H(M; \mathbb{F}_p), \tag{30}$$

the sum running over all the generalized fixed points of ψ. Now, as in the case of order p, if ψ had a noncontractible generalized fixed point, it would contribute $\dim_{\mathbb{F}_p} H(\mathcal{F}; \mathbb{F}_p) > 0$ to the sum, making it impossible for (30) and (29) to hold simultaneously.

For ψ of arbitrary integer order $d = p_1^{k_1} \cdots p_m^{k_m}$, we proceed by induction. We have already shown the base of induction. Now we suppose that the result is true for all orders having at most $m-1$ distinct prime divisors, and prove it for ψ of order d as above. Then $\psi_1 = \psi^{p_1^{k_1}}$ is of order $d/p_1^{k_1}$, which has exactly $m-1$ prime divisors, and hence by induction all the fixed points of ψ_1 are contractible. This implies that the order of the homotopy class of each fixed point of ψ divides $p_1^{k_1}$. In the same way, we obtain that this order also divides $p_j^{k_j}$ for all $1 \leq j \leq m$, and therefore it divides $\gcd(p_1^{k_1}, \dots, p_m^{k_m}) = 1$. Therefore each fixed point of ψ is contractible. This finishes the proof.

Acknowledgements We thank Leonid Polterovich and Dusa McDuff for inspiring discussions, Claude Viterbo for a helpful suggestion related to Remark 30, and Kai Cieliebak for an interesting question that prompted Remark 7. We thank Kaoru Ono and John Pardon for helpful communications. Finally, we thank the referee for a careful reading of our paper.

At the University of Montréal, Shelukhin was supported by an NSERC Discovery Grant, by the Fonds de recherche du Québec - Nature et technologies, and by the Fondation Courtois. Atallah was supported by the Fondation Courtois.

References

[1] **M Abouzaid**, *Symplectic cohomology and Viterbo's theorem*, from "Free loop spaces in geometry and topology" (J Latschev, A Oancea, editors), IRMA Lect. Math. Theor. Phys. 24, Eur. Math. Soc., Zürich (2015) 271–485 MR Zbl

[2] **M Audin**, *Torus actions on symplectic manifolds*, revised edition, Progr. Math. 93, Birkhäuser, Basel (2004) MR Zbl

[3] **P Biran**, **L Polterovich**, **D Salamon**, *Propagation in Hamiltonian dynamics and relative symplectic homology*, Duke Math. J. 119 (2003) 65–118 MR Zbl

[4] **A Borel**, *Seminar on transformation groups*, Annals of Mathematics Studies 46, Princeton Univ. Press (1960) MR Zbl

[5] **G E Bredon**, *Introduction to compact transformation groups*, Pure and Applied Mathematics 46, Academic, New York (1972) MR Zbl

[6] **L Buhovsky**, **V Humilière**, **S Seyfaddini**, *The action spectrum and C^0 symplectic topology*, Math. Ann. 380 (2021) 293–316 MR Zbl

[7] **E Çineli**, **V L Ginzburg**, **B Z Gürel**, *From pseudo-rotations to holomorphic curves via quantum Steenrod squares*, Int. Math. Res. Not. 2022 (2022) 2274–2297 MR

[8] **F Charest**, **C Woodward**, *Floer trajectories and stabilizing divisors*, J. Fixed Point Theory Appl. 19 (2017) 1165–1236 MR Zbl

[9] **W Chen**, *Group actions on* 4*–manifolds: some recent results and open questions*, from "Proceedings of the Gökova Geometry–Topology Conference 2009" (S Akbulut, D Auroux, T Önder, editors), International, Somerville, MA (2010) 1–21 MR Zbl

[10] **W Chen**, **S Kwasik**, *Symplectic symmetries of* 4*–manifolds*, Topology 46 (2007) 103–128 MR Zbl

[11] **R Chiang**, **L Kessler**, *Cyclic actions on rational ruled symplectic four-manifolds*, Transform. Groups 24 (2019) 987–1000 MR Zbl

[12] **R Chiang**, **L Kessler**, *Homologically trivial symplectic cyclic actions need not extend to Hamiltonian circle actions*, J. Topol. Anal. 12 (2020) 1047–1071 MR Zbl

[13] **K Cieliebak**, **K Mohnke**, *Symplectic hypersurfaces and transversality in Gromov–Witten theory*, J. Symplectic Geom. 5 (2007) 281–356 MR Zbl

[14] **M De Gosson**, **S De Gosson**, **P Piccione**, *On a product formula for the Conley–Zehnder index of symplectic paths and its applications*, Ann. Global Anal. Geom. 34 (2008) 167–183 MR Zbl

[15] **T Delzant**, *Sous-algèbres de dimension finie de l'algèbre des champs hamiltoniens*, preprint (1995) `http://www-irma.u-strasbg.fr/~delzant/preprint.html`

[16] **A Dress**, *Newman's theorems on transformation groups*, Topology 8 (1969) 203–207 MR Zbl

[17] **M Entov**, **L Polterovich**, *Calabi quasimorphism and quantum homology*, Int. Math. Res. Not. 2003 (2003) 1635–1676 MR Zbl

[18] **M Entov**, **L Polterovich**, *Rigid subsets of symplectic manifolds*, Compos. Math. 145 (2009) 773–826 MR Zbl

[19] **J W Fish**, *Target-local Gromov compactness*, Geom. Topol. 15 (2011) 765–826 MR Zbl

[20] **J W Fish**, **H Hofer**, *Lectures on polyfolds and symplectic field theory*, preprint (2018) arXiv 1808.07147

[21] **A Floer**, *Proof of the Arnol'd conjecture for surfaces and generalizations to certain Kähler manifolds*, Duke Math. J. 53 (1986) 1–32 MR Zbl

[22] **A Floer**, *Morse theory for fixed points of symplectic diffeomorphisms*, Bull. Amer. Math. Soc. 16 (1987) 279–281 MR Zbl

[23] **A Floer**, *Symplectic fixed points and holomorphic spheres*, Comm. Math. Phys. 120 (1989) 575–611 MR Zbl

[24] **A Floer**, *Witten's complex and infinite-dimensional Morse theory*, J. Differential Geom. 30 (1989) 207–221 MR Zbl

[25] **E E Floyd**, *On periodic maps and the Euler characteristics of associated spaces*, Trans. Amer. Math. Soc. 72 (1952) 138–147 MR Zbl

[26] **U Frauenfelder**, *The Arnold–Givental conjecture and moment Floer homology*, Int. Math. Res. Not. 2004 (2004) 2179–2269 MR Zbl

[27] **K Fukaya**, *Morse homotopy and its quantization*, from "Geometric topology" (W H Kazez, editor), AMS/IP Stud. Adv. Math. 2, Amer. Math. Soc., Providence, RI (1997) 409–440 MR Zbl

[28] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Lagrangian intersection Floer theory: anomaly and obstruction, II*, AMS/IP Stud. Adv. Math. 46, Amer. Math. Soc., Providence, RI (2009) MR Zbl

[29] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Technical details on Kuranishi structure and virtual fundamental chain*, preprint (2012) arXiv 1209.4410

[30] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Exponential decay estimates and smoothness of the moduli space of pseudoholomorphic curves*, preprint (2016) arXiv 1603.07026 To appear under Mem. Amer. Math. Soc.

[31] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Construction of Kuranishi structures on the moduli spaces of pseudo holomorphic disks, I*, from "Surveys in differential geometry 2017: celebrating the 50th anniversary of the Journal of Differential Geometry" (H-D Cao, J Li, R M Schoen, S-T Yau, editors), Surv. Differ. Geom. 22, International, Somerville, MA (2018) 133–190 MR Zbl

[32] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Kuranishi structures and virtual fundamental chains*, Springer (2020) MR Zbl

[33] **K Fukaya**, **K Ono**, *Arnold conjecture and Gromov–Witten invariant*, Topology 38 (1999) 933–1048 MR Zbl

[34] **V L Ginzburg**, **B Z Gürel**, *Conley conjecture for negative monotone symplectic manifolds*, Int. Math. Res. Not. 2012 (2012) 1748–1767 MR Zbl

[35] **V L Ginzburg**, **B Z Gürel**, *Hamiltonian pseudo-rotations of projective spaces*, Invent. Math. 214 (2018) 1081–1130 MR Zbl

[36] **V L Ginzburg**, **B Z Gürel**, *Conley conjecture revisited*, Int. Math. Res. Not. 2019 (2019) 761–798 MR Zbl

[37] **J González-Meneses**, *Basic results on braid groups*, Ann. Math. Blaise Pascal 18 (2011) 15–59 MR Zbl

[38] **M Gromov**, *Pseudo holomorphic curves in symplectic manifolds*, Invent. Math. 82 (1985) 307–347 MR Zbl

[39] **D Hein**, *The Conley conjecture for irrational symplectic manifolds*, J. Symplectic Geom. 10 (2012) 183–202 MR Zbl

[40] **H Hofer**, *On the topological properties of symplectic maps*, Proc. Roy. Soc. Edinburgh Sect. A 115 (1990) 25–38 MR Zbl

[41] **H Hofer**, **D A Salamon**, *Floer homology and Novikov rings*, from "The Floer memorial volume" (H Hofer, C H Taubes, A Weinstein, E Zehnder, editors), Progr. Math. 133, Birkhäuser, Basel (1995) 483–524 MR Zbl

[42] **H Hofer**, **K Wysocki**, **E Zehnder**, *sc-smoothness, retractions and new models for smooth spaces*, Discrete Contin. Dyn. Syst. 28 (2010) 665–788 MR Zbl

[43] **H Hofer**, **K Wysocki**, **E Zehnder**, *Applications of polyfold theory, I: The polyfolds of Gromov–Witten theory*, Mem. Amer. Math. Soc. 1179, Amer. Math. Soc., Providence, RI (2017) MR Zbl

[44] **H Hofer**, **K Wysocki**, **E Zehnder**, *Polyfold and Fredholm theory*, Ergebnisse der Math. (3) 72, Springer (2021) MR Zbl

[45] **W-y Hsiang**, *Cohomology theory of topological transformation groups*, Ergebnisse der Math. (3) 85, Springer (1975) MR Zbl

[46] **Y Karshon**, **J S Pearl**, *Shortening the Hofer length of Hamiltonian circle actions*, J. Symplectic Geom. 13 (2015) 209–259 MR Zbl

[47] **Y Kawamoto**, *On C^0–continuity of the spectral norm for symplectically non-aspherical manifolds*, Int. Math. Res. Not. 2022 (2022) 17187–17230 MR Zbl

[48] **A Kislev**, **E Shelukhin**, *Bounds on spectral norms and barcodes*, Geom. Topol. 25 (2021) 3257–3350 MR Zbl

[49] **J Kędra**, *Remarks on the flux groups*, Math. Res. Lett. 7 (2000) 279–285 MR Zbl

[50] **J Kędra**, **D Kotschick**, **S Morita**, *Crossed flux homomorphisms and vanishing theorems for flux groups*, Geom. Funct. Anal. 16 (2006) 1246–1273 MR Zbl

[51] **F Lalonde**, **D McDuff**, *The geometry of symplectic energy*, Ann. of Math. 141 (1995) 349–371 MR Zbl

[52] **R Leclercq**, **F Zapolsky**, *Spectral invariants for monotone Lagrangians*, J. Topol. Anal. 10 (2018) 627–700 MR Zbl

[53] **J Li**, **T-J Li**, **W Wu**, *Braid groups and symplectic mapping class groups of rational surfaces*, in preparation

[54] **G Liu**, *Associativity of quantum multiplication*, Comm. Math. Phys. 191 (1998) 265–282 MR Zbl

[55] **G Liu**, **G Tian**, *Floer homology and Arnold conjecture*, J. Differential Geom. 49 (1998) 1–74 MR Zbl

[56] **D McDuff**, *Symplectic diffeomorphisms and the flux homomorphism*, Invent. Math. 77 (1984) 353–366 MR Zbl

[57] **D McDuff**, *Hamiltonian S^1–manifolds are uniruled*, Duke Math. J. 146 (2009) 449–507 MR Zbl

[58] **D McDuff**, **D Salamon**, *J–holomorphic curves and symplectic topology*, 2nd edition, American Mathematical Society Colloquium Publications 52, Amer. Math. Soc., Providence, RI (2012) MR Zbl

[59] **D McDuff**, **D Salamon**, *Introduction to symplectic topology*, 3rd edition, Oxford Univ. Press (2017) MR Zbl

[60] **D McDuff**, **J Slimowitz**, *Hofer–Zehnder capacity and length minimizing Hamiltonian paths*, Geom. Topol. 5 (2001) 799–830 MR Zbl

[61] **M McLean**, *Local Floer homology and infinitely many simple Reeb orbits*, Algebr. Geom. Topol. 12 (2012) 1901–1923 MR Zbl

[62] **J W Morgan**, *The Smith conjecture*, from "The Smith conjecture" (J W Morgan, H Bass, editors), Pure Appl. Math. 112, Academic, Orlando, FL (1984) 3–6 MR Zbl

[63] **M H A Newman**, *A theorem on periodic transformations of spaces*, Q. J. Math., Oxf. Ser. 2 (1931) 1–8 Zbl

[64] **Y-G Oh**, *Construction of spectral invariants of Hamiltonian paths on closed symplectic manifolds*, from "The breadth of symplectic and Poisson geometry" (J E Marsden, T S Ratiu, editors), Progr. Math. 232, Birkhäuser, Boston, MA (2005) 525–570 MR Zbl

[65] **Y-G Oh**, *Spectral invariants, analysis of the Floer moduli space, and geometry of the Hamiltonian diffeomorphism group*, Duke Math. J. 130 (2005) 199–295 MR Zbl

[66] **Y-G Oh**, *Lectures on Floer theory and spectral invariants of Hamiltonian flows*, from "Morse theoretic methods in nonlinear analysis and in symplectic topology" (P Biran, O Cornea, F Lalonde, editors), NATO Sci. Ser. II Math. Phys. Chem. 217, Springer (2006) 321–416 MR Zbl

[67] **Y-G Oh**, *Symplectic topology and Floer homology, I: Symplectic geometry and pseudo-holomorphic curves*, New Mathematical Monographs 28, Cambridge Univ. Press (2015) MR Zbl

[68] **Y Ostrover**, *Calabi quasi-morphisms for some non-monotone symplectic manifolds*, Algebr. Geom. Topol. 6 (2006) 405–434 MR Zbl

[69] **J Pardon**, *An algebraic approach to virtual fundamental cycles on moduli spaces of pseudo-holomorphic curves*, Geom. Topol. 20 (2016) 779–1034 MR Zbl

[70] **S Piunikhin**, **D Salamon**, **M Schwarz**, *Symplectic Floer–Donaldson theory and quantum cohomology*, from "Contact and symplectic geometry" (C B Thomas, editor), Publ. Newton Inst. 8, Cambridge Univ. Press (1996) 171–200 MR Zbl

[71] **L Polterovich**, *The geometry of the group of symplectic diffeomorphisms*, Birkhäuser, Basel (2001) MR Zbl

[72] **L Polterovich**, *Growth of maps, distortion in groups and symplectic geometry*, Invent. Math. 150 (2002) 655–686 MR Zbl

[73] **L Polterovich**, **D Rosen**, *Function theory on symplectic manifolds*, CRM Monograph Series 34, Amer. Math. Soc., Providence, RI (2014) MR Zbl

[74] **L Polterovich**, **E Shelukhin**, *Autonomous Hamiltonian flows, Hofer's geometry and persistence modules*, Selecta Math. 22 (2016) 227–296 MR Zbl

