AGT

Algebraic & Geometric Topology

Volume 17 (2017)

Issue 2 (pages 645–1281)

msp

ALGEBRAIC & GEOMETRIC TOPOLOGY
msp.org/agt

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

The subscription price for 2017 is US $410/year for the electronic version, and $650/year (+$60, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Algebraic & Geometric Topology (ISSN 1472-2747 printed, 1472-2739 electronic) at Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published 6 times per year and continuously online. Periodical rate postage paid at Berkeley, CA 94704, 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.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Algebraic & Geometric Topology 17 (2017) 645–656

Indecomposable nonorientable PD_3–complexes

JONATHAN A HILLMAN

We show that the orientable double covering space of an indecomposable, nonorientable PD_3–complex has torsion-free fundamental group.

57P10; 57N10

One of the foundational results of Wall [12] on Poincaré duality complexes was the fact that there is a well-defined notion of connected sum for such complexes. In dimensions $n > 2$ the fundamental group of a connected sum of two PD_n–complexes is the free product of the groups of the summands. This notion is of particular interest when $n = 3$ for, by the well-known work of Kneser and Milnor, every closed orientable 3–manifold has an essentially unique factorization into indecomposable 3–manifolds. (The corresponding assertion for closed nonorientable 3–manifolds is slightly more complicated.) Moreover, such a 3–manifold is indecomposable with respect to connected sum if and only if its fundamental group is indecomposable with respect to free product. It is perhaps less widely known that Turaev [11] has shown that each of these results extends to the context of PD_3–complexes.

Indecomposable, orientable 3–manifolds are either aspherical, have finite fundamental group or have fundamental group $\mathbb{Z}$. This is no longer true for PD_3–complexes, although Crisp [3] has shown that (in the orientable case) the indecomposables are either aspherical or have virtually free fundamental group. There are examples of the latter kind with fundamental group neither finite nor $\mathbb{Z}$; see Hillman [9].

Let X be an indecomposable PD_3–complex, with fundamental group π and orientation character w. In [9] we showed that if $w \neq 1$ and π is virtually free then X is homotopy equivalent to $S^2 \tilde{\times} S^1$ or $\mathbb{RP}^2 \times S^1$, so $\pi \cong \mathbb{Z}$ or $\pi \cong \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$. In particular, $\pi^+ = \mathrm{Ker}(w)$ is torsion-free. We shall show that this remains true if $w \neq 1$ and π is *not* virtually free. This result is surely well-known for 3–manifolds. We give a short proof for this case in Section 2, which uses the "projective plane theorem" of Epstein [6] and a result from Hillman [9]. (The fact that $\mathbb{RP}^2$ does not bound provides a further restriction here which is not yet known in general.) Our main result is Theorem 6 in Section 3:

Theorem *Let X be an indecomposable, nonorientable* PD_3*–complex such that π has infinitely many ends. Then $\pi \cong \pi^+ \rtimes \mathbb{Z}/2\mathbb{Z}^-$ and π^+ is torsion-free, but not free.*

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.645

By passing to Sylow subgroups of the torsion in π, we may reduce potential counterexamples to special cases, which are eliminated by Lemmas 3, 4 and 5. The arguments are similar to those of [9].

1 Notation and major cited results

In order that this paper be reasonably self-contained we shall give here some of the notation and results used in [9].

Let X be a PD_3–complex, with fundamental group π and orientation character w, and let X^+ be the orientable covering space, with fundamental group $\pi^+ = \mathrm{Ker}(w)$. If $H \leq \pi$ then we shall write $H^+ = H \cap \pi^+$. It is convenient to say that such a subgroup H is *orientable* if $H = H^+$. (This usage depends upon the orientation character w.) Let $\mathbb{Z}/2\mathbb{Z}^-$ denote a subgroup of order two on which $w \neq 1$.

If G is a group, $|G|$, G' and ζG shall denote the order, commutator subgroup and centre of G, while if $H \leq G$ then $C_G(H)$ and $N_G(H)$ are the centralizer and normalizer, respectively. Let $F(r)$ be the free group of rank r.

If R is a ring, two finitely presentable left R–modules M and N are *stably isomorphic* if $M_1 \oplus R^a \cong N \oplus R^b$ for some $a, b \geq 0$. Let $[M]$ denote the stable isomorphism class of M.

A homomorphism $w\colon G \to \{\pm 1\}$ defines an anti-involution of $\mathbb{Z}[G]$ by $\overline{g} = w(g)g^{-1}$ for all $g \in G$. Tietze move considerations show that if A is any finite presentation matrix for the augmentation ideal I_G then the stable isomorphism class of the left $\mathbb{Z}[G]$–module J_G with presentation matrix the conjugate transpose $\bar{A}^{\mathrm{tr}}$ is well-defined [11].

A *graph of groups* $(\mathcal{G}, \Gamma)$ consists of a graph Γ with origin and target functions o and t from the set of edges $E(\Gamma)$ to the set of vertices $V(\Gamma)$, and a family $\mathcal{G}$ of groups G_v for each vertex v and subgroups $G_e \leq G_{o(e)}$ for each edge e, with monomorphisms $\phi_e\colon G_e \to G_{t(e)}$. (We shall usually suppress the maps ϕ_e from our notation.) In considering paths in Γ we shall not require that the edges be compatibly oriented.

The *fundamental group* of $(\mathcal{G}, \Gamma)$ is the group $\pi\mathcal{G}$ with presentation

$$\langle G_v, t_e \mid t_e g t_e^{-1} = \phi_e(g) \;\forall g \in G_e,\; t_e = 1 \;\forall e \in E(T)\rangle,$$

where T is some maximal tree for Γ. Different choices of maximal tree give isomorphic groups. We may assume that $(\mathcal{G}, \Gamma)$ is *reduced*: if an edge joins distinct vertices then the edge group is isomorphic to a proper subgroup of each of these vertex groups. The corresponding π–tree T is incompressible in the terminology of [5], so T and $\mathcal{G}$ are essentially unique, by [5, Proposition IV.7.4]. An edge e is a *loop isomorphism* at v if $o(e) = t(e) = v$ and the inclusions induce isomorphisms $G_e \cong G_v$.

Since fundamental groups of PD_n–complexes are $\mathbb{F}P_2$ [12], π is the fundamental group of a finite graph of groups $(\mathcal{G},\Gamma)$, where all vertex groups are finite or have one end and all edge groups are finite. (See [5, Theorem VI.6.3].) We may assume that π is indecomposable as a proper free product, by the splitting theorem, so $(\mathcal{G},\Gamma)$ is *indecomposable*: all edge groups are nontrivial. A graph of groups $(\mathcal{G},\Gamma)$ is *admissible* if it is reduced, all vertex groups are finite or one-ended groups and all edge groups are nontrivial finite groups.

Turaev gave the following characterization of the group pairs (π,w) which may be realized by finite PD_3–complexes [11]:

Theorem *Let π be a finitely presentable group and $w\colon \pi\to\{\pm1\}$ a homomorphism. Then there is a finite PD_3–complex K with $\pi_1(K)\cong\pi$ and $w_1(K)=w$ if and only if $[I_\pi]=[J_\pi]$.*

We wish to adapt the results from [9, Section 7] to the cases when π has infinitely many ends and $w\neq 1$. In particular, we use the following two results to control the possible edge groups:

(1) **Crisp's theorem** [3, Theorem 17] *If X is a PD_3–complex and $g\in\pi=\pi_1(X)$ has prime order p and infinite centralizer $C_\pi(g)$ then $p=2$, g is orientation-reversing and $C_\pi(g)$ has two ends.*

(2) **The normalizer condition** [10, Proposition 5.4.2] *A proper subgroup of a nilpotent group is properly contained in its normalizer.*

Note also that if G is a finite subgroup of π then the centralizer $C_\pi(G)$ has finite index in the normalizer $N_\pi(G)$.

The main result (Theorem 6 below) involves consideration of the finite groups with periodic cohomology, of period dividing 4. A finite group has cohomological period 2 if and only if it is cyclic, and has cohomological period 4 if and only if it is a product $B\times\mathbb{Z}/d\mathbb{Z}$ with $(|B|,d)=1$, where B is a generalized quaternionic group $\mathbb{Z}/a\mathbb{Z}\rtimes Q(2^i)$ (with a odd), an extended binary polyhedral group T_k^* (of order $2^3\cdot 3^k$), O_k^* (of order $2^4\cdot 3^k$) or $I^*=SL(2,5)$ (of order $2^3\cdot3\cdot5$) or a metacyclic group $\mathbb{Z}/a\mathbb{Z}\rtimes_{-1}\mathbb{Z}/2^e\mathbb{Z}$ (for some odd a and $e\geq 1$).

There seems to be no one reference with a complete proof of the above assertion. The six families of finite groups with periodic cohomology are determined in [1, pages 142–150]:

(1) $\mathbb{Z}/a\mathbb{Z}\rtimes\mathbb{Z}/b\mathbb{Z}$;

(2) $\mathbb{Z}/a\mathbb{Z}\rtimes(\mathbb{Z}/b\mathbb{Z}\times Q(2^i))$ for $i\geq 3$;

(3) $\mathbb{Z}/a\mathbb{Z} \rtimes (\mathbb{Z}/b\mathbb{Z} \times T_k^*)$ for $k \geq 1$;
(4) $\mathbb{Z}/a\mathbb{Z} \rtimes (\mathbb{Z}/b\mathbb{Z} \times O_k^*)$ for $k \geq 1$;
(5) $(\mathbb{Z}/a\mathbb{Z} \rtimes \mathbb{Z}/b\mathbb{Z}) \times \mathrm{SL}(2,p)$ for $p \geq 5$ prime;
(6) $\mathbb{Z}/a\mathbb{Z} \rtimes (\mathbb{Z}/b\mathbb{Z} \times \mathrm{TL}(2,p))$ for $p \geq 5$ prime.

Here a, b and the order of the quotient by the metacyclic subgroup $\mathbb{Z}/a\mathbb{Z} \rtimes \mathbb{Z}/b\mathbb{Z}$ are relatively prime. See [1, pages 142–150] for further details on the groups $\mathrm{TL}(2,p)$ (with $\mathrm{TL}(2,p)' \cong \mathrm{SL}(2,p)$, of index 2) and the actions in the semidirect products. If such a group G contains a semidirect product $\mathbb{Z}/m\mathbb{Z} \rtimes_\theta \mathbb{Z}/n\mathbb{Z}$, where θ has image of order k, then the cohomological period of G is a multiple of $2k$. (See [2, Exercise 6, page 159].) The class of groups of period dividing 4 follows on applying this criterion to the groups of the above list.

2 3–manifolds

The result is relatively easy (and no doubt well-known) in the case of irreducible 3–manifolds, as we may use the sphere theorem, as strengthened by Epstein [6].

Theorem 1 *Let M be an indecomposable, nonorientable 3–manifold with fundamental group π. If π has infinitely many ends then $\pi \cong \pi^+ \rtimes \mathbb{Z}/2\mathbb{Z}^-$ and π^+ is torsion-free, but not free.*

Proof Let $\mathcal{P}$ be a maximal set of pairwise nonparallel 2–sided projective planes in M. Then $\mathcal{P}$ is nonempty, since M is indecomposable and π has infinitely many ends. In particular, $\pi \cong \pi^+ \rtimes \mathbb{Z}/2\mathbb{Z}^-$, since the inclusion of a member of $\mathcal{P}$ splits $w = w_1(M)\colon \pi \to \mathbb{Z}/2\mathbb{Z}$. Let $\mathcal{P}^+$ be the preimage of $\mathcal{P}$ in M^+. Then $\mathcal{P}^+$ is a set of disjoint 2–spheres in M^+, and the components of $M^+ \setminus \mathcal{P}^+$ each double cover a component of $M \setminus \mathcal{P}$. Each such component of $M \setminus \mathcal{P}$ is indecomposable [6].

Suppose that $M \setminus \mathcal{P}$ has a component Y with virtually free fundamental group. Then the double DY is indecomposable (see [9, Lemma 2.4]), nonorientable and $\pi_1(DY)$ is virtually free. Moreover, $\pi_1(DY) \cong \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}^-$, since the inclusion of a boundary component of Y splits w. (See [9, Theorems 7.1 and 7.4].) But then $DY \cong \mathbb{RP}^2 \times S^1$, so $Y \cong \mathbb{RP}^2 \times [0,1]$. This is contrary to the hypothesis that the members of $\mathcal{P}$ are nonparallel. Thus the components of $M \setminus \mathcal{P}$ are punctured aspherical 3–manifolds.

Let Γ be the graph with vertex set $\pi_0(M \setminus \mathcal{P})$ and edge set $\mathcal{P}$, with an edge joining contiguous components. Then $\pi^+ \cong G * F(s)$, where G is a free product of PD_3–groups (corresponding to the fundamental groups of the components of $M \setminus \mathcal{P}$), and $s = \beta_1(\Gamma)$. Hence π^+ is torsion-free. □

We remark also that each component Y of $M \setminus \mathcal{P}$ has an even number of boundary components, since $\chi(\partial Y)$ is even (for any odd-dimensional manifold Y), by Poincaré duality. Thus the vertices of the graph Γ have even valence.

Example The canonical involution ι of the topological group $T^3 = \mathbb{R}^3/\mathbb{Z}^3$ has 8 isolated fixed points (the points of order 2). Let X be the complement of an equivariant open regular neighbourhood of the fixed point set, and let $M = D(X/\langle\iota\rangle)$. Then M is indecomposable and nonorientable, and $\pi \cong (\mathbb{Z}^3 * \mathbb{Z}^3 * F(7)) \rtimes \mathbb{Z}/2\mathbb{Z}^-$.

3 PD$_3$–complexes

Suppose now that X is an indecomposable PD$_3$–complex, with fundamental group π and orientation character w. Then π is finitely presentable, so $\pi \cong \pi\mathcal{G}$, where $(\mathcal{G}, \Gamma)$ is an admissible graph of groups.

Lemma 2 *Let X be an indecomposable, nonorientable* PD$_3$*–complex with $\pi = \pi_1(X) \cong \pi\mathcal{G}$, where $(\mathcal{G}, \Gamma)$ is an admissible graph of groups.*

(1) *If e is an edge with $G_{o(e)}$ or $G_{t(e)}$ infinite, then $G_e = \mathbb{Z}/2\mathbb{Z}^-$.*

(2) *If $X \not\simeq S^2 \tilde{\times} S^1$ then $\pi \cong \pi^+ \rtimes \mathbb{Z}/2\mathbb{Z}^-$.*

(3) *If all finite vertex groups are 2–groups then they are nonorientable and all edge groups are $\mathbb{Z}/2\mathbb{Z}^-$.*

Proof Suppose first that the vertex groups are all finite. Then $X \simeq S^2 \tilde{\times} S^1$ (if all the vertex groups are orientation-preserving) or $\mathbb{R}P^2 \times S^1$ (otherwise), by Theorems 7.1 and 7.4 of [9], respectively, so the lemma holds. Hence we may assume that $(\mathcal{G}, \Gamma)$ has at least one infinite vertex group G_v and at least one edge e with $o(e) = v$ or $t(e) = v$. If $w(g) = 1$ for some $g \in G_e$ of prime order then both $G^+_{o(e)}$ and $G^+_{t(e)}$ would be finite, by [3, Theorem 14]. But then G_v would be finite, contrary to hypothesis. Thus $G_e = \mathbb{Z}/2\mathbb{Z}^-$, and the inclusion of G_e into π splits w, so $\pi \cong \pi^+ \rtimes \mathbb{Z}/2\mathbb{Z}^-$.

Suppose that all finite subgroups are 2–groups. Let f be an edge such that the vertex groups $G_{o(f)}$ and $G_{t(f)}$ are finite. If $G_f = G_{o(f)}$ (or $G_{t(f)}$) then f must be a loop isomorphism, since $(\mathcal{G}, \Gamma)$ is reduced. But then $C_\pi(G_f)$ is infinite, so $G_f = \mathbb{Z}/2\mathbb{Z}^-$, by Crisp's theorem. Since $(\mathcal{G}, \Gamma)$ is reduced, f must be the only edge, contrary to the assumption that there is an infinite vertex group. Thus we may assume that $G_{o(f)}$ and $G_{t(f)}$ each properly contain G_f. Since $G_{o(f)}$ and $G_{t(f)}$ are 2–groups and hence nilpotent, $N_\pi(G_f)$ is infinite, by the normalizer condition. Since $C_\pi(G_f)$ has finite index in $N_\pi(G_e)$ we must have $G_f = \mathbb{Z}/2\mathbb{Z}^-$, by Crisp's theorem. Since Γ is connected it follows easily that every finite vertex group is nonorientable and every edge group is $\mathbb{Z}/2\mathbb{Z}^-$. □

The next two lemmas consider two parallel special cases, involving a prime p, which is odd or 2, respectively.

Lemma 3 *Let X be an indecomposable* PD_3*–complex with $\pi = \pi_1(X) \cong \kappa \rtimes W$, where κ is orientable and torsion-free, and W has order $2p$ for some odd prime p. Then X is orientable.*

Proof Suppose that X is not orientable. Then π and κ are infinite. Since π has a subgroup W of finite order > 2, we may assume that $\pi \cong \pi\mathcal{G}$, where $(\mathcal{G}, \Gamma)$ is an admissible graph of groups with $r \geq 1$ finite vertex groups and at least one edge. Let $s = \beta_1(\Gamma)$.

Each finite vertex group is mapped injectively by any projection from π onto W with kernel κ. If a vertex group G_v has prime order then every edge e with one vertex at v is a loop isomorphism, since $(\mathcal{G}, \Gamma)$ is reduced. But then Γ has just one vertex and $\pi \cong G_v \rtimes F$, which contradicts the hypothesis. Hence all finite vertex groups are isomorphic to W. If an edge e is a loop isomorphism then $G_e^+ \cong \mathbb{Z}/p\mathbb{Z}$ has infinite normalizer, contradicting Crisp's theorem. If there is an edge e with G_e of order p then both of the vertex groups $G_{o(e)}$ and $G_{t(e)}$ are finite, by Lemma 2. But then $[G_{o(e)} : G_e] = [G_{t(e)} : G_e] = 2$, so $N_\pi(G_e)$ is infinite, which again contradicts Crisp's theorem. Since the orientation character w factors through W it follows that every edge group is $\mathbb{Z}/2\mathbb{Z}^-$ and w is nontrivial on every vertex group.

Since each edge group is $\mathbb{Z}/2\mathbb{Z}^-$, w is nontrivial on each vertex group, so $\pi^+ = \pi\mathcal{G}^+$ is the fundamental group of a graph of groups $(\mathcal{G}^+, \Gamma)$ with the same underlying graph Γ, trivial edge groups and vertex groups G_v^+ for all $v \in V(\Gamma)$. Hence $\pi^+ \cong G * F(s) * P$, where G is a free product of orientable PD_3–groups and P is a free product of r copies of $\mathbb{Z}/p\mathbb{Z}$. We have $P \cong F(t) \rtimes \mathbb{Z}/p\mathbb{Z}$ for some $t \geq 0$. (In fact, $t = (p-1)(r-1)$, by a simple virtual Euler characteristic argument.)

Let $a \in \pi$ be such that $a^2 = 1$ and $w(a) = -1$, and let $\lambda \cong \kappa \rtimes \mathbb{Z}/2\mathbb{Z}^-$ be the subgroup generated by κ and a. Then λ is also the group of a PD_3–complex, since it has finite index in π. The involution of π^+ induced by conjugation by a maps each indecomposable factor which is not infinite cyclic to a conjugate of an isomorphic factor [7]. However, its behaviour on the free factor $F(s)$ may be more complicated.

Let $w\colon \mathbb{Z}[\pi] \to R = \mathbb{Z}[\langle a \rangle] = \mathbb{Z}[a]/(a^2 - 1)$ be the linear extension of the orientation character. Then $I_{\langle a \rangle} \cong \widetilde{\mathbb{Z}} = R/(a+1)$. We may factor out the action of π^+ on a $\mathbb{Z}[\pi]$–module by tensoring with R. The derived sequence of the functor $R \otimes_w -$ applied to the augmentation sequence

$$0 \to I_\pi \to \mathbb{Z}[\pi] \to \mathbb{Z}$$

gives an exact sequence

$$0 \to H_1(\pi; R) = \kappa/\kappa' \to R \otimes_w I_\pi \to R \to \mathbb{Z} \to 0.$$

The inclusion of $\langle a \rangle$ into π splits the epimorphism from $R \otimes_w I_\pi$ onto $I_{\langle a \rangle}$, so $R \otimes_w I_\pi \cong \kappa/\kappa' \oplus \widetilde{\mathbb{Z}}$.

Let γ be the normal subgroup of π generated by $G \cup F(s)$ and let H be the image of γ in κ/κ'. Then similar arguments show that

$$R \otimes_w I_\pi = H \oplus (R \otimes_w I_{\pi/\gamma}),$$
$$R \otimes_w I_\lambda = H \oplus (R \otimes_w I_{\lambda/\gamma}).$$

The groups P and its normal subgroup $F(t)$ have presentations

$$P = \langle b_i,\ 1 \le i \le r \mid b_i^p = 1\ \forall i \rangle$$

and

$$F(t) = \langle x_{i,j},\ 1 \le i \le r-1,\ 1 \le j \le p-1 \mid \rangle,$$

where $x_{i,j}$ has image $b_1^j b_{i+1}^{-j}$ in P for $1 \le i \le r-1$ and $1 \le j \le p-1$. (If $p = 2$ we shall write x_i instead of $x_{i,1}$ for $1 \le i \le r-1$.)

The quotient $\pi/\langle\langle G \rangle\rangle$ is the fundamental group of the (possibly *un*reduced) graph of groups $(\bar{\mathcal{G}}, \Gamma)$ with vertex groups W (or $\mathbb{Z}/2\mathbb{Z}^-$) and edge groups $\mathbb{Z}/2\mathbb{Z}^-$, obtained by replacing each infinite vertex group G_v of $(\mathcal{G}, \Gamma)$ by $G_v/G_v^+ = \mathbb{Z}/2\mathbb{Z}^-$. Thus if W is abelian (so has an unique element of order 2) then $\pi/\langle\langle G \rangle\rangle \cong (F(s) * P) \times \mathbb{Z}/2\mathbb{Z}^-$. Hence $\pi/\gamma \cong P \times \mathbb{Z}/2\mathbb{Z}^-$ and $\lambda/\gamma \cong F(t) \times \mathbb{Z}/2\mathbb{Z}^-$, so

$$R \otimes_w I_{\pi/\gamma} \cong (R/(p, a-1))^r \oplus \widetilde{\mathbb{Z}},$$
$$R \otimes_w I_{\lambda/\gamma} \cong (R/(a-1))^t \oplus \widetilde{\mathbb{Z}} = \mathbb{Z}^t \oplus \widetilde{\mathbb{Z}}.$$

The quotient ring $R/pR = \mathbb{F}_p[a]/(a^2-1)$ is semisimple, so p–torsion R–modules have unique factorizations as sums of simple modules. Since $I_\pi \otimes_w R$ and $I_\lambda \otimes_w R$ satisfy Turaev's criterion (and projective R–modules are $\mathbb{Z}$–torsion-free), the p–torsion submodule of $R \otimes_w I_\pi$ has the same numbers of summands of types $R/(p, a-1)$ and $R/(p, a+1)$, and similarly for $R \otimes_w I_\lambda$. Since $R \otimes_w I_{\lambda/\gamma}$ is p–torsion-free, the number of summands of types $R/(p, a-1)$ and $R/(p, a+1)$ in H must also be equal. On the other hand, $R \otimes_w I_{\pi/\gamma}$ has $r > 0$ summands of type $R/(p, a-1)$ and none of type $R/(p, a+1)$. These conditions are inconsistent, so π is not the group of a nonorientable PD_3–complex.

If W is not abelian then it has an unique conjugacy class of elements of order 2, and $\pi/\gamma \cong P \rtimes \mathbb{Z}/2\mathbb{Z}^-$ and $\lambda/\gamma \cong F(t) \rtimes \mathbb{Z}/2\mathbb{Z}^-$ have presentations

$$\langle a,\, b_i,\, 1 \le i \le r \mid a^2 = 1,\, b_i^p = 1,\, ab_ia = b_i^{-1}\ \forall i\rangle$$

and

$$\langle a,\, x_{i,j},\, 1 \le i \le r-1,\, 1 \le j \le p-1 \mid a^2 = 1,\, ax_{ij}a = x_{i,p-j}\ \forall i, j\rangle,$$

respectively. (In particular, $\lambda/\gamma \cong F(t/2) * \mathbb{Z}/2\mathbb{Z}^-$.) In this case,

$$R \otimes_w I_{\pi/\gamma} \cong (R/(p, a+1))^r \oplus \widetilde{\mathbb{Z}},$$
$$R \otimes_w I_{\lambda/\gamma} \cong R^{t/2} \oplus \widetilde{\mathbb{Z}}.$$

Consideration of the p–torsion submodules again shows that $R \otimes_w I_\pi$ and $R \otimes_w I_\lambda$ cannot both satisfy Turaev's criterion, and hence that π is not the group of a non-orientable PD_3–complex. Thus X must be orientable. $\square$

The case $p = 2$ involves slightly different calculations.

Lemma 4 *Let X be an indecomposable* PD_3*–complex with $\pi = \pi_1(X) \cong \kappa \rtimes W$, where κ is orientable and torsion-free, and W has order 4. Then X is orientable.*

Proof As in Lemma 3, we suppose that X is not orientable, so π and κ are infinite, and may assume that $\pi \cong \pi\mathcal{G}$, where $(\mathcal{G}, \Gamma)$ is an admissible graph of groups with $r \ge 1$ finite vertex groups and at least one edge. We continue with the notation P, γ, a and R from Lemma 3.

The inclusions of the edge groups split w, by Lemma 2. In this case, $W \cong (\mathbb{Z}/2\mathbb{Z})^2 = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}^-$ and has two orientation-reversing elements. Note that P is now a free product of r copies of $\mathbb{Z}/2\mathbb{Z}$.

The quotient π/γ is the group of a finite graph of groups with all vertex groups W and edge groups $\mathbb{Z}/2\mathbb{Z}^-$. Since P is a free product of cyclic groups, π/γ has a presentation

$$\langle a,\, b_i,\, 1 \le i \le r \mid a^2 = 1,\, b_i^2 = (aw_i)^2 = (aw_ib_i)^2 = 1\ \forall i\rangle,$$

where $w_i = 1$ and $w_i \in F(t)$ for $2 \le i \le r$. The free subgroup $F(t)$ has basis $\{x_i \mid 1 \le i \le r-1\}$, where x_i has image b_1b_{i+1} in P, and λ/γ has a presentation

$$\langle a,\, x_i,\, 1 \le i \le r-1 \mid a^2 = 1,\, ax_ia = x_ib_{i+1}w_{i+1}b_{i+1}w_{i+1}^{-1}\ \forall i\rangle.$$

In this case,

$$R \otimes_w I_{\pi/\gamma} \cong (R/(2, a-1))^r \oplus \widetilde{\mathbb{Z}},$$
$$R \otimes_w I_{\lambda/\gamma} \cong \mathbb{Z}^{r-1} \oplus \widetilde{\mathbb{Z}}.$$

Since $R/(2,a+1)=R/(2,a-1)$, torsion considerations do not appear to help. If $r>1$ we may instead compare the quotients by the $\mathbb{Z}$–torsion submodules, as in [9, Lemma 7.3], since finitely generated torsion-free R–modules are direct sums of copies of R, $\mathbb{Z}$ and $\widetilde{\mathbb{Z}}$, by [4, Theorem 74.3]. We again conclude that π is not the group of a nonorientable PD_3–complex.

The case when $p=2$ and $r=1$ requires a little more work. Let N be the R–module presented by the transposed conjugate of $\binom{2}{a-1}$. If $\{e,f\}$ is the standard basis for R^2 then $N=R^2/R(2e+(a+1)f)$. The $\mathbb{Z}$–torsion submodule of N is generated by the image of $(a-1)e$ and has order 2, but is not a direct summand. The quotient of N by its $\mathbb{Z}$–torsion submodule is generated by the images of e and $f-e$, and is a direct sum $\mathbb{Z}\oplus\widetilde{\mathbb{Z}}$. In particular, it has no free summand. It now follows easily that $H\oplus\widetilde{\mathbb{Z}}\oplus R/(2,a-1)$ is not stably isomorphic to $H\oplus\widetilde{\mathbb{Z}}\oplus N$. Therefore I_π and I_λ cannot both satisfy Turaev's criterion, so π is not the group of a nonorientable PD_3–complex. Thus X must be orientable. □

Our final lemma is needed to cope with three exceptional cases.

Lemma 5 *Let $G=H\rtimes\mathbb{Z}/2\mathbb{Z}$, where $H=T_1^*$, O_1^* or I^*. Suppose that every element of G divisible by 4 is in H. Then G has a subgroup W of order 6 such that $[W:W\cap H]=2$.*

Proof Let g be an element of order 2 whose image generates G/H.

Suppose first that $H=T_1^*$, with presentation

$$\langle x,\ y,\ z \mid x^2=(xy)^2=y^2,\ z^3=1,\ zxz^{-1}=y,\ zyz^{-1}=xy\rangle.$$

Then $\zeta T_1^*=\langle x^2\rangle$ has order 2. The outer automorphism group $\mathrm{Out}(T_1^*)$ is generated by the class of the involution ρ which sends x, y and z to y^{-1}, x^{-1} and z^2, respectively. (See [8, page 221].) Hence ρ preserves the subgroup S of order 3 generated by z.

If conjugation by g induces an inner automorphism of T_1^*, there is an $h\in T_1^*$ such that $gxg^{-1}=hxh^{-1}$ for all $x\in T_1^*$. Then $gh=hg$ and h^2 is central in T_1^*, so $(h^{-1}g)^2=h^2$ has order dividing 4. Therefore $h^{-1}g$ has order 2, by hypothesis.

Otherwise we may assume that there is an $h\in G^+$ such that $gxg^{-1}=h\rho(x)h^{-1}$ for all $x\in T_1^*$, so ρ is conjugation by $h^{-1}g$. Since ρ is an involution, $(h^{-1}g)^2$ is central in T_1^*. We again see that $h^{-1}g$ has order 2. In each case, $h^{-1}g$ normalizes S, so the subgroup W generated by S and $h^{-1}g$ has order 6, while $h^{-1}g\notin H$, so $[W:W\cap H]=2$.

The commutator subgroup of O_1^* is T_1^*. Since this is a characteristic subgroup, it is preserved by g. The group T_1^* is a nonnormal subgroup of I^*, of index 5. Since g acts as an involution on the set of conjugates of T_1^*, we may assume that it preserves T_1^*. In each case the lemma follows easily from its validity for $H = T_1^*$. □

We may now give our main result.

Theorem 6 *Let X be an indecomposable, nonorientable* PD_3*–complex such that $\pi = \pi_1(X)$ has infinitely many ends. Then:*

1. $\pi \cong \pi\mathcal{G}$, *where $(\mathcal{G}, \Gamma)$ is an admissible graph of groups with all vertex groups one-ended and all edge groups $\mathbb{Z}/2\mathbb{Z}^-$.*
2. $\pi \cong \pi^+ \rtimes \mathbb{Z}/2\mathbb{Z}^-$.
3. $\pi^+ \cong G * H$, *where G is a nontrivial free product of* PD_3*–groups and H is free. In particular, π^+ is torsion-free.*

Proof Let $\pi \cong \pi\mathcal{G}$, where $(\mathcal{G}, \Gamma)$ is an admissible graph of groups. At least one vertex group is infinite, for otherwise π has two ends, by [9, Theorems 7.1 and 7.4]. Hence $\pi^+ \cong G * H$, where G is a nontrivial free product of PD_3–groups and H is virtually free. Therefore π^+ is virtually torsion-free. Let κ be the intersection of the conjugates in π of a torsion-free subgroup of finite index in π^+, and let $\phi\colon \pi \to \pi/\kappa$ be the canonical projection. Then κ is orientable, torsion-free and of finite index, and w factors through π/κ.

If F is a finite subgroup then $\phi|_F$ is injective, and $\phi^{-1}(\phi(F))$ has finite index in π. Hence $\phi^{-1}(\phi(F))$ has a graph of groups structure in which all finite vertex groups are isomorphic to subgroups of F. In particular, if F is a nonorientable 2–group then at least one of these vertex groups is a nonorientable 2–group, so there is a $g \in F$ such that $g^2 = 1$ and $w(g) = -1$, by Lemma 2(3). Hence, if, moreover, F is cyclic, then it has order 2.

Assume that there is a nonorientable finite vertex group G_v. Then G_v has a nonorientable Sylow 2–subgroup $S(2)$, so there is a $g \in S(2)$ such that $g^2 = 1$ and $w(g) = -1$. The orientable subgroup G_v^+ has periodic cohomology, with period dividing 4, by [9, Theorems 4.3 and 4.6]. Moreover, every element of G_v divisible by 4 is in G_v^+, by the argument of the previous paragraph.

Let g be an element of order 2 whose image generates G_v/G_v^+. We may assume that $G_v^+ \cong B \times \mathbb{Z}/d\mathbb{Z}$, where B is either $\mathbb{Z}/a\mathbb{Z} \rtimes Q(2^i)$ (with a odd and $i \geq 3$), T_k^* or O_k^* (for some $k \geq 1$), I^* or $\mathbb{Z}/a\mathbb{Z} \rtimes_{-1} \mathbb{Z}/2^e\mathbb{Z}$ (with a odd and $e \geq 1$), as in the penultimate paragraph of Section 1 above. Suppose first that G_v^+ is not a 2–group.

Then it has a nontrivial subgroup S of order p for some odd prime p. If $d > 1$ we may assume that p divides d, and then S is characteristic in G_v^+. This is also the case if $G_v^+ \cong \mathbb{Z}/a\mathbb{Z} \rtimes Q(8)$ or $\mathbb{Z}/a\mathbb{Z} \rtimes_{-1} \mathbb{Z}/2^e\mathbb{Z}$ with a odd (so p divides a), or $G_v^+ \cong T_k^*$ or O_k^* with $k > 1$ (so $p = 3$). In these cases, S is normalized by g and the subgroup H generated by S and g has order $2p$. The remaining possibilities are that $G_v^+ \cong T_1^* \times \mathbb{Z}/d\mathbb{Z}$, $O_1^* \times \mathbb{Z}/d\mathbb{Z}$ or $I^* \times \mathbb{Z}/d\mathbb{Z}$. For these cases we appeal to Lemma 5 to see that G_v has a nonorientable subgroup W of order $2p$.

Since $\phi^{-1}\phi(W)$ has finite index in π, it is again the group of a nonorientable PD$_3$–complex. This complex has an indecomposable factor whose group has W as one of its finite vertex groups, so has fundamental group $\kappa \rtimes W$. But this factor is nonorientable, so contradicts Lemma 3.

Therefore we may assume that G_v^+ is a 2–group. If $S(2)^+ \neq 1$ (ie if G_v^+ is a nontrivial 2–group) it is cyclic or generalized quaternionic, so has an unique central element of order 2 (see [9, Lemma 2.1]). Hence G_v has a finite index subgroup $W \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}^-$. As before, passage to $\phi^{-1}\phi(W)$ leads to a contradiction, by Lemma 4.

Therefore all finite vertex groups are orientable. But the graph Γ is connected, and any edge connecting a finite vertex group to an infinite vertex group must be nonorientable, as in Lemma 2. Since there is at least one infinite vertex group there can be no finite vertex groups.

The second assertion follows from part (2) of Lemma 2, and $\pi^+ = \pi\mathcal{G}^+$ is the fundamental group of a graph of groups $(\mathcal{G}^+, \Gamma)$ with the same underlying graph Γ, trivial edge groups and vertex groups G_v^+ all PD$_3$–groups. Hence π^+ is torsion-free, but not free. □

As observed at the end of Section 2, when X is a 3–manifold and $(\mathcal{G}, \Gamma)$ is an admissible graph of groups such that $\pi = \pi\mathcal{G}$, all vertices of Γ have even valence. Can this observation be extended to the case of PD$_3$–complexes? Although there are indecomposable PD$_3$–complexes which are not homotopy equivalent to 3–manifolds [9; 12], it remains possible that every indecomposable, nonorientable PD$_3$–complex is homotopy equivalent to a 3–manifold.

Corollary 7.5 of [9] follows immediately from Crisp's theorem and Theorem 6. (The argument in [9] assumed that π is virtually free.) We restate it here:

Corollary 7 *Let X be a* PD$_3$*–complex and $g \in \pi = \pi_1(X)$ a nontrivial element of finite order. If $C_\pi(g)$ is infinite then g has order 2 and is orientation-reversing, and $C_\pi(g) = \langle g \rangle \times \mathbb{Z}$.* □

Question Are there any examples other than $\mathbb{R}P^2 \times S^1$ of indecomposable PD_3–complexes whose groups have a central element of order 2 with infinite centralizer?

Acknowledgements I would like to thank B Hanke for alerting me to the necessity of considering the present case. I would also like to thank the referee for detailed suggestions as to improving the exposition of this work.

References

[1] **A Adem**, **R J Milgram**, *Cohomology of finite groups*, 2nd edition, Grundl. Math. Wissen. 309, Springer, Berlin (2004) MR

[2] **K S Brown**, *Cohomology of groups*, Graduate Texts in Mathematics 87, Springer, New York (1982) MR

[3] **J Crisp**, *The decomposition of* 3*–dimensional Poincaré complexes*, Comment. Math. Helv. 75 (2000) 232–246 MR

[4] **C W Curtis**, **I Reiner**, *Representation theory of finite groups and associative algebras*, Pure and Applied Mathematics XI, Interscience, New York (1962) MR

[5] **W Dicks**, **M J Dunwoody**, *Groups acting on graphs*, Cambridge Studies in Advanced Mathematics 17, Cambridge University Press (1989) MR

[6] **D B A Epstein**, *Projective planes in* 3*–manifolds*, Proc. London Math. Soc. 11 (1961) 469–484 MR

[7] **N D Gilbert**, *Presentations of the automorphism group of a free product*, Proc. London Math. Soc. 54 (1987) 115–140 MR

[8] **J A Hillman**, *Four-manifolds, geometries and knots*, Geometry & Topology Monographs 5, Geometry & Topology Publications, Coventry (2002) MR Revised 2014

[9] **J A Hillman**, *Indecomposable* PD_3*–complexes*, Algebr. Geom. Topol. 12 (2012) 131–153 MR

[10] **D J S Robinson**, *A course in the theory of groups*, Graduate Texts in Mathematics 80, Springer, New York (1982) MR

[11] **V G Turaev**, *Three-dimensional Poincaré complexes: homotopy classification and splitting*, Mat. Sb. 180 (1989) 809–830 MR In Russian; translated in Math. USSR-Sb. 67 (1990) 261–282

[12] **C T C Wall**, *Poincaré complexes, I*, Ann. of Math. 86 (1967) 213–245 MR

School of Mathematics and Statistics F07, University of Sydney
Sydney NSW 2006, Australia

jonathan.hillman@sydney.edu.au

Received: 16 March 2014 Revised: 31 July 2016

Algebraic & Geometric Topology 17 (2017) 657–692

On the periodic v_2–self-map of A_1

PRASIT BHATTACHARYA
PHILIP EGGER
MARK MAHOWALD

The spectrum $Y := M_2(1) \wedge C\eta$ admits eight v_1-self-maps of periodicity 1. These eight self-maps admit four different cofibers, which we denote by $A_1[ij]$ for $i, j \in \{0, 1\}$. We show that each of these four spectra admits a v_2-self-map of periodicity 32.

55Q51

This paper is dedicated to the memory of Mark Mahowald (1931–2013)

1 Introduction

Convention Throughout this paper, we work in the stable homotopy category of spectra localized at the prime 2.

Let $K(n)$ be the n^{th} Morava K–theory. Let $\mathcal{C}_0$ be the category of 2–local finite spectra, $\mathcal{C}_n \subset \mathcal{C}_0$ the full subcategory of $K(n-1)$–acyclics and $\mathcal{C}_\infty$ the full subcategory of contractible spectra. Hopkins and Smith [8] showed that the $\mathcal{C}_n$ are thick subcategories of $\mathcal{C}_0$ (in fact, they are the only thick subcategories of $\mathcal{C}_0$), and they fit into a sequence

$$\mathcal{C}_0 \supset \mathcal{C}_1 \supset \cdots \supset \mathcal{C}_n \supset \cdots \supset \mathcal{C}_\infty.$$

We say a finite spectrum X is of type n if $X \in \mathcal{C}_n \setminus \mathcal{C}_{n+1}$.

A self-map $v\colon \Sigma^k X \to X$ of a finite spectrum X is called a *v_n–self-map* if

$$K(n)_*(v)\colon K(n)_*(X) \to K(n)_*(X)$$

is an isomorphism. For a finite spectrum X, a self-map $v\colon \Sigma^k X \to X$ can also be regarded as an element of $\pi_k(X \wedge DX)$, where DX is the Spanier–Whitehead dual of X.

For any ring spectrum E, let

$$\iota_{E*}\colon \pi_*(\_) \to E_*(\_)$$

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.657

denote the E–Hurewicz natural transformation. Let $k(n)$ denote the connective cover of $K(n)$. If $v\colon S^k \to X \wedge DX$ is a v_n–self-map, then $\iota_{k(n)*}(v) \in k(n)_*(X \wedge DX)$ has to be the image of $v_n^m \in k(n)_* \cong \mathbb{F}_2[v_n]$, for some positive integer m, under the map

$$k(n)_*\iota_{X\wedge DX}\colon k(n)_* \to k(n)_*(X \wedge DX),$$

where $\iota_{X\wedge DX}\colon S^0 \to X \wedge DX$ is the unit map. The value m is called the *periodicity* of the v_n–self-map v. We call v a *minimal v_n–self-map* for X if v is a v_n–self-map with minimal periodicity. An easy consequence of [8, Theorem 9] is that the periodicity of a minimal v_n–self-map is always a power of 2.

Hopkins and Smith showed, among other things, that every type-n spectrum admits a v_n–self-map, and the cofiber of a v_n–self-map is of type $n+1$. However, not much is known about the minimal periodicity of such v_n–self-maps.

The sphere spectrum S^0 is a type-0 spectrum with a v_0–self-map $2\colon S^0 \to S^0$. The cofiber of this v_0–self-map is the type-1 spectrum $M(1)$. The spectrum $M(1)$ is known to admit a unique minimal v_1–self-map of periodicity 4. The cofiber of this v_1–self-map is denoted by $M(1,4)$. In 2008, Behrens, Hill, Hopkins and the third author [1] showed that the minimal v_2–self-map on $M(1,4)$ has periodicity 32.

Instead of S^0, we can start with the type-0 spectrum $C\eta$, the cofiber of $\eta\colon S^1 \to S^0$. The spectrum $C\eta$ admits a nonzero v_0–self-map $2 \wedge 1_{C\eta}\colon C\eta \to C\eta$, with cofiber $M(1) \wedge C\eta := Y$. The type-1 spectrum Y admits eight minimal v_1–self-maps of periodicity 1. These eight maps are constructed by Davis and the third author [3] using stunted projective spaces. The cofiber of any of the v_1–self-maps is referred to as A_1. Though there are eight different v_1–self-maps, there are only four different homotopy types of the cofibers A_1; see [3, Proposition 2.1].

Let $A(1)$ be the subalgebra of the Steenrod algebra A generated by Sq^1 and Sq^2. It turns out that the cohomology of any homotopy type of A_1 is a free $A(1)$–module on one generator. However, different homotopy types of A_1 have different A–module structures, which are distinguished by the action of Sq^4. We depict the cohomologies of the four different spectra A_1 in Figure 1 where the square brackets represent an action of Sq^4, the curved lines represent an action of Sq^2, and the straight lines represent an action of Sq^1. The subalgebra $A(1)$ has four different A–module structures, each of which corresponds to a homotopy type of A_1. Any A–module structure on $A(1)$ has a nontrivial Sq^4 action on the generator in degree 1 forced by the Adem relations. However, there are choices for Sq^4 actions to be trivial or nontrivial on generators in degree 0 and degree 2, thus giving us four different A–module structures. We denote the different homotopy types of A_1 using the notation $A_1[ij]$ where i and j are the indicator functions for the action of Sq^4 on the generators in degree 0 and degree 2, respectively.

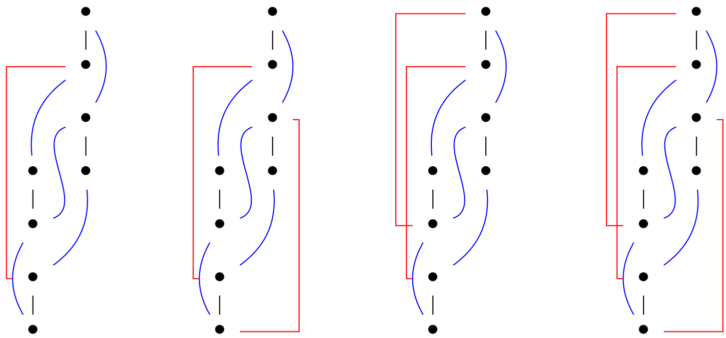


Figure 1: The A–module structures of $H^*(A_1[00])$, $H^*(A_1[10])$, $H^*(A_1[01])$ and $H^*(A_1[11])$

Remark 1.1 (determining A–module structure on Spanier–Whitehead duals) For every finite spectrum X, there is an isomorphism

$$H^*DX \cong DH^*X,$$

where we have Spanier–Whitehead duality on the left hand side and A–module duality on the right hand side. Thus, finding out the Spanier–Whitehead duality relations between the spectra $A_1[ij]$ boils down to finding the A–module duality relations between the A–modules depicted in Figure 1. The naïve guess is that dualizing these A–modules is equivalent to merely "flipping them upside down". However, this is not the case. For an A–module M and its dual DM, there is a pairing

$$\langle -,-\rangle\colon M \otimes DM \to \mathbb{F}_2$$

which is A–bilinear. Therefore, for elements $x, y \in M$ and $a \in A$, we have

$$\langle ax, y_*\rangle = \langle x, \chi(a)y_*\rangle,$$

where $\chi\colon A \to A$ is the antipode, and hence

$$(ax)_* = \sum_{\{g: ax = \chi(a)g\}} g_*.$$

Because $\chi(\mathrm{Sq}^1) = \mathrm{Sq}^1$ and $\chi(\mathrm{Sq}^2) = \mathrm{Sq}^2$, the naïve guess is correct when it comes to actions of Sq^1 and Sq^2. However, because we have $\chi(\mathrm{Sq}^4) = \mathrm{Sq}^4 + \mathrm{Sq}^3 Sq^1$, the naïve guess breaks down when considering the actions of Sq^4. Thus we find that $H^*(A_1[00])$ is dual to $H^*(A_1[11])$, while $H^*(A_1[10])$ and $H^*(A_1[01])$ are self-dual. It follows that the spectra $A_1[01]$ and $A_1[10]$ are Spanier–Whitehead self-dual, whereas $A_1[00]$ and $A_1[11]$ are Spanier–Whitehead dual to each other.

It is worth noting that A_1 is created in a way similar to $M(1,4)$, where $C\eta$ is analogous to S^0, and Y is analogous to $M(1)$. The minimal v_1–self-map of Y has periodicity 1, which is less than the periodicity of the minimal v_1–self-map on $M(1)$, which is 4. Hence, it is natural to ask if any of the four models of A_1 admit a v_2–self-map of periodicity less than that of $M(1,4)$.

In [3, Theorem 1.4(ii)], Davis and the third author claimed, incorrectly, that the periodicity of the minimal v_2–self-maps on $M(1,4)$ and the two self-dual models of A_1, namely $A_1[01]$ and $A_1[10]$, was 8. After successfully correcting the v_2–periodicity of $M(1,4)$ in [1], the v_2–periodicity of A_1 was called into question by the third author. He conjectured that the minimal v_2–self-map of A_1 should have periodicity 32, which is also the periodicity of the minimal v_2–self-map of $M(1,4)$.

The goal of this paper is to prove the following correction of [3, Theorem 1.4(ii)], as reported in Remark 1.4 of [1]:

Main Theorem *For all four models of A_1, the minimal v_2–self-map*

$$v\colon \Sigma^{192} A_1 \to A_1$$

has periodicity 32.

Notation 1.2 To lighten the notations, we use $\mathrm{Ext}_T^{s,t}(X)$ to denote $\mathrm{Ext}_T^{s,t}(H^*(X), \mathbb{F}_2)$, where T is a subalgebra of the Steenrod algebra A.

Notation 1.3 For any ring spectrum E, we denote the unit map by $\iota_E\colon S^0 \to E$. The unit map ι_E induces the Hurewicz natural transformation

$$\iota_{E*}\colon \pi_*(\_) \to E_*(\_)$$

as introduced earlier. When $E = A_1 \wedge DA_1$, we simply use $\iota\colon S^0 \to A_1 \wedge DA_1$ to denote the unit map. Let $i\colon S^0 \hookrightarrow A_1$ be the map that represents the inclusion of the bottom cell. Let $j\colon A_1 \wedge DA_1 \to A_1$ denote the map $1_{A_1} \wedge Di$. Given a map between two spectra $f\colon X \to Y$, the unit map ι_E induces a map in E–homology, which we denote by

$$E_*(f)\colon E_*X \to E_*Y,$$

and also a map of Adams spectral sequences, which we denote by

$$f_*^E\colon \mathrm{Ext}_A^{*,*}(E \wedge X) \to \mathrm{Ext}_A^{*,*}(E \wedge Y).$$

Outline

The proof of Main Theorem consists of two parts, namely

- the nonexistence part, where we eliminate the possibility of a v_2–self-map of A_1 of periodicity lower than 32,
- the existence part, where we show that there exists a v_2–self-map of A_1 of periodicity 32.

The proof makes use of several important differentials in the Adams spectral sequence that computes the homotopy groups of the spectrum *tmf*. As an A–module (see Hopkins and the third author [7]),

$$H^*(tmf) \cong A/\!/A(2),$$

where $A(2)$ is the subalgebra of A generated by Sq^1, Sq^2 and Sq^4. Therefore, by a change of rings formula, the E_2 page of that Adams spectral sequence simplifies to

$$E_2^{s,t} = \mathrm{Ext}^{s,t}_{A(2)}(S^0) \Rightarrow \pi_{t-s}(tmf). \tag{1.4}$$

The E_2 page is periodic with the periodicity generator $b_{3,0}^4$, which lives in bidegree $(s,t) = (8, 8+48)$. The periodicity generator $b_{3,0}^4$ and its square $b_{3,0}^8$ are not present in the E_∞ page of the above spectral sequence. There exist differentials

$$d_2(b_{3,0}^4) = e_0 r \quad \text{and} \quad d_3(b_{3,0}^8) = wgr \tag{1.5}$$

in the Adams spectral sequence computing tmf_*. But in that spectral sequence, $b_{3,0}^{16}$ is a nonzero permanent cycle which detects the periodicity generator $\Delta^8 \in \pi_{192}(tmf)$. All the details mentioned above are well documented by Henriques [6].

The unit map $\iota_{k(2)}\colon S^0 \to k(2)$ factors through *tmf* (see [1, Remark 1.3]): ie we have

$$\iota_{k(2)}\colon S^0 \xrightarrow{\iota_{tmf}} tmf \xrightarrow{r} k(2). \tag{1.6}$$

The map induced by r in homotopy

$$r_*\colon tmf_* \to k(2)_*$$

maps $\Delta^{8n} \mapsto v_2^{32n}$, which is why *tmf* can detect periodic v_2–self-maps. This can be observed through a map of Adams spectral sequences. Since

$$H^*(k(2)) \cong A/\!/E(Q_2)$$

(due to Lellmann [9]), by a change of rings formula, we have

$$E_2^{s,t} = \mathrm{Ext}^{s,t}_{E(Q_2)}(S^0) \Rightarrow \pi_{t-s}(k(2)).$$

The E_2 page is simply a polynomial algebra generated by v_2 in bidegree $(s,t) = (1, 1+6)$. The spectral sequence collapses due to sparseness, giving us the expected result $\pi_*(k(2)) = \mathbb{F}_2[v_2]$. The map $r\colon \mathit{tmf} \to k(2)$ induces a map of spectral sequences

$$\begin{tikzcd}
E_2 = \mathrm{Ext}^{s,t}_{A(2)}(S^0) \arrow[r, Rightarrow] \arrow[d, "r_*"'] & \pi_{t-s}(\mathit{tmf}) \arrow[d, "r_*"'] \\
E_2 = \mathrm{Ext}^{s,t}_{E(Q_2)}(S^0) \arrow[r, Rightarrow] & \pi_{t-s}(k(2))
\end{tikzcd}$$

which sends $b_{3,0}^{4n}$ to v_2^{8n} in the E_2 page, and therefore sends $b_{3,0}^{16n}$ to v_2^{32n} in the E_∞ page.

Next we study the commutative diagram of spectral sequences:

$$\begin{tikzcd}
& \mathit{tmf}_* \arrow[rr, "\mathit{tmf}_*\iota"] \arrow[dd] & & \mathit{tmf}_*(A_1 \wedge DA_1) \arrow[dd] \\
\mathrm{Ext}^{*,*}_{A(2)}(S^0) \arrow[ur, Rightarrow] \arrow[rr, "\iota_*^{\mathit{tmf}}"] \arrow[dd] & & \mathrm{Ext}^{*,*}_{A(2)}(A_1 \wedge DA_1) \arrow[ur, Rightarrow] \arrow[dd] & \\
& k(2)_* \arrow[rr, "k(2)_*\iota"] & & k(2)_*(A_1 \wedge DA_1) \\
\mathrm{Ext}^{*,*}_{E(Q_2)}(S^0) \arrow[ur, Rightarrow] \arrow[rr, "\iota_*^{k(2)}"'] & & \mathrm{Ext}^{*,*}_{E(Q_2)}(A_1 \wedge DA_1) \arrow[ur, Rightarrow] &
\end{tikzcd} \tag{1.7}$$

Since A_1 is a type-2 spectrum, Δ^8 has a nonzero image under the composite

$$\mathit{tmf}_* \xrightarrow{r_*} k(2)_* \xrightarrow{k(2)_*\iota} k(2)_*(A_1 \wedge DA_1).$$

Therefore, $\mathit{tmf}_*\iota(\Delta^{8n}) \in \mathit{tmf}_*(A_1 \wedge DA_1)$ is the lift of $k(2)_*\iota(v_2^{32n})$. Similarly, at the level of E_2 pages, we see that

$$\iota_*^{\mathit{tmf}}(b_{3,0}^{4n}) \in \mathrm{Ext}_{A(2)}(A_1 \wedge DA_1)$$

is the lift of $\iota_*^{k(2)}(v_2^{8n})$. In Section 3, we argue that the differentials in (1.5) induce a d_2 differential and a d_3 differential in the spectral sequence

$$\mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1) \Rightarrow \mathit{tmf}_*(A_1 \wedge DA_1),$$

supported by $\iota_*^{tmf}(b_{3,0}^4)$ and $\iota_*^{tmf}(b_{3,0}^8)$, respectively. This means that $k(2)_*\iota(v_2^8)$ and $k(2)_*\iota(v_2^{16})$ do not lift to $tmf_*(A_1 \wedge DA_1)$, thereby establishing the "nonexistence part" of Main Theorem.

The proof of the existence part of Main Theorem can roughly be divided into two parts:

- the lifting part, where we show that
$$\iota_*^{tmf}(b_{3,0}^{4n}) \in \mathrm{Ext}_{A(2)}^{8n,48n+8n}(A_1 \wedge DA_1)$$
lifts to an element $\widetilde{v_2^{8n}} \in \mathrm{Ext}_A^{8n,48n+8n}(A_1 \wedge DA_1)$ under the map
$$\iota_{tmf*}\colon \mathrm{Ext}_A^{*,*}(A_1 \wedge DA_1) \to \mathrm{Ext}_{A(2)}^{*,*}(A_1 \wedge DA_1),$$
- the survival part, where we show that $\widetilde{v_2^{32n}}$ is a nonzero permanent cycle in the Adams spectral sequence
$$E_2 = \mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \Rightarrow \pi_{t-s}(A_1 \wedge DA_1)$$
for all $n > 0$.

To achieve the lifting part, we use a Bousfield–Kan spectral sequence
$$E_1^{s,t,n} := \mathrm{Ext}_{A(2)}^{s-n,t}(H^*(X) \otimes \overline{A/\!/A(2)}^{\otimes n}, \mathbb{F}_2) \Rightarrow \mathrm{Ext}_A^{s,t}(H^*(X), \mathbb{F}_2),$$
which is also otherwise known as the algebraic *tmf* spectral sequence.

For the survival part of the argument, we show that the d_2 and d_3 differentials of (1.5) lift along the zigzag of spectral sequences:

(1.8)
$$\begin{tikzcd}
 & & & \pi_{t-s}(A_1 \wedge DA_1) \arrow[dd, "\iota_{tmf*}"] \\
 & & \mathrm{Ext}_A(A_1 \wedge DA_1) \arrow[ur, Rightarrow] \arrow[dd, "\iota_{tmf*}"] & \\
 & \pi_{t-s}(tmf) \arrow[rr, "\iota_*^{tmf}"'] & & tmf_{t-s}(A_1 \wedge DA_1) \\
\mathrm{Ext}_{A(2)}^{s,t}(S^0) \arrow[ur, Rightarrow] \arrow[rr, "\iota_*^{tmf}"'] & & \mathrm{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1) \arrow[ur, Rightarrow] &
\end{tikzcd}$$

Since $\widetilde{v_2^8}$ supports a d_2 differential and $\widetilde{v_2^{16}}$ supports a d_3 differential, $\widetilde{v_2^{32}}$ can only support a d_r differential for $r \geq 4$ by the Leibniz rule. There is another d_3 differential

(1.9)
$$d_3(v_2^{20}h_1) = g^6$$

in the Adams spectral sequence for $\pi_*(\mathit{tmf})$ which lifts along (1.8). The lifts of the differentials in (1.5) and (1.9), along with the multiplicative structure, allow us to deduce that there is no nonzero element in the E_4 page of

$$\mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \Rightarrow \pi_{t-s}(A_1 \wedge DA_1)$$

for $s \geq 36$ and $t - s = 191$. As a result, $\widetilde{v_2^{32}}$ is a nonzero permanent cycle, which detects a 32–periodic v_2–self-map of A_1.

Notation 1.10 Let T be any subalgebra of A, for example, $E(Q_2)$, $A(2)$ or A itself. Let X be any spectrum with a map $f\colon S^0 \to X$. Throughout the paper, we will denote any nonzero image of $a \in \mathrm{Ext}_T^{*,*}(S^0)$ under the map

$$f_*\colon \mathrm{Ext}_T^{*,*}(S^0) \to \mathrm{Ext}_T^{*,*}(X)$$

using the same notation.

Use of Bruner's Ext software

We will use Bruner's Ext software [2] for two purposes. Given any $A(2)$–module M which is finitely generated as an $\mathbb{F}_2$–vector space, the program can compute the groups $\mathrm{Ext}_{A(2)}^{s,t}(M, \mathbb{F}_2)$ to the extent of identifying generators in each bidegree within a finite range, determined by the user. Since we are interested in $\mathrm{Ext}_{A(2)}^{s,t}(X)$ for finite spectra X, such as $A_1 \wedge DA_1$, whose cohomology structures as $A(2)$–modules are known, this suits our task perfectly. The second purpose is the following: As any finite spectrum X is an S^0–module, $\mathrm{Ext}_{A(2)}^{*,*}(X)$ is a module over $\mathrm{Ext}_{A(2)}^{*,*}(S^0)$. Given an element $x \in \mathrm{Ext}_{A(2)}^{s,t}(X)$, the action of $\mathrm{Ext}_{A(2)}^{*,*}(S^0)$ can be computed using the `dolifts` functionality of the software.

One should also be aware that Main Theorem is by no means a consequence of the programming output. However, parts of the proof are reduced to pure algebraic computation, which can be performed using Bruner's program.

Organization of the paper

In Section 2, we use the May spectral sequence to compute $\mathrm{Ext}_{A(2)}^{*,*}(A_1)$. In particular, we establish a vanishing line of slope $\frac{1}{5}$, which will be useful for subsequent use of the algebraic *tmf* spectral sequence. In Section 3, we use the differentials in (1.5) to conclude that A_1 cannot admit a v_2–self-map of periodicity less than 32. We then use the algebraic *tmf* spectral sequence to lift the differentials in (1.5) along the zigzag (1.8), so that in the Adams spectral sequence

$$\mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \Rightarrow \pi_{t-s}(A_1 \wedge DA_1),$$

we have nonzero differentials $d_2(\widetilde{v_2^8})$ and $d_3(\widetilde{v_2^{16}})$. In Section 4, we use the algebraic *tmf* spectral sequence to lift the differential (1.9) along the zigzag (1.8). Finally, in Section 5, we complete the proof of Main Theorem.

In the Appendix, we provide a description of Bruner's Ext software to familiarize the readers with its usage. A summary of the output of the Bruner's program that is needed for some of the results in Section 5 is listed in the tables from the online supplement.

Acknowledgments Bhattacharya and Egger would like to thank Mark Behrens, Bob Bruner, Paul Goerss, Mike Hill and Mike Mandell for their invaluable assistance and encouragement throughout this project, as well as Irina Bobkova for some helpful discussions. We are greatly indebted to the anonymous referee for pointing out several subtle mistakes and gaps in our logical deductions, which ultimately led to the correct proof.

2 Computation of $\mathrm{Ext}^{s,t}_{A(2)}(A_1)$ and its vanishing line

J P May in his thesis [10] introduced a filtration of the Steenrod algebra called the May filtration, which induces a filtration of the cobar complex $C(\mathbb{F}_2, A_*, \mathbb{F}_2)$. This filtration gives a trigraded spectral sequence

$$E_1^{s,t,u} = \mathbb{F}_2[h_{i,j} : i \geq 1, j \geq 0] \Rightarrow \mathrm{Ext}^{s,t}_A(S^0), \quad |h_{i,j}| = (1, 2^j(2^i - 1), 2i - 1),$$

with differentials d_r of tridegree $(1, 0, 1-2r)$, which converges to the E_2 page of the Adams spectral sequence

$$E_2^{s,t} = \mathrm{Ext}^{s,t}_A(S^0) \Rightarrow \pi_{t-s}(S^0).$$

The element $h_{i,j}$ corresponds to the class $[\xi_i^{2^j}]$ in the cobar complex $C(\mathbb{F}_2, A_*, \mathbb{F}_2)$. We stick to the notation introduced by Tangora in his thesis [12]. For example, $h_{1,j}$ is abbreviated by h_j. Meanwhile, there are many elements $h_{i,j}$ that are not d_1–cycles in the May spectral sequence, however, even in these cases, the Leibniz rule means that $h_{i,j}^2$ will be d_1–cycles. To get around the awkwardness of talking about $h_{i,j}^2$ in later pages of the May spectral sequence, where $h_{i,j}$ may not even exist, Tangora uses $b_{i,j}$ to denote $h_{i,j}^2$ from the May E_2 page onwards.

One can use the same May filtration on the subalgebra $A(2)$ of A, to obtain a filtration on the cobar complex $C(\mathbb{F}_2, A(2)_*, \mathbb{F}_2)$. Thus we get a May spectral sequence with finitely many differentials

$$\mathbb{F}_2[h_0, h_1, h_2, h_{2,0}, h_{2,1}, h_{3,0}] \Rightarrow \mathrm{Ext}^{s,t}_{A(2)}(S^0),$$

all of which have been computed using techniques of [12]. The bigraded ring $\mathrm{Ext}^{s,t}_{A(2)}(S^0)$ is the Adams E_2 page for the homotopy groups of *tmf*.

We have obtained A_1 by a series of cofibrations

$$S^1 \xrightarrow{\eta} S^0 \to C\eta, \quad C\eta \xrightarrow{2} C\eta \to Y \quad \text{and} \quad \Sigma^2 Y \xrightarrow{v_1} Y \to A_1.$$

The maps 2, η and v_1 are detected by h_0, h_1 and $h_{2,0}$, respectively, in the May spectral sequence. Using the fact that cofiber sequences induce long exact sequences of E_1 pages of the May spectral sequence, we get that the E_1 page of the May spectral sequence converging to $\mathrm{Ext}^{s,t}_{A(2)}(A_1)$ is

$$\mathbb{F}_2[h_2, h_{2,1}, h_{3,0}] \Rightarrow \mathrm{Ext}^{s,t}_{A(2)}(A_1).$$

Alternatively, using a change of rings formula, we see that there is a quasi-isomorphism of cobar complexes

$$C(\mathbb{F}_2, A(2)_*, A(1)_*) \cong C(\mathbb{F}_2, (A(2)/\!\!/A(1))_*, \mathbb{F}_2).$$

Since, $C(\mathbb{F}_2, (A(2)/\!\!/A(1))_*, \mathbb{F}_2)$ is a quotient of $C(\mathbb{F}_2, A(2)_*, \mathbb{F}_2)$, the May filtration on $C(\mathbb{F}_2, A(2)_*, \mathbb{F}_2)$ induces a filtration on $C(\mathbb{F}_2, (A(2)/\!\!/A(1))_*, \mathbb{F}_2)$. As a result, we have a May spectral sequence

$$E_1^{s,t,u}(A_1) = \mathbb{F}_2[h_2, h_{2,0}, h_{3,0}] \Rightarrow \mathrm{Ext}^{s,t}_{A(2)}(A_1) \tag{2.1}$$

that is a module over the May spectral sequence for S^0,

$$E_1^{s,t,u}(S^0) = \mathbb{F}_2[h_0, h_1, h_2, h_{2,0}, h_{2,1}, h_{3,0}] \Rightarrow \mathrm{Ext}^{s,t}_{A(2)}(S^0). \tag{2.2}$$

The d_1 differentials in (2.2) come from the coproduct on $A(2)_*$. It is well known that $d_1(h_2) = 0$, $d_1(h_{2,1}) = h_1 h_2$ and $d_1(h_{3,0}) = h_0 h_{2,1} + h_2 h_{2,0}$. Under the quotient map

$$\mathbb{F}_2[h_0, h_1, h_2, h_{2,0}, h_{2,1}, h_{3,0}] \twoheadrightarrow \mathbb{F}_2[h_2, h_{2,1}, h_{3,0}],$$

all the images of the above differentials map to zero. Therefore, there are no d_1 differentials in (2.1).

One can use Nakamura's formula to compute higher May differentials. The operations Sq_i on the cobar complex of $C(\mathbb{F}_2, A_*, \mathbb{F}_2)$, defined by $\mathrm{Sq}_i(x) = x \cup_i x + \delta x \cup_{i+1} x$ (see [11]), satisfy

$$\mathrm{Sq}_0(h_{i,j}) = h_{i,j}^2, \quad \mathrm{Sq}_0(b_{i,j}) = b_{i,j}^2 \quad \text{and} \quad \mathrm{Sq}_1(h_{i,j}) = h_{i,j+1},$$

as well as Cartan's formulas (see [11, Propositions 4.4 and 4.5])

$$\mathrm{Sq}_0(xy) = \mathrm{Sq}_0(x)\mathrm{Sq}_0(y) \quad \text{and} \quad \mathrm{Sq}_1(xy) = \mathrm{Sq}_1(x)\mathrm{Sq}_0(y) + \mathrm{Sq}_0(x)\mathrm{Sq}_1(y),$$

whenever x and y are represented by elements in appropriate pages of the May spectral sequence. In particular, we have

$$\mathrm{Sq}_1(x^2) = 0$$

for every x. The differential δ in the cobar complex $C(\mathbb{F}_2, A_*, \mathbb{F}_2)$ satisfies the relation

$$\delta \mathrm{Sq}_i = \mathrm{Sq}_{i+1}\delta \tag{2.3}$$

for $i \geq 0$ (see [11, Lemma 4.1]), and is often called Nakamura's formula in the literature.

Since the May spectral sequence (2.2) is obtained by filtering the cobar complex, Nakamura's formula (2.3) helps to find differentials in (2.2). Furthermore, because the cobar complex $C(\mathbb{F}_2, (A(2)/\!/A(1))_*, \mathbb{F}_2)$ is a quotient of $C(\mathbb{F}_2, A(2)_*, \mathbb{F}_2)$, (2.3) can also help us to find differentials in (2.1).

Lemma 2.4 *In the May spectral sequence*

$$\mathbb{F}_2[h_2, h_{2,1}, h_{3,0}] \Rightarrow \mathrm{Ext}^{s,t}_{A(2)}(A_1),$$

we have the differentials

$$d_2(b_{2,1}) = h_2^3, \quad d_3(b_{3,0}) = h_2^2 h_{2,1} \quad \text{and} \quad d_4(b_{3,0}^2) = h_2 b_{2,1}^2,$$

and the spectral sequence collapses at E_5.

Proof In the May spectral sequence for S^0 (2.2), there is a differential

$$d_2(b_{2,1}) = h_2^3$$

which implies the corresponding d_2 differential in the May spectral sequence for A_1 (2.1). The element $b_{3,0}$ is represented by the element $[\xi_3|\xi_3]$ in the cobar complex $C(\mathbb{F}_2, A(2)_*, \mathbb{F}_2)$. Since $b_{3,0} = \mathrm{Sq}_0 h_{3,0}$, we apply Nakamura's formula (2.3) to obtain

$$\begin{aligned}
\mathrm{Sq}_1(d_1(h_{3,0})) &= \mathrm{Sq}_1(h_0 h_{2,1} + h_2 h_{2,0}) \\
&= h_0^2 h_{2,2} + h_1 h_{2,1}^2 + h_2^2 h_{2,1} + h_3 h_{2,0}^2 \\
&= h_2^2 h_{2,1}
\end{aligned}$$

in the May spectral sequence for A_1 (2.1). Therefore, it must be the case that, in the cobar complex $C(\mathbb{F}_2, (A(2)/\!/A(1))_*, \mathbb{F}_2)$,

$$\delta([\xi_3|\xi_3]) = [\xi_1^4|\xi_1^4|\xi_2^2] + \text{elements of higher May filtration.}$$

As a result, in (2.1), we have

$$d_3(b_{3,0}) = h_2^2 h_{2,1}.$$

Since $\mathrm{Sq}_0(b_{3,0}) = b_{3,0}^2$, we can apply Nakamura's formula (2.3) in a similar way to obtain

$$d_4(b_{3,0}^2) = h_2 b_{2,1}^2$$

in the May spectral sequence for S^0 (2.2) as well as A_1 (2.1).

For every r, we have that $E_r^{*,*,*}(A_1)$ is a differential graded module over $E_r^{*,*,*}(S^0)$. Since $b_{3,0}^4$ is a permanent cycle in (2.2), multiplication by $b_{3,0}^4$ commutes with differentials in (2.1). The elements of $E_5^{*,*,*}(A_1)$ that are not multiples of $b_{3,0}^4$ are permanent cycles by sparseness. Therefore, the elements of $E_5^{*,*,*}(A_1)$ that are multiples of $b_{3,0}^4$ are permanent cycles as well, and thus (2.1) collapses at the E_5 page. □

In Figure 2, the solid line of slope 1 represents multiplication by h_1, while the solid line of slope $\frac{1}{3}$ represents multiplication by h_2. The element $b_{3,0}^4$ is the periodicity generator of $\mathrm{Ext}_{A(2)}^{*,*}(A_1)$ and the solid lines in that part (right) are simply a repetition of the earlier pattern (left). This matches the output of Bruner's program [2] for $\mathrm{Ext}_{A(2)}^{s,t}(A_1)$, though different models of A_1 may have different hidden extensions some of which might not be detected in the May spectral sequence.

We have thus computed the E_∞ page of the May spectral sequence converging to $\mathrm{Ext}_{A(2)}^{s,t}(A_1)$. While Bruner's program [2] shows that different spectra have different hidden extensions, we are mainly interested in a vanishing line for $\mathrm{Ext}_{A(2)}^{s,t}(A_1)$, which will not be affected by these hidden extensions.