[75] **L Polterovich**, **E Shelukhin**, **V Stojisavljević**, *Persistence modules with operators in Morse and Floer theory*, Mosc. Math. J. 17 (2017) 757–786 MR Zbl

[76] **M Poźniak**, *Floer homology, Novikov rings and clean intersections*, from "Northern California symplectic geometry seminar" (Y Eliashberg, D Fuchs, T Ratiu, A Weinstein, editors), Amer. Math. Soc. Transl. Ser. 2 196, Amer. Math. Soc., Providence, RI (1999) 119–181 MR Zbl

[77] **I Mundet i Riera**, *Finite subgroups of Ham and Symp*, Math. Ann. 370 (2018) 331–380 MR Zbl

[78] **I Mundet i Riera**, *Finite group actions on homology spheres and manifolds with nonzero Euler characteristic*, J. Topol. 12 (2019) 744–758 MR Zbl

[79] **J Robbin**, **D Salamon**, *The Maslov index for paths*, Topology 32 (1993) 827–844 MR Zbl

[80] **Y Ruan**, *Virtual neighborhoods and pseudo-holomorphic curves*, Turkish J. Math. 23 (1999) 161–231 MR Zbl

[81] **Y Ruan**, **G Tian**, *A mathematical theory of quantum cohomology*, Math. Res. Lett. 1 (1994) 269–278 MR Zbl

[82] **Y Ruan**, **G Tian**, *A mathematical theory of quantum cohomology*, J. Differential Geom. 42 (1995) 259–367 MR Zbl

[83] **D Salamon**, *Lectures on Floer homology*, from "Symplectic geometry and topology" (Y Eliashberg, L Traynor, editors), IAS/Park City Math. Ser. 7, Amer. Math. Soc., Providence, RI (1999) 143–229 MR Zbl

[84] **D Salamon**, **E Zehnder**, *Morse theory for periodic solutions of Hamiltonian systems and the Maslov index*, Comm. Pure Appl. Math. 45 (1992) 1303–1360 MR Zbl

[85] **F Schmaschke**, *Abelianization and Floer homology of Lagrangians in clean intersection*, PhD thesis, Leipzig University (2016) `https://d-nb.info/1240847084/34`

[86] **M Schwarz**, *On the action spectrum for closed symplectically aspherical manifolds*, Pacific J. Math. 193 (2000) 419–461 MR Zbl

[87] **P Seidel**, π_1 *of symplectic automorphism groups and invertibles in quantum homology rings*, Geom. Funct. Anal. 7 (1997) 1046–1095 MR Zbl

[88] **P Seidel**, *Symplectic Floer homology and the mapping class group*, Pacific J. Math. 206 (2002) 219–229 MR Zbl

[89] **P Seidel**, *Fukaya categories and Picard–Lefschetz theory*, Eur. Math. Soc., Zürich (2008) MR Zbl

[90] **P Seidel**, *The equivariant pair-of-pants product in fixed point Floer cohomology*, Geom. Funct. Anal. 25 (2015) 942–1007 MR Zbl

[91] **P Seidel**, *Formal groups and quantum cohomology*, preprint (2019) arXiv 1910.08990 To appear in Geom. Topol.

[92] **P Seidel**, **N Wilkins**, *Covariant constancy of quantum Steenrod operations*, J. Fixed Point Theory Appl. 24 (2022) art. id. 52 MR Zbl

[93] **E Shelukhin**, *Pseudo-rotations and Steenrod squares*, J. Mod. Dyn. 16 (2020) 289–304 MR Zbl

[94] **E Shelukhin**, *Pseudo-rotations and Steenrod squares revisited*, Math. Res. Lett. 28 (2021) 1255–1261 MR Zbl

[95] **E Shelukhin**, *On the Hofer–Zehnder conjecture*, Ann. of Math. 195 (2022) 775–839 MR Zbl

[96] **E Shelukhin**, *Viterbo conjecture for Zoll symmetric spaces*, Invent. Math. 230 (2022) 321–373 MR Zbl

[97] **E Shelukhin**, **N Wilkins**, *Quantum Steenrod powers and Hamiltonian maps*, in preparation

[98] **E Shelukhin**, **J Zhao**, *The $\mathbb{Z}/p\mathbb{Z}$–equivariant product-isomorphism in fixed point Floer cohomology*, J. Symplectic Geom. 19 (2021) 1101–1188 MR Zbl

[99] **P A Smith**, *Transformations of finite period*, Ann. of Math. 39 (1938) 127–164 MR Zbl

[100] **P A Smith**, *Transformations of finite period, III: Newman's theorem*, Ann. of Math. 42 (1941) 446–458 MR Zbl

[101] **M Usher**, *Spectral numbers in Floer theories*, Compos. Math. 144 (2008) 1581–1592 MR Zbl

[102] **M Usher**, *Duality in filtered Floer–Novikov complexes*, J. Topol. Anal. 2 (2010) 233–258 MR Zbl

[103] **M Usher**, **J Zhang**, *Persistent homology and Floer–Novikov theory*, Geom. Topol. 20 (2016) 3333–3430 MR Zbl

[104] **C Viterbo**, *Symplectic topology as the geometry of generating functions*, Math. Ann. 292 (1992) 685–710 MR Zbl

[105] **K Wehrheim**, **C Woodward**, *Orientations for pseudoholomorphic quilts*, preprint (2015) arXiv 1503.07803

[106] **N Wilkins**, *A construction of the quantum Steenrod squares and their algebraic relations*, Geom. Topol. 24 (2020) 885–970 MR Zbl

[107] **N Wilkins**, *Quantum Steenrod squares and the equivariant pair-of-pants in symplectic cohomology*, J. Topol. Anal. 15 (2023) 1–56 MR Zbl

[108] **E Witten**, *Two-dimensional gravity and intersection theory on moduli space*, from "Surveys in differential geometry" (H B Lawson, Jr, S-T Yau, editors), Lehigh Univ., Bethlehem, PA (1991) 243–310 MR Zbl

[109] **Y Wu**, **X Liu**, *Symplectic group actions on homotopy elliptic surfaces*, Chinese Ann. Math. Ser. B 35 (2014) 873–884 MR Zbl

[110] **F Zapolsky**, *The Lagrangian Floer-quantum-PSS package and canonical orientations in Floer theory*, preprint (2015) arXiv 1507.02253

Department of Mathematics and Statistics, University of Montreal
Centre-Ville Montreal, QC, Canada

Department of Mathematics and Statistics, University of Montreal
Centre-Ville Montreal, QC, Canada

`marcelo.sarkis.atallah@umontreal.ca`, `egor.shelukhin@umontreal.ca`

Proposed: Yakov Eliashberg — Received: 20 February 2021
Seconded: Anna Wienhard, David Fisher — Revised: 14 September 2021

Geometry & Topology **27**:7 (2023) 2899–2936
DOI: 10.2140/gt.2023.27.2899
Published: 19 September 2023

A higher-rank rigidity theorem for convex real projective manifolds

ANDREW ZIMMER

For convex real projective manifolds we prove an analogue of the higher-rank rigidity theorem of Ballmann and Burns and Spatzier.

53C24; 20H10, 22E40, 37D40, 53C15

1 Introduction

A real projective structure on a d–manifold M is an open cover $M = \bigcup_\alpha U_\alpha$ along with coordinate charts $\varphi_\alpha : U_\alpha \to \mathbb{P}(\mathbb{R}^{d+1})$ such that each transition function $\varphi_\alpha \circ \varphi_\beta^{-1}$ coincides with the restriction of an element in $\mathrm{PGL}_{d+1}(\mathbb{R})$. A *real projective manifold* is a manifold equipped with a real projective structure.

An important class of real projective manifolds is the convex real projective manifolds, which are defined as follows. First, a subset $\Omega \subset \mathbb{P}(\mathbb{R}^{d+1})$ is called a *properly convex domain* if there exists an affine chart which contains it as a bounded convex open set. In this case, the *automorphism group of* Ω is

$$\mathrm{Aut}(\Omega) := \{g \in \mathrm{PGL}_{d+1}(\mathbb{R}) : g\Omega = \Omega\}.$$

If $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete subgroup that acts freely and properly discontinuously on Ω, then the quotient manifold $\Gamma\backslash\Omega$ is called a *convex real projective manifold*. Notice that local inverses to the covering map $\Omega \to \Gamma\backslash\Omega$ provide a real projective structure on the quotient. In the case when there exists a compact quotient, the domain Ω is called *divisible*. For more background see the expository papers by Benoist [7], Marquis [22] and Quint [25].

When $d \leq 3$, the structure of closed convex real projective d–manifolds is very well understood thanks to deep work of Benzécri [9], Goldman [16] and Benoist [6]. But, when $d \geq 4$, their general structure is mysterious.

We establish a dichotomy for convex real projective manifolds inspired by the theory of nonpositively curved Riemannian manifolds. In particular, a compact Riemannian manifold (M, g) with nonpositive curvature is said to have *higher rank* if every geodesic in the universal cover is contained in a totally geodesic subspace isometric to $\mathbb{R}^2$. Otherwise, (M, g) is said to have *rank one*. An important theorem of Ballmann [2] and Burns and Spatzier [11; 12] states that every compact irreducible Riemannian manifold with nonpositive curvature and higher rank is a locally symmetric space. This foundational result reduces many problems about nonpositively curved manifolds to the rank-one case. Further, rank-one manifolds possess very useful "weakly hyperbolic behavior" (see for instance Ballmann [1] and Knieper [20]).

In the context of convex real projective manifolds, the natural analogue of isometrically embedded copies of $\mathbb{R}^2$ are properly embedded simplices, see Section 2.6 below, which leads to a definition of higher rank:

Definition 1.1 (i) A properly convex domain $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ has *higher rank* if for every $p, q \in \Omega$ there exists a properly embedded simplex $S \subset \Omega$ with $\dim(S) \geq 2$ and $[p, q] \subset S$.

(ii) If a properly convex domain $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ does not have higher rank, then we say that Ω has *rank one*.

There are two basic families of properly convex domains with higher rank: reducible domains (see Section 2.4) and symmetric domains with real rank at least two.

A properly convex domain $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is called *symmetric* if there exists a semisimple Lie group $G \leq \mathrm{PGL}_d(\mathbb{R})$ which preserves Ω and acts transitively. In this case, the *real rank of* Ω is defined to be the real rank of G. Koecher and Vinberg characterized the irreducible symmetric properly convex domains and proved that G must be locally isomorphic to either

(i) $\mathrm{SO}(1, m)$ with $d = m + 1$,
(ii) $\mathrm{SL}_m(\mathbb{R})$ with $d = \frac{1}{2}(m^2 + m)$,
(iii) $\mathrm{SL}_m(\mathbb{C})$ with $d = m^2$,
(iv) $\mathrm{SL}_m(\mathbb{H})$ with $d = 2m^2 - m$, or
(v) $E_{6(-26)}$ with $d = 27$.

For details see Faraut and Korányi [15], Koecher [21] and Vinberg [28; 29]. Borel [10] proved that every semisimple Lie group contains a cocompact lattice, which implies that every symmetric properly convex domain is divisible.

We prove that these two families of examples are the only divisible domains with higher rank. In fact, we show that being symmetric with real rank at least two is equivalent to a number of other "higher rank" conditions. Before stating the main result we need a few more definitions.

Definition 1.2
- Given $g \in \mathrm{PGL}_d(\mathbb{R})$, let
$$\lambda_1(g) \geq \lambda_2(g) \geq \cdots \geq \lambda_d(g)$$
denote the absolute values of the eigenvalues of some (hence any) lift of g to $\mathrm{SL}_d^{\pm}(\mathbb{R}) := \{h \in \mathrm{GL}_d(\mathbb{R}) : \det h = \pm 1\}$.
- $g \in \mathrm{PGL}_d(\mathbb{R})$ is *proximal* if $\lambda_1(g) > \lambda_2(g)$. In this case, let $\ell_g^+ \in \mathbb{P}(\mathbb{R}^d)$ denote the eigenline of g corresponding to $\lambda_1(g)$.
- $g \in \mathrm{PGL}_d(\mathbb{R})$ is *biproximal* if g and g^{-1} are both proximal. In this case, define $\ell_g^- := \ell_{g^{-1}}^+$.

Next we define a distance on the boundary using projective line segments:

Definition 1.3 Given a properly convex domain $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ the (possibly infinite valued) *simplicial distance* on $\partial\Omega$ is defined by

$$s_{\partial\Omega}(x, y) = \inf\{k : \exists a_0, \dots, a_k \text{ with } x = a_0,\, y = a_k \text{ and } [a_j, a_{j+1}] \subset \partial\Omega \text{ for } 0 \leq j \leq k-1\}.$$

We will prove a characterization of higher rank in the context of convex real projective manifolds:

Theorem 1.4 (see Section 9) *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω. Then the following are equivalent*:

(i) *Ω is symmetric with real rank at least two.*

(ii) *Ω has higher rank.*

(iii) *The extreme points of Ω form a closed proper subset of $\partial\Omega$.*

(iv) *$[x_1, x_2] \subset \partial\Omega$ for every two extreme points $x_1, x_2 \in \partial\Omega$.*

(v) *$s_{\partial\Omega}(x, y) \leq 2$ for all $x, y \in \partial\Omega$.*

(vi) *$s_{\partial\Omega}(x, y) < +\infty$ for all $x, y \in \partial\Omega$.*

(vii) *Γ has higher rank in the sense of Prasad and Raghunathan* (see Section 8).

(viii) *For every $g \in \Gamma$ with infinite order, the cyclic group $g^{\mathbb{Z}}$ has infinite index in the centralizer of g in Γ.*

(ix) *Every $g \in \Gamma$ with infinite order has at least three fixed points in $\partial\Omega$.*

(x) $[\ell_g^+, \ell_g^-] \subset \partial\Omega$ *for every biproximal element $g \in \Gamma$.*

(xi) $s_{\partial\Omega}(\ell_g^+, \ell_g^-) < +\infty$ *for every biproximal element $g \in \Gamma$.*

M Islam [18] has recently defined and studied rank-one isometries of a properly convex domain. These are analogous to the classical definition of rank-one isometries of CAT(0) spaces (see [1]) and are defined as follows:

Definition 1.5 (Islam [18]) Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain. An element $g \in \mathrm{Aut}(\Omega)$ is a *rank-one isometry* if g is biproximal and $s_{\partial\Omega}(\ell_g^+, \ell_g^-) > 2$.

Remark 1.6 (1) When $g \in \mathrm{Aut}(\Omega)$ is a rank-one isometry, the properly embedded line segment $(\ell_g^+, \ell_g^-) \subset \Omega$ is preserved by g. Further, g acts by translations on (ℓ_g^+, ℓ_g^-) in the following sense: if H_Ω is the Hilbert metric on Ω, then there exists $T > 0$ such that

$$H_\Omega(g^n(x), x) = nT$$

for all $n \geq 0$ and $x \in (\ell_g^+, \ell_g^-)$.

(2) Islam [18, Proposition 6.3] also proved a weaker characterization of rank-one isometries: $g \in \mathrm{Aut}(\Omega)$ is a rank-one isometry if and only if g acts by translations on a properly embedded line segment $(a, b) \subset \Omega$ and $s_{\partial\Omega}(a, b) > 2$.

As an immediate consequence of Theorem 1.4:

Corollary 1.7 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω. Then the following are equivalent*:

(i) *Ω has rank one.*

(ii) *Γ contains a rank-one isometry.*

Islam has also established a number of remarkable results when the automorphism group contains a rank-one isometry; see [18] for details. For instance:

Corollary 1.8 (consequence of Theorem 1.4 and [18, Theorem 1.5]) *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω. If $d \geq 3$ and Ω is not symmetric with real rank at least two, then Γ is an acylindrically hyperbolic group.*

1.1 Outline of the proof of Theorem 1.4

The difficult part is showing that any one of conditions (ii)–(xi) implies that the domain is symmetric with real rank at least two.