Lemma 2.5 *The group* $\mathrm{Ext}_{A(2)}^{s,t}(A_1)$ *is zero if*

$$s > \tfrac{1}{5}(t-s) + 1,$$

and for $t-s \geq 29$, *it is zero if*

$$s > \tfrac{1}{5}(t-s).$$

In other words, there is a vanishing line

$$y = \tfrac{1}{5}x + 1.$$

Proof Of the three generators of the E_1 page, h_2 has slope $\frac{1}{3}$, $h_{2,1}$ has slope $\frac{1}{5}$ and $h_{3,0}$ has slope $\frac{1}{6}$. However, while $\mathrm{Ext}_{A(2)}^{s,t}(A_1)$ contains infinitely large powers of $h_{2,1}$ and $h_{3,0}$, it only contains powers up to 2 of h_2. Hence, the vanishing line of $\mathrm{Ext}_{A(2)}^{s,t}(A_1)$ must have slope $\frac{1}{5}$, determined by $b_{2,1}^2$. Now, since $h_2 b_{2,1}^2 = 0$, the vanishing line for stems greater than 29 is $y = \frac{1}{5}x$ and a glance at Figure 2 gives us the y–intercept of the overall vanishing line. □

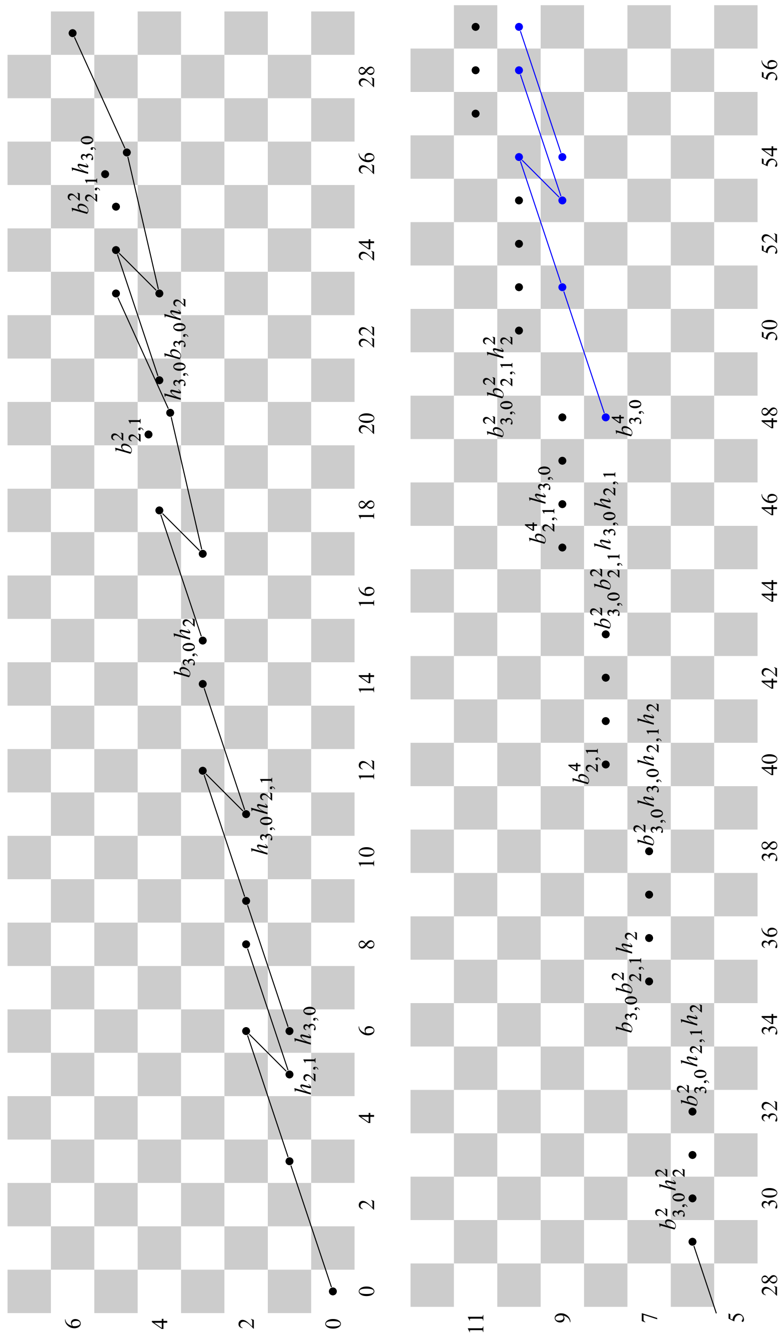


Figure 2: The E_∞ page of the May spectral sequence for $\mathrm{Ext}_{A(2)}(A_1)$

3 A d_2 and a d_3 differential

In this section, we first show that $b^4_{3,0}$ and $b^8_{3,0}$ in $\mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1)$ support a d_2 and a d_3 differential, respectively. Then we show that these differentials lift to $\mathrm{Ext}^{s,t}_{A}(A_1 \wedge DA_1)$ under the map of spectral sequences:

$$\begin{tikzcd}
\mathrm{Ext}^{s,t}_{A}(A_1 \wedge DA_1) \arrow[r, Rightarrow] \arrow[d, "\iota_{tmf*}"] & \pi_{t-s}(A_1 \wedge DA_1) \arrow[d, "\iota_{tmf*}"] \\
\mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1) \arrow[r, Rightarrow] & tmf_{t-s}(A_1 \wedge DA_1)
\end{tikzcd}$$

Some of the proofs in this section as well as in the subsequent sections use Bruner's program [2]. We provide the Appendix to help readers familiarize themselves with this software.

Lemma 3.1 *In the Adams spectral sequence*

$$E_2^{s,t} = \mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1) \Rightarrow tmf_{t-s}(A_1 \wedge DA_1),$$

we have $d_2(b^4_{3,0}) = e_0 r$ *and* $d_3(b^8_{3,0}) = wgr$.

Proof Recall the well known differentials (1.5) in the Adams spectral sequence

$$E_2^{s,t} = \mathrm{Ext}^{s,t}_{A(2)}(S^0) \Rightarrow tmf_{t-s}.$$

Using Bruner's program, we see that $e_0 r$ and wgr both have nonzero images in $\mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1)$. Hence, in the map of Adams spectral sequences

$$\begin{tikzcd}
E_2^{s,t} = \mathrm{Ext}^{s,t}_{A(2)}(S^0) \arrow[r, Rightarrow] \arrow[d] & tmf_{t-s} \arrow[d] \\
E_2^{s,t} = \mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1) \arrow[r, Rightarrow] & tmf_{t-s}(A_1 \wedge DA_1)
\end{tikzcd}$$

we have established that in the (abusive) Notation 1.3, we have

$$\begin{aligned}
\mathrm{Ext}^{s,t}_{A(2)}(S^0) &\xrightarrow{\iota_*^{tmf}} \mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1), \\
b^4_{3,0} &\mapsto b^4_{3,0}, \\
b^8_{3,0} &\mapsto b^8_{3,0}, \\
e_0 r &\mapsto e_0 r, \\
wgr &\mapsto wgr.
\end{aligned}$$

Therefore, the d_2 differential of (1.5) forces a d_2 differential

$$d_2(b_{3,0}^4) = e_0 r$$

in the Adams spectral sequence for $tmf_*(A_1 \wedge DA_1)$. By the Leibniz rule, $d_2(b_{3,0}^8) = 0$ and hence $b_{3,0}^8$ is nonzero in the E_3 page. The d_3 differential in (1.5) will force a nonzero d_3 differential

$$d_3(b_{3,0}^8) = wgr$$

in the Adams spectral sequence for $tmf_*(A_1 \wedge DA_1)$ as claimed, provided the image of wgr is nonzero in the E_3 page. Thus we have to show that there does not exist a differential of the form $d_2(x) = wgr$.

Using Bruner's program [2], we check that $wgr \in \mathrm{Ext}_{A(2)}^{19,95+19}(S^0)$ maps nontrivially to $\mathrm{Ext}_{A(2)}^{19,95+19}(A_1)$. Therefore if we have $d_2(x) = wgr$ in

$$\mathrm{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1) \Rightarrow tmf_{t-s}(A_1 \wedge DA_1),$$

then x must map to a nonzero element, say x', under the map

$$j_*\colon \mathrm{Ext}_{A(2)}^{17,96+17}(A_1 \wedge DA_1) \to \mathrm{Ext}_{A(2)}^{17,96+17}(A_1),$$

and we will have $d_2(x') = wgr$ in

$$\mathrm{Ext}_{A(2)}^{s,t}(A_1) \Rightarrow tmf_{t-s}(A_1).$$

There is exactly one generator of $\mathrm{Ext}_{A(2)}^{17,96+17}(A_1)$, and that generator is $b_{3,0}^4 \cdot y$ under the pairing

$$\mathrm{Ext}_{A(2)}^{8,48+8}(S^0) \otimes \mathrm{Ext}_{A(2)}^{9,48+9}(A_1) \to \mathrm{Ext}_{A(2)}^{17,96+17}(A_1).$$

It is clear that $d_2(y) = 0$ as $\mathrm{Ext}_{A(2)}^{11,47+11}(A_1) = 0$; see Figure 2. Thus using the Leibniz rule, we see that

$$d_2(b_{3,0}^4 y) = e_0 r \cdot y.$$

Using [2], we check that $e_0 r \cdot y = 0$. Therefore, wgr is nonzero in the E_3 page of the spectral sequence

$$\mathrm{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1) \Rightarrow tmf_{t-s}(A_1 \wedge DA_1),$$

and therefore

$$d_3(b_{3,0}^8) = wgr$$

in this spectral sequence. □

The fact that $v_2^{16} \in k(2)_*(A_1 \wedge DA_1)$ does not lift to $tmf_*(A_1 \wedge DA_1)$ implies that $v_2^{2^k} \in k(2)_*(A_1 \wedge DA_1)$ for $1 \le k \le 4$ does not lift to $tmf_*(A_1 \wedge DA_1)$. Indeed, suppose that for $k = 0, 1, 2$ or 3 the element $v_2^{2^k} \in k(2)_*(A_1 \wedge DA_1)$ lifts to an element

$$x \in tmf_*(A_1 \wedge DA_1),$$

then $x^{2^{4-k}}$ would be a lift of v_2^{16} as $A_1 \wedge DA_1$ is a ring spectrum. This would contradict Lemma 3.1. Since the unit map for $k(2)$ factors through the unit map of tmf (1.6), Lemma 3.1 implies the following:

Theorem 3.2 *The spectrum A_1 cannot admit a v_2–self-map of periodicity 16 or less.*

Next we describe an algebraic resolution which will allow us to lift the d_2 differential and the d_3 differential of Lemma 3.1 to the Adams spectral sequence

$$E_2^{s,t} = \operatorname{Ext}_A^{s,t}(A_1 \wedge DA_1) \Rightarrow \pi_{t-s}(A_1 \wedge DA_1).$$

We will briefly recall the resolution described in [1, Section 5], and how it is used to lift elements of Ext groups over $A(2)$ to Ext groups over A. Consider the A–module

$$A/\!/A(2) := A \otimes_{A(2)} \mathbb{F}_2,$$

and denote by $\overline{A/\!/A(2)}$ the kernel of the augmentation map

$$A/\!/A(2) \to \mathbb{F}_2.$$

When we consider the triangulated structure of the derived category of A–modules, we get maps

$$A/\!/A(2) \to \mathbb{F}_2 \to \overline{A/\!/A(2)}[1]$$

and a resulting diagram

$$\begin{tikzcd}
\mathbb{F}_2 \arrow[r] & \overline{A/\!/A(2)}[1] \arrow[r] & \overline{A/\!/A(2)}^{\otimes 2}[2] \arrow[r] & \cdots \\
A/\!/A(2) \arrow[u] & A/\!/A(2) \otimes \overline{A/\!/A(2)}[1] \arrow[u] & A/\!/A(2) \otimes \overline{A/\!/A(2)}^{\otimes 2}[2] \arrow[u] &
\end{tikzcd}$$

to which we shall apply the functor $\operatorname{Ext}_A^{s,t}(H^*(X) \otimes -, \mathbb{F}_2)$ to get a spectral sequence, which we shall refer to as the algebraic tmf spectral sequence to reflect the fact that $A/\!/A(2)$ is the cohomology of tmf. This spectral sequence will be trigraded, with E_1 page

$$\begin{aligned} E_1^{s,t,n} &= \operatorname{Ext}_A^{s,t}(H^*(X) \otimes A/\!/A(2) \otimes \overline{A/\!/A(2)}^{\otimes n}[n], \mathbb{F}_2) \\ &\cong \operatorname{Ext}_{A(2)}^{s-n,t}(H^*(X) \otimes \overline{A/\!/A(2)}^{\otimes n}, \mathbb{F}_2), \end{aligned}$$

which converges to

$$\operatorname{Ext}_A^{s,t}(H^*(X), \mathbb{F}_2).$$

For any element in the algebraic *tmf* spectral sequence in tridegree (s,t,n), we will refer to s as its Adams filtration, t as the internal degree and n as the algebraic *tmf* filtration. The differential d_r has tridegree $(1,0,r)$. It is shown in [4] that

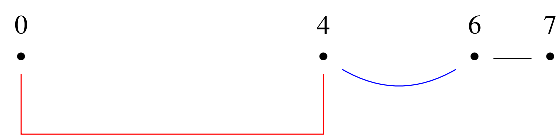
$$A/\!/A(2) \cong \bigoplus_{i \geq 0} H^*(\Sigma^{8i} bo_i),$$

where bo_i denotes the i^{th} bo–Brown–Gitler spectrum of [5]. As a result the E_1 page of the algebraic *tmf* spectral sequence simplifies to

$$E_1^{s,t,n} = \bigoplus_{i_1,\dots,i_n \geq 1} \operatorname{Ext}_{A(2)}^{s-n,t-8(i_1+\cdots+i_n)}(X \wedge bo_{i_1} \wedge \cdots \wedge bo_{i_n}) \Rightarrow \operatorname{Ext}_A^{s,t}(X).$$

We will attempt to exploit the relative sparseness of the E_1 page, especially its vanishing line properties, in the case when $X = A_1 \wedge DA_1$.

Remark 3.3 (the cellular structure of bo–Brown–Gitler spectra) The spectrum bo_0 is the sphere spectrum. The cohomology of the spectrum bo_1 as a module over the Steenrod algebra can be described through the following picture, with the generators labeled by cohomological degree:



where the straight line, curved line and square bracket describe the actions of Sq^1, Sq^2 and Sq^4, respectively. Note that the 4–skeleton of bo_1 is $C\nu$. Indeed, the bo_i fit together to form the following cofiber sequence

$$bo_{i-1} \to bo_i \to \Sigma^{4i} B(i),$$

where $B(i)$ is the i^{th} integral Brown–Gitler spectrum as described in [5]. Therefore for every $i \geq 1$, the 7–skeleton of bo_i is bo_1 and the 4–skeleton of bo_i is $C\nu$.

One can compute $\operatorname{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1 \wedge bo_i)$ from $\operatorname{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1)$ using the Atiyah–Hirzebruch spectral sequence or with Bruner's program [2].

Lemma 3.4 *The group*

$$\operatorname{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1 \wedge bo_{i_1} \wedge \cdots \wedge bo_{i_n})$$

is zero if $s > \frac{1}{5}((t-s)+6)$.

Proof We showed in Lemma 2.5 that $\mathrm{Ext}^{s,t}_{A(2)}(\mathrm{A}_1)$ has a vanishing line $s = \frac{1}{5}(t-s)$ for $t-s \geq 30$ and a vanishing line of $s = \frac{1}{5}(t-s)+1$ overall. The only generator of $\mathrm{Ext}^{s,t}_{A(2)}(A_1)$ with a slope greater than $\frac{1}{5}$ is h_2, so if we kill off h_2 by considering $\mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge C\nu)$ then the vanishing line is precisely $s = \frac{1}{5}(t-s)$.

As we mentioned in Remark 3.3, the 4–skeleton of any bo_i is $C\nu$ and the next cell is in dimension 6. So we can build bo_i by attaching finitely many cells of dimension at least 6 to $C\nu$. Hence by using the Atiyah–Hirzebruch spectral sequence and the fact that $\frac{1}{5}(x-6)+1 < \frac{1}{5}x$, one can see that the vanishing line of $A_1 \wedge bo_i$ is $s = \frac{1}{5}(t-s)$. One can build $A_1 \wedge bo_{i_1} \wedge \cdots \wedge bo_{i_n}$ from $A_1 \wedge bo_{i_1}$, iteratively using cofiber sequences, which depend on the cell structure of $bo_{i_2} \wedge \cdots \wedge bo_{i_n}$. Since we have already established that $\mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge bo_{i_1})$ has vanishing line $s = \frac{1}{5}(t-s)$ and that $bo_{i_2} \wedge \cdots \wedge bo_{i_n}$ is a connected spectrum, we conclude, using the Atiyah–Hirzebruch spectral sequence, that the vanishing line for $\mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge bo_{i_1} \wedge \cdots \wedge bo_{i_n})$ is $s = \frac{1}{5}(t-s)$.

However, DA_1 has cells in negative dimension, in fact the bottom cell is in dimension -6. Again by using the Atiyah–Hirzebruch spectral sequence, one concludes that the vanishing line for $\mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1 \wedge bo_{i_1} \wedge \cdots \wedge bo_{i_n})$ is

$$s = \tfrac{1}{5}(t-s+6)$$

for any $i_k \geq 1$, completing the proof. □

Corollary 3.5 *The group* $\mathrm{Ext}^{s,t}_{A}(A_1 \wedge DA_1)$ *is zero if*

$$s > \tfrac{1}{5}(t-s) + \tfrac{11}{5},$$

and for $t-s \geq 23$, *it is zero if*

$$s > \tfrac{1}{5}(t-s) + \tfrac{6}{5}.$$

The result is a straightforward consequence of Lemma 2.5, Lemma 3.4 and the algebraic *tmf* spectral sequence.

Lemma 3.6 *The element*

$$b_{3,0}^4 \in \mathrm{Ext}^{8,48+8}_{A(2)}(A_1 \wedge DA_1)$$

lifts to an element $\widetilde{v_2^8}$ *under the map*

$$\iota_{tmf*}\colon \mathrm{Ext}^{8,48+8}_{A}(A_1 \wedge DA_1) \to \mathrm{Ext}^{8,48+8}_{A(2)}(A_1 \wedge DA_1).$$

Proof Consider the algebraic *tmf* spectral sequence:

$$E_1^{s,t,n} = \bigoplus_{i_1 \geq 1, \ldots, i_n \geq 1} \mathrm{Ext}_{A(2)}^{s-n, t-8(i_1+\cdots+i_n)}(A_1 \wedge DA_1 \wedge bo_{i_1} \wedge \ldots bo_{i_n}) \Longrightarrow \mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1)$$

The element $b_{3,0}^4$ has tridegree $(s,t,n) = (8, 48+8, 0) = (8, 56, 0)$ in the above spectral sequence. The element $d_n(b_{3,0}^4)$ has tridegree $(9, 56, n)$ and hence belongs to

$$\mathrm{Ext}_{A(2)}^{9-n, 56-8(i_1+\cdots+i_n)}(A_1 \wedge DA_1 \wedge bo_{i_1} \wedge \cdots \wedge bo_{i_n})$$

for some $(i_1, \ldots, i_n)$ where $i_k \geq 1$. We will show that the above group is zero for all $n \geq 1$ and for all tuples $(i_1, \ldots, i_n)$ where $i_k \geq 1$.

By Lemma 3.4 the above group is zero if

$$\tfrac{1}{5}(56 - 8(i_1 + \cdots + i_n) - 9 + n + 6) < 9 - n, \tag{3.7}$$

which is trivially satisfied for $n > 4$.

For $n = 1$, (3.7) becomes

$$\tfrac{1}{5}(54 - 8i_1) < 8,$$

thus $i_1 > 1$, so it suffices to verify that

$$\mathrm{Ext}_{A(2)}^{8,48}(A_1 \wedge DA_1 \wedge bo_1) = 0.$$

For $n = 2$, (3.7) becomes

$$\tfrac{1}{5}(55 - 8(i_1 + i_2)) < 7,$$

thus $i_1 + i_2 > 2$, so it suffices to verify that

$$\mathrm{Ext}_{A(2)}^{7,40}(A_1 \wedge DA_1 \wedge bo_1 \wedge bo_1) = 0.$$

For $n = 3$, (3.7) becomes

$$\tfrac{1}{5}(56 - 8(i_1 + i_2 + i_3)) < 6,$$

thus $i_1 + i_2 + i_3 > 3$, so it suffices to verify that

$$\mathrm{Ext}_{A(2)}^{6,32}(A_1 \wedge DA_1 \wedge bo_1 \wedge bo_1 \wedge bo_1) = 0.$$

For $n = 4$, (3.7) becomes

$$\tfrac{1}{5}(57 - 8(i_1 + i_2 + i_3 + i_4)) < 5,$$

thus $i_1+i_2+i_3+i_4>4$, so it suffices to verify that

$$\mathrm{Ext}^{5,24}_{A(2)}(A_1\wedge DA_1\wedge bo_1\wedge bo_1\wedge bo_1\wedge bo_1)=0.$$

For all four models of A_1, Bruner's program [2] shows that all the groups we expected to be zero are in fact zero. □

Corollary 3.8 *For all $n\in\mathbb{N}$, the elements $b^{4n}_{3,0}\in\mathrm{Ext}^{8n,48n+8n}_{A(2)}(A_1\wedge DA_1)$ lift to an element $\widetilde{v_2^{8n}}\in\mathrm{Ext}^{8n,48n+8n}_{A}(A_1\wedge DA_1)$ under the map ι_{tmf*}.*

Proof Since $A_1\wedge DA_1$ is a ring spectrum, it follows that the map

$$\iota_{tmf*}\colon \mathrm{Ext}^{s,t}_A(A_1\wedge DA_1)\to\mathrm{Ext}^{s,t}_{A(2)}(A_1\wedge DA_1)$$

is a map of algebras. By Lemma 3.6, $b^4_{3,0}$ lifts and thus $b^{4n}_{3,0}$ lifts for every $n\in\mathbb{N}$. □

Remark 3.9 The lift of $\widetilde{v_2^{8n}}$ in Corollary 3.8 may not be unique. The indeterminacy in the choice of $\widetilde{v_2^{8n}}$ consists of elements of higher algebraic *tmf* filtration.

Lemma 3.10 *In the Adams spectral sequence*

$$E_2^{s,t}=\mathrm{Ext}^{s,t}_A(A_1\wedge DA_1)\Rightarrow\pi_{t-s}(A_1\wedge DA_1),$$

there is a d_2–differential

$$d_2(\widetilde{v_2^8})=e_0r+R$$

and a d_3–differential

$$d_3(\widetilde{v_2^{16}})=wgr+S$$

for some R and S in algebraic tmf filtration greater than zero.

Proof Recall that e_0r and wgr are elements in $\mathrm{Ext}^{*,*}_A(S^0)$ (see [12]), which maps nontrivially (see Lemma 3.1) under the composite

$$\mathrm{Ext}^{*,*}_A(S^0)\to\mathrm{Ext}^{*,*}_{A(2)}(S^0)\to\mathrm{Ext}^{*,*}_{A(2)}(A_1\wedge DA_1).$$

Therefore, by inspecting the commutative diagram

$$\begin{tikzcd}
\mathrm{Ext}^{*,*}_A(S^0) \arrow[r, "\iota_*"] \arrow[d, "\iota_{tmf*}"'] & \mathrm{Ext}^{*,*}_A(A_1\wedge DA_1) \arrow[d, "\iota_{tmf*}"] \\
\mathrm{Ext}^{*,*}_{A(2)}(S^0) \arrow[r, "\iota_*^{tmf}"] & \mathrm{Ext}^{*,*}_{A(2)}(A_1\wedge DA_1)
\end{tikzcd} \tag{3.11}$$

we see that e_0r and wgr are nonzero image in $\mathrm{Ext}^{*,*}_A(A_1\wedge DA_1)$. Since $\widetilde{v_2^8}$ and $\widetilde{v_2^{16}}$ are lifts of $b^4_{3,0}$ and $b^8_{3,0}$, respectively, the differentials of Lemma 3.1 force the differentials as claimed. □

4 Another d_3 differential

The goal of this section is to lift the d_3 differential (1.9) in the spectral sequence for tmf_* to a d_3 differential

$$d_3(\widetilde{v_2^{20}h_1}) = g^6$$

in the Adams spectral sequence

$$E_2^{s,t} = \mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \Rightarrow \pi_*(A_1 \wedge DA_1)$$

along the zigzag (1.8).

The element $g \in \mathrm{Ext}_A^{4,20+4}(S^0)$ is Tangora's name [12] for the element detected by $b_{2,1}^2$ in the May spectral sequence

$$\mathbb{F}_2[h_{i,j} : i > 0, j \geq 0] \Rightarrow \mathrm{Ext}_A^{s,t}(S^0).$$

In the literature, the same name is adopted for its image in $\mathrm{Ext}_{A(2)}^{4,20+4}(S^0)$.

Lemma 4.1 *In the Adams spectral sequence*

$$E_2^{s,t} = \mathrm{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1) \Rightarrow tmf_{t-s}(A_1 \wedge DA_1),$$

the element g^6 is hit by a d_3 differential

$$d_3(v_2^{20}h_1) = g^6.$$

Proof From the calculation in Lemma 2.4, it is clear that $g^6 = b_{2,1}^{12}$ has a nonzero image in $\mathrm{Ext}_{A(2)}^{24,120+24}(A_1)$. Since we have a factorization of maps

$$\mathrm{Ext}_{A(2)}^{24,120+24}(S^0) \to \mathrm{Ext}_{A(2)}^{24,120+24}(A_1 \wedge DA_1) \to \mathrm{Ext}_{A(2)}^{24,120+24}(A_1),$$

we have that g^6 must also be nonzero in the Adams E_2 page for $tmf_*(A_1 \wedge DA_1)$.

To show that it is also nonzero in the Adams E_3 page, we must exclude the possibility that $g^6 \in \mathrm{Ext}_{A(2)}^{24,120+24}(A_1 \wedge DA_1)$ might be hit by a d_2 differential

$$d_2(\hat{x}) = g^6$$

for some elements $\hat{x} \in \mathrm{Ext}_{A(2)}^{22,121+22}(A_1 \wedge DA_1)$. In such a case, $\hat{x}$ would have to map to a nonzero element $x \in \mathrm{Ext}_{A(2)}^{22,121+22}(A_1)$ and there would exist a differential

(4.2) $$d_2(x) = g^6$$

in the Adams spectral sequence

$$E_2^{s,t} = \mathrm{Ext}^{s,t}_{A(2)}(A_1) \Rightarrow \mathit{tmf}_{t-s}(A_1)$$

as $g^6 \neq 0 \in \mathrm{Ext}^{24,120+24}_{A(2)}(A_1)$. From the calculations of Lemma 2.4, there is exactly one possible nonzero $x \in \mathrm{Ext}^{22,121+22}_{A(2)}(A_1)$. Using Bruner's program [2] (see (A.2)) we see that this x is a multiple of $gb^4_{3,0}$ under the pairing

$$\mathrm{Ext}^{12,68+12}_{A(2)}(S^0) \otimes \mathrm{Ext}^{10,53+10}_{A(2)}(A_1) \to \mathrm{Ext}^{22,121+22}_{A(2)}(A_1), \qquad gb^4_{3,0} \otimes \overline{x} \mapsto x.$$

Clearly $d_2(\overline{x}) = 0$ as $\mathrm{Ext}^{12,52+12}_{A(2)}(A_1) = 0$, and hence by the Leibniz rule, we get

$$d_2(x) = ge_0r \cdot \overline{x}.$$

However, $ge_0r = 0$ in $\mathrm{Ext}^{14,67+14}_{A(2)}(S^0)$, therefore $d_2(x) = 0$. It follows that the d_2 differential in (4.2) cannot exist and g^6 is a nonzero element in the E_3 page of the spectral sequence

$$\mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1) \Rightarrow \mathit{tmf}_{t-s}(A_1 \wedge DA_1).$$

Thus the d_3 differential of (1.9) in Adams spectral sequence

$$\mathrm{Ext}^{s,t}_{A(2)}(S^0) \Rightarrow \mathit{tmf}_{t-s}$$

forces the d_3 differential

$$d_3(v_2^{20}h_1) = g^6$$

in the Adams spectral sequence for $\mathit{tmf}_*(A_1 \wedge DA_1)$ as claimed. □

Our next goal is to lift this d_3 differential to the Adams spectral sequence

$$\mathrm{Ext}^{s,t}_{A}(A_1 \wedge DA_1) \Rightarrow \pi_{t-s}(A_1 \wedge DA_1).$$

The main tool at our disposal is the algebraic *tmf* spectral sequence, described in Section 3.

Lemma 4.3 *The elements g^6 and $v_2^{20}h_1$ lift to $\mathrm{Ext}^{s,t}_{A}(A_1 \wedge DA_1)$ under the map*

$$\iota_{\mathit{tmf}*}\colon \mathrm{Ext}^{s,t}_{A}(A_1 \wedge DA_1) \to \mathrm{Ext}^{s,t}_{A(2)}(A_1 \wedge DA_1).$$

Proof In the proof of Lemma 4.1, we showed that g^6 is a nonzero element if $\mathrm{Ext}^{24,120+24}_{A(2)}(A_1 \wedge DA_1)$. Since g^6 is an element of $\mathrm{Ext}^{24,120+24}_{A}(S^0)$, from the commutative diagram

$$\begin{tikzcd}
\mathrm{Ext}_A^{*,*}(S^0) \arrow[r, "\iota_*"] \arrow[d, "\iota_{tmf*}"'] & \mathrm{Ext}_A^{*,*}(A_1 \wedge DA_1) \arrow[d, "\iota_{tmf*}"] \\
\mathrm{Ext}_{A(2)}^{*,*}(S^0) \arrow[r, "\iota_*^{tmf}"] & \mathrm{Ext}_{A(2)}^{*,*}(A_1 \wedge DA_1)
\end{tikzcd}$$

it easily follows that g^6 lifts to $\mathrm{Ext}_A^{24,120+24}(A_1 \wedge DA_1)$ under the map ι_{tmf*}.

It is known that $v_2^{20}h_1 = b_{3,0}^8 \cdot v_2^4 h_1$ under the pairing

$$\mathrm{Ext}_{A(2)}^{16,96+16}(S^0) \otimes \mathrm{Ext}_{A(2)}^{5,25+5}(S^0) \to \mathrm{Ext}_{A(2)}^{21,121+21}(S^0), \qquad b_{3,0}^8 \otimes v_2^4 h_1 \mapsto v_2^{20} h_1.$$

Therefore the same relation $v_2^{20}h_1 = b_{3,0}^8 \cdot v_2^4 h_1$ is true in $\mathrm{Ext}_{A(2)}^{21,121+21}(A_1 \wedge DA_1)$ as

$$\iota_*^{tmf}\colon \mathrm{Ext}_{A(2)}^{s,t}(S^0) \to \mathrm{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1)$$

is a map of algebras. From Corollary 3.8, we already know that $b_{3,0}^8$ lifts to

$$\widetilde{v_2^{16}} \in \mathrm{Ext}_A^{16,96+16}(A_1 \wedge DA_1).$$

Using the algebraic *tmf* spectral sequence

$$\begin{array}{c} E_1^{s,t,n} = \displaystyle\bigoplus_{i_1 \geq 1, \dots, i_n \geq 1} \mathrm{Ext}_{A(2)}^{s-n, t-8(i_1+\cdots+i_n)}(A_1 \wedge DA_1 \wedge bo_{i_1} \wedge \cdots \wedge bo_{i_n}) \\ \Downarrow \\ \mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \end{array}$$

and the vanishing lines established in Lemma 3.4, we see $v_2^4 h_1 \in \mathrm{Ext}_{A(2)}^{5,25+5}(A_1 \wedge DA_1)$ also has a lift

$$\widetilde{v_2^4 h_1} \in \mathrm{Ext}_A^{5,25+5}(A_1 \wedge DA_1).$$

Therefore,

$$\widetilde{v_2^{16}} \cdot \widetilde{v_2^4 h_1} \in \mathrm{Ext}_A^{21,121+21}(A_1 \wedge DA_1)$$

is a lift of $v_2^{20}h_1$, as

$$\iota_{tmf*}\colon \mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \to \mathrm{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1)$$

is a map of algebras. □

We will denote any lift of $v_2^{20}h_1$ by $\widetilde{v_2^{20}h_1} \in \mathrm{Ext}_A^{21,121+21}(A_1 \wedge DA_1)$. One should be aware that the choice of $\widetilde{v_2^{20}h_1}$ is not unique. The indeterminacy in the choice of $\widetilde{v_2^{20}h_1}$ consists of elements of higher algebraic *tmf* filtration. This does not cause problems later in the paper because of the following technical lemma.

Lemma 4.4 *Suppose that we have a nontrivial differential $d_r(x) = y$ in the Adams spectral sequence for a spectrum X,*

$$E_2^{s,t} = \mathrm{Ext}_A^{s,t}(X) \Rightarrow \pi_{t-s}(X).$$

If x has algebraic tmf filtration greater than zero, then so does y.

Proof If the algebraic *tmf* filtration of x is greater than zero then the map of spectral sequences

$$\begin{tikzcd} \mathrm{Ext}_A^{s,t}(X) \arrow[r, Rightarrow] \arrow[d, "\iota_{tmf*}"'] & \pi_{t-s}(X) \arrow[d, "\iota_{tmf*}"] \\ \mathrm{Ext}_{A(2)}^{s,t}(X) \arrow[r, Rightarrow] & tmf_{t-s}(X) \end{tikzcd}$$

sends x to 0. Therefore,

$$\begin{aligned} \iota_{tmf*}(y) &= \iota_{tmf*}(d_r(x)) \\ &= d_r(\iota_{tmf*}(x)) \\ &= 0, \end{aligned}$$

which means that the algebraic *tmf* filtration of y is greater than zero. □

Lemma 4.5 *In the Adams spectral sequence*

$$\mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \Rightarrow \pi_{t-s}(A_1 \wedge DA_1),$$

there exists a d_3 differential

$$d_3(\widetilde{v_2^{20}h_1}) = g^6.$$

Proof It is easy to check that Lemma 4.1, along with the map of Adams spectral sequences

$$\begin{tikzcd} E_2^{s,t} = \mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \arrow[r, Rightarrow] \arrow[d] & \pi_{t-s}(A_1 \wedge DA_1) \arrow[d] \\ E_2^{s,t} = \mathrm{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1) \arrow[r, Rightarrow] & tmf_{t-s}(A_1 \wedge DA_1) \end{tikzcd}$$

induced by ι_{tmf}, forces a d_3 differential (also see Remark 4.7)

$$d_3(\widetilde{v_2^{20}h_1}) = g^6 + R, \tag{4.6}$$

where R is an element of algebraic *tmf* filtration greater than zero. Studying the algebraic *tmf* spectral sequence for $A_1 \wedge DA_1$, using the vanishing lines of Lemma 3.4 and using the fact that (checked using Bruner's program)

$$\mathrm{Ext}_{A(2)}^{23,113+23}(A_1 \wedge DA_1 \wedge bo_1) = 0 \quad \text{and} \quad \mathrm{Ext}_{A(2)}^{22,106+22}(A_1 \wedge DA_1 \wedge bo_1 \wedge bo_1) = 0,$$

we conclude that R is in fact zero. □

Remark 4.7 Lemma 4.4 in particular eliminates the possibility of a differential of the form

$$d_r(S) = g^6,$$

where S is in the higher algebraic *tmf* filtration. This is needed for the conclusion of (4.6).

5 Proof of Main Theorem

Recall from Corollary 3.8 that there are candidates in the E_2 page of the Adams spectral sequence

$$E_2^{s,t} = \mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \Rightarrow \pi_{t-s}(A_1 \wedge DA_1), \tag{5.1}$$

denoted by $\widetilde{v_2^{8n}}$, that can detect an $8n$–periodic v_2–self-map. Since $\widetilde{v_2^8}$ supports a d_2 differential and $\widetilde{v_2^{16}}$ supports a d_3 differential (see Lemma 3.10), by the Leibniz formula $\widetilde{v_2^{32}}$ is a nonzero d_3–cycle. The only way $\widetilde{v_2^{32}}$ can fail to detect a 32–periodic v_2–self-map is by supporting a nonzero d_r differential for $r \geq 4$ in the Adams spectral sequence (5.1). So we look for candidates in the E_2 page of (5.1) that can potentially be the target of a nonzero d_r differential supported by $\widetilde{v_2^{32}}$ for $r \geq 4$. Such elements will live in $\mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1)$ with $t - s = 191$ and Adams filtration $s \geq 36$. We use the algebraic *tmf* spectral sequence to detect such candidates. The goal of this section is to argue that any such candidate is either zero or not present in the E_4 page of the spectral sequence (5.1).

From Section 3, we recall the algebraic *tmf* spectral sequence:

$$\begin{array}{c} E_1^{s,t,n} = \bigoplus_{i_1,\dots,i_n \geq 1} \mathrm{Ext}_{A(2)}^{s-n,t-8(i_1+\cdots+i_n)}(bo_{i_1} \wedge \cdots \wedge bo_{i_n} \wedge A_1 \wedge DA_1) \\ \Downarrow \\ \mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1) \end{array}$$

An easy consequence of the vanishing line established in Lemma 3.4 is the following.

Lemma 5.2 *The only potential contributors to* $\mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1)$ *for* $t-s=191$ *and* $s \geq 36$ *come from the following summands of the algebraic tmf* E_1 *page:*

$$\begin{aligned}\mathrm{Ext}_{A(2)}^{s,t}(A_1 \wedge DA_1) \oplus \bigoplus_{1\leq i\leq 3} &\mathrm{Ext}_{A(2)}^{s-1,t-8i}(A_1 \wedge DA_1 \wedge bo_i)\\ \oplus \bigoplus_{1\leq i\leq 2} &\mathrm{Ext}_{A(2)}^{s-2,t-8-8i}(A_1 \wedge DA_1 \wedge bo_1 \wedge bo_i)\\ \oplus\, &\mathrm{Ext}_{A(2)}^{s-3,t-24}(A_1 \wedge DA_1 \wedge bo_1 \wedge bo_1 \wedge bo_1).\end{aligned}$$

While the result holds for all models of A_1, the computations will be slightly different for different models, and so we will treat these models separately. Since $A_1[00]$ and $A_1[11]$ are Spanier–Whitehead dual to each other, we can treat the cases of $A_1[00]$ and $A_1[11]$ as one case. We will then have to treat the cases of the self-dual spectra $A_1[01]$ and $A_1[10]$ separately. The completeness of the tables in this section will be justified by the more detailed tables in the online supplement.

Notation 5.3 The elements of $E_1^{s,t,n}$, the E_1 page of the algebraic *tmf* spectral sequence for $A_1 \wedge DA_1$, which are nonzero permanent cycles, will detect nonzero elements of $\mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1)$. Therefore we place an element $x \in E_1^{s,t,n}$ in bidegree $(t-s-n, s+n)$. Thus the elements that may contribute to the same bidegree of $\mathrm{Ext}_A^{s,t}(A_1 \wedge DA_1)$ are placed together. With this arrangement any differential in the algebraic *tmf* spectral sequence will look like Adams d_1 differential. The generators of

$$E_1^{s,t,n} = \bigoplus_{i_1,\dots,i_n \geq 1} \mathrm{Ext}_{A(2)}^{s-n,t-8(i_1+\cdots+i_n)}(A_1 \wedge DA_1 \wedge bo_{i_1} \wedge \cdots \wedge bo_{i_n})$$

will be denoted by dots in the following manner (recall that $bo_0 = S^0$):

- elements with $n=0$ are denoted by a $\bullet$,
- elements with $n=1, i_1=1$ are denoted by a $\circ^1$,
- elements with $n=1, i_1=2$ are denoted by a $\circ^2$,
- elements with $n=2, i_1=1, i_2=1$ are denoted by a $\odot$,
- and N/A stands for "not applicable," ie coordinates of the table which are irrelevant to our arguments.

5.1 The case $A_1 = A_1[00]$ or $A_1 = A_1[11]$

We begin by laying out, in Table 1, the elements of the E_1 page of the algebraic *tmf* spectral sequence, in Notation 5.3. The table makes it clear that all elements

$s \setminus t-s$	189	190	191
40	0	0	0
39	0	$\langle \bullet\bullet \rangle := Y^0_{39}$	$\langle \bullet\bullet\bullet \rangle := X^0_{39}$
38	N/A	$\langle \bullet\bullet\bullet\bullet\bullet \rangle := Y^0_{38}$	$\langle \bullet\bullet\bullet \rangle := X^0_{38}$
37	N/A	$\langle \bullet\bullet\bullet\bullet\bullet \rangle$ $\langle \circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1 \rangle$	$\langle \bullet\bullet\bullet\bullet\bullet \rangle := X^0_{37}$ $\langle \circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1 \rangle := X^1_{37}$
36	N/A	N/A	$\langle \bullet\bullet\bullet \rangle := X^0_{36}$ $\langle \circ^1\circ^1 \rangle := X^1_{36}$ $\langle \odot\,\odot\,\odot\,\odot\,\odot\odot \rangle := X^{1,1}_{36}$

Table 1: E_1 page of the algebraic *tmf* spectral sequence for $\mathrm{Ext}^{s,t}_{A}(A_1 \wedge DA_1)$, where $A_1 = A_1[00]$ or $A_1 = A_1[11]$, stem 189–191.

$s \setminus t-s$	70	71
15	$\langle \bullet\bullet \rangle = g^{-6}Y^0_{39}$	$\langle \bullet\bullet\bullet \rangle = g^{-6}X^0_{39}$
14	$\langle \bullet\bullet\bullet\bullet\bullet \rangle = g^{-6}Y^0_{38}$	$\langle \bullet\bullet\bullet\bullet \rangle = g^{-6}X^0_{38}$
13	$\langle \bullet\bullet\bullet\bullet\bullet \rangle$ $\langle \circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1 \rangle$	$\langle \bullet\bullet\bullet\bullet\bullet\bullet\bullet \rangle = g^{-6}X^0_{37}$ $\langle \circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1 \rangle = g^{-6}X^1_{37}$
12	N/A	$\langle \circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1 \rangle = g^{-6}X^1_{36}$ $\langle \odot\,\odot\,\odot\,\odot\,\odot\odot \rangle = g^{-6}X^{1,1}_{36}$

Table 2: E_1 page of the algebraic *tmf* spectral sequence for $\mathrm{Ext}^{s,t}_{A}(A_1 \wedge DA_1)$, where $A_1 = A_1[00]$ or $A_1 = A_1[11]$, stem 70–71.

with $t-s = 191$, with the possible exception of those in X^0_{36}, are permanent cycles in the algebraic *tmf* spectral sequence. Our goal is to show that every linear combination of elements in $X^{i_1,\dots,i_n}_s$ is either absent or zero in the E_4 page of the Adams spectral sequence. Using Bruner's program (for details see Tables 1–4 from the online supplement), we observe that a lot of these elements are multiples of g^6 in the E_1 page of the algebraic *tmf* spectral sequence, which we record in Table 2.

Lemma 5.4 *Every element of*

$$X^0_{39} \oplus X^0_{38} \oplus X^0_{37} \oplus X^1_{37} \oplus X^1_{36} \oplus X^{1,1}_{36}$$

is present in the Adams E_2 page, but is either zero or absent in the Adams E_4 page.

Proof Tables 1–4 of the online supplement make clear that multiplication by g^6 surjects onto $X^0_{39} \oplus X^0_{38} \oplus X^0_{37} \oplus X^1_{37} \oplus X^1_{36} \oplus X^{1,1}_{36}$. Notice that for any

$$x = g^6 \cdot y \in X^0_{39} \oplus X^0_{38} \oplus X^0_{37} \oplus X^1_{37} \oplus X^1_{36} \oplus X^{1,1}_{36},$$

both x and y are nonzero permanent cycles in the algebraic *tmf* spectral sequence. Indeed, the target of any differential supported by y, must have algebraic *tmf* filtration greater than y and from Table 2 it is clear no such element is present in the appropriate bidegree. Hence y is present in the Adams E_2 page. The same argument holds for x.

Case 1 When $x = g^6 \cdot y \in X^0_{39} \oplus X^0_{38} \oplus X^1_{37} \oplus X^1_{36} \oplus X^{1,1}_{36}$, then both x and y are permanent cycles in the algebraic *tmf* spectral sequence as the differentials must increase algebraic *tmf* filtration. In fact these elements are permanent cycles in the Adams spectral sequence for either degree reasons or by Lemma 4.4. If y is a target of a differential in the algebraic *tmf* spectral sequence or an Adams d_2 differential, then y is zero in the E_3 page. Consequently, $x = g^6 \cdot y$ is zero in the E_3 page as well. If y is not a target of such differentials, then we have

$$d_3(\widetilde{v_2^{20}h_1} \cdot y) = \widetilde{v_2^{20}h_1} \cdot d_3(y) + d_3(\widetilde{v_2^{20}h_1}) \cdot y = g^6 \cdot y = x.$$

In either case, x is zero in the E_4 page.

Case 2 When $x = g^6 \cdot y \in X^0_{37}$ and y is a permanent cycle, then we can argue $x = g^6 \cdot y$ is zero in the E_4 page as we did in the previous cases. If

$$d_2(y) = y',$$

then y' must belong to $g^{-6}Y^0_{39}$. Since multiplication by g^6 is a bijection between $g^{-6}Y^0_{39}$ and Y^0_{39}, we get

$$d_2(x) = d_2(g^6 \cdot y) = g^6 \cdot d_2(y) + d_2(g^6) \cdot y = g^6 \cdot y' \neq 0.$$

Therefore, x is absent in the E_4 page. □

Thus we are left with the case when $x \in X^0_{36}$.

Lemma 5.5 *Every element of X^0_{36} is either zero or absent in the Adams E_4 page.*

Proof X^0_{36} is spanned by three generators $\{s_1, t_1, t_2\}$. Using Bruner's program, we explore the following relations in the E_1 page of the algebraic *tmf* spectral sequence:

$$\begin{aligned} s_1 &= b^4_{3,0} \cdot x_1, & Y^0_{38} &\ni e_0 r \cdot x_1 \neq 0, & & \\ t_1 &= b^4_{3,0} \cdot y_1 = b^8_{3,0} \cdot z_1, & e_0 r \cdot y_1 &= 0, & Y^0_{39} &\ni wgr \cdot z_1 \neq 0, \\ t_2 &= b^4_{3,0} \cdot y_2 = b^8_{3,0} \cdot z_2, & e_0 r \cdot y_2 &= 0, & Y^0_{39} &\ni wgr \cdot z_2 \neq 0, \end{aligned}$$

$s\downarrow\ t{-}s\rightarrow$	94	95	$s\downarrow\ t{-}s\rightarrow$	142	143
23	0	0	30	0	0
22	0	0	29	$\langle\bullet\bullet\bullet\bullet\bullet\rangle$	$\langle\bullet\bullet\bullet\bullet\bullet\rangle$
21	0	0	28	N/A	$\langle\bullet{=}x_1,\bullet{=}y_1,\bullet{=}y_2\rangle{:=}Z_{28}$ $\langle\circ^1\circ^1\rangle$
20	N/A	$\langle\bullet{=}z_1,\bullet{=}z_2\rangle{:=}Z_{20}$			

Table 3: E_1 page of the algebraic *tmf* spectral sequence for $\mathrm{Ext}_A^{s,t}(A_1\wedge DA_1)$, where $A_1=A_1[00]$ or $A_1=A_1[11]$.

and $wgr\cdot z_1$ and $wgr\cdot z_2$ are linearly independent. In Bruner's notation, $s_1=36_{64}$, $t_1=36_{65}$, $t_2=36_{66}$, $x_1=28_{32}$, $e_0r\cdot x_1=38_{25}$, $y_1=28_{33}$, $y_2=28_{34}$, $z_1=20_1$, $wgr\cdot z_1=39_1$, $z_2=20_2$ and $wgr\cdot z_2=39_2$; see Table 5 from the online supplement.

From Table 3, it is clear that any element in Z_{20} and Z_{28} are permanent cycles.

Case 1 If $x=\epsilon_1 s_1+\delta_1 t_1+\delta_2 t_2\neq 0$ in the Adams E_2 page with $\epsilon_1\neq 0$, then

$$d_2(x)=\epsilon_1 d_2(\widetilde{v_2^8}\cdot x_1)=\epsilon_1(e_0r\cdot x_1)\neq 0.$$

Thus x is not present in the E_4 page.

Case 2 If $x=\delta_1 t_1+\delta_2 t_2\neq 0$, then

$$d_2(x)=0.$$

If $x\neq 0$ in the Adams E_3 page, then

$$d_3(x)=\delta_1 d_3(\widetilde{v_2^{16}}\cdot z_1)+\delta_2 d_3(\widetilde{v_2^{16}}\cdot z_2)=wgr\cdot(\delta_1 z_1+\delta_2 z_2)\neq 0.$$

Thus x is not present in the E_4 page. □

This proves Main Theorem in the cases $A_1=A_1[00]$ or $A_1=A_1[11]$.

5.2 The case $A_1=A_1[01]$ or $A_1=A_1[10]$

A priori, $A_1[01]$ and $A_1[10]$ are two different spectra and we must therefore give two different proofs of Main Theorem. However, it turns out that Tables 4 and 5 are identical for $A_1[01]$ and $A_1[10]$, and therefore the exact same arguments will apply to both spectra. For $A_1[01]$, refer to Tables 6–9 of the online supplement, and for $A_1[10]$, refer to Tables 10–13 of the online supplement, to observe that most of the elements in Table 4 are multiples by g^6 of elements in Table 5.

$s\backslash t-s$	190	191
39	0	$\langle\bullet\rangle := X^0_{39}$
38	$\langle\bullet\bullet\bullet\bullet\rangle := Y^0_{38}$	$\langle\bullet\rangle := X^0_{38}$
37	$\langle\bullet\bullet\bullet\bullet\rangle$ $\langle\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1\rangle$	$\langle\bullet\bullet\bullet\bullet\bullet\rangle := X^0_{37}$ $\langle\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1\rangle := X^1_{37}$
36	N/A	$\langle\odot\odot\rangle := X^{1,1}_{36}$

Table 4: E_1 page of the algebraic *tmf* spectral sequence for $\mathrm{Ext}^{s,t}_A(A_1\wedge DA_1)$, where $A_1=A_1[01]$, stem 190–191.

$s\backslash t-s$	70	71
15	0	$\langle\bullet\rangle = g^{-6}X^0_{39}$
14	$\langle\bullet\bullet\bullet\bullet\rangle = g^{-6}Y^0_{38}$	$\langle\bullet\bullet\rangle = g^{-6}X^0_{38}$
13	$\langle\bullet\bullet\bullet\bullet\bullet\rangle$ $\langle\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1\rangle$	$\langle\bullet\bullet\bullet\bullet\bullet\bullet\bullet\rangle = g^{-6}X^0_{37}$ $\langle\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1\rangle = g^{-6}X^1_{37}$
12	N/A	$\langle\circ^1\,\circ^1\,\circ^1\,\circ^1\,\circ^1\circ^1\rangle$ $\langle\odot\odot\rangle = g^{-6}X^{1,1}_{36}$

Table 5: E_1 page of the algebraic *tmf* spectral sequence for $\mathrm{Ext}^{s,t}_A(A_1\wedge DA_1)$, where $A_1=A_1[01]$, stem 70–71.

Lemma 5.6 *All elements of*

$$X^0_{39}\oplus X^0_{38}\oplus X^0_{37}\oplus X^1_{37}\oplus X^{1,1}_{36} \tag{5.7}$$

are present in the Adams E_2 page, but are zero in the Adams E_4 page.

Proof Differentials in the algebraic *tmf* spectral sequence increase algebraic *tmf* filtration. Therefore, as Tables 4 and 5 make clear, all elements of (5.7) are permanent cycles in the algebraic *tmf* spectral sequence and are therefore present in the Adams E_2 page. Furthermore, all these elements are permanent cycles in the Adams spectral sequence, either for degree reasons or by Lemma 4.4.

Tables 6–13 of the online supplement make clear that multiplication by g^6 is surjective onto (5.7). Therefore, any element $x=g^6\cdot y$ in (5.7) which is not zero in the Adams E_3 page is a target of a d_3 differential

$$d_3(\widetilde{v_2^{20}h_1}\cdot y)=d_3(\widetilde{v_2^{20}h_1})\cdot y+\widetilde{v_2^{20}h_1}\cdot d_3(y)=g^6\cdot y=x,$$

hence zero in the E_4 page. □

Appendix: General remarks on the use of Bruner's program

Since many of our proofs relied on the output of Bruner's program, we append some facts about the program to justify our claims.

The program takes as input a graded module M over A (or $A(2)$) that is a finite dimensional $\mathbb{F}_2$–vector space and computes $\mathrm{Ext}_A^{s,t}(M,\mathbb{F}_2)$ (or $\mathrm{Ext}_{A(2)}^{s,t}(M,\mathbb{F}_2)$) for t in a user-defined range, and $0 \le s \le \mathtt{MAXFILT}$, where one has $\mathtt{MAXFILT} = 40$ by default. The structure of M as an A–module is encoded in a text file named `M`, placed in the directory `A/samples` in the way we will now describe.

The first line of the text file `M` consists of a positive integer n, the dimension of M as an $\mathbb{F}_2$–vector space, whose basis elements we will call $g_0, \dots, g_{n-1}$. The second line consists of an ordered list of integers $d_0, \dots, d_{n-1}$, which are the respective degrees of the g_i. Every subsequent line in the text file describes a nontrivial action of some Sq^k on some generator g_i. For instance, if we have

$$\mathrm{Sq}^k(g_i) = g_{j1} + \cdots + g_{jl},$$

we would encode this fact by writing the line

```
i k l j1 ...jl
```

followed by a new line. Every action not encoded by such a line is assumed to be trivial. To ensure that such a text file in fact represents an honest A–module, we must run the `newconsistency` script, which will alert us if:

- the text file contains a line

  ```
  i k l j1 ...jl
  ```

  and it turns out that one of the d_j is not equal to $d_i + k$, or
- the module taken as a whole fails to satisfy a particular Adem relation.

Example A.1 Consider the A–module given by Figure 3, where generators are depicted by dots and actions of Sq^1, Sq^2 and Sq^4 are depicted by straight lines, curved lines and square brackets, respectively.

Based on this picture, we get the text file in Figure 4, which we call `A1-00_def`. We go to the directory `A2` and run:

```
./newmodule A1-00 ../A/samples/A1-00_def
cd A1-00
```

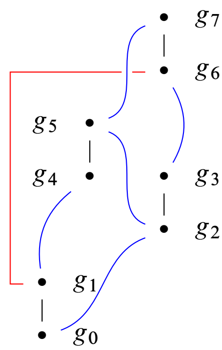


Figure 3: $H^* A_1[00]$ as an A–module

Now we are ready to compute. Running the script

```
./dims 0 250
```

will compute $\mathrm{Ext}^{s,t}_{A(2)}(A_1[00])$ for $0 \le s \le \texttt{MAXFILT} = 40$ and $0 \le t \le 250$. To see the Ext group, one runs

```
./report summary
./vsumm A1-00 > A1-00.tex
pdflatex A1-00.tex
```

to produce a pdf document `A1-00.pdf` as in the online supplement.

As this file makes apparent, the generators of the Ext group (as an $\mathbb{F}_2$ vector space) are stored in the computer with names such as s_g, where s is the Adams filtration of the generator, and g is some way of ordering all generators of filtration s. It should be emphasized that g is not the stem of the generator. In `A1-00.pdf` from the online supplement, for instance, the generator 1_2 is the second generator of filtration 1, but it is in stem 6. This file also tells us the action of the Hopf elements h_0 through h_3, so that in our example, h_2 multiplied by the generator 1_2 equals the generator 2_2.

By running

```
./display 0 A1-00_
```

to produce single-page pdf documents `A1-00_1.pdf`, `A1-00_2.pdf`, ..., it is also possible to see the Ext group in the visually more appealing form of a chart, as shown in `A1-00_1.pdf` from the online supplement.

The program is also capable of computing dual modules via the `dualizeDef` script, and tensor products via the `tensorDef` script. Both executables are conveniently located in

```
8

0 1 2 3 3 4 5 6

0 1 1 1
0 2 1 2
0 3 1 3
0 6 1 7
1 2 1 4
1 3 1 5
1 4 1 6
1 5 1 7
2 1 1 3
2 2 1 5
3 2 1 6
3 3 1 7
4 1 1 5
5 2 1 7
6 1 1 7
```

Figure 4: The text file `A/samples/A1-00_def`

the `A/samples` directory where we put our module definition text files. Thus, running

```
./dualizeDef A1-00_def DA1-00_def
./tensorDef A1-00_def DA1-00_def ADA1-00_def
```

produces the text file `ADA1-00_def`, with which we proceed in the same way as earlier with `A1-00_def`.

While `ADA1-00.pdf` only shows the action of the Hopf elements h_0 through h_3, the scripts `cocycle` and `dolifts` enable the user to input a specific generator and find the action of much of $\mathrm{Ext}^{s,t}_{A(2)}(S^0)$ on that specific generator. Let us do this with the generator $0_6 \in \mathrm{Ext}^{0,0}_{A(2)}(A_1[00] \wedge DA_1[00])$ by going to directory `A2` and running

```
./cocycle ADA1-00 0 6
```

which will create a subdirectory `A2/ADA1-00/0_6`. To find the action of all elements of $\mathrm{Ext}^{s,t}_{A(2)}(S^0)$ with $0 \leq s \leq 20$ on 0_6, we go back to directory `A2/ADA1-00` and run:

```
./dolifts 0 20 maps
```

Now `ADA1-00/0_6` will contain several text files, among them `brackets.sym` (which contains information about Massey products) and `Map.aug` (which contains information about the action of $\mathrm{Ext}^{s,t}_{A(2)}(S^0)$ on 0_6).

The generators of $\mathrm{Ext}^{s,t}_{A(2)}(S^0)$ are stored in the computer in the format s_g. Here we include a list of important elements of $\mathrm{Ext}^{s,t}_{A(2)}(S^0)$ and their s_g representations:

$$\begin{aligned} g &= 4_8 \in \mathrm{Ext}^{4,20+4}_{A(2)}(S^0) \\ b^4_{3,0} &= 8_{19} \in \mathrm{Ext}^{8,48+8}_{A(2)}(S^0) \\ e_0 r &= 10_{18} \in \mathrm{Ext}^{10,47+10}_{A(2)}(S^0) \\ b^8_{3,0} &= 16_{54} \in \mathrm{Ext}^{16,96+16}_{A(2)}(S^0) \\ wgr &= 19_{56} \in \mathrm{Ext}^{19,95+19}_{A(2)}(S^0) \\ v_2^{20} h_1 &= 21_{85} \in \mathrm{Ext}^{21,121+21}_{A(2)}(S^0) \\ g^6 &= 24_{90} \in \mathrm{Ext}^{24,120+24}_{A(2)}(S^0) \end{aligned}$$

We'd like to know what $s_g(0_6) \in \mathrm{Ext}_{A(2)}(A_1[00] \wedge DA_1[00])$ is in the notation of `ADA1-00.pdf`. Of course, $s_g(0_6)$ is in filtration s, so we only need to specify which of the generators in filtration s make up $s_g(0_6)$. If, for instance, we have

$$s_g(0_6) = s_{g1} + \cdots + s_{gn},$$

then `ADA1-00/0_6/Map.aug` will contain the lines:

```
s g1 g
s g2 g
  ⋮
s gn g
```

Now, in the Adams spectral sequence

$$\mathrm{Ext}^{s,t}_{A(2)}(S^0) \Rightarrow tmf_{t-s},$$

we have

$$d_2(b^4_{3,0}) = e_0 r = 10_{18} \in \mathrm{Ext}^{10,47+10}_{A(2)}(S^0) \quad \text{and} \quad d_3(b^8_{3,0}) = 19_{56} \in \mathrm{Ext}^{19,95+19}_{A(2)}(S^0).$$

It follows that if

$$10_{18}(0_6) = 10_x \in \mathrm{Ext}^{8,8+47}_{A(2)}(A_1 \wedge DA_1)$$

and

$$19_{56}(0_6) = 19_y \in \mathrm{Ext}^{19,19+95}_{A(2)}(A_1 \wedge DA_1),$$

then $b_{3,0}^4 \in \mathrm{Ext}_{A(2)}^{8,48+8}(A_1 \wedge DA_1)$ and $b_{3,0}^8 \in \mathrm{Ext}_{A(2)}^{16,96+16}(A_1 \wedge DA_1)$ support a d_2 differential and a d_3 differential, respectively. By doing the above steps for all four versions of A_1, and checking the respective `Map.aug` files, each contain lines

```
10 x 18
19 y 56
```

justifying the claim in Lemma 3.1.

Using the tools we have so far described, it is easy to verify the claim from the proof of Lemma 4.1, that for all four models of A_1 we have

$$gb_{3,0}^4 \cdot 10_3 = 22_7. \tag{A.2}$$

It is similarly easy to verify that if $A_1 = A_1[00]$ or $A_1 = A_1[11]$, we have

$$ge_0r \cdot 10_3 = 0,$$

while if $A_1 = A_1[01]$ or $A_1 = A_1[10]$, we have

$$ge_0r \cdot 10_3 = 24_0 = g^6.$$

Finally, in order to run the algebraic *tmf* spectral sequence, we will also need do computations involving the *bo*–Brown–Gitler spectra. We give the $A(2)$–module definitions for the cohomologies of bo_1 and bo_2 in `bo1_def` and `bo2_def` from the online supplement.

References

[1] **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

[2] **R R Bruner**, Ext *in the nineties*, from "Algebraic topology" (M C Tangora, editor), Contemp. Math. 146, Amer. Math. Soc., Providence, RI (1993) 71–90 MR

[3] **D M Davis**, **M Mahowald**, *v_1– and v_2–periodicity in stable homotopy theory*, Amer. J. Math. 103 (1981) 615–659 MR

[4] **D M Davis**, **M Mahowald**, *Ext over the subalgebra A_2 of the Steenrod algebra for stunted projective spaces*, from "Current trends in algebraic topology, Part 1" (R M Kane, S O Kochman, P S Selick, V P Snaith, editors), CMS Conf. Proc. 2, Amer. Math. Soc., Providence, RI (1982) 297–342 MR

[5] **P G Goerss**, **J D S Jones**, **M E Mahowald**, *Some generalized Brown–Gitler spectra*, Trans. Amer. Math. Soc. 294 (1986) 113–132 MR

[6] **A Henriques**, *The homotopy groups of tmf and its localizations*, from "Topological modular forms" (C L Douglas, J Francis, A G Henriques, M A Hill, editors), Math. Surveys Monogr. 201, Amer. Math. Soc., Providence, RI (2014) 189–205

[7] **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), Math. Surveys Monogr. 201, Amer. Math. Soc., Providence, RI (2014) 261–285 MR

[8] **M J Hopkins**, **J H Smith**, *Nilpotence and stable homotopy theory, II*, Ann. of Math. 148 (1998) 1–49 MR

[9] **W Lellmann**, *Connected Morava K–theories*, Math. Z. 179 (1982) 387–399 MR

[10] **J P May**, *The cohomology of restricted Lie algebras and of Hopf algebras: application to the Steenrod algebra*, PhD thesis, Princeton University (1964) MR Available at `http://search.proquest.com/docview/302273947`

[11] **O Nakamura**, *On the squaring operations in the May spectral sequence*, Mem. Fac. Sci. Kyushu Univ. Ser. A 26 (1972) 293–308 MR

[12] **M C Tangora**, *On the cohomology of the Steenrod algebra*, PhD thesis, Northwestern University (1966) MR Available at `http://search.proquest.com/docview/302217803`

Department of Mathematics, University of Notre Dame
106 Hayes-Healy Hall, Notre Dame, IN 46556, United States

Department of Mathematics, Pennsylvania State University
235 McAllister, University Park, PA 16802, United States

Department of Mathematics, Northwestern University
Evanston, IL 60208, United States

pbhattac@nd.edu, pee3@psu.edu, markmah@me.com

Received: 21 January 2015 Revised: 28 June 2016

Algebraic & Geometric Topology 17 (2017) 693–704

THH and base-change for Galois extensions of ring spectra

AKHIL MATHEW

We treat the question of base-change in THH for faithful Galois extensions of ring spectra in the sense of Rognes. Given a faithful Galois extension $A \to B$ of ring spectra, we consider whether the map $\mathrm{THH}(A) \otimes_A B \to \mathrm{THH}(B)$ is an equivalence. We reprove and extend positive results of Weibel and Geller, and McCarthy and Minasian, and offer new examples of Galois extensions for which base-change holds. We also provide a counterexample where base-change fails.

55P43; 13D03, 55P42

1 Introduction

Let R be an $\mathbb{E}_1$–ring spectrum. The *topological Hochschild homology* $\mathrm{THH}(R)$ of R is a spectrum constructed as the geometric realization of a certain cyclic object built from R, a homotopy-theoretic version of the Hochschild complex of an associative ring. Topological Hochschild homology has been studied in particular because of its connections with algebraic K–theory via the theory of trace maps. More generally, if R is an $\mathbb{E}_1$–algebra in A–modules for an $\mathbb{E}_\infty$–ring A, then one can define a relative version $\mathrm{THH}^A(R)$.

Weibel and Geller [15] showed that Hochschild homology for commutative rings satisfies an étale base-change result. Equivalently, if k is a commutative ring and if $A \to B$ is an étale morphism of commutative k–algebras with A flat over k, then there is a natural equivalence

$$B \otimes_A \mathrm{THH}^k(A) \simeq \mathrm{THH}^k(B).$$

Weibel and Geller's result also applies in the nonflat case, although it cannot be stated in this manner.

One can hope to generalize the Weibel–Geller result to the setting of ring spectra. This leads to the following general question:

Question Let $A \to B$ be a morphism of $\mathbb{E}_\infty$–ring spectra. When is the map

$$\mathrm{THH}(A) \otimes_A B \to \mathrm{THH}(B) \tag{1}$$

an equivalence?

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.693

Following Lurie, we will use the following definition of étaleness:

Definition 1.1 A morphism $A \to B$ of $\mathbb{E}_\infty$–ring spectra is *étale* if $\pi_0(A) \to \pi_0(B)$ is étale and the natural map $\pi_*(A) \otimes_{\pi_0(A)} \pi_0(B) \to \pi_*(B)$ is an isomorphism.

McCarthy and Minasian [11] consider this question for an étale morphism[1] of connective $\mathbb{E}_\infty$–rings and prove the analog of the Weibel–Geller theorem, ie that (1) is an equivalence (see [11, Lemma 5.7]). In fact, they prove the result more generally for any THH–*étale morphism* of connective $\mathbb{E}_\infty$–rings.

In the setting of structured ring spectra, however, there are additional morphisms of nonconnective ring spectra that have formal properties similar to those of étale morphisms, though they are not étale on homotopy groups. The *faithful Galois extensions* of Rognes [14] are key examples here.

This note is primarily concerned with the following analog of the Weibel–Geller and McCarthy–Minasian question:

Question Let $A \to B$ be a G–Galois extension of $\mathbb{E}_\infty$–ring spectra, with G finite. When is the comparison map (1) an equivalence?

We make two main observations here. Our first observation uses the fact that THH, like algebraic K–theory, is an invariant not only of ring spectra but of stable ∞–categories. We refer, for example, to Blumberg and Mandell [3] and Blumberg, Gepner and Tabuada [2] for a treatment of THH in this context. Using Galois descent, we observe that the map (1) is an equivalence if and only if the map $\mathrm{THH}(A) \to \mathrm{THH}(B)^{hG}$ is an equivalence. These maps are the comparison maps for the *Galois descent* problem in THH. Consequently, the results of Clausen, Mathew, Naumann and Noel [4] provide numerous examples in chromatic homotopy theory where (1) is an equivalence.

Our second observation is to reinterpret the base-change question for THH in terms of the formulation $\mathrm{THH}(R) \simeq S^1 \otimes R$ for $\mathbb{E}_\infty$–rings, due to McClure, Schwänzl and Vogt [12].

As a result, we obtain an example where (1) is not an equivalence.

Theorem 1.2 *There is a faithful G–Galois extension $A \to B$ of $\mathbb{E}_\infty$–ring spectra such that* (1) *is not an equivalence.*

[1] We note that McCarthy and Minasian use the word "étale" differently in their paper.

Our counterexample is very simple; it is the map $C^*(S^1;\mathbb{F}_p) \to C^*(S^1;\mathbb{F}_p)$ induced by the degree-p cover $S^1 \to S^1$.

We in fact pinpoint exactly what goes wrong from a categorical perspective, and why this phenomenon cannot happen in the étale setting, thus proving a variant of the Weibel–Geller–McCarthy–Minasian theorem in the nonconnective setting:

Theorem 1.3 *Let R be an $\mathbb{E}_\infty$–ring, and let $A \to B$ be an étale morphism of $\mathbb{E}_\infty$–R–algebras (possibly nonconnective). Then the natural map $\mathrm{THH}^R(A)\otimes_A B \to \mathrm{THH}^R(B)$ is an equivalence.*

The use of categorical interpretation of THH in proving such base-change theorems is not new; McCarthy and Minasian [11] use this interpretation in a different manner.

Acknowledgments I would like to thank John Rognes and the referee for several helpful comments. The author is supported by the NSF Graduate Research Fellowship under grant DGE-110640.

2 Categorical generalities

Let $\mathcal{C}$ be a cocomplete ∞–category, and let $x \in \mathcal{C}$. Given $x \in \mathcal{C}$, we can [5, Section 4.4.4] construct an object $S^1 \otimes x$.

Choose a basepoint $* \in S^1$. Then we have a diagram:

$$\begin{tikzcd} x \arrow[r] \arrow[d] & y \arrow[d] \\ S^1 \otimes x \arrow[r] & S^1 \otimes y \end{tikzcd} \tag{2}$$

As a result of this diagram, we have a natural map in $\mathcal{C}$,

$$(S^1 \otimes x) \sqcup_x y \to S^1 \otimes y. \tag{3}$$

In order for (3) to be an equivalence, for any object $z \in \mathcal{C}$, the square of spaces

$$\begin{tikzcd} \mathrm{Hom}(S^1, \mathrm{Hom}_{\mathcal{C}}(y,z)) \arrow[r] \arrow[d] & \mathrm{Hom}_{\mathcal{C}}(y,z) \arrow[d] \\ \mathrm{Hom}(S^1, \mathrm{Hom}_{\mathcal{C}}(x,z)) \arrow[r] & \mathrm{Hom}_{\mathcal{C}}(x,z) \end{tikzcd} \tag{4}$$

must be homotopy cartesian. This happens only in very special situations.

Proposition 2.1 *Let $f\colon X \to Y$ be a map of spaces. Then the diagram*

$$\begin{tikzcd} \mathrm{Hom}(S^1, X) \arrow[r] \arrow[d] & X \arrow[d] \\ \mathrm{Hom}(S^1, Y) \arrow[r] & Y \end{tikzcd} \tag{5}$$

is homotopy cartesian if and only if, for every point $p \in X$, the map from the connected component of X containing p to that of Y containing $f(p)$ is a homotopy equivalence.

Proof Without loss of generality, we may assume that X and Y are connected spaces. In this case, choosing compatible basepoints in X and Y, we get equivalences

$$\Omega X \simeq \mathrm{fib}(\mathrm{Hom}(S^1, X) \to X), \quad \Omega Y \simeq \mathrm{fib}(\mathrm{Hom}(S^1, Y) \to Y),$$

and the fact that (5) is homotopy cartesian now implies that $\Omega X \to \Omega Y$ is a homotopy equivalence. Since X and Y are connected, this implies that $X \to Y$ is a homotopy equivalence. □

Definition 2.2 We will say that a map of spaces $X \to Y$ is a *split covering space* if the equivalent conditions of Proposition 2.1 are met. In particular, $X \to Y$ is a covering space that is trivial on each connected component of Y.