One key idea is to construct and study special semigroups in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$ associated to each boundary face. This is accomplished as follows. First, motivated by a lemma of Benoist [5, Lemma 2.2], we consider a compactification of a subgroup of $\mathrm{PGL}_d(\mathbb{R})$:

Definition 1.9 Given a subgroup $G \leq \mathrm{PGL}_d(\mathbb{R})$ let

$$\overline{G}^{\mathrm{End}} \subset \mathbb{P}(\mathrm{End}(\mathbb{R}^d))$$

denote the closure of G in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$.

Next, for a dividing group, we introduce subsets of this compactification:

Definition 1.10 Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω. If $F \subset \partial\Omega$ is a boundary face and $V := \mathrm{Span}\, F \subset \mathbb{R}^d$, then define

$$\overline{\Gamma}_F^{\mathrm{End}} := \{T \in \overline{\Gamma}^{\mathrm{End}} : \mathrm{image}(T) \subset V\}$$

and

$$\overline{\Gamma}_{F,\star}^{\mathrm{End}} := \{T \in \overline{\Gamma}^{\mathrm{End}} : \mathrm{image}(T) = V \text{ and } \ker(T) \cap V = \{0\}\}.$$

We then prove:

Theorem 3.1 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω. If Ω is nonsymmetric, $F \subset \partial\Omega$ is a boundary face, $V := \mathrm{Span}\, F \subset \mathbb{R}^d$, and $\dim(V) \geq 2$, then:*

(a) *If $T \in \overline{\Gamma}_F^{\mathrm{End}}$, then $T(\Omega) \subset \overline{F}$.*
(b) *If $T \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}$, then $T(F)$ is an open subset of F.*
(c) *The set*

$$\{T|_V : T \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}\}$$

is a nondiscrete Zariski-dense semigroup in $\mathbb{P}(\mathrm{End}(V))$.

Using Theorem 3.1 we will show that any one of Theorem 1.4(ii)–(xi) implies that the domain is symmetric with real rank at least two. Here is a sketch of the argument: First suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain, $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω, and any one of Theorem 1.4(ii)–(xi) is true.

Then let $\mathcal{E}_\Omega \subset \partial\Omega$ denote the extreme points of Ω. We will show that there exists a boundary face $F \subset \partial\Omega$ such that

$$F \cap \bar{\mathcal{E}}_\Omega = \varnothing. \tag{1}$$

By choosing F minimally, we can also assume that $\bar{\mathcal{E}}_\Omega$ intersects every boundary face of strictly smaller dimension. As before, let $V := \operatorname{Span} F$. Then using (1) we show that $T|_V \in \operatorname{Aut}(F)$ for every $T \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}$. Therefore Theorem 3.1 implies that either Ω is symmetric or $\operatorname{Aut}(F)$ is a nondiscrete Zariski-dense subgroup of $\operatorname{PGL}(V)$. In the latter case, it is fairly easy to deduce that $\operatorname{PSL}(V) \subset \operatorname{Aut}(F)$, see Lemma 4.5 below, which is impossible. So Ω must be symmetric.

1.2 Outline of the paper

In Section 2 we recall some preliminary material. In Section 3 we prove Theorem 3.1. In Section 4 we prove the rigidity result mentioned in the previous subsection.

The rest of the paper is devoted to the proof of the various equivalences in Theorem 1.4. In Sections 5, 6, and 7 we prove some new results about the action of the automorphism group. In Section 8 we consider the rank of a group in the sense of Prasad and Raghunathan. Finally, in Section 9 we prove Theorem 1.4.

Acknowledgments

I would like to thank Ralf Spatzier and Mitul Islam for helpful conversations. I would also like to thank the referee for their very useful corrections and suggestions. This material is based upon work supported by the National Science Foundation under grants DMS-1904099, DMS-2105580, and DMS-2104381.

2 Preliminaries

2.1 Notation

Given a linear subspace $V \subset \mathbb{R}^d$, we let $\mathbb{P}(V) \subset \mathbb{P}(\mathbb{R}^d)$ denote its projectivization. In all other cases, given some object o, we will let $[o]$ be the projective equivalence class of o. For instance:

(i) If $v \in \mathbb{R}^d \setminus \{0\}$, let $[v]$ denote the image of v in $\mathbb{P}(\mathbb{R}^d)$.

(ii) If $\phi \in \operatorname{GL}_d(\mathbb{R})$, let $[\phi]$ denote the image of ϕ in $\operatorname{PGL}_d(\mathbb{R})$.

(iii) If $T \in \operatorname{End}(\mathbb{R}^d) \setminus \{0\}$, let $[T]$ denote the image of T in $\mathbb{P}(\operatorname{End}(\mathbb{R}^d))$.

We also identify $\mathbb{P}(\mathbb{R}^d) = \mathrm{Gr}_1(\mathbb{R}^d)$, so for instance if $x \in \mathbb{P}(\mathbb{R}^d)$ and $V \subset \mathbb{R}^d$ is a linear subspace, then $x \in \mathbb{P}(V)$ if and only if $x \subset V$.

Finally, given a subset X of $\mathbb{R}^d$ (respectively $\mathbb{P}(\mathbb{R}^d)$), we will let $\mathrm{Span}\, X \subset \mathbb{R}^d$ denote the smallest linear subspace containing X (respectively the preimage of X).

2.2 Convexity and line segments

A subset $C \subset \mathbb{P}(\mathbb{R}^d)$ is called *convex* if there exists an affine chart which contains it as a convex subset. A subset $C \subset \mathbb{P}(\mathbb{R}^d)$ is called *properly convex* if there exists an affine chart which contains it as a bounded convex subset. For convex subsets, we make some topological definitions:

Definition 2.1 Let $C \subset \mathbb{P}(\mathbb{R}^d)$ be a convex set. The *relative interior of* C, denoted by $\text{rel-int}(C)$, is the interior of C in its span and the *boundary of* C is $\partial C := \overline{C} \setminus \text{rel-int}(C)$.

A *line segment* in $\mathbb{P}(\mathbb{R}^d)$ is a connected subset of a projective line. Given two points $x, y \in \mathbb{P}(\mathbb{R}^d)$ there is no canonical line segment with endpoints x and y, but we will use the convention that if $C \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex set and $x, y \in \overline{C}$, then (when the context is clear) we will let $[x, y]$ denote the closed line segment joining x to y which is contained in $\overline{C}$. In this case, we will also let $(x, y) = [x, y] \setminus \{x, y\}$, $[x, y) = [x, y] \setminus \{y\}$, and $(x, y] = [x, y] \setminus \{x\}$.

2.3 Irreducibility

A subgroup $\Gamma \leq \mathrm{PGL}_d(\mathbb{R})$ is *irreducible* if $\{0\}$ and $\mathbb{R}^d$ are the only Γ–invariant linear subspaces of $\mathbb{R}^d$, and *strongly irreducible* if every finite-index subgroup is irreducible.

We will use the following observation several times:

Observation 2.2 *If $\Gamma \leq \mathrm{PGL}_d(\mathbb{R})$ is strongly irreducible, $x_1, \dots, x_k \in \mathbb{P}(\mathbb{R}^d)$, and*

$$V_1, \dots, V_k \subsetneq \mathbb{R}^d$$

are linear subspaces, then there exists $g \in \Gamma$ such that $g x_j \notin \mathbb{P}(V_j)$ for all $1 \leq j \leq k$.

Proof Let $G = \overline{\Gamma}^{\mathrm{Zar}}$ denote the Zariski closure of Γ in $\mathrm{PGL}_d(\mathbb{R})$ and let $G_0 \leq G$ denote the connected component of the identity of G (in the Zariski topology). Then $G_0 \cap \Gamma$ is a finite-index subgroup of Γ and hence G_0 is irreducible. So each set

$$\mathcal{O}_j = \{g \in G_0 : g x_j \notin \mathbb{P}(V_j)\}$$

is nonempty and Zariski open in G_0. Hence $\mathcal{O} = \bigcap_{j=1}^k \mathcal{O}_j$ is nonempty and Zariski open in G_0. Since $\Gamma \cap G_0$ is Zariski dense in G_0, there exists some $g \in \Gamma \cap \mathcal{O}$. □

2.4 Zariski closures

An open convex cone $C \subset \mathbb{R}^d$ is *reducible* if there exists a nontrivial vector space decomposition $\mathbb{R}^d = V_1 \oplus V_2$ and convex cones $C_1 \subset V_1$ and $C_2 \subset V_2$ such that $C = C_1 + C_2$. Otherwise, C is said to be *irreducible*. The preimage in $\mathbb{R}^d$ of a properly convex domain $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is the union of a cone and its negative; when this cone is reducible (respectively irreducible) we say that Ω is *reducible* (respectively *irreducible*).

Benoist determined the Zariski closures of discrete groups acting cocompactly on irreducible properly convex domains:

Theorem 2.3 (Benoist [5]) *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω. Then either*

(i) *Ω is symmetric, or*

(ii) *Γ is Zariski dense in $\mathrm{PGL}_d(\mathbb{R})$.*

2.5 The Hilbert distance

In this section we recall the definition of the Hilbert metric. But first some notation:

Given a projective line $L \subset \mathbb{P}(\mathbb{R}^d)$ and four distinct points $a, x, y, b \in L$ we define the *cross ratio* by

$$[a, x, y, b] = \frac{|x-b|\,|y-a|}{|x-a|\,|y-b|},$$

where $|\cdot|$ is some (any) norm in some (any) affine chart of $\mathbb{P}(\mathbb{R}^d)$ containing a, x, y and b.

Next, for $x, y \in \mathbb{P}(\mathbb{R}^d)$ distinct, let $L_{x,y} \subset \mathbb{P}(\mathbb{R}^d)$ denote the projective line containing x and y.

Definition 2.4 Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain. The *Hilbert distance* on Ω, denoted by H_Ω, is defined as follows: if $x, y \in \Omega$ are distinct, then

$$H_\Omega(x, y) = \tfrac{1}{2} \log[a, x, y, b],$$

where $\partial\Omega \cap L_{x,y} = \{a, b\}$ with the ordering a, x, y, b along $L_{x,y}$.

The following result is classical; see for instance [13, Section 28].

Proposition 2.5 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain. Then H_Ω is a complete $\mathrm{Aut}(\Omega)$–invariant metric on Ω which generates the standard topology on Ω. Moreover, if $p, q \in \Omega$, then there exists a geodesic joining p and q whose image is the line segment $[p, q]$.*

2.6 Properly embedded simplices

In this subsection we recall the definition of properly embedded simplices.

Definition 2.6 A subset $S \subset \mathbb{P}(\mathbb{R}^d)$ is a *simplex* if there exists $g \in \mathrm{PGL}_d(\mathbb{R})$ and $k \geq 0$ such that

$$gS = \{[x_1 : \cdots : x_{k+1} : 0 : \cdots : 0] \in \mathbb{P}(\mathbb{R}^d) : x_1 > 0, \ldots, x_{k+1} > 0\}.$$

In this case, we write $\dim(S) = k$ (notice that S is homeomorphic to $\mathbb{R}^k$).

Definition 2.7 Suppose that $A \subset B \subset \mathbb{P}(\mathbb{R}^d)$. Then A is *properly embedded in B* if the inclusion map $A \hookrightarrow B$ is a proper map (relative to the subspace topology).

By [23, Proposition 1.7], [17], or [26] the Hilbert metric on a simplex is isometric to a normed space, and so:

Observation 2.8 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $S \subset \Omega$ is a properly embedded simplex. Then (S, H_Ω) is quasi-isometric to $\mathbb{R}^{\dim S}$.*

2.7 Limits of linear maps

Every $T \in \mathbb{P}(\mathrm{End}(\mathbb{R}^d))$ induces a map

$$\mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(\ker T) \to \mathbb{P}(\mathbb{R}^d)$$

defined by $x \to T(x)$. We will frequently use:

Observation 2.9 *If $(T_n)_{n \geq 1}$ converges in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$ to $T \in \mathbb{P}(\mathrm{End}(\mathbb{R}^d))$, then*

$$T(x) = \lim_{n \to \infty} T_n(x)$$

for all $x \in \mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(\ker T)$. Moreover, the convergence is uniform on compact subsets of $\mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(\ker T)$.

2.8 The faces and extreme points of a properly convex domain

Definition 2.10 Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain. For $x \in \overline{\Omega}$ let $F_\Omega(x)$ denote the (*open*) *face* of x; that is,

$$F_\Omega(x) = \{x\} \cup \{y \in \overline{\Omega} : \exists \text{ an open line segment in } \overline{\Omega} \text{ containing } x \text{ and } y\}.$$

If $x \in \partial\Omega$ and $F_\Omega(x) = \{x\}$, then x is called an *extreme point of* Ω. Finally, let

$$\mathcal{E}_\Omega \subset \partial\Omega$$

denote the set of all extreme points.

These subsets have some basic properties:

Observation 2.11 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain.*

(i) *If $x \in \Omega$, then $F_\Omega(x) = \Omega$.*

(ii) *$F_\Omega(x)$ is open in its span.*

(iii) *$y \in F_\Omega(x)$ if and only if $x \in F_\Omega(y)$ if and only if $F_\Omega(x) = F_\Omega(y)$.*

(iv) *If $y \in \partial F_\Omega(x)$, then $F_\Omega(y) \subset \partial F_\Omega(x)$ and $F_\Omega(y) = F_{F_\Omega(x)}(y)$.*

(v) *If $x, y \in \overline{\Omega}$ and $z \in (x, y)$, then*

$$(p, q) \subset F_\Omega(z)$$

for all $p \in F_\Omega(x)$ and $q \in F_\Omega(y)$.

Proof These are all simple consequences of convexity. □

We will also use results about the action of the automorphism group:

Proposition 2.12 [19, Proposition 5.6] *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain, $p_0 \in \Omega$, and $(g_n)_{n \geq 1}$ is a sequence in* $\operatorname{Aut}(\Omega)$ *such that*

(i) $g_n(p_0) \to x \in \partial\Omega$,

(ii) $g_n^{-1}(p_0) \to y \in \partial\Omega$, *and*

(iii) g_n *converges in* $\mathbb{P}(\operatorname{End}(\mathbb{R}^d))$ *to* $T \in \mathbb{P}(\operatorname{End}(\mathbb{R}^d))$.

Then $\operatorname{image} T \subset \operatorname{Span} F_\Omega(x)$, $\mathbb{P}(\ker T) \cap \Omega = \varnothing$, *and* $y \in \mathbb{P}(\ker T)$.

In the case of "nontangential" convergence we can say more:

Proposition 2.13 [19, Proposition 5.7] *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain, $p_0 \in \Omega$, $x \in \partial\Omega$, $(p_n)_{n \geq 1}$ is a sequence in $[p_0, x)$ converging to x, and $(g_n)_{n \geq 1}$ is a sequence in* $\operatorname{Aut}(\Omega)$ *such that*

$$\sup_{n \geq 1} H_\Omega(g_n(p_0), p_n) < +\infty.$$

If g_n converges in $\mathbb{P}(\operatorname{End}(\mathbb{R}^d))$ to $T \in \mathbb{P}(\operatorname{End}(\mathbb{R}^d))$, then

$$T(\Omega) = F_\Omega(x),$$

and hence $\operatorname{image} T = \operatorname{Span} F_\Omega(x)$.