Observe that the base-change of a split covering space is still a split covering space.

Corollary 2.3 *Suppose $x \to y$ is a morphism in $\mathcal{C}$ as above. Then the natural map $(S^1 \otimes x) \sqcup_x y \to S^1 \otimes y$ is an equivalence if and only if, for every object $z \in \mathcal{C}$, the induced map of spaces $\mathrm{Hom}_{\mathcal{C}}(y, z) \to \mathrm{Hom}_{\mathcal{C}}(x, z)$ is a split cover.*

Proof Our map is an equivalence if and only if (4) is homotopy cartesian for each $z \in \mathcal{C}$. By Proposition 2.1, we get the desired claim. □

We now give this class of morphisms a name.

Definition 2.4 A morphism $x \to y$ in an ∞–category $\mathcal{C}$ is said to be *strongly* 0–*cotruncated* if, for every $z \in \mathcal{C}$, the map $\mathrm{Hom}_{\mathcal{C}}(y, z) \to \mathrm{Hom}_{\mathcal{C}}(x, z)$ is a split covering space.

Corollary 2.3 states that $x \to y$ has the property that $(S^1 \otimes x) \sqcup_x y \to S^1 \otimes y$ is an equivalence if and only if the map is strongly 0–cotruncated.

For passage to a relative setting, we will find the following useful:

Proposition 2.5 *Let $\mathcal{C}$ be a cocomplete ∞–category, let $a \in \mathcal{C}$, and let $x \to y$ be a morphism in $\mathcal{C}_{a/}$. If $x \to y$ is strongly 0–cotruncated when regarded as a morphism in $\mathcal{C}$, then it is strongly 0–cotruncated when regarded as a morphism in $\mathcal{C}_{a/}$.*

Proof Suppose $a \to z$ is an object of $\mathcal{C}_{a/}$. Then we have

$$\mathrm{Hom}_{\mathcal{C}_{a/}}(y,z) = \mathrm{fib}(\mathrm{Hom}_{\mathcal{C}}(y,z) \to \mathrm{Hom}_{\mathcal{C}}(a,z)),$$
$$\mathrm{Hom}_{\mathcal{C}_{a/}}(x,z) = \mathrm{fib}(\mathrm{Hom}_{\mathcal{C}}(x,z) \to \mathrm{Hom}_{\mathcal{C}}(a,z)).$$

Since $\mathrm{Hom}_{\mathcal{C}}(y,z) \to \mathrm{Hom}_{\mathcal{C}}(x,z)$ is a split cover, it follows easily that the same holds after taking homotopy fibers over the basepoint in $\mathrm{Hom}_{\mathcal{C}}(a,z)$. In fact, we can assume without loss of generality that $\mathrm{Hom}_{\mathcal{C}}(x,z)$ is connected, in which case $\mathrm{Hom}_{\mathcal{C}}(y,z)$ is a disjoint union $\bigsqcup_S \mathrm{Hom}_{\mathcal{C}}(x,y)$. Taking fibers over the map to $\mathrm{Hom}_{\mathcal{C}}(a,z)$ preserves the disjoint union as desired, so the map on fibers is a split cover. $\square$

3 $\mathbb{E}_\infty$–ring spectra

We let CAlg denote the ∞–category of $\mathbb{E}_\infty$–ring spectra. The construction THH in this case can be interpreted (by [12]) as tensoring with S^1; that is, we have

$$\mathrm{THH}(A) \simeq S^1 \otimes A, \quad A \in \mathrm{CAlg}.$$

If one works in a relative setting, under an $\mathbb{E}_\infty$–ring R, then $\mathrm{THH}^R(A) \simeq S^1 \otimes A$, where the tensor product is computed in $\mathrm{CAlg}_{R/}$.

Given a morphism in $\mathrm{CAlg}_{R/}$, $A \to B$, we can use the setup of the previous section and obtain a morphism

$$\mathrm{THH}^R(A) \otimes_A B \to \mathrm{THH}^R(B),$$

which is a special case of (3). The base-change problem for THH asks when this is an equivalence.

By Corollary 2.3, this is equivalent to the condition that the morphism $A \to B$ in $\mathrm{CAlg}_{R/}$ should be strongly 0–cotruncated. We can now prove Theorem 1.3 from the introduction, which we restate for convenience.

Theorem *Let R be an $\mathbb{E}_\infty$–ring and let $A \to B$ be an étale morphism* (as in Definition 1.1) *in $\mathrm{CAlg}_{R/}$. Then the natural morphism $\mathrm{THH}^R(A) \otimes_A B \to \mathrm{THH}^R(B)$ is an equivalence.*

This is closely related to [15, Theorem 0.1] and includes it in the case of a flat extension $R \to A$ of discrete $\mathbb{E}_\infty$–rings. For connective $\mathbb{E}_\infty$–rings, this result is [11, Lemma 5.7] (who treat more generally the case of a THH–étale morphism).

Proof Given an étale morphism $A \to B$ in $\mathrm{CAlg}_{R/}$, we need to argue that it is strongly 0–cotruncated. By Proposition 2.5, we may reduce to the case where $R = S^0$. Given $C \in \mathrm{CAlg}$, we have a homotopy cartesian square

$$\begin{array}{ccc} \mathrm{Hom}_{\mathrm{CAlg}}(B, C) & \longrightarrow & \mathrm{Hom}_{\mathrm{Ring}}(\pi_0 B, \pi_0 C) \\ \downarrow & & \downarrow \\ \mathrm{Hom}_{\mathrm{CAlg}}(A, C) & \longrightarrow & \mathrm{Hom}_{\mathrm{Ring}}(\pi_0 A, \pi_0 C) \end{array}$$

by eg [8, Section 7.5]. Here Ring is the category of rings. Since the right vertical map is a map of discrete spaces and therefore a split covering, it follows that $\mathrm{Hom}_{\mathrm{CAlg}}(B, C) \to \mathrm{Hom}_{\mathrm{CAlg}}(A, C)$ is a split covering, as desired. □

We also note in passing that the étale descent theorem has a partial converse in the setting of *connective* $\mathbb{E}_\infty$–rings. We note that this rules out nonalgebraic Galois extensions.

Corollary 3.1 *Let $A \to B$ be a morphism of connective $\mathbb{E}_\infty$–rings which is almost of finite presentation* [8, Section 7.2.4]. *Suppose the map* $\mathrm{THH}(A) \otimes_A B \to \mathrm{THH}(B)$ *is an equivalence. Then $A \to B$ is étale.*

Proof Indeed, B defines a 0–cotruncated object (Definition 5.1) of $\mathrm{CAlg}_{A/}$ and it is well known that this, combined with the fact that B is almost of finite presentation, implies that B is étale. We reproduce the argument for the convenience of the reader.

In fact, since B is 0–cotruncated, one finds that for any B–module M, the space of maps[2] $\mathrm{Hom}_{\mathrm{CAlg}_{A//B}}(B, B \oplus M)$ is homotopy discrete, where the $\mathbb{E}_\infty$–ring $B \oplus M$ is given the square-zero multiplication. Replacing M by ΣM, it follows that

$$\mathrm{Hom}_{\mathrm{CAlg}_{A//B}}(B, B \oplus M) \simeq \Omega \mathrm{Hom}_{\mathrm{CAlg}_{A//B}}(B, B \oplus \Sigma M)$$

is actually contractible. Thus the cotangent complex $L_{B/A}$ vanishes, which implies that B is étale over A by [6, Lemma 8.9]. The connectivity is used in this last step. □

The above argument also appears in [14, Section 9.4], where it is shown that a map $A \to B$ which is 0–cotruncated as in Definition 5.1 below (which Rognes calls

[2]For an ∞–category $\mathcal{C}$ and a morphism $x \to y$, we let $\mathcal{C}_{x//y}$ denote $(\mathcal{C}_{x/})_{/y}$, where $y \in \mathcal{C}_{x/}$ via the given morphism.

formally symmetrically étale, and which has been called THH–*étale* in [11]) has to have vanishing cotangent complex (which is called TAQ–*étale)*; see [14, Lemma 9.4.4]. The key point is that in the connective setting, TAQ–étaleness plus a weak finiteness condition is enough to imply étaleness. This entirely breaks down when one works with nonconnective $\mathbb{E}_\infty$–ring spectra.

4 Connection with descent

In this section, we will show that the question of base-change in THH is equivalent to a descent-theoretic question. We will then use some of the descent results of [4] to obtain examples where base-change for THH holds. Let $A \to B$ be a faithful G–Galois extension of $\mathbb{E}_\infty$–rings for G a finite group.

To begin with, we will need to recall a fact about Galois descent.

Proposition 4.1 (see [13, Chapter 6], [1, Theorem 2.8] or [10, Theorem 9.4], for example) *If $A \to B$ is a faithful G–Galois extension, then we have an equivalence of symmetric monoidal ∞–categories*

$$\mathrm{Mod}(A) \simeq \mathrm{Mod}(B)^{hG},$$

where the left adjoint is extension of scalars along $A \to B$ and the right adjoint is given by taking homotopy fixed points.

We can restate the above equivalence in the following manner:

Corollary 4.2 *Let $\mathrm{Fun}(BG, \mathrm{Sp})$ be the symmetric monoidal ∞–category of G–spectra equipped with a G–action. Then we have a natural equivalence*

$$\mathrm{Mod}_{\mathrm{Fun}(BG,\mathrm{Sp})}(B) \simeq \mathrm{Mod}_{\mathrm{Sp}}(A)$$

given by taking homotopy fixed points.

Proof This follows from Proposition 4.1 using the fact that the construction of forming modules in a symmetric monoidal ∞–category is compatible with homotopy limits of symmetric monoidal ∞–categories. □

Let $\mathcal{C} = \mathrm{Fun}(BG, \mathrm{CAlg})$ be the ∞–category of $\mathbb{E}_\infty$–algebras equipped with a G–action, so that B defines an object of $\mathcal{C}$. We have therefore have natural equivalences of ∞–categories

$$\mathcal{C}_{B/} \simeq \mathrm{CAlg}(\mathrm{Fun}(BG, \mathrm{Sp}))_{B/} \simeq \mathrm{CAlg}(\mathrm{Mod}_{\mathrm{Fun}(BG,\mathrm{Sp})}(B)) \simeq \mathrm{CAlg}(\mathrm{Mod}(A)), \tag{6}$$

where the last equivalence is given by taking homotopy fixed points. We now obtain:

Proposition 4.3 *For a faithful G–Galois extension $A \to B$, the following two statements are equivalent:*

- $\mathrm{THH}(A) \otimes_A B \to \mathrm{THH}(B)$ *is an equivalence.*
- $\mathrm{THH}(A) \to \mathrm{THH}(B)$ *is a faithful G–Galois extension.*
- *The map* $\mathrm{THH}(A) \simeq (\mathrm{THH}(A) \otimes_A B)^{hG} \to \mathrm{THH}(B)^{hG}$ *is an equivalence.*

Proof In this case, the maps $B \to \mathrm{THH}(A) \otimes_A B \to \mathrm{THH}(B)$ that we obtain are G–equivariant, as they are natural in the $\mathbb{E}_\infty$-A–algebra B. Therefore, the map $\mathrm{THH}(A) \otimes_A B \to \mathrm{THH}(B)$ is naturally a morphism in $\mathrm{CAlg}(\mathrm{Fun}(BG, \mathrm{Sp}))_{B/}$. By (6), the map is an equivalence if and only if it induces an equivalence on homotopy fixed points.

Finally, if the map $\mathrm{THH}(A) \otimes_A B \to \mathrm{THH}(B)$ is an equivalence, then the morphism $\mathrm{THH}(A) \to \mathrm{THH}(B)$ is a base-change of the faithful G–Galois extension $A \to B$ and is thus a faithful G–Galois extension itself. Conversely, if $\mathrm{THH}(A) \to \mathrm{THH}(B)$ is a faithful G–Galois extension, then the descent map $\mathrm{THH}(A) \to \mathrm{THH}(B)^{hG}$ is an equivalence. □

In particular, the map $A \to B$ is strongly 0–cotruncated if and only if one has *Galois descent* for THH along the map $A \to B$. In [4], we give a general criterion for proving descent in telescopically localized THH.

Theorem 4.4 [4] *Suppose $A \to B$ is a G–Galois extension such that the map $K_0(B) \otimes \mathbb{Q} \to K_0(A) \otimes \mathbb{Q}$ induced by restriction of scalars is surjective. Fix an implicit prime p and a height n. Fix a weakly additive* (see [4, Definition 3.11]) *invariant E of κ–compact small idempotent-complete A–linear ∞–categories taking values in a presentable stable ∞–category. Then the natural morphisms*

$$L_n^f E(\mathrm{Perf}(A)) \to L_n^f E(\mathrm{Perf}(B))^{hG} \to (L_n^f E(\mathrm{Perf}(B)))^{hG}$$

are equivalences, where L_n^f denotes finitary L_n–localization. In particular, one can take $E = K$, THH *or TC.*

As a result, we can prove that the base-change map is an equivalence in a large class of "chromatic" examples of Galois extensions.

Theorem 4.5 *Suppose $A \to B$ is a faithful G–Galois extension of $\mathbb{E}_\infty$–rings. Assume that for every prime p, the localization $A_{(p)}$ is L_n^f–local for some $n = n(p)$. Suppose the map $K_0(B) \otimes \mathbb{Q} \to K_0(A) \otimes \mathbb{Q}$ is surjective (or equivalently has image containing the unit). Then the base-change map* $\mathrm{THH}(A) \otimes_A B \to \mathrm{THH}(B)$ *is an equivalence.*

Proof To check that the map $\mathrm{THH}(A)\otimes_A B \to \mathrm{THH}(B)$ is an equivalence, it suffices to localize at p, so we may assume A and B are p–local, and therefore L_n^f–local. Since L_n^f is a smashing localization, it follows that all THH terms in sight are automatically L_n^f–localized. In this case, the result follows by combining Proposition 4.3 and Theorem 4.4. □

Example 4.6 Most classes of examples of faithful Galois extensions in chromatic homotopy theory satisfy the conditions of Theorem 4.4. We refer to [4, Section 5] for a detailed treatment. For example:

(1) The C_2–Galois extension $KO \to KU$ or the C_{p-1}–Galois extension $L \to \widehat{KU}_p$.

(2) The G–Galois extension $E_n^{hG} \to E_n$ if G is a finite subgroup of the extended Morava stabilizer group (see [4, Appendix B] by Meier, Naumann and Noel).

(3) Any Galois extension of $\mathrm{TMF}[1/n]$, $\mathrm{Tmf}_0(n)$ or related spectra.

It follows that the comparison map in THH is an equivalence for these Galois extensions.

5 A counterexample

In this section, we will give an example over $\mathbb{F}_p$ where the comparison (or equivalently descent) map for THH is not an equivalence. We begin with a useful weakening of Definition 2.4.

Definition 5.1 A morphism $x \to y$ in an ∞–category $\mathcal{C}$ is said to be 0*–cotruncated* if, for every $z \in \mathcal{C}$, the map $\mathrm{Hom}_{\mathcal{C}}(y,z) \to \mathrm{Hom}_{\mathcal{C}}(x,z)$ is a covering space (ie has discrete homotopy fibers over any basepoint). An object $x \in \mathcal{C}$ is said to be 0*–cotruncated* if $\mathrm{Hom}_{\mathcal{C}}(x,z)$ is discrete for any $z \in \mathcal{C}$.

The condition that $x \to y$ should be cotruncated is equivalent to the statement that $y \in \mathcal{C}_{x/}$ should define a 0–cotruncated object. Note that an object $x \in \mathcal{C}$ is 0–cotruncated if and only if the natural map $x \to S^1 \otimes x$ is an equivalence.

In the setting of $\mathbb{E}_\infty$–ring spectra, étale morphisms are far from the only examples of 0–cotruncated morphisms. For example, any faithful G–Galois extension in the sense of Rognes [14] is 0–cotruncated. This is essentially [14, Lemma 9.2.6]. However, we show that faithful Galois extensions need not be *strongly* 0–cotruncated. Equivalently, base-change for THH can fail for them.

Proof of Theorem 1.2 Consider the degree-p map $S^1 \to S^1$, which is a $\mathbb{Z}/p$–torsor. Let k be a separably closed field of characteristic p. For a space X, we let $C^*(X;k) = F(X_+;k)$ denote the $\mathbb{E}_\infty$–rings of k–valued cochains on X. The induced map of $\mathbb{E}_\infty$–rings $\phi: C^*(S^1;k) \to C^*(S^1;k)$ is a faithful $\mathbb{Z}/p$–Galois extension of $\mathbb{E}_\infty$–ring spectra. This follows from [14, Proposition 5.6.3(a)] together with the criterion for the faithfulness via vanishing of the Tate construction [14, Proposition 6.3.3]. See also [10, Theorem 7.13].

We will show, nonetheless, that ϕ does not satisfy base-change for THH, or equivalently that it is not strongly 0–cotruncated. It suffices to show this in $\mathrm{CAlg}_{k/}$ in view of Proposition 2.5.

By p–adic homotopy theory [9] (see also [7], which does not assume $k = \overline{\mathbb{F}}_p$), the natural map

$$S^1 \to \mathrm{Hom}_{\mathrm{CAlg}_{k/}}(C^*(S^1;k), k)$$

exhibits $\mathrm{Hom}_{\mathrm{CAlg}_{k/}}(C^*(S^1;k), k)$ as the p–adic completion of S^1. In particular, $\mathrm{Hom}_{\mathrm{CAlg}_{k/}}(C^*(S^1;k), k) \simeq K(\mathbb{Z}_p, 1)$ and the map given by precomposition with ϕ

$$\mathrm{Hom}_{\mathrm{CAlg}_{k/}}(C^*(S^1;k), k) \xrightarrow{\phi^*} \mathrm{Hom}_{\mathrm{CAlg}_{k/}}(C^*(S^1;k), k),$$

is identified with the multiplication by p map $K(\mathbb{Z}_p, 1) \to K(\mathbb{Z}_p, 1)$. In particular, while this is a covering map, it is *not* a split covering map, so that ϕ is not strongly 0–cotruncated. □

The use of cochain algebras in providing such counterexamples goes back to an idea of Mandell [11, Example 3.5], who gives an example of a morphism of $\mathbb{E}_\infty$–ring spectra with trivial cotangent complex (ie is TAQ–étale) which is not THH–étale. Namely, Mandell shows that if $n > 1$ then the map $C^*(K(\mathbb{Z}/p, n); \mathbb{F}_p) \to \mathbb{F}_p$ has trivial cotangent complex.

We close by observing that it is the fundamental group that it is at the root of these problems.

Proposition 5.2 *Let X be a simply connected, pointed space and let $A \to B$ be a faithful G–Galois extension of $\mathbb{E}_\infty$–rings. In this case, the map of $\mathbb{E}_\infty$–rings*

$$(X \otimes A) \otimes_A B \to X \otimes B$$

is an equivalence.

In particular, one does have base-change for higher topological Hochschild homology (ie where $X = S^n$ with $n > 1$).

Proof Following the earlier reasoning, it suffices to show that whenever $C \in \mathrm{CAlg}$, the square

$$\begin{tikzcd} \mathrm{Hom}(X, \mathrm{Hom}_{\mathrm{CAlg}}(B, C)) \arrow[r] \arrow[d] & \mathrm{Hom}_{\mathrm{CAlg}}(B, C) \arrow[d] \\ \mathrm{Hom}(X, \mathrm{Hom}_{\mathrm{CAlg}}(A, C)) \arrow[r] & \mathrm{Hom}_{\mathrm{CAlg}}(A, C) \end{tikzcd}$$

is homotopy cartesian. However, this follows since $\mathrm{Hom}_{\mathrm{CAlg}}(B, C) \to \mathrm{Hom}_{\mathrm{CAlg}}(A, C)$ is a covering space, and X is simply connected. □

References

[1] **R Banerjee**, *Galois descent for real spectra*, J. Homotopy Relat. Struct. (online publication February 2016)

[2] **A J Blumberg**, **D Gepner**, **G Tabuada**, *A universal characterization of higher algebraic K–theory*, Geom. Topol. 17 (2013) 733–838 MR

[3] **A J Blumberg**, **M A Mandell**, *Localization theorems in topological Hochschild homology and topological cyclic homology*, Geom. Topol. 16 (2012) 1053–1120 MR

[4] **D Clausen**, **A Mathew**, **N Naumann**, **J Noel**, *Descent in algebraic K–theory and a conjecture of Ausoni–Rognes*, preprint (2016) arXiv

[5] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton University Press (2009) MR

[6] **J Lurie**, *Derived algebraic geometry, VII: Spectral schemes*, preprint (2011) `http://www.math.harvard.edu/~lurie/papers/DAG-VII.pdf`

[7] **J Lurie**, *Derived algebraic geometry, XIII: Rational and p–adic homotopy theory*, preprint (2011) `http://math.harvard.edu/~lurie/papers/DAG-XIII.pdf`

[8] **J Lurie**, *Higher algebra*, unpublished (2014) `http://www.math.harvard.edu/~lurie/papers/HA.pdf`

[9] **M A Mandell**, *E_∞ algebras and p–adic homotopy theory*, Topology 40 (2001) 43–94 MR

[10] **A Mathew**, *The Galois group of a stable homotopy theory*, Adv. Math. 291 (2016) 403–541 MR

[11] **R McCarthy**, **V Minasian**, *HKR theorem for smooth S–algebras*, J. Pure Appl. Algebra 185 (2003) 239–258 MR

[12] **J McClure**, **R Schwänzl**, **R Vogt**, $\mathrm{THH}(R) \cong R \otimes S\mathrm{Sp}1$ *for E_∞ ring spectra*, J. Pure Appl. Algebra 121 (1997) 137–159 MR

[13] **L Meier**, *United elliptic homology*, PhD thesis, Universät Bonn (2012) `http://www.math.uni-bonn.de/~grk1150/DISS/dissertation-meier.pdf`

[14] **J Rognes**, *Galois extensions of structured ring spectra: stably dualizable groups*, Mem. Amer. Math. Soc. 898, Amer. Math. Soc., Providence, RI (2008) MR

[15] **C A Weibel**, **S C Geller**, *Étale descent for Hochschild and cyclic homology*, Comment. Math. Helv. 66 (1991) 368–388 MR

Department of Mathematics, Harvard University
1 Oxford Street, Cambridge, MA 02138, United States

amathew@math.harvard.edu

Received: 30 January 2015 Revised: 25 June 2016

Algebraic & Geometric Topology 17 (2017) 705–740

Thickness, relative hyperbolicity, and randomness in Coxeter groups

JASON BEHRSTOCK
MARK F HAGEN
ALESSANDRO SISTO

APPENDIX WRITTEN JOINTLY WITH PIERRE-EMMANUEL CAPRACE

For right-angled Coxeter groups W_Γ, we obtain a condition on Γ that is necessary and sufficient to ensure that W_Γ is *thick* and thus not relatively hyperbolic. We show that Coxeter groups which are not thick all admit canonical minimal relatively hyperbolic structures; further, we show that in such a structure, the peripheral subgroups are both parabolic (in the Coxeter group-theoretic sense) and strongly algebraically thick. We exhibit a polynomial-time algorithm that decides whether a right-angled Coxeter group is thick or relatively hyperbolic. We analyze random graphs in the Erdős–Rényi model and establish the asymptotic probability that a random right-angled Coxeter group is thick.

In the joint appendix, we study Coxeter groups in full generality, and we also obtain a dichotomy whereby any such group is either strongly algebraically thick or admits a minimal relatively hyperbolic structure. In this study, we also introduce a notion we call *intrinsic horosphericity*, which provides a dynamical obstruction to relative hyperbolicity which generalizes thickness.

05C80, 20F55, 20F65

Introduction

The notion of relative hyperbolicity was introduced by Gromov [38], then developed by Farb [35]. This notion is both sufficiently general to include many important classes of groups, including all (uniform and nonuniform) lattices in rank-one semisimple Lie groups, yet is sufficiently restrictive that it allows for powerful geometric, algebraic and algorithmic results to be proven; see Arzhantseva, Minasyan and Osin [1], Druțu [27], Druțu and Sapir [30] and Farb [35]. Further, relatively hyperbolicity admits numerous geometric, topological and dynamical formulations which are all equivalent; see eg Bowditch [12], Dahmani [21], Druțu and Sapir [29], Osin [44], Sisto [45; 46] and Yaman [48].

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.705

Let G be a finitely generated group and $\mathcal{P}$ a finite collection of proper subgroups of G. The group G is *hyperbolic relative to the subgroups* $\mathcal{P}$ if collapsing the left cosets of $\mathcal{P}$ to finite-diameter sets, in any (hence every) word metric on G, yields a δ–hyperbolic space, and if the collection $\mathcal{P}$ satisfies the *bounded coset property* which, roughly speaking, requires that in the δ–hyperbolic metric space obtained as above, any pair of quasigeodesics with the same endpoints travels through the collapsed cosets in approximately the same manner. The subgroups in $\mathcal{P}$ are called *peripheral subgroups*. We say a group is *relatively hyperbolic* when there is some collection of subgroups for which this holds. A collection $\mathcal{P}$ of peripheral subgroups of the relatively hyperbolic group G is *minimal* if, for any other relatively hyperbolic structure $(G, \mathcal{Q})$ on G, each $P \in \mathcal{P}$ is conjugate into some $Q \in \mathcal{Q}$. Relatively hyperbolic groups do not always admit minimal structures; see Behrstock, Druţu and Mosher [5, Theorem 6.3]. We will follow the convention of requiring the subgroups to be proper, which rules out the trivial case of G being hyperbolic relative to itself. Note also that a group G is hyperbolic relative to hyperbolic subgroups if and only if G is hyperbolic.

We will also be interested in the notion of *thickness* which was introduced by Behrstock, Druţu and Mosher [5] as a powerful geometric obstruction to relative hyperbolicity which holds in many interesting cases, including most mapping class groups, right-angled Artin groups, lattices in higher-rank semisimple Lie groups, and elsewhere. Thickness is defined inductively: At the base level, *thick of order* 0, it is characterized by linear divergence. Roughly, a group is *thick of order* n if it is a "network of left cosets of subgroups" which are thick of lower orders. This essentially means that the union of these cosets is the entire space, and any two points in the space can be connected by a sequence of these cosets which successively intersect along infinite-diameter subsets; the precise definition appears in Section 1.2. Thickness has proven to be an important invariant for obtaining upper bounds on divergence, and we shall utilize this below; cf Behrstock and Charney [3], Behrstock and Druţu [4], Behrstock and Hagen [7], Brock and Masur [13] and Sultan [47]. In a relatively hyperbolic group, any thick subgroup must be contained inside a peripheral subgroup; see [5, Corollary 7.9, Theorem 4.1]. This fact yields the useful application that any relatively hyperbolic structure in which the peripheral subgroups are thick is a minimal relatively hyperbolic structure; see [29, Theorem 1.8] and [5, Corollary 4.7].

In this paper, we study thickness and relative hyperbolicity in the setting of Coxeter groups. One reason to do so is that Coxeter groups have many interesting properties, making them a standard testing ground in geometric group theory. For example, these groups are known to act properly on $\mathrm{CAT}(0)$ cube complexes (see Niblo and Reeves [43]), which allows them to be studied using the tools of $\mathrm{CAT}(0)$ geometry. In particular, this connects them to the study of thickness of cubulated groups initiated in [7].

We first specialize to the case of right-angled Coxeter groups, the class of which is diverse; for instance, each right-angled Artin group is a finite-index subgroup of a right-angled Coxeter group; see Davis and Januszkiewicz [24]. The right-angled Coxeter group W_Γ is generated by involutions indexed by vertices of the finite simplicial graph Γ; the relations are commutation relations corresponding to edges. We prove that, for every right-angled Coxeter group, either it is thick or it admits a canonical relatively hyperbolic structure in which the peripheral subgroups are thick:

Theorem I (right-angled Coxeter groups are thick or relatively hyperbolic) *Let $\mathcal{T}$ be the class consisting of the finite simplicial graphs Λ such that W_Λ is strongly algebraically thick. Then for any finite simplicial graph Γ, either $\Gamma \in \mathcal{T}$ or there exists a collection $\mathbb{J}$ of induced subgraphs of Γ such that $\mathbb{J} \subset \mathcal{T}$, W_Γ is hyperbolic relative to the collection $\{W_J : J \in \mathbb{J}\}$, and this relatively hyperbolic structure is minimal.*

One application of this theorem is to the quasi-isometric classification of Coxeter groups. As thickness is a quasi-isometric invariant, this provides a way to distinguish the thick Coxeter groups from many other groups. A more refined classification also follows from this result using the theorem which states that the quasi-isometric image of a group which is hyperbolic relative to thick peripheral subgroups is also hyperbolic relative to thick peripheral subgroups, each of which is quasi-isometric to one of the peripherals in the source; see [5, Corollary 4.8] and [27]. Prior to this application of Theorem I, major methods of classifying right-angled Coxeter groups included using classification theorems in right-angled Artin groups (ie Behrstock and Neumann [9], Behrstock, Januszkiewicz and Neumann [8] and Bestvina, Kleiner and Sageev [10]) in conjunction with results about commensurability between right-angled Artin and Coxeter groups (for instance, results in Davis and Januszkiewicz [24]) and, for some hyperbolic right-angled Coxeter groups, applying a result in Crisp and Paoluzzi [20].

Additionally, Theorem I provides an effective classification theorem because $\mathcal{T}$ can be characterized combinatorially as follows:

Theorem II (combinatorial characterization of thick right-angled Coxeter groups) *Let $\mathcal{T}$ be the class of finite simplicial graphs whose corresponding right-angled Coxeter groups are strongly algebraically thick. Then $\mathcal{T}$ is the smallest class of graphs satisfying the following conditions:*

1. *$K_{2,2} \in \mathcal{T}$, where $K_{2,2}$ is the complete bipartite graph on two sets of two elements, ie a 4–cycle.*
2. *Let $\Gamma \in \mathcal{T}$ and let $\Lambda \subset \Gamma$ be an induced subgraph which is not a clique. Then the graph obtained from Γ by coning off Λ is in $\mathcal{T}$.*

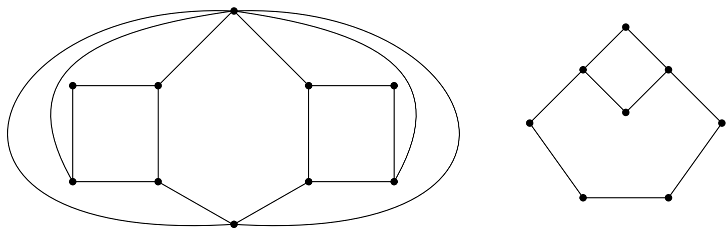


Figure 1: A graph in $\mathcal{T}$ (left) and a graph not in $\mathcal{T}$ (right)

(3) *Let $\Gamma_1, \Gamma_2 \in \mathcal{T}$, and suppose there exists a graph Γ which is not a clique and which arises as a subgraph of each of the Γ_i. Then the union Λ of Γ_1 and Γ_2 along Γ is in $\mathcal{T}$, and so is any graph obtained from Λ by adding any collection of edges joining vertices in $\Gamma_1 - \Gamma$ to vertices of $\Gamma_2 - \Gamma$.*

Theorems I and II together imply that any thick right-angled Coxeter group is strongly algebraically thick. A special case of this is that W_Γ is thick of order 0 if and only if it is the product of two infinite right-angled Coxeter groups; see Proposition 2.11, which generalizes a result of Dani and Thomas [22, Theorem 4.1].

Figure 1 illustrates examples of graphs in and not in $\mathcal{T}$. See also Remark 2.8. The right-angled Coxeter groups with polynomial divergence constructed by Dani and Thomas [22] are strongly algebraically thick; this was shown in [loc. cit.] and can also be verified either by observing that the corresponding graphs are in $\mathcal{T}$, or by combining the fact that they have subexponential divergence with Theorem I and the exponential divergence of any relatively hyperbolic group.

An important consequence of the above characterization of the class $\mathcal{T}$ is that it allows thickness/relative hyperbolicity to be detected algorithmically:

Theorem III (polynomial algorithm for relative hyperbolicity; Theorem 4.1) *There exists a polynomial-time algorithm to decide if a given graph is in $\mathcal{T}$, and hence whether a given right-angled Coxeter group is (strongly algebraically) thick or relatively hyperbolic.*

Random graphs

We consider right-angled Coxeter groups on random graphs in the Erdős–Rényi model [31]: $G(n, p(n))$ is the class of graphs on n vertices with the probability measure corresponding to independently declaring each pair of vertices to be adjacent with probability $p(n)$. The results of this section are summarized in Figure 2.

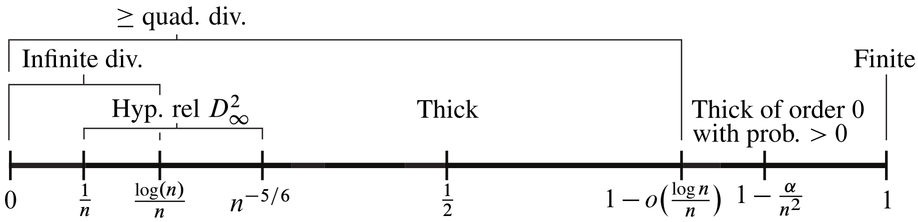


Figure 2: The results of Section 3 illustrated on the same spectrum of densities as addressed conjecturally in Figure 4. Each listed property occurs aas at the given density, unless the specific asymptotic probability is stated.

An important result of Erdős and Rényi states that a random graph is asymptotically almost surely (aas) connected when $p(n)$ grows more quickly that $(\log n)/n$, and is aas disconnected when $p(n) = o((\log n)/n)$. This implies that for slowly growing $p(n)$, when $\Gamma \in G(n, p(n))$, the right-angled Coxeter group W_Γ is aas a nontrivial free product, and hence relatively hyperbolic. In light of Theorem I, it is natural to wonder if there are densities at which a random right-angled Coxeter group is relatively hyperbolic but not a free product. The following gives a positive answer to this question; the technical terms in this theorem will be defined in Section 3.

Theorem IV (low density, Theorem 3.4) *Suppose $p(n)n \to \infty$ and $p(n)^6 n^5 \to 0$. For $\Gamma \in G(n, p(n))$, the group W_Γ is aas hyperbolic relative to a nonempty collection of $D_\infty \times D_\infty$ subgroups; the same holds for $W_{\Gamma'}$, where $\Gamma' \subseteq \Gamma$ is the giant component of Γ.*

Intuitively, the probability of thickness should increase with the growth rate of $p(n)$, up to the point where Γ is aas sufficiently dense that W_Γ is either finite or virtually cyclic. The following result confirms this intuition.

Theorem V (high density, Theorem 3.9) *Suppose that $(1-p(n))n^2 \to \alpha \in [0,\infty)$. Then for $\Gamma \in G(n, p(n))$, the group W_Γ is*

(1) *finite with probability tending to $\beta = e^{-\alpha/2}$,*
(2) *virtually $\mathbb{Z}$ with probability tending to $\gamma = \frac{1}{2}\alpha e^{-\alpha/2}$,*
(3) *virtually $\mathbb{Z}^k$ for $k \geq 2$, and thus thick of order 0, with probability tending to $1-(\beta+\gamma)$.*

The following describes the situation at a natural choice of "intermediate" $p(n)$:

Theorem VI (intermediate density) *For $\Gamma \in G\left(n, \frac{1}{2}\right)$, the group W_Γ is aas thick.*

We conjecture that for all $p \in (0,1)$, the group W_Γ is aas thick for $\Gamma \in G(n,p)$.[1] This conjecture is strongly supported by computer experiments; for example, for $n = 200$ and for each of several values of p, we tested 50 random graphs and found *all* to correspond to thick right-angled Coxeter groups. For any given $p \in (0,1)$, we expect the strategy used in the proof of Theorem VI will work. However, there are two serious complications to implementing this strategy for any particular p: first, combinatorially, the requisite set-up may be more intricate, and second, establishing the base case of the induction is likely to be computationally prohibitive for some values of p, since it involves checking all graphs of a size depending on p for membership in $\mathcal{T}$.

One of our motivations for our study of random Coxeter groups was the results of Charney and Farber [18] on hyperbolicity of random right-angled Coxeter groups. More recently, results have been obtained about cohomological properties of such random groups by Davis and Kahle [25]. Together with our results, this represents the beginning of a systematic study of random Coxeter groups.

General Coxeter groups

In the appendix, we generalize Theorems I and II to all Coxeter groups. Accordingly, we recommend reading the first part of the appendix, Section A.1, concurrently with Section 2 in order to see how the results on thickness versus relative hyperbolicity for right-angled Coxeter groups generalize to arbitrary Coxeter groups, as well as the limitations of the generalization. In the latter vein, as shown by the example in Remark 2.9, there is no characterization of strongly algebraically thick Coxeter groups that are not right-angled purely in terms of the underlying graph of the free Coxeter diagram.

Theorem I generalizes as follows:

Theorem VII (minimal relatively hyperbolic structures for Coxeter groups) *Let (W, S) be a Coxeter system. Then there is a (possibly empty) collection $\mathcal{J}$ of subsets of S enjoying the following properties:*

(i) *The parabolic subgroup W_J is strongly algebraically thick for every $J \in \mathcal{J}$.*

(ii) *W is relatively hyperbolic with respect to $\mathcal{P} = \{W_J \mid J \in \mathcal{J}\}$.*

In particular, $\mathcal{P}$ is a minimal relatively hyperbolic structure for W.

Theorem II takes the following form for general Coxeter groups. Note that thickness is now described using a class of labeled graphs instead of a class of graphs.

[1] While this paper was circulating as a preprint, a resolution of a strong form of this conjecture was obtained by Behrstock, Falgas-Ravry, Hagen and Susse [6].

Theorem VIII (classification of thick Coxeter groups) *The class $\mathbb{T}$ of Coxeter systems (W,S) for which W is strongly algebraically thick is the smallest class satisfying:*

(1) *$\mathbb{T}$ contains the class $\mathbb{T}_0$ of all irreducible* affine *Coxeter systems (W,S) with S of cardinality at least 3, as well as all Coxeter systems of the form $(W, S_1\cup S_2)$ with W_{S_1} and W_{S_2} irreducible nonspherical and $[W_{S_1}, W_{S_2}]=1$.*

(2) *Suppose $(W, S\cup s)$ has the properties that $s^{\perp}$ is nonspherical and (W_S, S) belongs to $\mathbb{T}$. Then $(W, S\cup s)$ belongs to $\mathbb{T}$.*

(3) *Suppose (W,S) has the property that there exist $S_1, S_2\subseteq S$ with $S_1\cup S_2=S$, $(W_{S_1},S_1),(W_{S_2},S_2)\in\mathbb{T}$ and $W_{S_1\cap S_2}$ nonspherical. Then $(W,S)\in\mathbb{T}$.*

We also introduce the notion, which we feel will be of independent interest, of an *intrinsically horospherical* group, ie one for which every proper isometric action of Γ on a proper hyperbolic geodesic metric space fixes a unique point at infinity. Any group G admits a collection of maximal intrinsically horospherical subgroups, and any relatively hyperbolic structure on G has the property that every maximal intrinsically horospherical subgroup is conjugate into a peripheral subgroup. We show that any thick group is intrinsically horospherical. In the case of Coxeter groups, we say more:

Corollary IX *Let (W,S) be a Coxeter system. Then the following conditions are equivalent:*

(I) *(W,S) is in $\mathbb{T}$.*

(II) *W is strongly algebraically thick.*

(III) *W is intrinsically horospherical.*

(IV) *W is not relatively hyperbolic with respect to any family of proper subgroups.*

(V) *W is not relatively hyperbolic with respect to any family of proper Coxeter-parabolic subgroups.*

Outline

In Section 1, we discuss background on Coxeter groups, thickness and divergence. Sections 2, 3 and 4 are devoted to right-angled Coxeter groups: In the second section, we treat Theorems I and II. In the third section, we study right-angled Coxeter groups presented by random graphs, dealing in particular with Theorems IV, V and VI. In the fourth section, we produce an algorithm for testing whether a given graph is in $\mathcal{T}$. We also include source code containing an implementation of a refined version of this algorithm; this program is needed for a computation in the proof of Theorem VI. (This source code is available from the authors' web pages and on the arXiv.) In the appendix, we study arbitrary Coxeter groups and introduce the notion of intrinsic horosphericity; in particular, we prove Theorems VII and VIII and Corollary IX.

Acknowledgments Hagen and Sisto thank the organizers of the conference Geometric and Analytic Group Theory (Ventotene 2013). We thank Kaia Behrstock for her help making Figure 4. Finally, we are grateful to Tim Susse, Ha-Young Shin and the referees for several helpful comments and corrections.

Hagen was supported by the National Science Foundation under Grant Number NSF 1045119. Behrstock was supported as an Alfred P Sloan Fellow and by the National Science Foundation under Grant Number NSF 1006219.

1 Preliminaries

In this section, we review definitions and facts related to Coxeter groups, divergence and thick metric spaces. A comprehensive discussion of Coxeter groups can be found in [23]. The notion of divergence used here is due to Gersten [36]. Our consideration of divergence in the setting of Coxeter groups was motivated largely by the discussion in [22] and, to some extent, by questions about divergence in cubulated groups (of which Coxeter groups are examples) raised in [7]. Thick spaces and groups were introduced in [5], and we also refer to results of [4].

1.1 Background on Coxeter groups

Throughout this paper, we confine our discussion to finitely generated Coxeter groups. A *Coxeter group* is a group of the form

$$\langle S \mid (st)^{m_{st}} : s,t \in S\rangle,$$

where each $m_{ss}=1$, and for $s\neq t$, either $m_{st}\geq 2$ or there is no relation between s and t of this form. Also, $m_{st}=m_{ts}$ for each $s,t\in S$. The pair (W,S) is a *Coxeter system.*

The Coxeter group W is *reducible* if there are nonempty sets $S_1,S_2\subset S$ such that $S=S_1\sqcup S_2$, and for all $s_1\in S_2, s_2\in S_2$, we have $m_{s_1s_2}=2$. If W is not reducible, then it is *irreducible*. The Coxeter system (W,S) is said to be *(ir-)reducible* if W has the corresponding property.

To the Coxeter system (W,S), we associate a bilinear form $\langle -,-\rangle$ on $\mathbb{R}[S]$ defined by $\langle s,t\rangle=-\cos(\pi/m_{st})$ when there is a relation $(st)^{m_{st}}$, and $\langle s,t\rangle=-1$ otherwise. It is well known that this bilinear form is positive definite if and only if W is finite, in which case the Coxeter system (W,S) is *spherical*. Otherwise, (W,S) is nonspherical (or *aspherical*). If the bilinear form is positive semidefinite and (W,S) is irreducible, then there is a short exact sequence $\mathbb{Z}^n\to W\to W_0$, where $n+1=|S|$ and W_0 is a finite Coxeter group. In this case, the Coxeter system (W,S) is *(irreducible) affine*.

For any $J \subset S$, the subgroup $W_J := \langle J \rangle \subset W$ is a *parabolic* subgroup. Evidently, W_J is again a Coxeter group and (W_J, J) a Coxeter system. The subset J is *spherical, irreducible, affine, etc.* if the Coxeter system (W_J, J) has the same property.

Right-angled Coxeter groups If each relation in the above presentation has the form $(st)^2$, then W is a *right-angled Coxeter group*. In this case, let Γ be the graph with vertex set S and with an edge joining $s, t \in S$ if and only if $(st)^2 = 1$, ie if and only if the involutions s and t commute. Then W decomposes as a graph product: the underlying graph is Γ, and the vertex groups are the subgroups $\langle s \rangle \cong \mathbb{Z}_2$ and $s \in S$.

Conversely, given a finite simplicial graph Γ with vertex set S and edge set $\mathcal{E}$, there is a right-angled Coxeter group

$$W_\Gamma := \langle S \mid s^2, (st)^2 : s, t \in S, (s,t) \in \mathcal{E} \rangle.$$

For example, if Γ is disconnected, then W_Γ is isomorphic to the free product of the parabolic subgroups generated by the vertex sets of the various components, while if Γ decomposes as a nontrivial join, then W_Γ is isomorphic to the product of the parabolic subgroups generated by the factors of the join. For $J \subset S$, the parabolic subgroup $W_J \leq W_\Gamma$ is isomorphic to the right-angled Coxeter group W_Λ, where Λ is the subgraph of Γ induced by J.

Finally, we remark that if W_Γ is a right-angled Coxeter group, then there exists a CAT(0) cube complex $\widetilde{X}_\Gamma$ on which W_Γ acts properly discontinuously and cocompactly. This CAT(0) cube complex is the *Davis complex* X_Γ, which is obtained from the universal cover of the presentation complex of W_Γ by collapsing bigons to edges, noting that each remaining 2–cell is a 2–cube, and then iteratively attaching a k–cube whenever its vertex set is contained in the $(k-1)$–skeleton, for $k \geq 3$; see [23] for details. We will make use of the existence of such a CAT(0) cube complex in the proof of Proposition 2.11.

1.2 Background on divergence and thickness

Given functions $f, g \colon \mathbb{R}_+ \to \mathbb{R}_+$, we write $f \preceq g$ if for some $K \geq 1$, we have $f(s) \leq Kg(Ks + K) + Ks + K$ for all $s \in \mathbb{R}_+$, and $f \asymp g$ if $f \preceq g$ and $g \preceq f$.

Definition 1.1 (divergence) Let (M, d) be a geodesic metric space, let $\delta \in (0,1)$ and $\gamma \geq 0$, and let $f \colon \mathbb{R}_+ \to \mathbb{R}_+$ be given by $f(r) = \delta r - \gamma$. Given $a, b, c \in M$ with $d(c, \{a, b\}) = r > 0$, let $\mathrm{div}_f(a, b; c) = \inf\{|P|\}$, where P varies over all paths in M joining a to b and avoiding the ball of radius $f(r)$ about c. If no such path exists, $\operatorname{div}_f(a,b;c)=\infty$. The *divergence function* $\operatorname{Div}_f^M\colon \mathbb{R}_+\to\mathbb{R}_+$ of M is then defined by

$$\operatorname{Div}_f^M(s)=\sup\{\operatorname{div}_f(a,b;c):d(a,b)\le s\}.$$

Note that M has finite divergence if and only if M has one end.

Given a function $g\colon \mathbb{R}_+\to\mathbb{R}_+$, we say that M has *divergence of order at most* g if for some f as above, $\operatorname{Div}_f^M(s)\preccurlyeq g(s)$. Much of the interest in divergence comes from the fact that the divergence function of M is a quasi-isometry invariant in the following sense: if M_1 and M_2 are quasi-isometric geodesic metric spaces and $\operatorname{Div}_f^{M_1}\asymp g$, then $\operatorname{Div}_{f'}^{M_2}\asymp g$ for some f'. In particular, the divergence of a finitely generated group is well defined up to the relation $\asymp$. A group has linear divergence if and only if it does not have cut-points in any asymptotic cone. Such spaces are called *wide*; see [2; 28].

One family of metric spaces which are particularly amenable to divergence computations is the family of *thick* spaces, as introduced in [5]. Thickness is a quasi-isometrically invariant notion, and this family of spaces is partitioned into quasi-isometrically invariant subclasses by their *order of thickness*, which is a nonnegative integer. In the present paper, we work with a refinement of the notion of thickness which is tuned for the study of finitely generated groups:

Definition 1.2 (strongly algebraically thick [4]) A finitely generated group G is said to be *strongly algebraically thick of order* 0 if it is *wide*. For $n\ge 1$, the finitely generated group G is *strongly algebraically thick of order at most n* if there exists a finite collection $\mathcal{H}$ of subgroups such that:

(1) Each $H\in\mathcal{H}$ is strongly algebraically thick of order at most $n-1$.
(2) $\langle\bigcup_{H\in\mathcal{H}}H\rangle$ has finite index in G.
(3) There exists $C\ge 0$ such that for all $H,H'\in\mathcal{H}$, there is a sequence $H=H_1,\dots,H_k=H'$ with each $H_i\in\mathcal{H}$ such that for all $i\le k$, the intersection $H_i\cap H_{i+1}$ is infinite and the C–neighborhood of $H_i\cap H_{i+1}$ (with respect to some fixed word metric on G) is path-connected.
(4) For all $H\in\mathcal{H}$, any two points in H can be connected in the C–neighborhood of H by a (C,C)–quasigeodesic.

G is *strongly algebraically thick of order n* if G is strongly algebraically thick of order at most n but is not strongly algebraically thick of order at most $n-1$.

As shown in [4], if G is strongly algebraically thick of order n, then G, with any word metric, is a (strongly) thick metric space. In the present paper, we are particularly interested in the following consequences of strong algebraic thickness:

Proposition 1.3 (upper bound on divergence [4, Corollary 4.17]) *Let G be a finitely generated group that is strongly algebraically thick of order n. Then the divergence function of G is of order at most s^{n+1}.*

Proposition 1.4 (nonrelative hyperbolicity [5, Corollary 7.9]) *Let G be strongly algebraically thick. Then G is not hyperbolic relative to any collection of proper subgroups.*

Note that the above establishes that the divergence function of thick groups is qualitatively different from that of relatively hyperbolic groups, as the latter class has divergence functions which are at least exponential; cf [45, Theorem 1.3].

2 Hyperbolicity relative to thick subgroups: the right-angled case

In this section, Γ will denote a finite simplicial graph and W_Γ will denote the associated right-angled Coxeter group. We will postpone proofs of most of the results of this section to the appendix, where we will consider them in the context of arbitrary Coxeter groups. We focus on the right-angled case here, both for the benefit of readers specifically interested in the right-angled case and because these groups are cocompactly cubulated, which allow for more refined results, such as those in Proposition 2.11 and in Section 3.

We will adopt the following:

Convention 2.1 When we say *graph*, we will always mean a finite simplicial graph (ie no multiedges or monogons). Graphs will often be denoted by Greek letters. When we say Λ is a subgraph of Γ, or when we write $\Lambda \subset \Gamma$, we will mean the *full induced subgraph*; ie a pair of vertices of Λ spans an edge in Λ if and only if they span one in Γ.

We begin by defining the class of graphs $\mathcal{T}$ that we discussed briefly in the introduction.

Definition 2.2 (new graphs from old) If Γ is a graph and $\Lambda \subset \Gamma$, then we say that the graph Γ' is obtained by *coning off* Λ if the graph Γ' can be obtained from Γ by adding one new vertex along with edges between that vertex and each vertex of Λ. Given two graphs Γ_1 and Γ_2 with isomorphic subgraphs Γ, we say the *union of* Γ_1 *and* Γ_2 *along* Γ is the graph obtained by taking the disjoint union of the graphs Γ_1 and Γ_2 and identifying the corresponding Γ subgraphs of Γ_i by the given isomorphism taking one of the Γ subgraphs to the other. Given two graphs Γ_1 and Γ_2 with isomorphic subgraphs Γ, we say that a graph Γ' is a *generalized union of* Γ_1 *and* Γ_2 *along* Γ if Γ' can be obtained from the associated union by adding a collection of edges between vertices of $\Gamma_1 \setminus \Gamma$ and vertices of $\Gamma_2 \setminus \Gamma$.

Definition 2.3 (thick graphs) The set of *thick graphs*, $\mathcal{T}$, is the smallest set of graphs satisfying the following conditions:

(1) $K_{2,2} \in \mathcal{T}$.

(2) If $\Gamma \in \mathcal{T}$ and $\Lambda \subset \Gamma$ is any induced subgraph of diameter greater than one, then the graph obtained by *coning off* Λ is in $\mathcal{T}$.

(3) Let $\Gamma_1, \Gamma_2 \in \mathcal{T}$ with both Γ_i containing an isomorphic subgraph, Γ, which is not a clique. Then any generalized union of the Γ_i along Γ is in $\mathcal{T}$.

When W is a right-angled Coxeter group, there are no irreducible affine Coxeter systems (W, S) with S of cardinality at least 3. In particular, it is straightforward to check that a right-angled Coxeter group is defined by a graph in $\mathcal{T}$ if and only if the group is in the class of right-angled Coxeter groups $\mathbb{T}$ which is defined at the beginning of Section A.1. The next result is thus a consequence of Proposition A.2.

Theorem 2.4 *For each $\Gamma \in \mathcal{T}$, the right-angled Coxeter group W_Γ is strongly algebraically thick.*

The main result of this section is the following, which provides an effective classification theorem with our explicit description of $\mathcal{T}$.

Theorem 2.5 *Let Γ be a graph. The right-angled Coxeter group W_Γ satisfies exactly one of the following:*

- *it is strongly algebraically thick and $\Gamma \in \mathcal{T}$, or*
- *it is hyperbolic relative to a (possibly empty) minimal collection $\mathbb{A}$ of parabolic subgroups for which each $W_\Lambda \in \mathbb{A}$ is strongly algebraically thick and with each such $\Lambda \in \mathcal{T}$.*

If a group is hyperbolic relative to the empty collection of subgroups, then it is hyperbolic; hence if $\mathbb{A}$ is empty, then W_Γ is hyperbolic.

Theorem 2.5 can now be proven considering the collection of all maximal subgraphs of Γ that belong to $\mathcal{T}$ and checking that conditions (RH1)–(RH3) of [15, Theorem A′] hold. We postpone the proof of this to the appendix.

Remark 2.6 An alternative way to prove Theorem 2.5 is to define $\mathcal{T}$ to be the set of finite graphs whose corresponding right-angled Coxeter groups are thick. It would then suffice to establish the following statements about induced subgraphs J_1, J_2 of Γ belonging to $\mathcal{T}$:

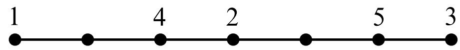


Figure 3: A length-6 geodesic in Γ shows that $\Gamma \in \mathcal{F}$.

(1) If $J_1 \cap J_2$ is aspherical, then the subgraph induced by $J_1 \cup J_2$ belongs to $\mathcal{T}$.

(2) If $v \in \Gamma - J_1$ and the link of v in J_1 is nonempty and aspherical, then $J_1 \cup \{v\} \in \mathcal{T}$.

(3) Joins of aspherical subgraphs belong to $\mathcal{T}$.

Our explicit definition of $\mathcal{T}$ allows us to characterize thick right-angled Coxeter groups, as we do now.

Corollary 2.7 *W_Γ is strongly algebraically thick if and only if $\Gamma \in \mathcal{T}$.*

Proof If W_Γ is strongly algebraically thick, then Γ is not relatively hyperbolic by [5, Corollary 7.9]. Thus by Theorem 2.5, we must have $W_\Gamma \in \mathcal{T}$. In the other direction: by Theorem 2.4, if $\Gamma \in \mathcal{T}$, then W_Γ is strongly algebraically thick. □

Remark 2.8 From Corollary 2.7, we know that all right-angled Coxeter groups which are wide have corresponding graphs in $\mathcal{T}$. As we shall see in Proposition 2.11, these graphs all decompose as nontrivial joins, and thus in particular, the number of squares in these graphs is linear in the number of vertices. In the case of right-angled Coxeter groups which are thick of order 1, it was proven in [22] that each vertex in the corresponding graph is contained in a square; hence in that case as well, the number of squares is at least linear in the number of vertices.

Accordingly, it is natural to expect that a graph in $\mathcal{T}$ contains "many" squares relative to the number of vertices it contains. However, this is not the case in general. Indeed, for all sufficiently large $N \in \mathbb{N}$, the set of graphs in $\mathcal{T}$ containing at most N squares is infinite. We define a class of graphs $\mathcal{F}$ consisting of graphs Γ such that $\Gamma \in \mathcal{T}$ and Γ contains vertices $v_1, \dots, v_5$ for which $d(v_i, v_{i+1}) \geq 3$ for each i. If $\Gamma \in \mathcal{F}$, then the graph obtained by joining v_i and v_{i+1} by a path of length 2 is also in $\mathcal{F}$, and it has the same number of squares as Γ and strictly more vertices. Any element of $\mathcal{T}$ of diameter at least 6 is in $\mathcal{F}$, since it has an induced subgraph which is in $\mathcal{F}$, namely, the path of length 6 (as shown in Figure 3).

The claim now follows for some N since $\mathcal{T}$ contains graphs of arbitrarily large diameter, as we shall now show. Begin with a graph $\Gamma_0 \in \mathcal{T}$ of diameter $d \geq 3$ with the additional property that some vertex v_0 of Γ_0 lies at distance d from nonadjacent vertices u_0 and w_0 (for example, the graph in Figure 1 (left)). Form Γ_1 from Γ_0 by adding two new vertices u_1 and w_1, each joined by an edge to u_0 and w_0. By Theorem 2.4, $\Gamma_1 \in \mathcal{T}$. By construction, the distance in Γ_1 from each of u_1 and w_1 to v_0 is $d+1$, so the diameter has increased. Finally, the triple v_0, u_1, w_1 shows that Γ_1 has the property needed to repeat this procedure. Hence, the existence of graphs in $\mathcal{T}$ of arbitrarily large diameter follows by induction.

Remark 2.9 (Theorem 2.4 does not hold for general Coxeter groups) Given a (not necessarily right-angled) Coxeter system (W, S), there is a naturally associated labeled graph Γ, the *free Coxeter diagram*, with vertex set S and an edge labeled $n \geq 2$ joining vertices s and t that satisfy a relation $(st)^n = 1$. Note that since $m_{ss} = 1$ for all $s \in S$, this graph is simplicial. Furthermore, if (W, S) is right angled, then all labels are 2, and Γ is the graph considered above.

If the Coxeter group W is not right-angled, the thickness of W cannot be characterized by a purely graph-theoretic property of the free Coxeter diagram. Indeed, there exists a hyperbolic Coxeter group W whose free Coxeter diagram is a 4–cycle: Consider the Coxeter system determined by the presentation

$$W = \langle s, t, u, v \mid s^2, t^2, u^2, v^2, (st)^n, (su)^2, (uv)^2, (tv)^2 \rangle,$$

with $n \geq 3$. The labeled graph Γ is a 4–cycle, with the edge joining s, t labeled $n \geq 3$ and all other edges labeled 2. However, the group W is a Fuchsian group, being generated by reflections in the sides of a 4–gon in $\mathbb{H}^2$ with angles $\frac{\pi}{2}$, $\frac{\pi}{2}$, $\frac{\pi}{2}$ and $\frac{\pi}{n}$. Being hyperbolic, W cannot be thick.

Combining the upper bound on divergence of strongly thick spaces given in [4, Corollary 4.17], the fact that relatively hyperbolic groups have exponential divergence (see eg [45, Theorem 1.3]) and Theorem 2.5, we obtain:

Corollary 2.10 *Let Γ be a connected graph. Then the divergence function of W_Γ is either exponential or bounded above by a polynomial.*

2.1 Characterizing thickness of order 0

As it turns out, the class $\mathcal{T}_0$ of graphs Γ for which W_Γ is wide admits a simple description as we shall see below. The triangle-free case of this result was previously established using different techniques in [22, Theorem 4.1]. We note that since there exist wide Coxeter groups which are not products (for instance the 3–3–3 triangle group), the following result does not generalize beyond the right-angled case.

Proposition 2.11 *$\mathcal{T}_0$ is the set of graphs of the form $(\Gamma_1 \star \Gamma_2) \star K$, where Γ_1 and Γ_2 are aspherical and K is a (possibly empty) clique.*

Proof If Γ decomposes as in the statement of the proposition, then W_Γ decomposes as the product of infinite subgroups $(W_{\Gamma_1} \times W_{\Gamma_2}) \times \mathbb{Z}_2^{|K|}$, whence W_Γ has linear divergence and is therefore wide, ie $\Gamma \in \mathcal{T}_0$. Conversely, suppose that W_Γ has linear divergence, and let $\widetilde{X}_\Gamma$ be the Davis complex (see [23]). Then $\widetilde{X}_\Gamma$ is a CAT(0) cube complex on which W_Γ acts properly and cocompactly by isometries. Each hyperplane H of $\widetilde{X}_\Gamma$ is regarded as being labeled by a pair $(v, g) \in \Gamma^{(0)} \times W_\Gamma$, where gvg^{-1} acts as an inversion in the hyperplane H.

Recall that W_Γ acts *essentially*, in the sense of [17], on $\widetilde{X}_\Gamma$ if for each hyperplane H, the two components of $\widetilde{X}_\Gamma - H$ each contain points in some W_Γ–orbit which are arbitrarily far from H. A hyperplane without this property is called *inessential*.

Suppose that the action of W_Γ on $\widetilde{X}_\Gamma$ is *essential*. Then since W_Γ is wide, it contains no rank-one isometry of $\widetilde{X}_\Gamma$, and hence the rank-rigidity theorem of [17] implies that there exist unbounded convex subcomplexes $\widetilde{Y}$ and $\widetilde{Y}'$ such that $\widetilde{X}_\Gamma = \widetilde{Y} \times \widetilde{Y}'$. It follows that the link of the vertex in $\widetilde{X}_\Gamma$ decomposes as the join of aspherical subgraphs. But this link is exactly Γ, and hence Γ has the desired form.

Now we may assume W_Γ is not acting essentially on $\widetilde{X}_\Gamma$. Thus, by definition, there exists an inessential hyperplane $H_{(v,1)}$, and it is easy to see that every generator must commute with v. Indeed, if $H_{(w,1)}$ and $H_{(v,1)}$ are disjoint hyperplanes, then $\langle v, w\rangle\{H_{(w,1)}\}$ contains hyperplanes arbitrarily far from $H_{(v,1)}$ in each of its half-spaces. Let K be the clique in Γ whose vertices label such inessential hyperplanes. Then $\Gamma = \Gamma' \star K$, where Γ' is an aspherical set whose vertices label essential hyperplanes of $\widetilde{X}_\Gamma$. This provides the desired decomposition of Γ' as the join of aspherical subsets. $\square$

3 Random right-angled Coxeter groups

We now consider the right-angled Coxeter group W_Γ, where Γ is a random graph in the following sense. Let $p\colon \mathbb{N} \to [0,1]$ be a function such that $p(n)\binom{n}{2}$ has a limit in $\mathbb{R} \cup \{\infty\}$ as $n \to \infty$. A random graph on n vertices is formed by declaring each pair of vertices to span an edge, independently of other pairs, with probability $p = p(n)$. In other words, we define $G(n,p)$ to be the probability space consisting of simplicial graphs with n vertices where, for each graph Γ on n vertices, $\mathbb{P}(\Gamma) = p^E(1-p)^{\binom{n}{2}-E}$, where E is the number of edges in Γ. This model of random graphs was introduced by Gilbert in [37], and is both contemporaneous with and very similar to the Erdős–Rényi model of random graphs first studied in [31; 32]. For a survey of more recent results on random graphs, see [19].

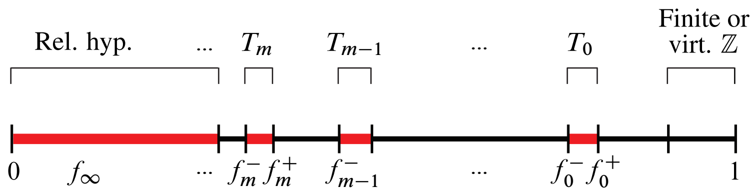


Figure 4: Prevalence of thickness along the "spectrum" of densities $p(n)$, if the answer to Question is positive; bold intervals are where, conjecturally, W_Γ is aas thick of a specified order.

Since the assignment $\Gamma \mapsto W_\Gamma$ of a finite simplicial graph to the corresponding right-angled Coxeter group is bijective [42], it is sensible to define "generic" properties of right-angled Coxeter groups with reference to the above model of random graphs. More precisely, if $\mathcal{P}$ is some property of right-angled Coxeter groups and $\mathcal{G}$ is a class of finite simplicial graphs such that W_Γ has the property $\mathcal{P}$ if and only if $\Gamma \in \mathcal{G}$, then we say that W_Γ satisfies $\mathcal{P}$ *asymptotically almost surely (aas)* if $\mathbb{P}(\Gamma \in \mathcal{G} \cap G(n,p)) \to 1$ as $n \to \infty$. We emphasize that the notion of asymptotically almost surely depends on the choice of probability function p even though it is customary to not explicitly mention p in the notation.

The following question describes the authors' best guess regarding the behavior of thickness and relative hyperbolicity for random right-angled Coxeter groups. In this section, we will provide both theorems and computations that motivate this picture, but we lead with it to contextualize the theorems that follow.

Question Let T_m be the set of graphs Γ for which W_Γ is thick of order $m \geq 0$, and denote by T_∞ the set of graphs for which W_Γ is hyperbolic relative to proper subgroups. Do there exist functions $f_m^-, f_m^+ \colon \mathbb{N} \to [0,1]$, for $m \geq 0$, such that $f_m^- = o(f_m^+)$, $f_m^+ = O(f_{m-1}^-)$ and

$$\lim_{n\to\infty} \mathbb{P}\big(\Gamma \in T_m \mid \Gamma \in G(n,p(n))\big) = \begin{cases} 0 & \text{if } p(n)/f_m^-(n) \to 0, \\ 1 & \text{if } p(n)/f_m^-(n) \to \infty \text{ and } p(n)/f_m^+(n) \to 0, \end{cases}$$

for all $m \geq 0$? Similarly, does there exist f_∞ such that W_Γ is asymptotically almost surely relatively hyperbolic when $\Gamma \in G(n,p(n))$ and $p = o(f_\infty)$?

The situation that would occur in the event of a positive answer to Question is illustrated heuristically in Figure 4. Given $p_1, p_2 \colon \mathbb{N} \to [0,1]$, we place p_1 to the left of p_2 in the picture of $[0,1]$ if and only if $p_1 = o(p_2)$. Compare also Figure 2, which summarizes the results of this section.

In the interval where W_Γ is aas relatively hyperbolic, it is interesting to speculate whether the order of thickness of the peripheral subgroups might be determined by $p(n)$, especially in view of Theorem 3.4, which we will see below. In other words, one could

a	n	Prop. thick	a	n	Prop. thick
1.95	2000	0.53	3	4000	0.5
1.95	2100	0.515	3	5000	0
1.95	4000	0	4	4000	1
2	2000	0.8	4	10000	1
2	2500	0.46	5	4000	1
2	3000	0.19	5	10000	1
2	4000	0.025	10	4000	1
2.5	2500	1	10	10000	1
2.5	3000	0.53			
2.5	4000	0			

Table 1: Experimental proportion of $\Gamma \in G(n, (a\log n)/n)$ that are thick. For each a, this proportion tends to 0 as $n \to \infty$ by Theorem 3.4 but, as illustrated, may do so quite slowly.

sensibly ask if there are functions $g_m^{\pm}$ such that W_Γ is aas hyperbolic relative to groups that are thick of order n for p between g_m^- and g_m^+, and if there is a function g_∞ such that W_Γ is aas hyperbolic — ie hyperbolic relative to hyperbolic subgroups — when $p = o(g_\infty)$. In fact, Charney and Farber have established that we can take $g_\infty(n) = n^{-1}$: when $np(n) \to 0$, the group W_Γ is aas hyperbolic, and if $p(n) \to 0$ and $p(n)n \to \infty$, then aas W_Γ is not hyperbolic [18]. However, identifying the functions g_m appears to be an open question.

The results in this section are summarized in Figure 2. These results are consistent with a positive answer to Question, but there are significant "gaps" in the spectrum about which nothing is presently known.

Remark 3.1 (thickness and connectivity) If Γ is disconnected, then W_Γ splits as a nontrivial free product and is therefore not thick. Hence the function f_∞ from Question, if it exists, must satisfy $\log n/(nf_\infty) \to 0$, by Theorem 3.4 (as shown in Figure 2), since $(\log^6 n)/n \to 0$. In other words, there are densities at which Γ is aas connected but W_Γ is not aas thick. However, the convergence to 0 of the proportion of random graphs at density $O((\log n)/n)$ is quite slow. This is illustrated in Table 1, which shows data selected from the output of many computer experiments;[2] for correctly chosen $a > 0$, even at $n = 10000$ it is not yet clear that W_Γ is not aas thick at density $(a\log n)/n$.

[2] Source code available from the authors and at arXiv.

3.1 Behavior at low densities

In the next theorem, we collect a few facts about random right-angled Coxeter groups. Recall from [23, Theorem 8.7.4] that W_Γ is one-ended provided Γ has no separating clique.

Theorem 3.2 *W_Γ asymptotically almost surely decomposes as a nontrivial free product if and only if there exists $\epsilon > 0$ such that $p(n) < ((1-\epsilon)\log n)/n$. Hence, if $p(n) < ((1-\epsilon)\log n)/n$, then the divergence of W_Γ is aas infinite.*

If there exists $\epsilon > 0$ such that $p(n) > ((1+\epsilon)\log n)/n$, and there exists $k \in \mathbb{N}$ such that $n^k p(n)^{k^2} \to 0$, then aas Γ has no separating clique, and hence W_Γ is aas one-ended and has finite-divergence function.

Proof W_Γ admits a nontrivial free product decomposition if and only if Γ is disconnected, and $\log n/n$ is the threshold for $p(n)$ above which connectedness occurs aas and below which disconnectedness occurs aas; see [32].

Let $K_n = K_n(\Gamma)$ equal 1 or 0 according to whether Γ is disconnected. For $0 \le j \le n$, let $K_n^j(\Gamma) = \sum_\Lambda K_{n-j}(\Gamma - \Lambda)$, where Λ varies over the size-j subgraphs of Γ. Then $\mathbb{E}(K_n^j) = \binom{n}{j}\mathbb{E}(K_{n-j})p^{\binom{j}{2}}$ is an upper bound for the expected number of separating j–simplices, and the expected number of separating simplices in Γ is therefore bounded by

$$\sum_{j=0}^{n-2}\binom{n}{j}\mathbb{E}(K_{n-j})p^{\binom{j}{2}}.$$

Now, for $p(n) > (1+\epsilon)(\log(n))/n$ and $p = o(1)$, Theorem 1 of [31] implies that $\sum_{j\le k}\binom{n}{j}\mathbb{E}(K_{n-j})p^{\binom{j}{2}}$ tends to 0 for any fixed k. If $p(n)$ is sufficiently small to ensure that aas all cliques in Γ have size $O(1)$, ie if there exists k such that $\binom{n}{k}p^{\binom{k}{2}} \to 0$, then the preceding sum bounds the limiting expected number of separating cliques of any size, and the proof is complete. $\square$

Because of the hypothesis that $n^k p(n)^{k^2} \to 0$ for some $k \in \mathbb{N}$, the second assertion of Theorem 3.2 says nothing about how many ends W_Γ aas has when $\Gamma \in G(n,p)$ and $p \neq o(1)$. This should be expected in light of Theorem 3.9 below, which shows that if $p(n) \to 1$ sufficiently quickly, the random right-angled Coxeter group W_Γ will have 2 or 0 ends with positive probability. However, it is likely possible to improve the second assertion to show that W_Γ is aas one-ended for a wider range of p, provided we still have $p \not\to 1$ as $n \to \infty$, using the fact that aas all cliques in Γ have size in $O(\log n)$ provided $p \not\to 1$, by an application of Markov's inequality.

Indeed, under the assumptions that $p(n) \geq 5(\log(n))/n$ and $p \not\to 1$, it is proven in [34, Lemma 4.1] that linearly many edges must be removed to disconnect Γ; thus the bound on the size of cliques, as noted above, implies that there are no separating cliques. It would be interesting to know if this last comment can be improved to hold when $p(n) \geq (1+\epsilon)(\log(n))/n$ and $p \not\to 1$.