Proposition 5.7 in [19] is stated differently, so we provide the proof:

Proof Proposition 2.12 implies $T(\Omega) \subset F_\Omega(x)$, so we have to prove $T(\Omega) \supset F_\Omega(x)$. Fix $y \in F_\Omega(x)$. Then we can pick a sequence $(y_n)_{n\geq 1}$ in $[p_0, y)$ such that

$$\sup_{n\geq 1} H_\Omega(y_n, p_n) < \infty.$$

Thus

$$\sup_{n\geq 1} H_\Omega(g_n^{-1}(y_n), p_0) < \infty.$$

So there exists $n_j \to \infty$ such that the limit

$$q := \lim_{j\to\infty} g_{n_j}^{-1}(y_{n_j})$$

exists in Ω. Notice that $q \notin \mathbb{P}(\ker T)$ by Proposition 2.12 and so the "moreover" part of Observation 2.9 implies that

$$T(q) = \lim_{n\to\infty} g_n(q) = \lim_{j\to\infty} g_{n_j}(q) = \lim_{j\to\infty} g_{n_j}(g_{n_j}^{-1}(y_{n_j})) = \lim_{j\to\infty} y_{n_j} = y.$$

Since y was arbitrary, $F_\Omega(x) \subset T(\Omega)$. □

2.9 Proximal elements

In this section we recall some basic properties of proximal elements. For more background we refer the reader to [8].

Definition 2.14 Suppose that $F: M \to M$ is a C^1 map of a manifold M. Then a fixed point $x \in M$ of F is *attractive* if $|\lambda| < 1$ for every eigenvalue λ of $d(F)_x : T_x M \to T_x M$.

A straightforward calculation provides a characterization of proximality:

Observation 2.15 *Suppose that $g \in \mathrm{PGL}_d(\mathbb{R})$ and x is a fixed point of the g action on $\mathbb{P}(\mathbb{R}^d)$. Then the following are equivalent*:

(i) *x is an attractive fixed point of g.*

(ii) *g is proximal and $x = \ell_g^+$.*

Next we explain the global dynamics of a proximal element.

Definition 2.16 If $g \in \mathrm{PGL}_d(\mathbb{R})$ is proximal, then define $H_g^- \in \mathrm{Gr}_{d-1}(\mathbb{R}^d)$ to be the unique g–invariant linear hyperplane with

$$\ell_g^+ \oplus H_g^- = \mathbb{R}^d.$$

If g is biproximal, then also define $H_g^+ := H_{g^{-1}}^-$.

When $g \in \mathrm{PGL}_d(\mathbb{R})$ is proximal, H_g^- is usually called the repelling hyperplane of g. This is motivated by the following observation:

Observation 2.17 *If $g \in \mathrm{PGL}_d(\mathbb{R})$ is proximal, then*

$$T_g := \lim_{n \to \infty} g^n$$

exists in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. Moreover, $\mathrm{image}\, T_g = \ell_g^+$, $\ker T_g = H_g^-$, *and*

$$\mathrm{image}\, T_g \oplus \ker T_g = \mathbb{R}^d.$$

Hence

$$\ell_g^+ = \lim_{n \to \infty} g^n x$$

for all $x \in \mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(H_g^-)$.

Observation 2.18 *Suppose $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain. If $g \in \mathrm{Aut}(\Omega)$ is proximal, then ℓ_g^+ is an extreme point of $\partial\Omega$ and $\mathbb{P}(H_g^-) \cap \partial\Omega = \varnothing$.*

Proof Proposition 2.12 implies that $\ell_g^+ \in \partial\Omega$ and $\mathbb{P}(H_g^-) \cap \partial\Omega = \varnothing$. Let $F = F_\Omega(\ell_g^+)$ and $V = \mathrm{Span}\, F$. Then $g(V) = V$. Let $\bar{g} \in \mathrm{GL}_d(\mathbb{R})$ be a lift of $g \in \mathrm{PGL}_d(\mathbb{R})$ and let $h \in \mathrm{GL}(V)$ denote the element obtained by restricting $\bar{g}$ to V. Notice that h is proximal since $\ell_g^+ \subset V$. Further $[h] \in \mathrm{Aut}(F)$ and $h(\ell_g^+) = \ell_g^+$. Since $\mathrm{Aut}(F)$ acts properly on F and $\ell_g^+ \in F$, the cyclic group

$$[h]^{\mathbb{Z}} \le \mathrm{Aut}(F) \le \mathrm{PGL}(V)$$

must be relatively compact. This implies that every eigenvalue of h has the same absolute value. Then, since h is proximal, V must be one-dimensional and so $F = \{\ell_g^+\}$. Thus ℓ_g^+ is an extreme point. □

The following result can be viewed as a converse to Observation 2.17 and will be used to construct proximal elements.

Proposition 2.19 *Suppose that $(g_n)_{n \ge 1}$ is a sequence in $\mathrm{PGL}_d(\mathbb{R})$ and*

$$T := \lim_{n \to \infty} g_n$$

exists in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. If $\dim(\mathrm{image}\, T) = 1$ and

$$\mathrm{image}\, T \oplus \ker T = \mathbb{R}^d,$$

then, for n sufficiently large, g_n is proximal and

$$\mathrm{image}\, T = \lim_{n \to \infty} \ell_{g_n}^+.$$

Proof Since $g_n \to T$ in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$,

$$\lim_{n\to\infty} g_n(x) = T(x) = \operatorname{image} T \in \mathbb{P}(\mathbb{R}^d)$$

for all $x \in \mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(\ker T)$. Moreover, the convergence is uniform on compact subsets of $\mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(\ker T)$.

By assumption,

$$\operatorname{image} T \notin \mathbb{P}(\ker T),$$

so we can find a compact neighborhood U of $\operatorname{image} T$ in $\mathbb{P}(\mathbb{R}^d)$ such that U is homeomorphic to a closed ball and

$$U \cap \mathbb{P}(\ker T) = \varnothing.$$

Then, by passing to a tail, we can assume that $g_n(U) \subset U$ for all n. So, by the Brouwer fixed-point theorem, each g_n has a fixed point $x_n \in U$. Since U can be chosen arbitrarily small,

$$\operatorname{image} T = \lim_{n\to\infty} x_n.$$

We claim that, for n large, x_n is an attractive fixed point of g_n. By Observation 2.15 this will finish the proof. Let $f_n : \mathbb{P}(\mathbb{R}^d) \to \mathbb{P}(\mathbb{R}^d)$ be the diffeomorphism induced by g_n, that is $f_n(x) = g_n(x)$ for all x. Then, since each g_n acts by projective linear transformations, we see that the f_n converge locally uniformly in the C^∞ topology on $\mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(\ker T)$ to the constant map $f \equiv \operatorname{image} T$. So, fixing a Riemannian metric on $\mathbb{P}(\mathbb{R}^d)$, we have

$$\lim_{n\to\infty} \|d(f_n)_{x_n}\| = 0.$$

Hence, for n large, x_n is an attractive fixed point of g_n. □

2.10 Rank-one isometries

In this section we state a characterization of rank-one isometries established in [18]:

Theorem 2.20 (Islam [18, Proposition 6.3]) *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $\gamma \in \mathrm{Aut}(\Omega)$. If*

$$\inf_{p\in\Omega} H_\Omega(\gamma(p), p) > 0$$

and γ fixes two points $x, y \in \partial\Omega$ with $s_{\partial\Omega}(x, y) > 2$, then:

(i) *γ is biproximal and $\{\ell_\gamma^+, \ell_\gamma^-\} = \{x, y\}$. In particular, γ is a rank-one isometry.*

(ii) *The only points fixed by γ in $\partial\Omega$ are ℓ_γ^+ and ℓ_γ^-.*

(iii) *If $w \in \partial\Omega$, then*

$$(\ell_\gamma^+, w) \cup (w, \ell_\gamma^-) \subset \Omega.$$

(iv) *If $z \in \partial\Omega \setminus \{\ell_\gamma^\pm\}$, then*

$$s_{\partial\Omega}(\ell_\gamma^\pm, z) = \infty.$$

Remark 2.21 Notice that (iv) is a consequence of (iii).

3 A semigroup associated to a boundary face

Theorem 3.1 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω. If Ω is nonsymmetric, $F \subset \partial\Omega$ is a boundary face, $V := \mathrm{Span}\, F$, and $\dim(V) \geq 2$, then:*

(a) *If $T \in \overline{\Gamma}_F^{\mathrm{End}}$, then $T(\Omega) \subset \overline{F}$.*

(b) *If $T \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}$, then $T(F)$ is an open subset of F.*

(c) *The set*

$$\{T|_V : T \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}\}$$

is a nondiscrete Zariski-dense semigroup in $\mathbb{P}(\mathrm{End}(V))$.

The proof of Theorem 3.1 will follow from a series of lemmas, many of which hold in greater generality.

For the rest of the section fix a properly convex domain $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ and a subgroup $\Gamma \leq \mathrm{Aut}(\Omega)$. Notice that we are not (currently) assuming that Ω is irreducible, that Γ is discrete, or that Γ acts cocompactly on Ω.

Observation 3.2 (a) *If $T \in \overline{\Gamma}^{\mathrm{End}}$, then $\mathbb{P}(\ker T) \cap \Omega = \varnothing$.*

(b) *If $S, T \in \overline{\Gamma}^{\mathrm{End}}$ and $\mathrm{image}\, T \setminus \ker S \neq \varnothing$, then $S \circ T \in \overline{\Gamma}^{\mathrm{End}}$.*

Proof Part (a) follows immediately from Proposition 2.12.

For part (b), fix $S, T \in \overline{\Gamma}^{\mathrm{End}}$ with $\mathrm{image}\, T \setminus \ker S \neq \varnothing$. By hypothesis $S \circ T$ is a well-defined element of $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. To show that $S \circ T \in \overline{\Gamma}^{\mathrm{End}}$, fix sequences $(g_n)_{n\geq 1}$ and $(h_n)_{n\geq 1}$ in Γ such that

$$S = \lim_{n\to\infty} g_n \quad \text{and} \quad T = \lim_{n\to\infty} h_n$$

in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. Then, since $S \circ T \neq 0$,

$$S \circ T = \lim_{n\to\infty} g_n h_n$$

in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. So $S \circ T \in \overline{\Gamma}^{\mathrm{End}}$. □

Lemma 3.3 *If $F \subset \partial\Omega$ is a boundary face and $T \in \overline{\Gamma}_F^{\mathrm{End}}$, then $T(\Omega) \subset \overline{F}$.*

Proof Suppose $T \in \overline{\Gamma}_F^{\mathrm{End}}$. Then there exists a sequence $(g_n)_{n\geq 1}$ in Γ such that

$$T = \lim_{n\to\infty} g_n$$

in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. Since $\mathbb{P}(\ker T) \cap \Omega = \varnothing$,

$$T(p) = \lim_{n\to\infty} g_n(p) \in \overline{\Omega}$$

for all $p \in \Omega$. So $T(\Omega) \subset \overline{\Omega}$. Since $\mathrm{image}(T) \subset V$,

$$T(\Omega) \subset \mathbb{P}(V) \cap \overline{\Omega} = \overline{F}. \qquad \square$$

Lemma 3.4 *If $F \subset \partial\Omega$ is a boundary face and $T \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}$, then $T(F)$ is an open subset of F.*

Proof By definition and Observation 3.2

$$(\Omega \cup F) \cap \mathbb{P}(\ker T) \subset (\Omega \cup \mathbb{P}(V)) \cap \mathbb{P}(\ker T) = \varnothing.$$

So T induces a continuous map on $\Omega \cup F$. Since $F \subset \overline{\Omega}$, the previous lemma implies that

$$T(F) \subset \overline{T(\Omega)} \subset \overline{F}.$$

Since $V \cap \ker T = \{0\}$, $T(F)$ is an open subset of $\mathbb{P}(V)$. So

$$T(F) \subset \text{rel-int}(\overline{F}) = F. \qquad \square$$

Lemma 3.5 *If $F \subset \partial\Omega$ is a boundary face, then the set*

$$\{T|_V : T \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}\}$$

is a semigroup in $\mathbb{P}(\mathrm{End}(V))$.

Proof Fix $T_1, T_2 \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}$. Then

$$\mathrm{image}\, T_2 \setminus \ker T_1 = V \setminus \ker T_1 = V \setminus \{0\} \neq \varnothing,$$

and so $T_1 \circ T_2 \in \overline{\Gamma}^{\mathrm{End}}$ by Observation 3.2.

We first show $\ker(T_1 \circ T_2) \cap V = \{0\}$. Suppose $v \in \ker(T_1 \circ T_2) \cap V$. Then $T_2(v) \in \ker T_1$. But $\mathrm{image}\, T_2 = V$ and $\ker T_1 \cap V = \{0\}$, so $T_2(v) = 0$ and so $v \in \ker T_2 \cap V = \{0\}$. So $v = 0$, and thus

$$\{0\} = \ker(T_1 \circ T_2) \cap V. \tag{2}$$

Next, by definition,

$$\mathrm{image}(T_1 \circ T_2) \subset \mathrm{image}\, T_1 = V.$$

So by (2) and dimension counting

$$\operatorname{image}(T_1 \circ T_2) = V.$$

Thus $T_1 \circ T_2 \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}$.

Since image $T_2 = V$

$$T_1|_V \circ T_2|_V = (T_1 \circ T_2)|_V,$$

so

$$(T_1 \circ T_2)|_V \in \{T|_V : T \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}\}.$$

Then, since $T_1, T_2 \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}$ were arbitrary, we see that

$$\{T|_V : T \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}\}$$

is a semigroup in $\mathbb{P}(\mathrm{End}(V))$. □

The next lemma requires a definition.

Definition 3.6 A point $x \in \partial\Omega$ is a *conical limit point of* Γ if there exist $p_0 \in \Omega$, a sequence $(p_n)_{n\ge1}$ in $[p_0, x)$ with $p_n \to x$, and a sequence $(\gamma_n)_{n\ge1}$ in Γ with

$$\sup_{n\ge1} H_\Omega(\gamma_n(p_0), p_n) < +\infty.$$

Notice that if Γ acts cocompactly on Ω then every boundary point is a conical limit point.

Lemma 3.7 *Suppose $x \in \partial\Omega$ is a conical limit point of Γ, $F = F_\Omega(x)$, $V = \mathrm{Span}\, F$, and $\dim(V) = k$. If $k \ge 2$ and the image of $\Gamma \hookrightarrow \mathrm{PGL}\big(\bigwedge^k \mathbb{R}^d\big)$ is strongly irreducible (eg Γ is Zariski dense in $\mathrm{PGL}_d(\mathbb{R})$), then there exists a sequence $(g_n)_{n\ge1}$ in Γ with:*

(i) *$g_n \to T$ in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$, where $T \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}$.*

(ii) *$g_1|_V, g_2|_V, \dots$ are pairwise distinct elements of $\mathbb{P}(\mathrm{Lin}(V, \mathbb{R}^d))$.*

Proof By hypothesis there exist $p_0 \in \Omega$, a sequence $(p_n)_{n\ge1}$ in $[p_0, x)$ with $p_n \to x$, and a sequence $(\gamma_n)_{n\ge1}$ in Γ with

$$\sup_{n\ge1} H_\Omega(\gamma_n(p_0), p_n) < +\infty.$$

After passing to a subsequence we can suppose that the limit

$$S = \lim_{n\to\infty} \gamma_n$$

exists in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. Then, by Proposition 2.13,

$$\operatorname{image} S = \mathrm{Span}\, F = V,$$

and so $S \in \overline{\Gamma}_F^{\mathrm{End}}$. By passing to another subsequence we can suppose that

$$V_\infty = \lim_{n\to\infty} \gamma_n^{-1} V$$

exists in $\mathrm{Gr}_k(\mathbb{R}^d)$.

Let $V = \mathrm{Span}\{v_1, \dots, v_k\}$, $V_\infty = \mathrm{Span}\{u_1, \dots, u_k\}$, and $\ker S = \mathrm{Span}\{s_1, \dots, s_{d-k}\}$, and let $W_1 = [u_1 \wedge \dots \wedge u_k]$ and

$$W_2 = \left\{\alpha \in \bigwedge^k \mathbb{R}^d : \alpha \wedge s_1 \wedge \dots \wedge s_{d-k} = 0\right\}.$$

Since the image of $\Gamma \hookrightarrow \mathrm{PGL}\left(\bigwedge^k \mathbb{R}^d\right)$ is strongly irreducible, Observation 2.2 implies that there exists some $\phi \in \Gamma$ such that $\phi[v_1 \wedge \dots \wedge v_k] \notin W_1 \cup W_2$. Equivalently, $\ker S \cap \phi V = \{0\}$ and $\phi V \neq V_\infty$.

Define $g_n := \gamma_n \phi$. Then

$$T := S \circ \phi = \lim_{n\to\infty} g_n$$

exists in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. Further, $\operatorname{image} T = \operatorname{image} S = V$ and

$$\ker T \cap V = \phi^{-1}(\ker S \cap \phi V) = \{0\},$$

so $T \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}$. Also, since $T(V) = V$,

$$V = T(V) = \lim_{n\to\infty} g_n V.$$

Next we claim that $g_n V \neq V$ for n sufficiently large. Notice that $g_n V = V$ if and only if $g_n^{-1} V = V$ if and only if $\gamma_n^{-1} V = \phi V$. But $\gamma_n^{-1} V \to V_\infty$ and $\phi V \neq V_\infty$, so $g_n V \neq V$ for n sufficiently large.

Finally, since $g_n V \to V$ and $g_n V \neq V$ for n sufficiently large, we can pass to a subsequence so that $V, g_1 V, g_2 V, \dots$ are pairwise distinct subspaces. Thus $g_1|_V, g_2|_V, \dots$ must be pairwise distinct. □

Lemma 3.8 *Suppose* $x \in \partial\Omega$ *is a conical limit point of* Γ, $F = F_\Omega(x)$, $V = \mathrm{Span}\, F$, *and* $\dim(V) = k$. *If* $k \geq 2$ *and the image of* $\Gamma \hookrightarrow \mathrm{PGL}(\wedge^k \mathbb{R}^d)$ *is strongly irreducible (eg* Γ *is Zariski dense in* $\mathrm{PGL}_d(\mathbb{R})$*), then the set*

$$\{T|_V : T \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}\}$$

is nondiscrete in $\mathbb{P}(\mathrm{End}(V))$.

Proof Let $T \in \overline{\Gamma}_{F,\star}^{\mathrm{End}}$ and $(g_n)_{n\geq 1}$ be as in the previous lemma. Since $g_1|_V, g_2|_V, \dots$ are pairwise distinct and each $g_n|_V$ is determined by its values on any set of $\dim V + 1$ points in general position, after passing to a subsequence we can find a point $x_0 \in F$ such that $g_1(x_0), g_2(x_0), \dots$ are pairwise distinct.

Since $x_0 \in F$ and $\mathbb{P}(\ker T) \cap F = \varnothing$,

$$T(x_0) = \lim_{n\to\infty} g_n(x_0).$$

Since $g_1(x_0), g_2(x_0), \dots$ are pairwise distinct, by passing to another sequence we can assume that $g_n(x_0) \neq T(x_0)$ for all n. Then, for each n there exists a unique projective line L_n containing $T(x_0)$ and $g_n(x_0)$. By passing to a subsequence we can suppose that L_n converges to a projective line L. Then let $W \subset \mathbb{R}^d$ be the two-dimensional linear subspace with $L = \mathbb{P}(W)$.

Fix some $W' \in \mathrm{Gr}_k(\mathbb{R}^d)$ with $W \subset W'$ and suppose that $V = \mathrm{Span}\{v_1, \dots, v_k\}$, $W' = \mathrm{Span}\{w_1, \dots, w_k\}$, and $\ker T = \mathrm{Span}\{t_1, \dots, t_{d-k}\}$. Let

$$U = \left\{\alpha \in \textstyle\bigwedge^k \mathbb{R}^d : \alpha \wedge t_1 \wedge \cdots \wedge t_{d-k} = 0\right\}.$$

Since the image of $\Gamma \hookrightarrow \mathrm{PGL}\big(\bigwedge^k \mathbb{R}^d\big)$ is strongly irreducible, Observation 2.2 implies that there exists $\varphi \in \Gamma$ such that $\varphi[v_1 \wedge \cdots \wedge v_k] \notin U$ and $\varphi[w_1 \wedge \cdots \wedge w_k] \notin U$. Hence $\ker T \cap \varphi V = \{0\}$ and $\ker T \cap \varphi W = \{0\}$.

Notice that $T\varphi T = \lim_{n\to\infty} g_n \varphi g_n$ is in $\overline{\Gamma}^{\mathrm{End}}_{F,\star}$. Then replacing $(g_n)_{n\geq 1}$ with a tail, we can assume that

$$S_n := T\varphi g_n \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}$$

for all n.

We claim that the set

$$\{S_n(x_0) : n \geq 0\} \subset F$$

is infinite. For this calculation we fix an affine chart $\mathbb{A}$ of $\mathbb{P}(\mathbb{R}^d)$ which contains $\overline{\Omega}$. We then identify $\mathbb{A}$ with $\mathbb{R}^{d-1}$ so that $T(x_0) = 0$ and

$$\mathbb{A} \cap L = \{(t, 0, \dots, 0) : t \in \mathbb{R}\}.$$

Since $\ker T \cap \varphi V = \{0\}$, in these coordinates the map $T\varphi$ is smooth in a neighborhood of $0 = T(x_0)$. Further, since $\ker T \cap \varphi W = \{0\}$, in these coordinates

$$d(T\varphi)_0(1, 0, \dots, 0) \neq 0.$$

Now, since $L_n \to L$ and $g_n(x_0) \to T(x_0)$ in these coordinates,

$$g_n(x_0) = (t_n, 0, \dots, 0) + \mathrm{o}(|t_n|)$$

for some sequence $(t_n)_{n\geq 1}$ converging to 0. Then, in these coordinates,

$$\begin{aligned} S_n(x_0) = T\varphi g_n(x_0) &= T\varphi((t_n, 0, \dots, 0) + \mathrm{o}(|t_n|)) \\ &= T\varphi T(x_0) + t_n d(T\varphi)_0(1, 0, \dots, 0) + \mathrm{o}(|t_n|). \end{aligned}$$

Since $d(T\varphi)_0(1,0,\dots,0) \neq 0$ and $t_n \to 0$, we see that the set $\{S_n(x_0) : n \geq 0\}$ is infinite.

Finally, since $S_n|_V \to T\varphi T|_V$, this implies that

$$\{S_n|_V : n \geq 0\} \cup \{T\varphi T|_V\}$$

is nondiscrete in $\mathbb{P}(\mathrm{End}(V))$. $\square$

Lemma 3.9 *Suppose $x \in \partial\Omega$ is a conical limit point of Γ, $F = F_\Omega(x)$, $V = \mathrm{Span}\, F$, and $\dim(V) = k$. If $k \geq 2$ and Γ is Zariski dense in $\mathrm{PGL}_d(\mathbb{R})$, then*

$$\{T|_V : T \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}\}$$

is Zariski dense in $\mathbb{P}(\mathrm{End}(V))$.

Proof Let Z_0 be the Zariski closure of

$$\{T|_V : T \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}\}$$

in $\mathbb{P}(\mathrm{End}(V))$.

Lemma 3.7 implies that $\overline{\Gamma}^{\mathrm{End}}_{F,\star}$ is nonempty, so fix $T \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}$. Then define

$$Z_1 = \{g \in \mathrm{PGL}_d(\mathbb{R}) : \mathrm{rank}(T \circ g|_V) < \dim(V)\}.$$

Notice that Z_1 is a proper Zariski-closed set in $\mathrm{PGL}_d(\mathbb{R})$ since $\mathrm{rank}(T) = \dim(V)$. Also define

$$Z_2 = \{g \in \mathrm{PGL}_d(\mathbb{R}) : T \circ g|_V \in Z_0\}.$$

Notice that Z_2 is a Zariski-closed subset of $\mathrm{PGL}_d(\mathbb{R})$.

We claim that $\Gamma \subset Z_1 \cup Z_2$. If $g \in \Gamma \setminus Z_1$, then $\mathrm{rank}(T \circ g|_V) = \dim V$ and

$$\mathrm{image}(T \circ g|_V) \subset \mathrm{image}\, T = V.$$

So $(T \circ g)(V) = V$, which implies that $T \circ g \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}$, and hence that $g \in Z_2$. So $\Gamma \subset Z_1 \cup Z_2$.

Then, since Z_1 is a proper Zariski closed subset of $\mathrm{PGL}_d(\mathbb{R})$ and Γ is Zariski dense in $\mathrm{PGL}_d(\mathbb{R})$, we see that $Z_2 = \mathrm{PGL}_d(\mathbb{R})$. Therefore

$$Z_0 \supset \{T \circ g|_V : g \in Z_2\} = \{T \circ g|_V : g \in \mathrm{PGL}_d(\mathbb{R})\} \supset \mathrm{PGL}(V),$$

since $\mathrm{image}\, T = V$. Thus $Z_0 = \mathbb{P}(\mathrm{End}(V))$. $\square$

Proof of Theorem 3.1 Parts (a) and (b) follow from Lemmas 3.3 and 3.4, respectively. Since Γ acts cocompactly on Ω, every point in $\partial\Omega$ is a conical limit point, and

Theorem 2.3 implies that Γ is Zariski dense in $\mathrm{PGL}_d(\mathbb{R})$. So part (c) follows from Lemmas 3.3, 3.8, and 3.9. $\square$

4 The main rigidity theorem

Recall that $\mathcal{E}_\Omega \subset \partial\Omega$ denotes the set of extreme points of a properly convex domain Ω. In this section we prove the following rigidity result:

Theorem 4.1 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex divisible domain and there exists a boundary face $F \subset \partial\Omega$ such that*

$$F \cap \bar{\mathcal{E}}_\Omega = \varnothing.$$

Then Ω is symmetric with real rank at least two.

The rest of the section is devoted to the proof of the theorem, so suppose $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ satisfies the hypothesis of the theorem. Then let $\Gamma \leq \mathrm{Aut}(\Omega)$ be a discrete group acting cocompactly on Ω.

We assume, for a contradiction, that Ω is not symmetric with real rank at least two.

Lemma 4.2 *It holds that Ω is not symmetric.*

Proof If Ω were symmetric, then by assumption it would have real rank one. Then, by the characterization of symmetric convex divisible domains, Ω coincides with the unit ball in some affine chart. Therefore $\mathcal{E}_\Omega = \partial\Omega$, which is impossible since there exists a boundary face $F \subset \partial\Omega$ such that

$$F \cap \bar{\mathcal{E}}_\Omega = \varnothing. \qquad \square$$

Now we fix a boundary face $F \subset \partial\Omega$, where

$$\bar{\mathcal{E}}_\Omega \cap F = \varnothing$$

and if $F' \subset \partial\Omega$ is a face with $\dim F' < \dim F$ then

$$\bar{\mathcal{E}}_\Omega \cap F' \neq \varnothing.$$

Then define $V := \mathrm{Span}\, F$.

Lemma 4.3 *If $T \in \overline{\Gamma}^{\mathrm{End}}_{F,\star}$, then the map*

$$F \to \mathbb{P}(V), \quad p \mapsto T(p),$$

is in $\mathrm{Aut}(F)$.

Proof Notice that $T|_V \in \mathrm{PGL}(V)$ since $T(V) \subset V$ and $\ker T \cap V = \{0\}$. So we just have to show that $T(F) = F$. Theorem 3.1(b) says that $T(F) \subset F$, and so we just have to show that $F \subset T(F)$.

Fix $y \in F$. Since the set $T(\bar{F}) \cap F$ is closed in F, there exists $x_0 \in T(\bar{F}) \cap F$ such that

$$H_F(y, x_0) = \min_{x \in T(\bar{F}) \cap F} H_F(y, x).$$

Since $T|_V \in \mathrm{PGL}(V)$, the set $T(F)$ is open in F. So we either have $y = x_0 \in T(F)$ or $x_0 \in T(\partial F)$. Suppose for a contradiction that $x_0 \in T(\partial F)$. Then let $x_0' \in \partial F$ be the point where $T(x_0') = x_0$. Next, let $F' \subset \partial F$ be the face of x_0'. Then $\dim F' < \dim F$, so

$$\bar{\mathcal{E}}_\Omega \cap F' \neq \varnothing.$$

Thus we can find $z \in F'$ and a sequence $(z_n)_{n \geq 1}$ in $\mathcal{E}_\Omega$ such that $z_n \to z$. Since $z \in F'$, there exists an open line segment L in $\bar{F}$ which contains z and x_0'. Then $T(L)$ is an open line segment in $\bar{F}$ since $T|_V \in \mathrm{PGL}(V)$. So, since $T(x_0') \in F$, we also have $T(z) \in F$, and since

$$T \in \overline{\Gamma}^{\mathrm{End}}_{F,\star} \subset \overline{\Gamma}^{\mathrm{End}},$$

there exists a sequence $(g_n)_{n \geq 1}$ in Γ such that $g_n \to T$ in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. Now note that $z \notin \mathbb{P}(\ker T)$ since $\ker T \cap V = \{0\}$. So by the "moreover" part of Observation 2.9,

$$T(z) = \lim_{n \to \infty} g_n(z_n) \in F.$$

However, $g_n(z_n) \in \mathcal{E}_\Omega$, and so

$$T(z) \in \bar{\mathcal{E}}_\Omega \cap F = \varnothing.$$

Thus we have a contradiction. Hence $y = x_0 \in T(F)$, and since $y \in F$ was arbitrary we have $F \subset T(F)$. □

Lemma 4.4 $\mathrm{Aut}(F)$ *is nondiscrete and Zariski dense in* $\mathrm{PGL}(V)$.

Proof This follows immediately from Lemma 4.3 and Theorem 3.1(c). □

Lemma 4.5
$$\mathrm{PSL}(V) \subset \mathrm{Aut}(F).$$

Proof Let $\mathrm{Aut}_0(F)$ denote the connected component of the identity in $\mathrm{Aut}(F)$ and let $\mathfrak{g} \subset \mathfrak{sl}(V)$ denote the Lie algebra of $\mathrm{Aut}_0(F)$. Then $\mathfrak{g} \neq \{0\}$ since $\mathrm{Aut}(F)$ is closed and nondiscrete. Also $\mathrm{Aut}_0(F)$ is normalized by $\mathrm{Aut}(F)$, and so

$$\mathrm{Ad}(g)\mathfrak{g} = \mathfrak{g}$$

for all $g \in \mathrm{Aut}(F)$. Then, since $\mathrm{Aut}(F)$ is Zariski dense in $\mathrm{PGL}(V)$, we see that

$$\mathrm{Ad}(g)\mathfrak{g} = \mathfrak{g}$$

for all $g \in \mathrm{PGL}(V)$. Since the representation $\mathrm{Ad}\colon \mathrm{PGL}(V) \to \mathrm{GL}(\mathfrak{sl}(V))$ is irreducible, we must have $\mathfrak{g} = \mathfrak{sl}(V)$. Thus $\mathrm{Aut}_0(F) = \mathrm{PSL}(V)$. □

Proof of Theorem 4.1 The previous lemma immediately implies a contradiction: fix $x \in F$, then

$$\mathbb{P}(V) \supset F \supset \mathrm{Aut}(F) \cdot x \supset \mathrm{PSL}(V) \cdot x = \mathbb{P}(V).$$

So $F = \mathbb{P}(V)$, which contradicts the fact that Ω is properly convex. □

5 Density of biproximal elements

In this section we prove a density result for the attracting and repelling fixed points of biproximal elements. To state the result we need one definition: if $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$, then the *limit set of* Γ is

$$\mathcal{L}_\Omega(\Gamma) = \bigcup_{p \in \Omega} \overline{\Gamma \cdot p} \cap \partial\Omega.$$

Equivalently, a point $x \in \partial\Omega$ is in $\mathcal{L}_\Omega(\Gamma)$ if and only if there exist $p \in \Omega$ and a sequence $(\gamma_n)_{n \geq 1}$ in Γ such that $\gamma_n(p) \to x$.