Theorem 3.3 *If for some $\epsilon > 0$, we have $1 - p(n) \geq (1+\epsilon)(\log n)/n$, then W_Γ is not thick of order 0, and hence has at least quadratic divergence, aas.*

Proof Let Γ' be the *complement* of Γ, ie the graph with the same vertex set as Γ, but with each pair of vertices adjacent if and only if they are nonadjacent in Γ. Observe that Γ decomposes as a nontrivial join if and only if Γ' is disconnected. Moreover, note that if $\Gamma \in \mathcal{G}(n,p)$, then $\Gamma' \in \mathcal{G}(n, 1-p)$. Hence if $1 - p(n) \geq (1+\epsilon)(\log(n))/n$ for some $\epsilon > 0$, then Γ' is asymptotically almost surely connected; ie Γ is asymptotically almost surely not a nontrivial join for such $p(n)$. In this case, we thus have that W_Γ is not thick of order 0 and hence has superlinear divergence. By [17, Corollary B], since W_Γ acts cocompactly on its Davis complex, it contains a periodic rank-one geodesic, and thus by [40, Proposition 3.3], the divergence of W_Γ is at least quadratic. □

Theorem 3.4 *If $p(n)n \to \infty$ and $p(n)^6 n^5 \to 0$, then the following holds asymptotically almost surely: Γ has a component Γ' such that $W_{\Gamma'}$ is hyperbolic relative to a nonempty collection of proper subgroups each isomorphic to $D_\infty \times D_\infty$. Hence W_Γ is aas hyperbolic relative to a nonempty collection of proper $D_\infty \times D_\infty$ subgroups, at least one of which is not a proper free factor of W_Γ.*

Remark 3.5 Of greatest interest are densities $p(n)$ growing faster than $(\log n)/n$ but slower than $n^{-1/6}$. At such densities, Theorem 3.2 and Theorem 3.4 together ensure that W_Γ is asymptotically almost surely one-ended and hyperbolic relative to $D_\infty \times D_\infty$ subgroups.

Proof of Theorem 3.4 Since $pn \to \infty$, [33] together with [11, Theorem 2.2(ii)] implies that aas Γ has a *giant component* Γ' containing a positive proportion $\alpha \in (0,1)$ of the vertices, and every other component Γ_i has no more than $O(\log n)$ vertices. It suffices to show that, a.a.s, Γ' contains $K_{2,2}$ as an induced proper subgraph and Γ does not contain $K_{2,3}$. Indeed, the second assertion together with Lemma 3.8 implies that every element of $\mathcal{T}$ arising as an induced subgraph of Γ' is isomorphic to $K_{2,2}$. The first assertion, together with Theorem 2.5, will then complete the proof.

$K_{2,3}$ is aas absent Since $p(n)^6 n^5 \to 0$ as $n \to \infty$ by hypothesis, Corollary 5 of [32] implies that, aas, Γ, and therefore Γ', does not contain $K_{2,3}$.

An induced $K_{2,2}$ aas appears in Γ' Let $v_1, \dots, v_4$ be distinct vertices in the random size-n graph Γ, and let the random variable $I(v_1, \dots, v_4)$ take the value 1 or 0 according to whether or not $\{v_1, \dots, v_4\}$ is the vertex set of an induced $K_{2,2}$ in Γ. The random variable $S_n = \sum_{v_1,v_2,v_3,v_4} I(v_1, \dots, v_4)$ counts each induced $K_{2,2}$ in Γ 24 times, reflecting the eight automorphisms of $K_{2,2}$ and the three ways of choosing which pairs of vertices in $K_{2,2}$ will be nonadjacent. Since there are $\binom{n}{4}$ such quadruples, and each forms an induced copy of $K_{2,2}$ exactly when there is some permutation $\sigma\colon \{1,2,3,4\} \to \{1,2,3,4\}$ such that $v_{\sigma(i)}$ is adjacent to $v_{\sigma(i)+1}$ for each i, and the remaining two possible edges are absent, we have $\mathbb{E}(S_4) = 24\binom{n}{4} p^4(1-p)^2$.

Let $N \in \mathbb{N}$ and let $\epsilon \in (0,1)$. The preceding discussion shows that since $p(n)n \to \infty$, there exists $N_1 \in \mathbb{N}$ such that $\mathbb{E}(S_n) \geq N/\epsilon$ for all $n \geq N_1$. The proof of Theorem 4.1 of [18] shows that since $pn \to \infty$ and $(1-p)n^2 \to \infty$,

$$\frac{\mathbb{E}(S_n)^2}{\mathbb{E}(S_n^2)} \to 1,$$

so there exists $N_2 \in \mathbb{N}$ such that

$$\frac{\mathbb{E}(S_n)^2}{\mathbb{E}(S_n^2)} > 1 - \epsilon$$

for $n \geq N_2$. The Paley–Zygmund inequality implies that for all $n \geq \max\{N_1, N_2\}$,

$$\begin{aligned}\mathbb{P}(S_n \geq N) &\geq \mathbb{P}(S_n \geq \epsilon\,\mathbb{E}(S_n)) \\ &\geq (1-\epsilon)^2 \frac{\mathbb{E}(S_n)^2}{\mathbb{E}(S_n^2)} > (1-\epsilon)^3.\end{aligned}$$

This implies that for each $N \in \mathbb{N}$, we have $\lim_n \mathbb{P}(S_n < N) = 0$. Lemma 3.7 below states that aas, every component of Γ is either a tree or equal to Γ', so it suffices to find squares in Γ. We have shown that $\mathbb{P}(S_n < 48) \to 0$ as $n \to \infty$, so Γ' aas contains at least two induced copies of $K_{2,2}$. □

Remark 3.6 The fact that W_Γ is hyperbolic relative to $D_\infty \times D_\infty$ subgroups that are not free factors can be seen slightly more easily as follows. First we produce induced $K_{2,2}$ subgraphs in Γ and verify that Γ aas does not contain $K_{2,3}$, as in the proof of Theorem 3.4. Then we observe that by Theorem 5.16 of [11], Γ aas has no component which is a 4–cycle. Theorem 3.4 is, of course, a stronger conclusion since it rules out the possibility that $W_{\Gamma'}$ is hyperbolic and every 4–cycle lies in a unicyclic component that is not a 4–cycle.

Lemma 3.7 *Let $\Gamma \in G(n, p(n))$, with $p(n)$ satisfying the hypotheses of Theorem 3.4. Asymptotically almost surely, each component of Γ is either the giant component or a tree.*

Proof of Lemma 3.7 This follows immediately from [11, Theorem 6.10(iii)] and [11, Theorem 2.2(ii)]. □

Lemma 3.8 *If $\Lambda \in \mathcal{T}$, then either $\Lambda \cong K_{2,2}$ or Λ contains $K_{2,3}$.*

Proof Since Λ must contain the join of two subgraphs of diameter at least 2, we have that $|\Lambda^0| \geq 4$ and either $\Lambda \cong K_{2,2}$ or $|\Lambda| \geq 5$. In the latter case, suppose that each maximal join in Λ is isomorphic to $K_{2,2}$ and let $\Lambda_0 \subset \Lambda$ be such a join. Then no two nonadjacent vertices in Λ_0 have a common adjacent vertex, since otherwise Λ_0 would extend to a copy of $K_{2,3}$. Hence $\Lambda \cong K_{2,2}$, a contradiction. □

3.2 Behavior at high densities

Charney–Farber showed in [18] that a random right-angled Coxeter group on n vertices is aas finite when $(1 - p(n))n^2 \to 0$ as $n \to \infty$. The following description of random right-angled Coxeter groups for rapidly growing $p(n)$ generalizes this result.

Theorem 3.9 *Suppose $(1 - p(n))n^2 \to \alpha$ as $n \to \infty$ for some $\alpha \in [0, \infty)$, and let the random variable M_n count the number of "missing edges" in $\Gamma \in \mathcal{G}(n, p)$, ie the number of pairs of distinct vertices that are not joined by an edge. Then $M_n = O(1)$ aas, and the following hold:*

(1) *With probability tending to $e^{-\alpha/2}$, $M_n = 0$ and the group W_Γ is finite.*
(2) *With probability tending to $\frac{1}{2}\alpha e^{-\alpha/2}$, $M_n = 1$ and the group W_Γ is virtually $\mathbb{Z}$ and thus hyperbolic.*
(3) *With probability tending to $1 - (1 + \frac{1}{2}\alpha)e^{-\alpha/2}$, $M_n \geq 2$ and the group W_Γ is virtually $\mathbb{Z}^{M_n}$, and is thus thick of order 0 and has linear divergence.*

Proof **Finite and virtually $\mathbb{Z}$** If $M_n = 0$, then Γ is a complete graph, so $W_\Gamma \cong \mathbb{Z}_2^n$ is finite. Conversely, if W_Γ is finite, then since any two nonadjacent vertices together generate a subgroup isomorphic to D_∞, we see that $M_n = 0$. Similarly, W_Γ is virtually $\mathbb{Z}$ if and only if $M_n = 1$.

For $k \geq 0$, we have

$$\mathbb{P}(M_n = k) = \binom{\binom{n}{2}}{k}(1 - p(n))^k p^{\binom{n}{2}-k},$$

and

$$p(n)^{\binom{n}{2}-k} \sim e^{-\alpha/2}.$$

Hence $\mathbb{P}(M_n = 0) \to e^{-\alpha/2}$, while $\mathbb{P}(M_n = 1) \sim \binom{n}{2}(\alpha/n^2)e^{-\alpha/2} \to \frac{1}{2} - \alpha e^{-\alpha/2}$. This establishes the first two assertions.

Thick of order 0 For each vertex $v \in \Gamma$, let I_v be 1 or 0 according to whether or not v belongs to exactly one missing edge, so that $\mathbb{P}(I_v = 1) = \mathbb{E}(I_v) = n(1-p(n))p(n)^{n-2}$. Let $E_n = \sum_v I_v$ count the number of vertices belonging to exactly one missing edge, and observe that $\mathbb{E}(E_n) = n^2(1-p(n))p(n)^{n-2} \sim \alpha$.

Similarly, let J_v be 1 or 0 according to whether or not v belongs to at least one missing edge, and let $F_n = \sum_v J_v$ count the vertices appearing in at least one missing edge. Note that $\mathbb{P}(J_v = 1) = \mathbb{E}(J_v) = 1 - p(n)^{n-1}$. Hence

$$\begin{aligned} \mathbb{E}(F_n) &= n(1-p(n)^{n-1}) \\ &= n\Big[1 - \Big(1 - \frac{\alpha}{n^2}\Big)^{n-1}\Big] \\ &= \frac{\alpha n(n-1)}{n^2} + o(1) \sim \alpha. \end{aligned}$$

Since $F_n \geq E_n$, and $\mathbb{E}(F_n - E_n) \to 0$, aas $F_n = E_n$. In other words, aas every vertex occurs in at most one missing edge. Therefore, aas there are pairwise-distinct vertices $v_1, \dots, v_k, w_1, \dots, w_k$ such that v_i and w_i are not adjacent for all i, and every other pair of vertices spans an edge. This implies that W_Γ is virtually the product of k copies of D_∞.

The above argument shows that aas $M_n = \frac{1}{2}E_n$. For distinct vertices v and w, we have

$$\mathbb{P}(I_v I_w = 1) = (n-1)^2 p^{2n-5}(1-p)^2 + p^{2n-4}(1-p),$$

from which a computation shows that $\mathbb{E}(M_n) \to \frac{1}{8}\alpha(\alpha+1)$. It follows from Markov's inequality that $M_n = O(1)$ aas. □

3.3 Constant-density behavior

In this section, we prove:

Theorem 3.10 *For $\Gamma \in G\big(n, \frac{1}{2}\big)$, the group W_Γ is aas thick.*

The following lemma isolates the most crucial estimates we will use in the proof of the theorem.

Lemma 3.11 *Let $\pi_n = \mathbb{P}\big(\Gamma \notin \mathcal{T} \mid \Gamma \in G\big(n, \frac{1}{2}\big)\big)$. Then the following hold:*

(1) $\pi_{2n} \leq \pi_n^2 + f(n)$, *where* $f(n) = 2n \sum_{i=0}^{n} \binom{n}{i} 2^{-n-\binom{i}{2}}$.

(2) $\pi_{2n} \leq \pi_n^2 + 2\pi_n(1-\pi_n)(nc(n)/2^n t(n)) + (1-\pi_n)^2$, *where $c(n)$ is the number of cliques in the disjoint union of all $\mathcal{T}$–graphs on n vertices (with the 0–clique counted once), and $t(n)$ is the total number of $\mathcal{T}$–graphs on n vertices.*

(3) $\pi_{n+1} \leq \pi_n + f(n)$.

Proof Let $\Gamma \in G\big(2n, \frac{1}{2}\big)$ and let $A \sqcup B$ be a partition of $\Gamma^{(0)}$ into sets of size n. For $v \in B$, we denote by $\mathrm{Link}_A(v)$ the set of vertices in A adjacent to v. Note that if $\Gamma \notin \mathcal{T}$, then one of the following holds:

- (i) The subgraphs generated by A and B are not in $\mathcal{T}$.
- (ii) There exists $v \in B$ [or $v \in A$] such that $\mathrm{Link}_A(v)$ [or $\mathrm{Link}_B(v)$] is a (possibly empty) clique.

To establish this dichotomy, first we assume (i) does not hold, and hence without loss of generality, we may assume the subgraph generated by A is in $\mathcal{T}$. If additionally, (ii) does not hold, we show this yields $\Gamma \in \mathcal{T}$, which is a contradiction. Condition (ii) implies that for each vertex v of B, the set $\mathrm{Link}_A(v)$ is nonempty and has diameter exceeding 1. Now, for each $v \in B$ we have that the subgraph Γ_v of Γ generated by $A \cup \{v\}$ is in $\mathcal{T}$ since it is obtained by coning off a set of diameter at least 2 and applying Definition 2.3(2). Also, for each $v, v' \in B$, since the graphs Γ_v and $\Gamma_{v'}$ are both thick and their intersection is the thick graph generated by A, we see that the graph generated by $A \cup \{v, v'\}$, which is the generalized union of Γ_v and $\Gamma_{v'}$, is thus thick by Definition 2.3(3). Thus, by adding one vertex from B at a time in the above way we see that $\Gamma \in \mathcal{T}$.

Next, we claim that $\mathbb{P}((\mathrm{i})) = \pi_n^2$. Indeed, since in the construction of Γ, edges joining pairs of vertices in A are added independently of those joining vertices in B, the events "A generates a subgraph in $\mathcal{T}$" and "B generates a subgraph in $\mathcal{T}$" are independent. Moreover, the subgraphs of Γ generated by A and B are in $G\big(n, \frac{1}{2}\big)$. It follows that (i) occurs with probability π_n^2, whence

$$\pi_{2n} \le \pi_n^2 + \mathbb{P}((\mathrm{ii})).$$

We finally show that $\mathbb{P}((\mathrm{ii})) \le f(n)$. To this end, let $\mathcal{V}$ be the number of vertices of B whose links in A are (possibly empty) cliques. Then $\mathbb{P}((\mathrm{ii})) \le 2\,\mathbb{P}(\mathcal{V} > 0)$ and $\mathbb{P}(\mathcal{V} > 0) \le \mathbb{E}(\mathcal{V})$. The initial factor of 2 reflects the fact that we are assuming that $A \in \mathcal{T}$ and counting vertices in B whose links in A are cliques; (ii) could just as easily occur with the roles of A and B reversed.

For each $v \in B$, if $\mathrm{Link}_A(v)$ has k vertices, then it is generated by one of $\binom{n}{k}$ subsets of A. Each such subset is a clique with probability $2^{-\binom{k}{2}}$, and such a subset generates $\mathrm{Link}_A(v)$ with probability $2^{-k}2^{k-n} = 2^{-n}$, reflecting the fact that the k vertices of the putative link must be adjacent to v, and the $n-k$ remaining vertices of A must not. Summing over k yields the probability that $\mathrm{Link}_A(v)$ is a clique, so $\mathbb{E}(\mathcal{V}) = n\sum_{k=0}^{n}\binom{n}{k}2^{-n-\binom{k}{2}}$, and (1) follows.

To establish (2), write $\Gamma^{(0)} = A \sqcup B$ as above. If $\Gamma \notin \mathcal{T}$, then one of the following holds:

(a) The subgraphs generated by A and B are both not in $\mathcal{T}$. This event occurs with probability π_n^2.

(b) Exactly one of the subgraphs generated by A and B belongs to $\mathcal{T}$. In this case, suppose that A generates a subgraph in $\mathcal{T}$. This subgraph is among the $t(n)$ graphs of its size in $\mathcal{T}$, and as above, B must contain a vertex v whose link in A generates one of the $c(n)$ possible cliques. There are n choices for this vertex, and each has a given clique as its link with probability at most 2^{-n}. Hence this situation occurs with probability at most $2\pi_n(1-\pi_n)nc(n)2^{-n}t(n)^{-1}$.

(c) The subgraphs generated by A and B both belong to $\mathcal{T}$. In this case, it must be true that some vertex in A has link in B a clique (or vice versa), but we do not use this fact; we just note that the probability of this event is certainly at most $(1-\pi_n)^2$.

Finally, to establish (3), regard the size-$(n+1)$ graph Γ as the subgraph of Γ generated by $A \sqcup \{v\}$, with v a vertex. If $\Gamma \notin \mathcal{T}$, then either $A \notin \mathcal{T}$ or the link of v is a clique. The claim now follows by arguing as in the proof of (1). Note that in this case, since the two parts are not symmetric and we are looking at the link of only one point rather than n, this removes a factor of $2n$ from the second term in the sum, and actually establishes the stronger fact that $\pi_{n+1} \le \pi_n + f(n)/2n$. □

Remark 3.12 The relation between the first two parts of the above lemma are as follows. In the language of conditional probability, to prove Lemma 3.11(1), we use the fact that

$$\pi_{2n} \le \mathbb{P}[A, B \notin \mathcal{T}] + \mathbb{P}[\text{(ii)}].$$

Whereas, for Lemma 3.11(2) we exploited the following:

$$\pi_{2n} \le \mathbb{P}[A, B \notin \mathcal{T}] + 2\,\mathbb{P}[A \in \mathcal{T}, B \notin \mathcal{T}] \cdot \mathbb{P}[\text{(ii)}_B \mid A \in \mathcal{T}, B \notin \mathcal{T}] + \mathbb{P}[A, B \in \mathcal{T}],$$

where (ii)_B is the same as (ii) except that we require only the condition on links of vertices of B. We then sum over these probabilities to yield Lemma 3.11(2).

We will make use of the following estimate:

Lemma 3.13 *Let X_n be a binomial random variable with mean $\frac{1}{2}\cdot n$ and variance $\frac{1}{4}\cdot n$. Then for all $M \le \frac{1}{2}n$, we have*

$$\mathbb{P}(X_n \le M) \le \exp\Big(-\frac{n}{2} + 2M - \frac{2M^2}{n}\Big).$$

Proof Viewing X_n as the sum of n Bernoulli trials, this follows from Hoeffding's inequality [39]. □

Lemma 3.14 *The function f of Lemma 3.11 has the following properties:*

(1) $f(n) \xrightarrow{n} 0$ *exponentially, and in particular,* $\sum_{n\geq 0} f(n) < \infty$.

(2) $f(n) < 0.03760$ *for all* $n \geq 18$.

Proof Let $M = \lfloor n^{a/b} \rfloor$ for natural numbers $a < b$, and define (I) and (II) by writing

$$f(n) = 2n\Bigg[\underbrace{\sum_{i=0}^{M}\binom{n}{i}2^{-n-\binom{i}{2}}}_{\text{(I)}} + \underbrace{\sum_{i=M+1}^{n}\binom{n}{i}2^{-n-\binom{i}{2}}}_{\text{(II)}}\Bigg].$$

For each n,

$$\text{(I)} \leq 2^{-n}\sum_{i=0}^{M}\binom{n}{i} = \mathbb{P}(X_n \leq M),$$

where X_n is a binomial random variable with mean $n\cdot\frac{1}{2}$. From Lemma 3.13, we have, for $M \leq n/2$,

$$\begin{aligned}\text{(I)} &\leq \exp\left[-\frac{n}{2} + 2M - \frac{2M^2}{n}\right] \\ &\leq e^{-n/2}e^{2\lfloor n^{a/b}\rfloor}e^{-2\lfloor n^{a/b}\rfloor^2/n} := g(n,M).\end{aligned}$$

We also have

$$\begin{aligned}\text{(II)} &\leq 2^{-n-\binom{M}{2}}\sum_{i=M+1}^{n}\binom{n}{i} \\ &\leq 2^{-n-\binom{M}{2}}\left(2^n - \sum_{i=0}^{M}\binom{n}{i}\right) \\ &\leq 2^{-\binom{M}{2}} \leq 2^{-n^{a/b}(n^{a/b}-1)/2}.\end{aligned}$$

Suppose now that a and b also satisfy $2a/b > 1$. Then the lemma follows from summing the above estimates: $f(n)$ decays exponentially and is hence summable. This establishes the first assertion.

The second assertion requires a refinement of one of the above bounds. Let $a = 2$ and $b = 3$, and let $M = \lfloor n^{a/b} \rfloor$, X_n and the expressions (I) and (II) be as above. As before, we have

$$\text{(II)} \leq 2^{-n^{2/3}(n^{2/3}-1)/2}.$$

We need to estimate (I) more carefully when $n \geq 18$. We thus write

$$\begin{aligned}(\mathrm{I}) &\leq 2^{-n}\left(\sum_{i=0}^{5}\binom{n}{i}2^{-\binom{i}{2}}\right) + 2^{-\binom{6}{2}}\,\mathbb{P}(X_n \leq \lfloor n^{2/3}\rfloor) \\ &\leq 2^{-n}\left(\sum_{i=0}^{5}\binom{n}{i}2^{-\binom{i}{2}}\right) + 2^{-\binom{6}{2}}g(n, \lfloor n^{2/3}\rfloor) := h(n).\end{aligned}$$

The second inequality is an application of Lemma 3.13, justified by the fact that $n^{2/3} < n/2$ for $n \geq 18$. Hence

$$f(n) \leq 2n \cdot h(n) + 2n \cdot 2^{-n^{2/3}(n^{2/3}-1)/2}.$$

The second term is strictly decreasing for $n \geq 8$, as can be seen by differentiating, and takes a value less than $3.09 \cdot 10^{-5}$ at $n = 18$. Next, a straightforward computation gives

$$g(n, \lfloor n^{2/3}\rfloor) \leq \exp\left(-\frac{n}{2} + 2n^{2/3} - 2n^{1/3} + 4n^{-1/3} - \frac{2}{n}\right),$$

which is decreasing for $n \geq 12$ and, for $n = 18$, yields

$$2n \cdot 2^{-\binom{6}{2}} \cdot g(n, \lfloor n^{2/3}\rfloor) \leq 0.00273.$$

The remaining term can be shown by direct differentiation to decrease for $n \geq 5$, and takes the value 0.3484 at $n = 18$. Combining the above shows that $f(n) \leq 3.09 \cdot 10^{-5} + 0.00273 + 0.03484 = 0.03760$ for $n \geq 18$. □

Remark 3.15 As we will see in the proof of Theorem 3.10, any bound sharper than around $f(18) \leq 0.06045$ is sufficient.

Proof of Theorem 3.10 The idea of the proof is to use Lemma 3.11(1) and the fact that f is small to get convergence to 0 of a subsequence of (π_n). We then use this in order to show that (π_n) converges to 0, and then we apply Lemma 3.11(3) and the summability of f.

Accumulation at 0 implies convergence to 0 For each n and k, Lemma 3.11(3) yields

$$\pi_{n+k} \leq \pi_n + \sum_{i=0}^{k-1} f(i+n) < \pi_n + \sum_{i=n}^{\infty} f(i).$$

Suppose that 0 is an accumulation point of (π_n). Then for each $\epsilon > 0$, we can choose n so that $\pi_n < \frac{1}{2}\epsilon$ and $\sum_{i=n}^{\infty} f(n) < \frac{1}{2}\epsilon$. The latter inequality follows from summability of f, ie from Lemma 3.14(1). Hence for all k, we have $\pi_{n+k} < \epsilon$, ie $\pi_n \xrightarrow{n} 0$.

Nonaccumulation at 0 implies convergence to 1 Suppose now that the subsequence $(\pi_{k\cdot 2^m})_{m\in\mathbb{N}}$ does not have 0 as an accumulation point for some $k\in\mathbb{N}$. Then we claim that $(\pi_{k\cdot 2^m})$ converges to 1. Indeed, consider the smallest accumulation point π of the sequence, and suppose that it is the limit of the subsequence $(\pi_{k\cdot 2^{m_i}})_{i\in\mathbb{N}}$. We have to show $\pi=1$. By Lemma 3.11(1) and the fact that f converges to 0, we get that any accumulation point π' of $(\pi_{k\cdot 2^{m_i+1}})$ satisfies $\pi'\le\pi^2$. As we also have $\pi\le\pi'$, we get $\pi\le\pi^2$, so that $\pi=1$.

A subsequence bounded away from 1 It is thus sufficient to show that the subsequence $(\pi_{k\cdot 2^m})_{m\in\mathbb{N}}$ is bounded away from 1 for some $k\in\mathbb{N}$. In fact, if this is the case, then $(\pi_{k\cdot 2^m})_{m\in\mathbb{N}}$ does not converge to 1, hence it must have 0 as an accumulation point, and hence (π_n) converges to 0 as required. Suppose that for some k, we have $m_0\in\mathbb{N}$ and constants $\alpha,\beta\in[0,1)$ such that $f(k\cdot 2^m)\le\beta$ for all $m\ge m_0$, and $\pi_{k\cdot 2^{m_0}}\le\alpha$. Suppose, moreover, that $\alpha^2+\beta<\alpha$. Then $\pi_{k\cdot 2^{m_0+1}}<\alpha$ by Lemma 3.11(1), and by induction and the same lemma, we have $\pi_{k\cdot 2^m}<\alpha$ for all $m\ge m_0$.

Let $k=9$ and $m_0=1$. The computer program in the online supplement returned the following data:

- $t(9)=14853635863$,
- $c(9)=683846354560$,
- $\pi_9=1-t(9)/2^{\binom{9}{2}}\approx 0.78385$.

Together with Lemma 3.11(2), this implies

$$\pi_{18}\le\alpha:=\Bigl(1-\frac{t(9)}{2^{36}}\Bigr)^2+\Bigl(\frac{t(9)}{2^{36}}\Bigr)^2+2\Bigl(1-\frac{t(9)}{2^{36}}\Bigr)\cdot\frac{t(9)}{2^{36}}\cdot\frac{9\cdot c(9)}{512\cdot t(9)}\approx 0.93537.$$

Lemma 3.14(2) gives $f(n)\le\beta=0.03760$ for all $n\ge 18$. The above discussion, together with the fact that these values satisfy $\alpha^2+\beta<\alpha$, implies that $(\pi_{9\cdot 2^m})$ is bounded away from 1, whence $\pi_n\xrightarrow{n}0$; ie Γ is aas in $\mathcal{T}$. □

4 Detecting thickness algorithmically

In this section, we exhibit a polynomial-time algorithm for deciding whether a finite graph is in $\mathcal{T}$. The construction of the algorithm presented in this section prioritized simplicity over speed. We also provide a C++ implementation of a simple algorithm to compute the constants needed in the proof of Theorem 3.10. The main part of this computer program implements the algorithm for deciding if a given right-angled Coxeter group is thick.

Theorem 4.1 *There exists an algorithm which decides, in polynomial time, whether a graph Γ is in $\mathcal{T}$. Hence the problem of deciding whether a right-angled Coxeter group admits a relatively hyperbolic structure is soluble in polynomial time.*

Proof The second assertion follows from the first by Theorem 2.5. The algorithm takes as input the finite simplicial graph Γ on n vertices and decides whether $\Gamma \in \mathcal{T}$. For ease of exposition, we provide an algorithm which admits an easy description, but we note that there are more efficient algorithms; in particular, the code in the online supplement contains an implementation of a more efficient algorithm for the same task. The steps are:

(1) Make a list $\mathcal{M}$ of all induced $K_{2,2}$ subgraphs of Γ. The running time is in $O(n^4)$ and $|\mathcal{M}|$ is in $O(n^4)$.

(2) Make a list $\mathcal{N}$ of pairs of nonadjacent vertices. The running time is in $O(n^2)$ and $|\mathcal{N}|$ is in $O(n^2)$.

(3) Perform a *union subroutine*; ie for each pair $M, M' \in \mathcal{M}$, determine whether $M \cap M'$ contains some $(v, v') \in \mathcal{N}$. If so, modify $\mathcal{M}$ by removing M and M', and adding the subgraph induced by $M \cup M'$. The running time of a union subroutine is in $O(n^{11})$.

(4) Perform a *coning subroutine*; ie for each $M \in \mathcal{M}$ and each vertex v, determine whether there exists $(w, w') \in \mathcal{N}$ such that $w, w' \in M$ and both are adjacent to v. If so, replace M by the subgraph generated by $M \cup \{v\}$. The running time of a coning subroutine is in $O(n^7)$.

(5) If $\mathcal{M}$ did not change during the coning and union subroutines, then we are finished: the graph is thick if and only if $|\mathcal{M}| = 1$, and the unique element of $\mathcal{M}$ is Γ.

(6) If $\mathcal{M}$ changed, then return to step (2).

The number of union subroutines that modify $\mathcal{M}$ is in $O(n^4)$ since each such union subroutine decreases $|\mathcal{M}|$. The number of coning subroutines that modify $\mathcal{M}$ is in $O(n^5)$ since each such subroutine increases the size of some subgraph in $\mathcal{M}$. Hence the total running time is in $O(n^{15})$. □

4.1 Computing $t(9)$ and $c(9)$

To obtain the values used in the proof of Theorem 3.10, one can use the C++ program in the online supplement, which takes a single command line argument, namely the number n of vertices. We have also checked the computations by hand up to $n = 6$

beyond which they become infeasible. The reader seeking to reproduce our computer computation for $n = 9$ should be aware that the program requires being run for several days with typical 2013 hardware.

The efficiency of the program can be significantly improved. However, we decided to keep the code as simple as possible. Source code for a much more efficient, albeit more complex, version of this program can be obtained from the authors.

Appendix: Generalizing to all Coxeter groups
by J Behrstock, P-E Caprace, M F Hagen and A Sisto

All Coxeter groups considered here are assumed finitely generated. In this appendix, we generalize Theorems I and II to Coxeter groups which are not necessarily right angled. Further considerations are contained in Section A.3.

We can summarize the main result in this appendix as follows.

Theorem A.1 (minimal relatively hyperbolic structures) *Let (W, S) be a Coxeter system. Then there is a (possibly empty) collection $\mathcal{J}$ of subsets of S enjoying the following properties:*

(i) *The parabolic subgroup W_J is strongly algebraically thick for every $J \in \mathcal{J}$.*
(ii) *If $J \neq S$ for all $J \in \mathcal{J}$, then W is hyperbolic relative to $\mathcal{P} = \{W_J \mid J \in \mathcal{J}\}$.*

In particular, $\mathcal{P}$ is a minimal relatively hyperbolic structure for W.

A.1 Thick Coxeter groups

We consider the class $\mathbb{T}$ of Coxeter systems (W, S) defined as follows.

(1) $\mathbb{T}$ contains the class $\mathbb{T}_0$ of all irreducible *affine* Coxeter systems (W, S) with S of cardinality at least 3, as well as all Coxeter systems of the form $(W, S_1 \cup S_2)$ with W_{S_1} and W_{S_2} irreducible nonspherical and $[W_{S_1}, W_{S_2}] = 1$.
(2) Suppose that $(W, S \cup s)$ is such that $s^\perp$ is nonspherical and (W_S, S) belongs to $\mathbb{T}$. Then $(W, S \cup s)$ belongs to $\mathbb{T}$.
(3) Suppose that (W, S) is such that there exist $S_1, S_2 \subseteq S$ with $S_1 \cup S_2 = S$, $(W_{S_1}, S_1), (W_{S_2}, S_2) \in \mathbb{T}$ and $W_{S_1 \cap S_2}$ nonspherical. Then $(W, S) \in \mathbb{T}$.

Proposition A.2 *For $(W, S) \in \mathbb{T}$, the Coxeter group W is strongly algebraically thick.*

The proof requires the following subsidiary fact.

Lemma A.3 *Let (W, S) be a Coxeter system. Let $s \in S$ and set $K = S \setminus \{s\}$. Then the group $\langle W_K \cup sW_K s\rangle$ has index at most 2 in W.*

Proof The group $\langle W_K \cup sW_K s\rangle$ is a reflection subgroup whose fundamental domain for its action on the Cayley graph of (W, S) contains at most two chambers, namely the base vertex 1 and the unique vertex s–adjacent to it, see [26]. □

Proof of Proposition A.2 If (W, S) is in $\mathbb{T}_0$ then the group W is either virtually abelian of rank at least 2 or a direct product of two infinite (Coxeter) groups. In particular, W is wide and, hence, strongly algebraically thick of order 0.

Let $(W, S \cup \{s\})$ be of the form described in item (2) of the definition of $\mathbb{T}$. Lemma A.3 then implies that W contains the group $\langle W_S \cup sW_S s\rangle$ with index at most 2. Therefore W is strongly algebraically thick, being an algebraic network with respect to the pair of strongly thick groups $\{W_S, sW_S s\}$.

Finally, let (W, S) be as in item (3) of the definition of $\mathbb{T}$. Then W is strongly algebraically thick, being an algebraic network with respect to the pair of strongly thick groups $\{W_{S_1}, W_{S_2}\}$. □

A.2 Proof of minimal relatively hyperbolic structures theorem

We will use the following criterion for relative hyperbolicity of Coxeter groups, which corrects [14, Theorem A], where a hypothesis on the peripheral subgroups was missing.

Theorem A.4 [15, Theorem A′] *Let (W, S) be a Coxeter system and $\mathcal{J}$ a collection of proper subsets of S. Then W is hyperbolic relative to $\{W_J \mid J \in \mathcal{J}\}$ if and only if the following conditions hold:*

(RH1) *For each irreducible affine subset $K \subseteq S$ of cardinality at least 3, there exists $J \in \mathcal{J}$ such that $K \subseteq J$. Similarly, given any pair of irreducible nonspherical subsets $K_1, K_2 \subseteq S$ with $[K_1, K_2] = 1$, there exists $J \in \mathcal{J}$ such that $K_1 \cup K_2 \subseteq J$.*

(RH2) *For all $J_1, J_2 \in \mathcal{J}$ with $J_1 \neq J_2$, the intersection $J_1 \cap J_2$ is spherical.*

(RH3) *For each $J \in \mathcal{J}$ and each irreducible nonspherical $K \subseteq J$, we have $K^\perp \subseteq J$.*

We are now ready to prove Theorem A.1. We will give an explicit description of $\mathcal{J}$:

Theorem A.5 *Let (W, S) be a Coxeter system and let $\mathcal{J}$ be the (possibly empty) collection of all* maximal *subsets $J \subseteq S$ such that $(W_J, J) \in \mathbb{T}$. Then we have:*

(i) *The parabolic subgroup W_J is strongly algebraically thick for every $J \in \mathcal{J}$.*

(ii) *If $\mathcal{J} \neq \{S\}$, then W is hyperbolic relative to $\mathcal{P} = \{W_J \mid J \in \mathcal{J}\}$.*

In particular, $\mathcal{P}$ is a minimal relatively hyperbolic structure for W.

Proof By Moussong's characterization of hyperbolic Coxeter groups [41, Theorem 17.1] (and the fact that S is finite), $\mathcal{J}$ is not empty if and only if W is not hyperbolic, which we assume from now on.

By Proposition A.2, (i) holds.

We are now left to show that $\mathcal{J}$ satisfies the three conditions (RH1)–(RH3) from Theorem A.4.

It is clear that $\mathcal{J}$ satisfies (RH1).

If $J_1, J_2 \in \mathcal{J}$ are distinct, then $W_{J_1 \cap J_2}$ must be spherical. In fact, if it was nonspherical, then we would have $J_1 \cup J_2 \in \mathcal{J}$, contradicting the maximality of either J_1 or J_2. So $\mathcal{J}$ satisfies (RH2).

Let K be a nonspherical subgraph of some $J \in \mathcal{J}$. We have to show that $K^\perp$ is contained in J as well. Indeed, if there was an element $s \in K^\perp \backslash J$, then $J \cup \{s\}$ would be in $\mathbb{T}$, contradicting the maximality of J.