Theorem 5.1 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a strongly irreducible group. If $x, y \in \mathcal{L}_\Omega(\Gamma)$ are extreme points of Ω and $(x, y) \subset \Omega$, then there exists a sequence of biproximal elements $(g_n)_{n \geq 1}$ in Γ such that*

$$\lim_{n \to \infty} \ell^+_{g_n} = x \quad \textit{and} \quad \lim_{n \to \infty} \ell^-_{g_n} = y.$$

Before proving the theorem we state and prove one corollary:

Corollary 5.2 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group that acts cocompactly on Ω. If $x, y \in \partial\Omega$ are extreme points and $(x, y) \subset \Omega$, then there exists a sequence of biproximal elements $(g_n)_{n \geq 1}$ in Γ such that*

$$\lim_{n \to \infty} \ell^+_{g_n} = x \quad \textit{and} \quad \lim_{n \to \infty} \ell^-_{g_n} = y.$$

Proof A result of Vey [27, Theorem 5] implies that Γ is strongly irreducible and Proposition 2.13 implies that $\partial\Omega = \mathcal{L}_\Omega(\Gamma)$, so Theorem 5.1 implies the corollary. □

Proof of Theorem 5.1 By definition there exist $p \in \Omega$ and a sequence $(\gamma_n)_{n\geq 1}$ in Γ such that $\gamma_n(p) \to x$. Passing to a subsequence, we can suppose the limits

$$T^+ = \lim_{n\to\infty} \gamma_n \quad \text{and} \quad T^- = \lim_{n\to\infty} \gamma_n^{-1}$$

exist in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. By Proposition 2.12

$$\text{image}\, T^+ \subset \text{Span}\, F_\Omega(x) = \text{Span}\{x\} = x,$$

and so image $T^+ = x$. Proposition 2.12 also implies that $\mathbb{P}(\ker T^-) \cap \Omega = \varnothing$ and $x \in \mathbb{P}(\ker T^-)$. Notice that $y \notin \mathbb{P}(\ker T^-)$ since $(x, y) \subset \Omega$.

Similarly, we can find a sequence $(\phi_n)_{n\geq 1}$ in Γ such that the limits

$$S^+ = \lim_{n\to\infty} \phi_n \quad \text{and} \quad S^- = \lim_{n\to\infty} \phi_n^{-1}$$

exist in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$, image $S^+ = y$, and $x \notin \mathbb{P}(\ker S^-)$.

Fix some $x' \in$ image T^- and $y' \in$ image S^-. Since Γ is strongly irreducible, by Observation 2.2 there exists $h \in \Gamma$ such that:

(i) $h(y') \notin \mathbb{P}(\ker T^+)$; hence, $h(\text{image}\, S^-) \not\subset \ker T^+$.

(ii) $hS^-(x) \notin \mathbb{P}(\ker T^+)$.

(iii) $h(x') \notin \mathbb{P}(\ker S^+)$; hence, $h(\text{image}\, T^-) \not\subset \ker S^+$.

(iv) $hT^-(y) \notin \mathbb{P}(\ker S^+)$.

Then consider $g_n = \gamma_n \circ h \circ \phi_n^{-1}$. By our choice of h, we have $T^+ \circ h \circ S^- \neq 0$ and hence

$$T^+ \circ h \circ S^- = \lim_{n\to\infty} g_n$$

in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. Notice that $\text{image}(T^+ \circ h \circ S^-) = \text{image}\, T^+ = x$ and, by our choice of h,

$$x \notin \mathbb{P}(\ker(T^+ \circ h \circ S^-)).$$

So

$$\text{image}(T^+ \circ h \circ S^-) + \ker(T^+ \circ h \circ S^-) = x + \ker(T^+ \circ h \circ S^-) = \mathbb{R}^d,$$

and hence, by Proposition 2.19, g_n is proximal for n sufficiently large and $\ell^+_{g_n} \to x$. By similar reasoning g_n^{-1} is proximal for n sufficiently large and $\ell^-_{g_n} = \ell^+_{g_n^{-1}} \to y$. □

6 North–south dynamics

In this section we prove a stronger version of Theorem 5.1 for pairs of extreme points in the limit set whose simplicial distance is greater than two.

Theorem 6.1 *Suppose $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is strongly irreducible. Assume $x, y \in \mathcal{L}_\Omega(\Gamma)$ are extreme points of Ω and $s_{\partial\Omega}(x, y) > 2$. If $A, B \subset \overline{\Omega}$ are neighborhoods of x and y, then there exists $g \in \Gamma$ with*

$$g(\overline{\Omega} \setminus B) \subset A \quad \text{and} \quad g^{-1}(\overline{\Omega} \setminus A) \subset B.$$

Remark 6.2 Theorem 6.1 is an analogue of a result for CAT(0) spaces; see Chapter 3 and specifically Theorem 3.4 of [3].

Before proving the theorem we state and prove one corollary:

Corollary 6.3 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω. Assume $x, y \in \partial\Omega$ are extreme points and $s_{\partial\Omega}(x, y) > 2$. If $A, B \subset \overline{\Omega}$ are neighborhoods of x and y, then there exists $g \in \Gamma$ with*

$$g(\overline{\Omega} \setminus B) \subset A \quad \text{and} \quad g^{-1}(\overline{\Omega} \setminus A) \subset B.$$

Proof A result of Vey [27, Theorem 5] implies that Γ is strongly irreducible and Proposition 2.13 implies that $\partial\Omega = \mathcal{L}_\Omega(\Gamma)$, so Theorem 6.1 implies the corollary. □

Lemma 6.4 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain, $\gamma \in \mathrm{Aut}(\Omega)$ is biproximal, and $s_{\partial\Omega}(\ell_\gamma^+, \ell_\gamma^-) > 2$. If $A, B \subset \overline{\Omega}$ are neighborhoods of ℓ_γ^+ and ℓ_γ^-, then there exists $N \geq 0$ such that*

$$\gamma^n(\overline{\Omega} \setminus B) \subset A \quad \text{and} \quad \gamma^{-n}(\overline{\Omega} \setminus A) \subset B$$

for all $n \geq N$.

Proof Observation 2.17 implies that

$$\ell_\gamma^+ = \lim_{n\to\infty} \gamma^n(x) \tag{3}$$

for all $x \in \mathbb{P}(\mathbb{R}^d) - \mathbb{P}(H_g^-)$ and the convergence is locally uniform.

We claim that

$$\mathbb{P}(H_g^-) \cap \overline{\Omega} = \{\ell_g^-\}.$$

Proposition 2.12 implies that $\{\ell_g^-\} \subset \mathbb{P}(H_g^-) \cap \overline{\Omega}$ and that $\Omega \cap \mathbb{P}(H_g^-) = \varnothing$. So if $y \in \mathbb{P}(H_g^-) \cap \overline{\Omega}$ then $[y, \ell_g^-] \subset \mathbb{P}(H_g^-) \cap \overline{\Omega}$, and hence $[y, \ell_g^-] \subset \partial\Omega$. Then, by Theorem 2.20(ii), we have $y = \ell_g^+$. So $\mathbb{P}(H_g^-) \cap \overline{\Omega} \subset \{\ell_g^-\}$ and the claim is established.

Then, by the locally uniform convergence in (3), there exists $N_1 > 0$ such that

$$\gamma^n(\overline{\Omega} \setminus B) \subset A$$

for all $n \geq N_1$.

Repeating the same argument with γ^{-1} shows that there exists $N_2 > 0$ such that

$$\gamma^{-n}(\overline{\Omega} \setminus A) \subset B$$

for all $n \geq N_2$.

Then $N = \max\{N_1, N_2\}$ satisfies the conclusion of the lemma. □

Proof of Theorem 6.1 By Theorem 5.1 there exists a sequence of biproximal elements $(g_n)_{n\geq 1}$ in Γ such that

$$\lim_{n\to\infty} \ell^+_{g_n} = x \quad \text{and} \quad \lim_{n\to\infty} \ell^-_{g_n} = y.$$

Since $s_{\partial\Omega}(x, y) > 2$ we may pass to a tail of $(g_n)_{n\geq 1}$ and assume that

$$s_{\partial\Omega}(\ell^+_{g_n}, \ell^-_{g_n}) > 2$$

for all n.

Next, fix n sufficiently large that $\ell^+_{g_n} \in A$ and $\ell^-_{g_n} \in B$. Then, by Lemma 6.4, there exists $m \geq 0$ such that

$$g_n^m(\overline{\Omega} \setminus B) \subset A \quad \text{and} \quad g_n^{-m}(\overline{\Omega} \setminus A) \subset B,$$

so $g = g_n^m$ satisfies the theorem. □

7 Fixed points and centralizers

In this section we prove the following result, connecting the number of boundary fixed points of an element with the size of its centralizer:

Theorem 7.1 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group that acts cocompactly on Ω. If $g \in \Gamma$ has infinite order then the following are equivalent*:

- (i) *There exist two distinct points $x, y \in \partial\Omega$ fixed by g with $s_{\partial\Omega}(x, y) < +\infty$.*
- (ii) *g fixes at least three points in $\partial\Omega$.*
- (iii) *The cyclic group $g^{\mathbb{Z}}$ has infinite index in its centralizer.*

Corollary 7.2 *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group that acts cocompactly on Ω. If $g \in \Gamma$ is biproximal, then the following are equivalent*:

(i) $[\ell_g^+, \ell_g^-] \subset \partial\Omega$.

(ii) $s_{\partial\Omega}(\ell_g^+, \ell_g^-) < +\infty$.

(iii) *g has at least three fixed points in $\partial\Omega$.*

(iv) *The cyclic group $g^{\mathbb{Z}}$ has infinite index in its centralizer.*

We will first recall some results established in [19], then prove the theorem and corollary.

7.1 Maximal abelian subgroups and minimal translation sets

Theorem 7.3 (Islam and Zimmer [19, Theorem 1.6]) *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group that acts cocompactly on Ω. If $A \leq \Gamma$ is a maximal abelian subgroup of Γ then there exists a properly embedded simplex $S \subset \Omega$ such that*

(i) *S is A–invariant*,

(ii) *A acts cocompactly on S, and*

(iii) *A fixes each vertex of S.*

Moreover, A has a finite-index subgroup isomorphic to $\mathbb{Z}^{\dim(S)}$.

Remark 7.4 The above result is a special case of [19, Theorem 1.6], which holds in the more general case when $\Gamma \leq \mathrm{Aut}(\Omega)$ is a naive convex cocompact subgroup.

Definition 7.5 Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $g \in \mathrm{Aut}(\Omega)$. Define the *minimal translation length of g* to be

$$\tau_\Omega(g) := \inf_{x \in \Omega} H_\Omega(x, g(x))$$

and the *minimal translation set of g* to be

$$\mathrm{Min}_\Omega(g) = \{x \in \Omega : H_\Omega(g(x), x) = \tau_\Omega(g)\}.$$

Cooper, Long and Tillmann [14] showed that the minimal translation length of an element can be determined from its eigenvalues:

Proposition 7.6 [14, Proposition 2.1] *If $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $g \in \mathrm{Aut}(\Omega)$, then*

$$\tau_\Omega(g) = \tfrac{1}{2} \log \frac{\lambda_1(g)}{\lambda_d(g)}.$$

Remark 7.7 Recall that

$$\lambda_1(g) \geq \lambda_2(g) \geq \cdots \geq \lambda_d(g)$$

denote the absolute values of the eigenvalues of some (and hence any) lift of g to $\mathrm{SL}_d^{\pm}(\mathbb{R}) := \{h \in \mathrm{GL}_d(\mathbb{R}) : \det h = \pm 1\}$.

As a consequence of Proposition 7.6, we observe the following:

Observation 7.8 *If $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain, $p_0 \in \Omega$, and $g \in \mathrm{Aut}(\Omega)$, then*

$$\lim_{n\to\infty} \frac{1}{n} H_\Omega(g^n(p_0), p_0) = \tau_\Omega(g).$$

Proof Proposition 7.6 implies that $\tau_\Omega(g^n) = n\tau_\Omega(g)$, and hence

$$\liminf_{n\to\infty} \frac{1}{n} H_\Omega(g^n(p_0), p_0) \geq \tau_\Omega(g).$$

For the other inequality, fix $\epsilon > 0$ and $q \in \Omega$ with $H_\Omega(g(q), q) < \tau_\Omega(g) + \epsilon$. Then

$$\begin{aligned}
\limsup_{n\to\infty} & \frac{H_\Omega(g^n(p_0), p_0)}{n} \\
&\leq \limsup_{n\to\infty} \frac{H_\Omega(g^n(q), q) + 2H_\Omega(p_0, q)}{n} \\
&\leq \limsup_{n\to\infty} \frac{H_\Omega(g^n(q), g^{n-1}(q)) + \cdots + H_\Omega(g(q), q) + 2H_\Omega(p_0, q)}{n} \\
&= \limsup_{n\to\infty} H_\Omega(g(q), q) + \frac{2H_\Omega(p_0, q)}{n} < \tau_\Omega(g) + \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, the proof is complete. $\square$

Next, given a group G and an element $g \in G$, let $C_G(g)$ denote the centralizer of g in G. Then given a subset $X \subset G$, define

$$C_G(X) = \bigcap_{x\in X} C_G(x).$$

Theorem 7.9 (Islam and Zimmer [19, Theorem 1.10]) *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain, $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group that acts cocompactly on Ω, and $A \leq \Gamma$ is an abelian subgroup. Then*

$$\mathrm{Min}_\Omega(A) := \bigcap_{a\in A} \mathrm{Min}_\Omega(a)$$

is nonempty and $C_\Gamma(A)$ acts cocompactly on the convex hull of $\mathrm{Min}_\Omega(A)$ in Ω.

Remark 7.10 The above result is a special case of [19, Theorem 1.9], which holds in the more general case when $\Gamma \leq \mathrm{Aut}(\Omega)$ is a naive convex cocompact subgroup.

Proposition 7.11 *Suppose that $S \subset \mathbb{P}(\mathbb{R}^d)$ is a simplex. If $g \in \mathrm{Aut}(S)$ fixes every vertex of S, then $\mathrm{Min}_S(g) = S$.*

Proof See for instance [19, Proposition 7.3]. □

Observation 7.12 *Suppose $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group. If $g \in \Gamma$ is biproximal and $(\ell_g^+, \ell_g^-) \subset \Omega$, then $g^{\mathbb{Z}}$ has finite index in $C_\Gamma(g)$.*

Proof First notice that $C_\Gamma(g)$ preserves (ℓ_g^+, ℓ_g^-). Since $\mathrm{Aut}(\Omega)$ acts properly on Ω and $\Gamma \leq \mathrm{Aut}(\Omega)$ is discrete, we see that $C_\Gamma(g)$ acts properly on (ℓ_g^+, ℓ_g^-). Then $g^{\mathbb{Z}}$ has finite index in $C_\Gamma(g)$ since $g^{\mathbb{Z}}$ acts cocompactly on (ℓ_g^+, ℓ_g^-). □

7.2 Proof of Theorem 7.1

Fix a maximal abelian subgroup $A \leq \Gamma$ which contains g. Then, by Theorem 7.3, there exists $S \subset \Omega$ such that

- S is a properly embedded simplex,
- A acts cocompactly on S,
- A fixes every vertex of S, and
- A has a finite-index subgroup isomorphic to $\mathbb{Z}^{\dim(S)}$.

Since g has infinite order, $\dim(S) \geq 1$.

We consider a number of cases and prove that in each case (i), (ii), and (iii) are either all true or all false.

Case 1 Assume $\dim(S) \geq 2$. Then clearly (i), (ii), and (iii) are all true.

Case 2 Assume $\dim(S) = 1$. Let v^+ and v^- be the vertices of S and fix some $p_0 \in S$. Then, after possibly relabeling, we can assume that

$$\lim_{n \to \pm\infty} g^n(p_0) = v^{\pm}.$$

Case 2(a) Assume $s_{\partial\Omega}(v^+, v^-) > 2$. Then Theorem 2.20 implies that g is a rank-one isometry and $v^{\pm} = \ell_g^{\pm}$. Theorem 2.20 also implies that v^+ and v^- are the only fixed points of g in $\partial\Omega$ and $s_{\partial\Omega}(v^+, v^-) = \infty$. Hence (i) and (ii) are false. Observation 7.12 implies that $g^{\mathbb{Z}}$ has finite index in $C_\Gamma(g)$ and hence (iii) is false.

Case 2(b) Assume $s_{\partial\Omega}(v^+, v^-) = 2$. Then, by definition, (i) is true. Fix $y_0 \in \partial\Omega$ such that $[v^+, y_0] \cup [y_0, v^-]$.

Pick a sequence $n_j \to \infty$ such that the limits

$$T^\pm := \lim_{j\to\infty} g^{\pm n_j}$$

exist in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. Then Proposition 2.12 implies that $v^\mp \in \mathbb{P}(\ker T^\pm)$ and $\mathbb{P}(\ker T^\pm) \cap \Omega = \varnothing$. This implies that $v^\pm \notin \mathbb{P}(\ker T^\pm)$ since $(v^+, v^-) \subset \Omega$. Also, g commutes with $T^\pm$ and hence $g\mathbb{P}(\ker T^\pm) = \mathbb{P}(\ker T^\pm)$.

Passing to a further sequence, we can suppose that $g^{\pm n_j}(y_0) \to y^\pm$. Then

$$[v^+, y^\pm] \cup [y^\pm, v^-] \subset \partial\Omega$$

and so, since $(v^+, v^-) \subset \Omega$, $y^\pm$ must be distinct from v^+ and v^-. Since $g^{\pm n_j}(x) \to v^\pm$ for all $x \in \mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(\ker T^\pm)$, we must have $y \in \mathbb{P}(\ker T^+ \cap \ker T^-)$. Thus the set

$$C := \partial\Omega \cap \mathbb{P}(\ker T^+ \cap \ker T^-)$$

is nonempty. Then g has a fixed point $y \in C$ since C is g–invariant, closed, and convex, so g has at least three fixed points in $\partial\Omega$ and (ii) is true.

Recall that $v^\mp \in \mathbb{P}(\ker T^\pm)$ and $\mathbb{P}(\ker T^\pm) \cap \Omega = \varnothing$; hence,

$$[v^+, y] \cup [y, v^-] \subset \partial\Omega.$$

Let S' be the open simplex with vertices v^+, v^- and y. Since $(v^+, v^-) \subset \Omega$ we have $S' \subset \Omega$. In particular,

$$H_{S'}(p, q) \geq H_\Omega(p, q) \tag{4}$$

for all $p, q \in S'$. Since $p_0 \in (v^-, v^+) \subset S' \subset \Omega$, Observation 7.8 implies that

$$\tau_\Omega(g) = \lim_{n\to\infty} \frac{H_\Omega(g^n(p_0), p_0)}{n} = \lim_{n\to\infty} \frac{H_{S'}(g^n(p_0), p_0)}{n} = \tau_{S'}(g).$$

Then, by (4) and Proposition 7.11,

$$S' = \mathrm{Min}_{S'}(g) \subset \mathrm{Min}_\Omega(g).$$

Now we claim that $g^{\mathbb{Z}}$ has infinite index in $C_\Gamma(g)$. Theorem 7.9 implies that there is a compact set $K \subset \Omega$ such that

$$S' \cup (v^+, v^-) \subset C_\Gamma(g) \cdot K.$$

Further, $g^{\mathbb{Z}}$ preserves (v^+, v^-), so it is enough to show that

$$\sup_{p\in S'} H_\Omega(p, (v^+, v^-)) = \infty.$$

Fix $(p_n)_{n\geq 1}$ in S' converging to y. Since $(v^+, v^-) \subset \Omega$ and $[v^+, y] \cup [y, v^-] \subset \partial\Omega$, Observation 2.11 implies that the faces $F_\Omega(v^+)$, $F_\Omega(v^-)$, and $F_\Omega(y)$ are all distinct. Then, by the definition of the Hilbert metric,

$$\lim_{n\to\infty} H_\Omega(p_n, (v^+, v^-)) = \infty.$$

Thus $g^{\mathbb{Z}}$ has infinite index in $C_\Gamma(g)$ and so (iii) is true.

7.3 Proof of Corollary 7.2

Theorem 7.1 implies that (ii) $\Longrightarrow$ (iii) $\Longleftrightarrow$ (iv), and by definition (i) $\Longrightarrow$ (ii). Finally, by Observation 7.12, (iv) $\Longrightarrow$ (i).

8 Rank in the sense of Prasad and Raghunathan

In this section we consider the rank of a group in the sense of [24].

Definition 8.1 (Prasad and Raghunathan) Suppose that Γ is an abstract group. For $i \geq 0$ let $A_i(\Gamma) \subset \Gamma$ be the subset of elements whose centralizer contains a free abelian group of rank at most i as a subgroup of finite index. Next define $r(\Gamma)$ to be the minimal $i \in \{0, 1, 2, \dots\} \cup \{\infty\}$ such that there exist $\gamma_1, \dots, \gamma_m \in \Gamma$ with

$$\Gamma \subset \bigcup_{j=1}^{m} \gamma_j A_i(\Gamma).$$

Then the *Prasad–Raghunathan rank of* Γ is defined to be

$$\operatorname{rank}_{\mathrm{PR}}(\Gamma) := \sup\{r(\Gamma^*) : \Gamma^* \text{ is a finite-index subgroup of } \Gamma\}.$$

Prasad and Raghunathan computed the rank of lattices in semisimple Lie groups, which implies:

Theorem 8.2 [24, Theorem 3.9] *Suppose that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain. If Ω is symmetric with real rank r and $\Gamma \leq \operatorname{Aut}(\Omega)$ is a discrete group acting cocompactly on Ω, then* $\operatorname{rank}_{\mathrm{PR}}(\Gamma) = r$.

As a corollary to Selberg's lemma we get a lower bound on the Prasad–Raghunathan rank:

Corollary 8.3 *If $\Gamma \leq \mathrm{PGL}_d(\mathbb{R})$ is a finitely generated infinite group*, $\operatorname{rank}_{\mathrm{PR}}(\Gamma) \geq 1$.

Proof By Selberg's lemma, there exists a finite-index torsion-free subgroup $\Gamma^* \leq \Gamma$. Notice that every element of $A_0(\Gamma^*)$ has finite order and hence $A_0(\Gamma^*) = \{\mathrm{id}\}$. Then, since Γ^* is infinite,

$$\mathrm{rank}_{\mathrm{PR}}(\Gamma) \geq r(\Gamma^*) \geq 1. \qquad \square$$

In this section we will show that the existence of a rank-one isometry implies that the Prasad–Raghunathan rank is one.

Proposition 8.4 *Suppose $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is a properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a finitely generated strongly irreducible discrete group. If there exists a biproximal element $g \in \Gamma$ with $(\ell_g^+, \ell_g^-) \subset \Omega$, then*

$$\mathrm{rank}_{\mathrm{PR}}(\Gamma) = 1.$$

Remark 8.5 The proof of Proposition 8.4 is a simple modification of Ballmann and Eberlein's proof [4] of the analogous statement for CAT(0) groups.

The rest of the section is devoted to the proof of Proposition 8.4, so suppose $\Omega \subset \mathbb{P}(\mathbb{R}^d)$, $\Gamma \leq \mathrm{Aut}(\Omega)$, and $g \in \Gamma$ satisfy the hypothesis of the proposition. By Corollary 8.3 it is enough to fix a finite-index subgroup $\Gamma^* \subset \Gamma$ and show that $r(\Gamma^*) \leq 1$. Also, by replacing g with a sufficiently large power, we may assume that $g \in \Gamma^*$.

Lemma 8.6 *Suppose that $x_1, x_2 \in \partial\Omega$ and $(x_1, x_2) \subset \Omega$. If $A, B \subset \partial\Omega$ are open sets with $\bar{A} \cap \bar{B} = \varnothing$, then we can find disjoint neighborhoods V_1 and V_2 of x_1 and x_2 such that for each $\varphi \in \mathrm{Aut}(\Omega)$ at least one of the following occurs*:

(i) $\varphi(V_1) \cap A = \varnothing$.

(ii) $\varphi(V_1) \cap B = \varnothing$.

(iii) $\varphi(V_2) \cap A = \varnothing$.

(iv) $\varphi(V_2) \cap B = \varnothing$.

Proof The following argument is essentially the proof of Lemma 3.10 in [4].

Fix a distance $d_{\mathbb{P}}$ on $\mathbb{P}(\mathbb{R}^d)$ induced by a Riemannian metric. Then, for each n and $j = 1, 2$, let $V_{j,n}$ be a neighborhood of x_j whose diameter with respect to $d_{\mathbb{P}}$ is less than $1/n$.

Suppose for a contradiction that the lemma is false. Then, for each n, there exists $\varphi_n \in \mathrm{Aut}(\Omega)$ such that

$$\varphi_n(V_{j,n}) \cap A \neq \varnothing \quad \text{and} \quad \varphi_n(V_{j,n}) \cap B \neq \varnothing \tag{5}$$

for $j = 1, 2$. By passing to a subsequence, we can suppose that

$$T := \lim_{n\to\infty} \varphi_n$$

exists in $\mathbb{P}(\mathrm{End}(\mathbb{R}^d))$. Then

$$T(u) = \lim_{n\to\infty} \varphi_n(u)$$

for all $u \in \mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(\ker T)$. Moreover, the convergence is uniform on compact subsets of $\mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(\ker T)$.

Proposition 2.12 implies that $\mathbb{P}(\ker T) \cap \Omega = \varnothing$. Then, since $(x_1, x_2) \subset \Omega$, it is impossible for both x_1 and x_2 to be contained in $\mathbb{P}(\ker T)$. So, after possibly relabelling, we may assume that $x_1 \notin \mathbb{P}(\ker T)$.

By (5) there exist sequences $a_n, b_n \in \partial\Omega$ converging to x_1 such that $\varphi_n(a_n) \in A$ and $\varphi_n(b_n) \in B$. Then, since $x_1 \notin \mathbb{P}(\ker T)$,

$$T(x_1) = \lim_{n\to\infty} \varphi_n(a_n) \in \bar{A} \quad \text{and} \quad T(x_1) = \lim_{n\to\infty} \varphi_n(b_n) \in \bar{B}.$$

So $T(x_1) \in \bar{A} \cap \bar{B} = \varnothing$, which is a contradiction. □

Lemma 8.7 $r(\Gamma^*) \leq 1$.

Proof The following argument is essentially the proof of Theorem 3.1 in [4].

Since Γ is strongly irreducible Γ^* is also strongly irreducible, so, by Observation 2.2, there exists $\phi \in \Gamma^*$ such that

$$\phi\ell_g^+, \phi\ell_g^-, \ell_g^+ \text{ and } \ell_g^-$$

are all distinct. Then $h := \phi g \phi^{-1}$ is biproximal, $\ell_h^\pm = \phi\ell_g^\pm$, and

$$(\ell_h^+, \ell_h^-) = \phi(\ell_g^+, \ell_g^-) \subset \Omega.$$

Fix open neighborhoods $A, B \subset \partial\Omega$ of ℓ_h^+ and ℓ_h^- such that $\bar{A} \cap \bar{B} = \varnothing$. Then let $V_1, V_2 \subset \partial\Omega$ be neighborhoods of ℓ_g^+ and ℓ_g^- such that A, B, V_1 and V_2 satisfy Lemma 8.6.

By further shrinking each V_j, we can assume that each $\partial\Omega \setminus V_j$ is homeomorphic to a closed ball.

Next, let $U_1 \subset V_1$ be a closed neighborhood of ℓ_g^+ such that, if $x \in U_1$ and $y \in \partial\Omega \setminus V_1$, then $s_{\partial\Omega}(x, y) > 2$. Such a choice is possible by Theorem 2.20(ii). In a similar fashion, let $U_2 \subset V_2$ be a closed neighborhood of ℓ_g^- such that, if $x \in U_2$ and $y \in \partial\Omega \setminus V_2$, then $s_{\partial\Omega}(x, y) > 2$.

By further shrinking each U_j, we can assume that each U_j is homeomorphic to a closed ball.

By Observation 2.18, each $\ell_g^\pm$ and $\ell_h^\pm$ is an extreme point of Ω. Furthermore, by Theorem 2.20(iii),

$$s_{\partial\Omega}(\ell_g^\pm, \ell_h^\pm) = \infty = s_{\partial\Omega}(\ell_g^\pm, \ell_h^\mp).$$

So, by Theorem 6.1, there exist $\varphi_1, \varphi_2, \psi_1, \psi_2 \in \Gamma^*$ such that

(i) $\varphi_1(\partial\Omega \setminus A) \subset U_1$ and $\varphi_1^{-1}(\partial\Omega \setminus U_1) \subset A$,

(ii) $\psi_1(\partial\Omega \setminus A) \subset U_2$ and $\psi_1^{-1}(\partial\Omega \setminus U_2) \subset A$,

(iii) $\varphi_2(\partial\Omega \setminus B) \subset U_1$ and $\varphi_2^{-1}(\partial\Omega \setminus U_1) \subset B$,

(iv) $\psi_2(\partial\Omega \setminus B) \subset U_2$ and $\psi_2^{-1}(\partial\Omega \setminus U_2) \subset B$.

We claim that

$$\Gamma^* = \varphi_1^{-1} A_1(\Gamma^*) \cup \psi_1^{-1} A_1(\Gamma^*) \cup \varphi_2^{-1} A_1(\Gamma^*) \cup \psi_2^{-1} A_1(\Gamma^*).$$

Fix $\gamma \in \Gamma^*$. By construction, at least one of the four possibilities in Lemma 8.6 must occur.

Case 1 Assume $\gamma(V_1) \cap A = \varnothing$. Then

$$\varphi_1\gamma(U_1) \subsetneq \varphi_1\gamma(V_1) \subset \varphi_1(\partial\Omega \setminus A) \subset U_1, \tag{6}$$

so, by the Brouwer fixed-point theorem, $\varphi_1\gamma$ has a fixed point in $x \in U_1$ (recall that U_1 is homeomorphic to a closed ball). Further,

$$(\varphi_1\gamma)^{-1}(\partial\Omega \setminus V_1) \subset (\varphi_1\gamma)^{-1}(\partial\Omega \setminus U_1) \subset \gamma^{-1}(A) \subset \partial\Omega \setminus V_1,$$

so $\varphi_1\gamma$ also has a fixed point in $y \in \partial\Omega \setminus V_1$. Now, by construction, $s_{\partial\Omega}(x, y) > 2$. So, by Theorem 2.20(i), either

$$\inf_{p\in\Omega} H_\Omega(\varphi_1\gamma(p), p) = 0$$

or $\varphi_1\gamma$ is biproximal with

$$\{x, y\} = \{\ell^+_{\varphi_1\gamma}, \ell^-_{\varphi_1\gamma}\}.$$

In the latter case, $(\ell^+_{\varphi_1\gamma}, \ell^-_{\varphi_1\gamma}) \subset \Omega$, and so $\varphi_1\gamma \in A_1(\Gamma)$ by Observation 7.12. Thus we have reduced to showing that

$$\inf_{p\in\Omega} H_\Omega(\varphi_1\gamma(p), p) > 0.$$

Assume for a contradiction that

$$\inf_{p \in \Omega} H_\Omega(\varphi_1\gamma(p), p) = 0.$$

Then, by Proposition 7.6, we have

$$\lambda_1(\varphi_1\gamma) = \lambda_2(\varphi_1\gamma) = \cdots = \lambda_d(\varphi_1\gamma).$$

Since x and y are eigenlines of $\varphi_1\gamma$, this implies that $\varphi_1\gamma$ fixes every point of the line (x, y). Then, since $\mathrm{Aut}(\Omega)$ acts properly on Ω and Γ^* is discrete, the group

$$K = \{(\varphi_1\gamma)^n : n \in \mathbb{Z}\}$$

is finite. So $(\varphi_1\gamma)^N = \mathrm{id}$ for some large N. Then (6) implies that

$$U_1 = (\varphi_1\gamma)^N(U_1) \subsetneq U_1.$$

So we have a contradiction, and hence

$$\inf_{p \in \Omega} H_\Omega(\varphi_1\gamma(p), p) > 0$$

and so $\varphi_1\gamma \in A_1(\Gamma^*)$.

Case 2 Assume $\gamma(V_1) \cap B = \varnothing$. Then arguing as in Case 1 shows that $\varphi_2\gamma \in A_1(\Gamma^*)$.

Case 3 Assume $\gamma(V_2) \cap A = \varnothing$. Then arguing as in Case 1 shows that $\psi_1\gamma \in A_1(\Gamma^*)$.

Case 4 Assume $\gamma(V_2) \cap B = \varnothing$. Then arguing as in Case 1 shows that $\psi_2\gamma \in A_1(\Gamma^*)$.

Since $\gamma \in \Gamma^*$ was arbitrary,

$$\Gamma^* = \varphi_1^{-1}A_1(\Gamma^*) \cup \psi_1^{-1}A_1(\Gamma^*) \cup \varphi_2^{-1}A_1(\Gamma^*) \cup \psi_2^{-1}A_1(\Gamma^*).$$

Hence $r(\Gamma^*) \leq 1$. □

9 Proof of Theorem 1.4

Suppose for the rest of the section that $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ is an irreducible properly convex domain and $\Gamma \leq \mathrm{Aut}(\Omega)$ is a discrete group that acts cocompactly on Ω. We will show that the following conditions are equivalent:

(i) Ω is symmetric with real rank at least two.

(ii) Ω has higher rank.

(iii) The extreme points of Ω form a closed proper subset of $\partial\Omega$.

$$
\begin{tikzcd}
(\mathrm{vi}) \arrow[r, Rightarrow, "\text{Lemma 9.4}"] & (\mathrm{viii}) \arrow[r, Leftrightarrow, "\text{Theorem 7.1}"] & (\mathrm{ix}) \arrow[r, Rightarrow, "\text{Corollary 7.2}"] & (\mathrm{x}) \arrow[r, Leftrightarrow, "\text{Corollary 7.2}"] \arrow[dll, Rightarrow, "\text{Lemma 9.5}"'] & (\mathrm{xi}) \\
(\mathrm{v}) \arrow[u, Rightarrow, "\text{definition}"'] & (\mathrm{iv}) \arrow[ul, Rightarrow, "\text{definition}"'] \arrow[d, Rightarrow, "\text{Lemma 9.3}"] & & & \\
(\mathrm{ii}) \arrow[u, Rightarrow, "\text{Lemma 9.2}"'] & (\mathrm{xii}) \arrow[dl, Rightarrow, "\text{Theorem 4.1}"'] & (\mathrm{iii}) \arrow[l, Rightarrow, "\text{definition}"'] & & \\
(\mathrm{i}) \arrow[u, Rightarrow, "\text{Lemma 9.1}"'] \arrow[urr, Rightarrow, "\text{Lemma 9.1}"] \arrow[rrr, Rightarrow, "\text{Theorem 8.2}"] & & & (\mathrm{vii}) \arrow[uuu, Rightarrow, "\text{Proposition 8.4}"] &
\end{tikzcd}
$$

Figure 1: The proof of Theorem 1.4.

(iv) $[x_1, x_2] \subset \partial\Omega$ for every two extreme points $x_1, x_2 \in \partial\Omega$.

(v) $s_{\partial\Omega}(x, y) \le 2$ for all $x, y \in \partial\Omega$.

(vi) $s_{\partial\Omega}(x, y) < +\infty$ for all $x, y \in \partial\Omega$.

(vii) Γ has higher rank in the sense of Prasad and Raghunathan.

(viii) For every $g \in \Gamma$ with infinite order, the cyclic group $g^{\mathbb{Z}}$ has infinite index in the centralizer $C_\Gamma(g)$ of g in Γ.

(ix) Every $g \in \Gamma$ with infinite order has at least three fixed points in $\partial\Omega$.

(x) $[\ell_g^+, \ell_g^-] \subset \partial\Omega$ for every biproximal element $g \in \Gamma$.

(xi) $s_{\partial\Omega}(\ell_g^+, \ell_g^-) < +\infty$ for every biproximal element $g \in \Gamma$.

(xii) There exists a boundary face $F \subset \partial\Omega$ such that

$$F \cap \bar{\mathcal{E}}_\Omega = \varnothing.$$

We verify all the implications shown in Figure 1. First notice that (iii) $\Longrightarrow$ (xii), (iv) $\Longrightarrow$ (vi), and (v) $\Longrightarrow$ (vi) are by definition. The implication (i) $\Longrightarrow$ (vii) is due to Prasad and Raghunathan; see Theorem 8.2 above. Proposition 8.4 implies that (vii) $\Longrightarrow$ (x). Theorem 7.1 implies that (viii) $\Longleftrightarrow$ (ix). Corollary 7.2 implies that (ix) $\Longrightarrow$ (x) and (x) $\Longleftrightarrow$ (xi). Theorem 4.1 implies that (xii) $\Longrightarrow$ (i). The remaining implications in Figure 1 are given as lemmas below.

Lemma 9.1 (i) $\Longrightarrow$ (ii) *and* (i) $\Longrightarrow$ (iii).

Proof These implications follow from direct inspection of the short list of irreducible symmetric properly convex domains. □

Lemma 9.2 (ii) $\Longrightarrow$ (v).

Proof Suppose $x, y \in \partial\Omega$. If $[x,y] \subset \partial\Omega$, then $s_{\partial\Omega}(x,y) \le 1$. If $(x,y) \subset \Omega$, then there exists a properly embedded simplex $S \subset \Omega$ with $\dim(S) \ge 2$ and $(x,y) \subset S$. Then

$$s_{\partial\Omega}(x,y) \le s_{\partial S}(x,y) \le 2.$$

Since $x, y \in \partial\Omega$ were arbitrary, we see that (v) holds. □

Lemma 9.3 (iv) $\Longrightarrow$ (xii).

Proof Fix a boundary face $F \subset \partial\Omega$ of maximal dimension. We claim that

$$\bar{\mathcal{E}}_\Omega \cap F = \varnothing.$$

Otherwise, there exists $x \in F$ and a sequence $x_n \in \mathcal{E}_\Omega$ such that $x_n \to x \in F$. Now fix an extreme point $y \in \partial\Omega \setminus \bar{F}$. Then, by hypothesis, $[x_n, y] \subset \partial\Omega$ for all n, so $[x,y] \subset \partial\Omega$.

Fix $z \in (x,y) \subset \partial\Omega$ and let C denote the convex hull of y and F. By Observation 2.11,

$$\partial\Omega \supset F_\Omega(z) \supset \text{rel-int}(C).$$

Then

$$\dim F_\Omega(z) > \dim F,$$

which is a contradiction. So we must have $\bar{\mathcal{E}}_\Omega \cap F = \varnothing$, and hence (xii) holds. □

Lemma 9.4 (vi) $\Longrightarrow$ (viii).

Proof By Theorem 7.3 every infinite-order element $g \in \Gamma$ preserves a properly embedded simplex $S \subset \Omega$ with $\dim(S) \ge 1$. Hence g fixes the vertices $v_1, \dots, v_k$ of S. By hypothesis $s_{\partial\Omega}(v_1, v_2) < +\infty$ and hence, by Theorem 7.1, $g^{\mathbb{Z}}$ has infinite index in the centralizer $C_\Gamma(g)$. □

Lemma 9.5 (x) $\Longrightarrow$ (iv).

Proof We prove the contrapositive: if there exist extreme points $x, y \in \partial\Omega$ with $(x,y) \subset \Omega$, then there exists a biproximal element $g \in \Gamma$ with $(\ell_g^+, \ell_g^-) \subset \Omega$. If such x and y exist, then by Theorem 5.1 there exist biproximal elements $g_n \in \Gamma$ with $\ell_{g_n}^+ \to x$ and $\ell_{g_n}^- \to y$. Then, for n large, we must have $(\ell_{g_n}^+, \ell_{g_n}^-) \subset \Omega$. □

References

[1] **W Ballmann**, *Axial isometries of manifolds of nonpositive curvature*, Math. Ann. 259 (1982) 131–144 MR Zbl

[2] **W Ballmann**, *Nonpositively curved manifolds of higher rank*, Ann. of Math. 122 (1985) 597–609 MR Zbl

[3] **W Ballmann**, *Lectures on spaces of nonpositive curvature*, DMV Seminar 25, Birkhäuser, Basel (1995) MR Zbl

[4] **W Ballmann**, **P Eberlein**, *Fundamental groups of manifolds of nonpositive curvature*, J. Differential Geom. 25 (1987) 1–22 MR Zbl

[5] **Y Benoist**, *Convexes divisibles, II*, Duke Math. J. 120 (2003) 97–120 MR Zbl

[6] **Y Benoist**, *Convexes divisibles, IV: Structure du bord en dimension* 3, Invent. Math. 164 (2006) 249–278 MR Zbl

[7] **Y Benoist**, *A survey on divisible convex sets*, from "Geometry, analysis and topology of discrete groups" (L Ji, K Liu, L Yang, S-T Yau, editors), Adv. Lect. Math. 6, International, Somerville, MA (2008) 1–18 MR Zbl

[8] **Y Benoist**, **J-F Quint**, *Random walks on reductive groups*, Ergebnisse der Math. (3) 62, Springer (2016) MR Zbl

[9] **J-P Benzécri**, *Sur les variétés localement affines et localement projectives*, Bull. Soc. Math. France 88 (1960) 229–332 MR Zbl

[10] **A Borel**, *Compact Clifford–Klein forms of symmetric spaces*, Topology 2 (1963) 111–122 MR Zbl

[11] **K Burns**, **R Spatzier**, *Manifolds of nonpositive curvature and their buildings*, Inst. Hautes Études Sci. Publ. Math. 65 (1987) 35–59 MR Zbl

[12] **K Burns**, **R Spatzier**, *On topological Tits buildings and their classification*, Inst. Hautes Études Sci. Publ. Math. 65 (1987) 5–34 MR Zbl

[13] **H Busemann**, **P J Kelly**, *Projective geometry and projective metrics*, Academic, New York (1953) MR Zbl

[14] **D Cooper**, **D D Long**, **S Tillmann**, *On convex projective manifolds and cusps*, Adv. Math. 277 (2015) 181–251 MR Zbl

[15] **J Faraut**, **A Korányi**, *Analysis on symmetric cones*, Clarendon, New York (1994) MR Zbl

[16] **W M Goldman**, *Convex real projective structures on compact surfaces*, J. Differential Geom. 31 (1990) 791–845 MR Zbl

[17] **P de la Harpe**, *On Hilbert's metric for simplices*, from "Geometric group theory, I" (G A Niblo, M A Roller, editors), London Math. Soc. Lecture Note Ser. 181, Cambridge Univ. Press (1993) 97–119 MR Zbl

[18] **M Islam**, *Rank-one Hilbert geometries*, preprint (2019) arXiv 1912.13013

[19] **M Islam**, **A Zimmer**, *A flat torus theorem for convex co-compact actions of projective linear groups*, J. Lond. Math. Soc. 103 (2021) 470–489 MR Zbl

[20] **G Knieper**, *The uniqueness of the measure of maximal entropy for geodesic flows on rank* 1 *manifolds*, Ann. of Math. 148 (1998) 291–314 MR Zbl

[21] **M Koecher**, *The Minnesota notes on Jordan algebras and their applications*, Lecture Notes in Math. 1710, Springer (1999) MR Zbl

[22] **L Marquis**, *Around groups in Hilbert geometry*, from "Handbook of Hilbert geometry" (A Papadopoulos, M Troyanov, editors), IRMA Lect. Math. Theor. Phys. 22, Eur. Math. Soc., Zürich (2014) 207–261 MR Zbl

[23] **R D Nussbaum**, *Hilbert's projective metric and iterated nonlinear maps*, Mem. Amer. Math. Soc. 391, Amer. Math. Soc., Providence, RI (1988) MR Zbl

[24] **G Prasad**, **M S Raghunathan**, *Cartan subgroups and lattices in semi-simple groups*, Ann. of Math. 96 (1972) 296–317 MR Zbl

[25] **J-F Quint**, *Convexes divisibles (d'après Yves Benoist)*, from "Séminaire Bourbaki 2008/2009", Astérisque 332, Soc. Math. France, Paris (2010) Exposé 999, 45–73 MR Zbl

[26] **C Vernicos**, *On the Hilbert geometry of convex polytopes*, from "Handbook of Hilbert geometry" (A Papadopoulos, M Troyanov, editors), IRMA Lect. Math. Theor. Phys. 22, Eur. Math. Soc., Zürich (2014) 111–125 MR Zbl

[27] **J Vey**, *Sur les automorphismes affines des ouverts convexes saillants*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. 24 (1970) 641–665 MR Zbl

[28] **E B Vinberg**, *The theory of homogeneous convex cones*, Trudy Moskov. Mat. Obšč. 12 (1963) 303–358 MR Zbl In Russian; translated in Trans. Mosc. Math. Soc. 12 (1963) 340–403

[29] **E B Vinberg**, *Structure of the group of automorphisms of a homogeneous convex cone*, Trudy Moskov. Mat. Obšč. 13 (1965) 56–83 MR Zbl In Russian; translated in Trans. Mosc. Math. Soc. r3 (1967) 63–93

Department of Mathematics, Louisiana State University
Baton Rouge, LA, United States

Current address: *Department of Mathematics, University of Wisconsin, Madison*
Madison, WI, United States

amzimmer2@wisc.edu

Proposed: John Lott Received: 23 February 2021
Seconded: Bruce Kleiner, Anna Wienhard Revised: 18 January 2022

Guidelines for Authors

Submitting a paper to Geometry & Topology

Papers must be submitted using the upload page at the GT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Geometry & Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in GT is available on the GT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeX package, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with production by sending an email to graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

GEOMETRY & TOPOLOGY

Volume 27 Issue 7 (pages 2497–2936) 2023