We have now shown the peripherals are in $\mathbb{T}$ and hence thick by Proposition A.2. Thus, as noted in the introduction, minimality now follows from [5, Corollary 4.7]. $\square$

A.3 Intrinsic horosphericity and further corollaries

We say that a discrete group Γ is *(intrinsically) horospherical* if every proper isometric action of Γ on a proper hyperbolic geodesic metric space fixes a unique point at infinity. In particular, the group Γ cannot be virtually cyclic, and every element of infinite order acts as a parabolic isometry in any such Γ–action. As one may expect, thickness and horosphericity are related properties (compare Theorem 4.1 from [5]):

Proposition A.6 *Every strongly algebraically thick group is intrinsically horospherical.*

The proof requires the following result, which follows from the exact same arguments as the proof of Lemma 3.25 in [28].

Lemma A.7 *Let H be a finitely generated group (endowed with its word metric with respect to a finite generating set), (X, d) a metric space and $\mathfrak{q}\colon H \to X$ a map which is Lipschitz up to an additive constant. Given $h \in H$, if the map $\mathbb{Z} \to X, n \mapsto \mathfrak{q}(h^n)$ is a Morse quasigeodesic in X, then h is a Morse element in H.* $\square$

Lemma A.8 *Let H be a group acting properly by isometries on a proper Gromov hyperbolic metric space X. Assume that H has a unique fixed point ξ at infinity of X. Then every infinite subgroup of H has ξ as its unique fixed point at infinity.*

Proof The hypotheses imply that H does not contain any hyperbolic isometry. From Proposition 5.5 in [16], it follows that every subgroup of H either has a bounded orbit or has a unique fixed point at infinity of X. The desired conclusion follows since the H–action on X is proper. □

Proof of Proposition A.6 We argue by induction on the order of thickness. In the base case, let H be a finitely generated group which is wide. Suppose that H acts properly by isometries on a proper Gromov hyperbolic metric space X. H can not contain a hyperbolic isometry since otherwise, Lemma A.7 implies that some asymptotic cone of H has cut-points, which would contradict the assumption that H is wide. Since H is infinite and the H–action on X is proper, it follows from [16, Proposition 5.5] that H fixes a unique point at infinity of X. This proves that strongly algebraically thick groups of order 0 are intrinsically horospherical.

The inductive step is given by the following observation. Let G be an infinite group which is an M–algebraic network with respect to a finite collection $\mathcal{H}$ of subgroups. If each subgroup in $\mathcal{H}$ is intrinsically horospherical, then so is G.

Indeed, let G act properly by isometries on a proper Gromov hyperbolic metric space X. Then each group $H \in \mathcal{H}$ has a unique fixed point ξ_H at infinity of X. Given $H, H' \in \mathcal{H}$, there is a sequence $H = H_1, \dots, H_N = H'$ in $\mathcal{H}$ in which any two consecutive groups have an infinite intersection; see Definition 5.2 in [5]. From Lemma A.8, we deduce that $\xi_H = \xi_{H_1} = \cdots = \xi_{H_n} = \xi_{H'}$. Hence all groups in $\mathcal{H}$ have the same fixed point at infinity, say ξ. By the definition of an algebraic network, this point ξ must be fixed by a finite-index subgroup of G. Thus the G–orbit of ξ is finite.

If this orbit has exactly one point, then G fixes ξ (and no other point at infinity of X), and we are done. If this orbit contains exactly two points, then G is virtually cyclic and hence does not contain any intrinsically horospherical subgroups, which is absurd. If $|G\xi| \geq 3$, then it follows from [38, Proposition-Definition 8.2.L] that G has bounded orbits in X, contradicting the assumption that G is infinite and acts properly. □

Notice that the converse to Proposition A.6 does not hold in general: indeed, horospherical groups include all amenable groups that are not virtually cyclic. In particular, infinite locally finite groups are examples of horospherical groups that are not strongly algebraically thick. By Zorn's lemma, every intrinsically horospherical subgroup of Γ is contained in a maximal one. It is thus a natural question to determine all the maximal intrinsically horospherical subgroups. Theorem A.1 yields the answer to this question when Γ is a Coxeter group.

Corollary A.9 *Let W be a Coxeter group. Then the maximal intrinsically horospherical subgroups of W are parabolic subgroups (in the sense of Coxeter group theory) with respect to any Coxeter generating set. Those parabolic subgroups are precisely the conjugates of the elements of the set $\mathcal{P}$ afforded by Theorem A.1.*

Proof Every strongly algebraically thick group is intrinsically horospherical by Proposition A.6. Moreover, a subgroup of W properly containing a conjugate of an element of $\mathcal{P}$ cannot be intrinsically horospherical by Theorem A.1. Thus the elements of $\mathcal{P}$ are indeed maximal horospherical subgroups. Since W is relatively hyperbolic with respect to $\mathcal{P}$, every intrinsically horospherical subgroup is conjugate to a subgroup of an element of $\mathcal{P}$. □

Corollary A.10 *Let (W, S) be a Coxeter system. Then the following conditions are equivalent:*

(i) *(W, S) is in $\mathbb{T}$.*

(ii) *W is strongly algebraically thick.*

(iii) *W is intrinsically horospherical.*

(iv) *W is not relatively hyperbolic with respect to any family of proper subgroups.*

(v) *W is not relatively hyperbolic with respect to any family of proper Coxeter-parabolic subgroups.*

(vi) *For every collection $\mathcal{J}$ of subsets of S satisfying* (RH1)–(RH3), *we have $S \in \mathcal{J}$.*

Proof The implication (i) $\implies$ (ii) is the content of Proposition A.2. The implication (ii) $\implies$ (iii) follows from Proposition A.6. The implication (iii) $\implies$ (iv) is straightforward. Property (iv) trivially implies (v). That (v) is equivalent to (vi) follows from Theorem A.4. Applying Theorem A.5, we get that (v) implies (i). □

References

[1] **G Arzhantseva**, **A Minasyan**, **D Osin**, *The SQ-universality and residual properties of relatively hyperbolic groups*, J. Algebra 315 (2007) 165–177 MR

[2] **J A Behrstock**, *Asymptotic geometry of the mapping class group and Teichmüller space*, Geom. Topol. 10 (2006) 1523–1578 MR

[3] **J Behrstock**, **R Charney**, *Divergence and quasimorphisms of right-angled Artin groups*, Math. Ann. 352 (2012) 339–356 MR

[4] **J Behrstock**, **C Druţu**, *Divergence, thick groups, and short conjugators*, Illinois J. Math. 58 (2014) 939–980 MR

[5] **J Behrstock**, **C Druţu**, **L Mosher**, *Thick metric spaces, relative hyperbolicity, and quasi-isometric rigidity*, Math. Ann. 344 (2009) 543–595 MR

[6] **J Behrstock**, **V Falgas-Ravry**, **M F Hagen**, **T Susse**, *Global structural properties of random graphs*, Int. Math. Res. Not. (online publication December 2016)

[7] **J Behrstock**, **M F Hagen**, *Cubulated groups: thickness, relative hyperbolicity, and simplicial boundaries*, Groups Geom. Dyn. 10 (2016) 649–707 MR

[8] **J A Behrstock**, **T Januszkiewicz**, **W D Neumann**, *Quasi-isometric classification of some high dimensional right-angled Artin groups*, Groups Geom. Dyn. 4 (2010) 681–692 MR

[9] **J A Behrstock**, **W D Neumann**, *Quasi-isometric classification of graph manifold groups*, Duke Math. J. 141 (2008) 217–240 MR

[10] **M Bestvina**, **B Kleiner**, **M Sageev**, *The asymptotic geometry of right-angled Artin groups, I*, Geom. Topol. 12 (2008) 1653–1699 MR

[11] **B Bollobás**, *Random graphs*, 2nd edition, Cambridge Studies in Advanced Mathematics 73, Cambridge University Press (2001) MR

[12] **B H Bowditch**, *Relatively hyperbolic groups*, Internat. J. Algebra Comput. 22 (2012) 1250016, 1–66 MR

[13] **J Brock**, **H Masur**, *Coarse and synthetic Weil–Petersson geometry: quasi-flats, geodesics and relative hyperbolicity*, Geom. Topol. 12 (2008) 2453–2495 MR

[14] **P-E Caprace**, *Buildings with isolated subspaces and relatively hyperbolic Coxeter groups*, Innov. Incidence Geom. 10 (2009) 15–31 MR

[15] **P-E Caprace**, *Erratum to* [14], Innov. Incidence Geom. 14 (2015) 77–79 MR

[16] **P-E Caprace**, **K Fujiwara**, *Rank-one isometries of buildings and quasi-morphisms of Kac–Moody groups*, Geom. Funct. Anal. 19 (2010) 1296–1319 MR

[17] **P-E Caprace**, **M Sageev**, *Rank rigidity for* $\mathrm{CAT}(0)$ *cube complexes*, Geom. Funct. Anal. 21 (2011) 851–891 MR

[18] **R Charney**, **M Farber**, *Random groups arising as graph products*, Algebr. Geom. Topol. 12 (2012) 979–995 MR

[19] **F Chung**, *Random graphs, a whirlwind tour of*, from "Computational complexity: Theory, techniques, and applications" (R A Meyers, editor), Springer, New York (2012) 2536–2548 MR

[20] **J Crisp**, **L Paoluzzi**, *Commensurability classification of a family of right-angled Coxeter groups*, Proc. Amer. Math. Soc. 136 (2008) 2343–2349 MR

[21] **F Dahmani**, *Les groupes relativement hyperboliques et leurs bords*, PhD thesis, Université Louis Pasteur (2003) MR

[22] **P Dani**, **A Thomas**, *Divergence in right-angled Coxeter groups*, Trans. Amer. Math. Soc. 367 (2015) 3549–3577 MR

[23] **M W Davis**, *The geometry and topology of Coxeter groups*, London Math. Soc. Monogr. Ser. 32, Princeton University Press (2008) MR

[24] **M W Davis**, **T Januszkiewicz**, *Right-angled Artin groups are commensurable with right-angled Coxeter groups*, J. Pure Appl. Algebra 153 (2000) 229–235 MR

[25] **M W Davis**, **M Kahle**, *Random graph products of finite groups are rational duality groups*, J. Topol. 7 (2014) 589–606 MR

[26] **V V Deodhar**, *A note on subgroups generated by reflections in Coxeter groups*, Arch. Math. (Basel) 53 (1989) 543–546 MR

[27] **C Druţu**, *Relatively hyperbolic groups: geometry and quasi-isometric invariance*, Comment. Math. Helv. 84 (2009) 503–546 MR

[28] **C Druţu**, **S Mozes**, **M Sapir**, *Divergence in lattices in semisimple Lie groups and graphs of groups*, Trans. Amer. Math. Soc. 362 (2010) 2451–2505 MR

[29] **C Druţu**, **M Sapir**, *Tree-graded spaces and asymptotic cones of groups*, Topology 44 (2005) 959–1058 MR

[30] **C Druţu**, **M V Sapir**, *Groups acting on tree-graded spaces and splittings of relatively hyperbolic groups*, Adv. Math. 217 (2008) 1313–1367 MR

[31] **P Erdős**, **A Rényi**, *On random graphs, I*, Publ. Math. Debrecen 6 (1959) 290–297 MR

[32] **P Erdős**, **A Rényi**, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. 5 (1960) 17–61 MR

[33] **P Erdős**, **A Rényi**, *On the evolution of random graphs*, Bull. Inst. Internat. Statist. 38 (1961) 343–347 MR

[34] **V Falgas-Ravry**, **T Kittipassorn**, **D Korándi**, **S Letzter**, **B P Narayanan**, *Separating path systems*, J. Comb. 5 (2014) 335–354 MR

[35] **B Farb**, *Relatively hyperbolic groups*, Geom. Funct. Anal. 8 (1998) 810–840 MR

[36] **S M Gersten**, *Quadratic divergence of geodesics in* CAT(0) *spaces*, Geom. Funct. Anal. 4 (1994) 37–51 MR

[37] **E N Gilbert**, *Random graphs*, Ann. Math. Statist. 30 (1959) 1141–1144 MR

[38] **M Gromov**, *Hyperbolic groups*, from "Essays in group theory" (S M Gersten, editor), Math. Sci. Res. Inst. Publ. 8, Springer, New York (1987) 75–263 MR

[39] **W Hoeffding**, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. 58 (1963) 13–30 MR

[40] **M Kapovich**, **B Leeb**, 3*–manifold groups and nonpositive curvature*, Geom. Funct. Anal. 8 (1998) 841–852 MR

[41] **G Moussong**, *Hyperbolic Coxeter groups*, PhD thesis, The Ohio State University (1988) MR Available at `http://search.proquest.com/docview/303686972`

[42] **B Mühlherr**, *Automorphisms of graph-universal Coxeter groups*, J. Algebra 200 (1998) 629–649 MR

[43] **G A Niblo**, **L D Reeves**, *The geometry of cube complexes and the complexity of their fundamental groups*, Topology 37 (1998) 621–633 MR

[44] **D V Osin**, *Relatively hyperbolic groups: intrinsic geometry, algebraic properties, and algorithmic problems*, Mem. Amer. Math. Soc. 843, Amer. Math. Soc., Providence, RI (2006) MR

[45] **A Sisto**, *On metric relative hyperbolicity*, preprint (2012) arXiv

[46] **A Sisto**, *Projections and relative hyperbolicity*, Enseign. Math. 59 (2013) 165–181 MR

[47] **H Sultan**, *The asymptotic cone of Teichmüller space and thickness*, Algebr. Geom. Topol. 15 (2015) 3071–3106 MR

[48] **A Yaman**, *A topological characterisation of relatively hyperbolic groups*, J. Reine Angew. Math. 566 (2004) 41–89 MR

The Graduate Center and Lehman College, CUNY
New York, NY 10016, United States

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge
Wilberforce Rd., Cambridge, CB3 0WB, United Kingdom

Departement Mathematik HG G 28, ETH
Rämistrasse 101, CH-8092 Zürich, Switzerland

Departement de Mathematiques, Universite Catholique de Louvain - IRMP
Chemin du Cyclotron 2, boîte L7.01.02, 1348 Louvain-la-Neuve, Belgium

jason@math.columbia.edu, mfh26@cam.ac.uk, sisto@math.ethz.ch, pe.caprace@uclouvain.be

Received: 23 March 2015 Revised: 5 October 2016

Algebraic & Geometric Topology 17 (2017) 741–763

On RO(G)–graded equivariant "ordinary" cohomology where G is a power of $\mathbb{Z}/2$

JOHN HOLLER
IGOR KRIZ

We compute the complete RO(G)–graded coefficients of "ordinary" cohomology with coefficients in $\mathbb{Z}/2$ for $G = (\mathbb{Z}/2)^n$. As an important intermediate step, we identify the ring of coefficients of the corresponding geometric fixed point spectrum, revealing some interesting algebra. This is a first computation of its kind for groups which are not cyclic p–groups.

55N91

1 Introduction

The notion of a cohomology theory graded by elements of the real representation ring (RO(G)–graded cohomology) is a key concept of equivariant stable homotopy theory of a finite or compact Lie group G. Like much of stable homotopy theory, perhaps one of the first known examples was K-theory. Atiyah and Singer [4] introduced equivariant K-theory of a compact Lie group G and proved that it is naturally RO(G)–graded. In fact, Bott periodicity identifies many of the "dimensions" in RO(G), and relates others to "twistings" (see Karoubi [7] and, for a more recent treatment, Freed, Hopkins and Teleman [9]). Pioneered by Adams and Greenlees [10], the general RO(G)–graded stable homotopy theory found firm foundations in the fundamental book of Lewis, May and Steinberger [22].

Despite the clear importance of the concept, beyond K-theory, calculations of RO(G)–graded cohomology are few and far in between. Perhaps the most striking case is "ordinary" RO(G)–graded cohomology. Bredon [5] discovered $\mathbb{Z}$–graded G–equivariant cohomology associated with a *coefficient system* which is "ordinary" in the sense that the cohomology of a point is concentrated in a single dimension. It was later discovered (Lewis, May and McClure [20]) that such a theory becomes RO(G)–graded when the coefficient system enjoys the structure of a *Mackey functor* (see Dress [8]), which means that it allows building in an appropriate concept of *transfer*. Strikingly, the RO(G)–graded coefficients were not known in any single nontrivial case.

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.741

Complete calculations of $\mathrm{RO}(\mathbb{Z}/2)$–graded coefficients, however, are important in Real-oriented stable homotopy theory, because they exhibit the analogy with the complex-oriented case. Real orientation was, once again, discovered first by Atiyah [3] in the case of K-theory, and was subsequently extended to cobordism by Landweber [19]. $\mathrm{RO}(\mathbb{Z}/2)$–graded cohomology with coefficients in the Burnside ring Mackey functor was calculated by Stong [21]. A systematic pursuit of real-oriented homotopy theory was started by Araki [2], and developed further by Hu and Kriz [13] with many calculations, including a complete calculation of the $\mathrm{RO}(G)$–graded coefficients of Landweber's Real cobordism spectrum. In the process, Hu and Kriz [13] also calculated the $\mathrm{RO}(\mathbb{Z}/2)$–graded ordinary cohomology of the "constant" Mackey functors $\mathbb{Z}$ and $\mathbb{Z}/2$ (ie the Mackey functors uniquely extending the constant coefficient systems). A major development was the work of Hill, Hopkins and Ravenel [12], who partially extended the calculations of [13] to $\mathbb{Z}/(2^k)$ (with special interest in $k=3$), and applied this to solving the Kervaire–Milnor problem by showing the nonexistence of manifolds of Kervaire invariant 1 in dimensions > 126. A still more complete calculation of $\mathrm{RO}(G)$–graded ordinary equivariant cohomology of the constant Mackey functors for $G=\mathbb{Z}/(2^k)$ was more recently given in Hu and Kriz [14].

Still, no calculations of $\mathrm{RO}(G)$–graded cohomology beyond K-theory were known for groups other than where G is a primary cyclic group. In a spin-off of their joint solution with Ormsby [16] of Thomason's homotopy limit problem for Hermitian K-theory, Hu and Kriz [15] computed the $\mathrm{RO}(G)$–graded coefficients of topological Hermitian cobordism, which has $G=\mathbb{Z}/2\times\mathbb{Z}/2$. However, this is a rather special case, where many periodicities occur.

The purpose of the present paper is to calculate the $\mathrm{RO}(G)$–graded coefficients of the ordinary equivariant cohomology of the "constant" $\mathbb{Z}/2$ Mackey functor for $G=(\mathbb{Z}/2)^n$. There are several reasons to focus on this case. The group $(\mathbb{Z}/2)^n$ has an exceptionally simply described real representation ring, thus eliminating the need to handle representation-theoretical exceptions such as distinguishing between real and complex (let alone, quaternionic) representations. The coefficients $\mathbb{Z}/2$ are more convenient than $\mathbb{Z}$, since they eliminate the need to consider extensions. Despite all this, the complete answer is complicated, however, and in general, we are only able to present it in the form of the cohomology of an n–stage chain complex.

Our method is based on *isotropy separation*, a term coined by Greenlees and May [11], to mean considering separately the contributions of subgroups of G. An isotropy separation spectral sequence was developed in Abram and Kriz [1], but we use a different spectral sequence here. The reason is that in [1], we are not concerned with $\mathrm{RO}(G)$–graded coefficients, but rather with computing the complete $\mathbb{Z}$–graded coefficients of equivariant complex cobordism of a finite abelian group G as a ring.

Based on generalizing the method of Kriz [17] in the case of $G = \mathbb{Z}/p$, in the case of $\mathbb{Z}$–graded equivariant complex cobordism, one can set up a spectral sequence of rings which collapses to E^2 in a single filtration degree. This means that the complete ring structure can be recovered, which is a special property of complex cobordism. It is worth mentioning that the spectral sequence of [1] contains many "completed" (for example, uncountable) terms.

The case of ordinary RO(G)–graded equivariant cohomology is quite different, however, in that the spectral sequence fails to collapse to a single degree. Even for $G = \mathbb{Z}/p$, we observe that a part of the coefficients are in filtration degree 0 and a part in filtration degree -1 (graded homologically). This caused us to give up, at least for now, calculating the complete ring structure, and use a spectral sequence which is more amenable to calculations instead.

Another key ingredient in our computation is the concept of *geometric fixed points* of an RO(G)–graded equivariant cohomology theory. This concept was introduced (using a different terminology) by tom Dieck [6], who calculated the geometric fixed points of equivariant complex cobordism. As far as we know, the term geometric fixed points was coined by Greenlees and May, and is recorded in Lewis, May, Steinberger and McClure [22]. Unlike actual fixed points, the geometric fixed point coefficients are *periodic* with respect to all nontrivial irreducible real representations of G. Thus, instead of RO(G)–graded, the geometric fixed points are, again, only $\mathbb{Z}$–graded. This is a big advantage in expressing the answer. Note that the ring RO$((\mathbb{Z}/2)^n)$ is huge: it is the free abelian group on 2^n generators! On the downside, the term "geometric" fails to carry the expected implications in the case of ordinary equivariant cohomology: we know of no geometry that would help calculating them. Still, in the case $G = (\mathbb{Z}/2)^n$, a complete calculation of the geometric fixed point ring of $H\mathbb{Z}/2$ is possible using spectral sequence methods. This is our Theorem 2.

The main method of this paper is, basically, setting up another spectral sequence which enables the calculation of the coefficients of $H\mathbb{Z}/2_{(\mathbb{Z}/2)^n}$ by investigating how they differ from the coefficients of the geometric fixed points. There results a spectral sequence, which, in a fairly substantial range of RO(G)–graded dimensions, collapses to E^2 in degree 0. More precisely, the range is, graded homologically, suspensions by elements of RO(G) where summands of nontrivial irreducible representations occur with *nonpositive* coefficients. Alternately, graded cohomologically, this is the range of suspensions by actual representations, possibly minus a trivial representation. (As it turns out, however, in this case, when the trivial representation has a negative coefficient, the cohomology group is 0.) In this case, we can both recover the complete ring structure, since the ring embeds into the ring of geometric fixed points tensored with RO(G). We also have a nice concise formula for the Poincaré series in this case (Theorem 5).

In the case of completely general $\mathrm{RO}(G)$–dimension with $G=(\mathbb{Z}/2)^n$, we are only able to give a spectral sequence in n filtration degrees, which collapses to E^2 and calculates the $\mathrm{RO}(G)$–graded coefficient group of $H\mathbb{Z}/2_G$. Thus, this gives an algebraically defined chain complex whose homology are the desired groups (Theorem 7). We give an example of a complete calculation of the Poincaré series of the $\mathrm{RO}(G)$–graded coefficients of $H\mathbb{Z}/2_{\mathbb{Z}/2\times\mathbb{Z}/2}$ (the case $n=2$), which clearly shows that the answer gets complicated, and additional complications arise for $n\geq 3$.

The present paper is organized as follows: In Section 2, we introduce the necessary conventions and notation. In Section 3, we compute the geometric fixed points. In Section 4, we compute the coefficients in dimensions involving elements of $\mathrm{RO}(G)$ where nontrivial irreducible representations have nonpositive coefficients (graded homologically). In Section 5, we calculate the chain complex computing the complete $\mathrm{RO}(G)$–graded coefficients of $H\mathbb{Z}/2_G$ for $G=(\mathbb{Z}/2)^n$. In Section 6, we treat the example of $n=2$. The authors apologize to the readers for not stating their theorems in the introduction. Even in the prettiest cases, the theorems involve quite a lot of notation and technical prerequisites. We prefer to state them properly in the text.

Recent developments: odd primes, and hyperplane arrangements While this paper was under review, several developments took place. A generalization of the present result to $(\mathbb{Z}/p)^n$ for p an odd prime was found by Holler. The authors also found out that the ring described in Theorem 2 is a previously known object in algebraic geometry, related to a certain compactification of complements of hyperplane arrangements referred to as *the reciprocal plane*.

More concretely, for a set $S=\{z_\alpha\}$ of equations of hyperplanes through 0 in an affine space $\mathrm{Spec}(F[u_1,\dots,u_n])$ of a field F, one considers the subring R_S of

$$\left(\prod_{\alpha\in S} z_\alpha\right)^{-1} F[u_1,\dots,u_n] \tag{1}$$

generated by the elements z_α^{-1} (which correspond to our elements x_α). The ring was first described by Terao [24], and a particularly nice presentation was found by Proudfoot and Speyer [23]. In the case of an odd prime p, one deals analogously with the subring Ξ_S of

$$\left(\prod_{\alpha\in S} z_\alpha\right)^{-1} F[u_1,\dots,u_n]\otimes_F \Lambda(du_1,\dots,du_n) \tag{2}$$

generated by z_α^{-1} and $d\log(z_\alpha)$, which are topologically in dimensions 2 and 1, respectively. The analogues of the constructions of [23; 24] in this graded-commutative case, and the reciprocal plane compactification, were recently worked out by S Kriz [18].

Our emphasis is quite different form the authors of [23; 24], who, doing classical algebraic geometry, were mostly interested in characteristic 0. Their arguments, however, work in general. The ring described in Theorem 2 (and its $\mathbb{Z}/p$ analogue discovered by Holler, ie the geometric fixed point ring of $H\mathbb{Z}/p_G$ where $G=(\mathbb{Z}/p)^n$) is related to the hyperplane arrangement of *all* hyperplanes through 0 in the n–dimensional affine space over $\mathbb{Z}/p$. It follows, however, from the descriptions of [23; 24; 18] that for a subset S' of a hyperplane arrangement S, the ring $R_{S'}$ (resp. $\Xi_{S'}$) is a subring of R_S (resp. Ξ_S). It follows in turn that for *every* hyperplane arrangement in $G=(\mathbb{Z}/p)^n$, the $\mathbb{Z}$–graded part of the coefficient ring of the spectrum

$$\bigwedge_{\alpha\in S} S^{\infty\alpha}\wedge H\mathbb{Z}/p_G$$

is R_S for $p=2$, and Ξ_S for any odd prime p.

2 Conventions and notation

Throughout this paper, let $G=(\mathbb{Z}/2)^n$. Then the real representation ring of G is canonically identified as

$$\mathrm{RO}(G)=\mathbb{Z}[G^*],$$

where $G^*=\mathrm{Hom}(G,\mathbb{Z}/2)$. Recall [22] that for $H\subseteq G$, we have the family $\mathcal{F}[H]$ consisting of all subgroups $K\subset G$ with $H\not\subseteq K$. (In the case of $H=G$, we see that $\mathcal{F}[G]$ is simply the family $\mathcal{P}$ of proper subgroups of G.) Recall further that for any family $\mathcal{F}$ (a set of subgroups of G closed under subconjugation, which is the same as closed under subgroups, as G is commutative), we have a cofibration sequence

$$E\mathcal{F}_+\to S^0\to\widetilde{E\mathcal{F}},$$

where $E\mathcal{F}$ is a G–CW-complex whose K–fixed point set is contractible when $K\in\mathcal{F}$ and empty otherwise. For our choice of G, we may then choose a model

$$\widetilde{E\mathcal{F}[H]}=\bigwedge_{\substack{\alpha\in G^*\\ \alpha|H\neq 0}} S^{\infty\alpha}. \tag{3}$$

Here $S^{\infty\alpha}$ is the direct limit of $S^{n\alpha}$ with respect to the inclusions

$$S^0\to S^\alpha \tag{4}$$

given by sending the non-basepoint to 0. The other construction we use is the family $\mathcal{F}(H)$ of all subgroups of a subgroup $H\subseteq G$. We will write simply

$$EG/H=E\mathcal{F}(H).$$

The cardinality of a finite set S will be denoted by $|S|$. We will also adopt a convention from [13] where, for an $\mathrm{RO}(G)$–graded spectrum E, the $\mathbb{Z}$–indexed coefficients (= homotopy groups) of E are denoted by E_*, while the $\mathrm{RO}(G)$–indexed coefficients will be denoted by $E_\star$. As is customary, we will also denote by $S(V)$ the unit sphere of a representation V, while by S^V we denote the 1–point compactification of V. The $\mathrm{RO}(G)$–graded dimension of a homogeneous element $x \in E_\star$ will be denoted by $|x|$.

3 The geometric fixed points

In this section, we compute the coefficients of the geometric fixed point spectrum $\Phi^G H\mathbb{Z}/2$. We have

$$\Phi^G H\mathbb{Z}/2 = (\widetilde{E\mathcal{F}[G]} \wedge H\mathbb{Z}/2)^G. \tag{5}$$

By (3), suspension of $H\mathbb{Z}/2$ by any nontrivial irreducible real representation of G gives an isomorphism on coefficients, so the coefficients $\left(\Phi^G \Sigma^? H\mathbb{Z}/2\right)_*$ are only $\mathbb{Z}$–graded, not $\mathrm{RO}(G)$–graded. More specifically, we have a cofibration sequence

$$EG/\mathrm{Ker}(\alpha)_+ \to S^0 \to S^{\infty\alpha}, \tag{6}$$

so smashing over all nontrivial 1–dimensional representations α, using (3), we may represent

$$\widetilde{E\mathcal{F}[G]} \wedge H\mathbb{Z}/2$$

as the iterated cofiber of a (2^n-1)–dimensional cube of the form

$$H\mathbb{Z}/2 \wedge \bigwedge_{0\neq\alpha\in G^*} (EG/\mathrm{Ker}(\alpha)_+ \to S^0). \tag{7}$$

Taking coefficients in (7) then gives a spectral sequence converging to $\Phi^G H\mathbb{Z}/2_*$. Now also note that

$$EG/H_1 \times \cdots \times EG/H_k \simeq EG/(H_1 \cap \cdots \cap H_k). \tag{8}$$

From this, we can calculate the spectral sequence associated with the iterated cofiber of the cube (7). Let us grade the spectral sequence homologically, so the term $H\mathbb{Z}/2_*$, which equals $\mathbb{Z}/2$, is in $E^1_{0,0}$. The rest of the E^1–term is then given as

$$E^1_{p,*} = \bigoplus_{S\in\mathcal{S}_p} \mathrm{Sym}_{\mathbb{Z}/2}\big(\big(G/\textstyle\bigcap\{\mathrm{Ker}(\alpha) \mid \alpha \in S\}\big)^*\big) \cdot y_S, \tag{9}$$

where $\mathcal{S}_p$ is the set of all subsets of $G^* \smallsetminus \{0\}$ of cardinality p. (The last factor y_S of (9) is only a generator written to distinguish the summands.) Now the E^2–term can also be calculated using the following:

Lemma 1 *Consider the differential* ∂ *on*

$$Q_n = \mathbb{Z}/2\{y_S \mid S \subseteq (\mathbb{Z}/2)^n \smallsetminus \{0\}\}$$

given by

$$\partial(y_S) = \sum_{\substack{s \in S \\ \langle S \smallsetminus \{s\} \rangle = \langle S \rangle}} y_{S \smallsetminus \{s\}}. \tag{10}$$

Then the homology is the $\mathbb{Z}/2$*–vector space (freely) generated by a set* F_n *described inductively as follows:*

$$F_1 = \{y_\varnothing, y_{\{(1)\}}\},$$
$$F_n = F_{n-1} \cup \{y_{S \cup \{x\}} \mid S \in F_{n-1}, x \in (\mathbb{Z}/2)^{n-1} \times \{1\}\}.$$

In other words, F_n *consists of the basis elements* y_S *where* S *are all the* $\mathbb{Z}/2$*–linearly independent (in* G^**) subsets in (not necessarily reduced) row echelon form with respect to reversed order of columns (so the first pivot is in the last possible column etc).*

Proof Consider a differential on Q_n given by

$$d(y_S) = \sum_{s \in S} y_{S \smallsetminus \{s\}}. \tag{11}$$

Then the homology is 0 for $n > 0$ and $\mathbb{Z}/2$ for $n = 0$. Now consider an increasing filtration on Q_n by making the filtration degree $\gamma(S)$ of a basis element y_S equal to $\operatorname{rank}\langle S \rangle$, the rank of the $\mathbb{Z}/2$–vector space generated by S. Then the E^1–term is what we are trying to calculate.

On the other hand, in the answer $C = \mathbb{Z}/2(F_n)$ suggested in the statement of the lemma (which, note, consists of elements of E^1), the formula for d^1 is the same as the formula (11) for d. We claim that

$$H_*(C, d) = 0. \tag{12}$$

To see this, note that for any fixed nonempty set S in row echelon form, the subcomplex C_S generated by $y_{S'}$ subsets of $S' \subseteq S$ is just a tensor product of copies of

$$\mathbb{Z}/2 \xrightarrow{\cong} \mathbb{Z}/2, \tag{13}$$

and hence satisfies

$$H_*(C_S, d) = 0.$$

On the other hand, C for $n > 0$ is a sum of the complexes C_S where S ranges over *maximal* linearly independent subsets of $(\mathbb{Z}/2)^n$ in row echelon form (ie those which have exactly n elements), while the intersection of any subset of those complexes

$$C_{S_1\cap\cdots\cap S_k} = C_{S_1}\cap\cdots\cap C_{S_k}$$

has zero homology because

$$(1,0,\ldots,0)\in S_1\cap\cdots\cap S_k$$

and hence $S_1\cap\cdots\cap S_k\neq\varnothing$. This implies (12).

Now the statement follows by induction on n using comparison theorems for spectral sequences. More concretely, if we denote by $C'\subset C$ the subcomplex generated by linearly independent subsets S with $|S|<n$, and $Q'\subset Q_n$ the subcomplex generated by sets S which span a subspace of dimension $<n$, then the induction hypothesis (given that an intersection of vector subspaces is a vector subspace), shows that the embedding $C\subset Q_n$ restricts to a quasi-isomorphism

$$C'\subset Q'. \tag{14}$$

Since the homologies of both C and Q_n are 0, we see that the homomorphism on degree n subcomplexes must induce an isomorphism on homology, thus implying that the degree n part of the E^1–term of our spectral sequence for Q_n is just the degree n part of C (which is, of course, isomorphic to $\mathbb{Z}/2$). □

Now by Lemma 1, the E^2–term of the spectral sequence of the cube (7) is

$$E^2=\bigoplus_{S\in F_n}\mathrm{Sym}_{\mathbb{Z}/2}\big(\big(G/\textstyle\bigcap\{\mathrm{Ker}(\alpha)\mid\alpha\in S\}\big)^*\big)\cdot y_S \tag{15}$$

(where we make the identification $G^*\cong(\mathbb{Z}/2)^n$).

Now consider, for $0\neq\alpha\colon G\to\mathbb{Z}/2$, the map

$$f_\alpha\colon \Phi^{G/\mathrm{Ker}(\alpha)}H\mathbb{Z}/2_*=\Phi^{G/\mathrm{Ker}(\alpha)}(H\mathbb{Z}/2)_*^{\mathrm{Ker}(\alpha)}\to\Phi^G H\mathbb{Z}/2_*. \tag{16}$$

It is fairly obvious that for $n=1$ the spectral sequence associated with the (1–dimensional) cube (7) collapses to E^1 and that in fact

$$\Phi^{G/\mathrm{Ker}\,\alpha}H\mathbb{Z}/2_*=\mathbb{Z}/2[x_\alpha], \tag{17}$$

where in the spectral sequence, the element x_α is filtration degree 1 and is represented by the set $\{(1)\}$ if we make the identification $G/\mathrm{Ker}(\alpha)\cong\mathbb{Z}/2$. We will also denote the image under (16)

$$f_\alpha(x_\alpha)\in\Phi^G H\mathbb{Z}/2$$

by x_α.

Theorem 2 *We have*

(18) $\Phi^G_* H\mathbb{Z}/2 = \mathbb{Z}/2[x_\alpha \mid \alpha \in G^* \smallsetminus \{0\}] \,/ (x_\alpha x_\beta + x_\alpha x_\gamma + x_\beta x_\gamma \mid \alpha+\beta+\gamma = 0),$

where the classes x_α *are in dimension* 1*.*

Before proving the theorem, it is useful to record the following algebraic fact:

Proposition 3 *Let* $\{\alpha_1, \dots, \alpha_k\}$ *be a minimal* $\mathbb{Z}/2$*–linearly dependent subset of* $G^* \smallsetminus \{0\}$*, where* $k \geq 3$*. Then the ring* R_n *on the right-hand side of* (18) *satisfies*

$$\sigma_{k-1}(x_{\alpha_1}, \dots, x_{\alpha_k}) = 0. \tag{19}$$

(Here σ_i *denotes the* i^{th} *elementary symmetric polynomial.)*

Proof We will proceed by induction on k. For $k=3$, this is by definition. Suppose $k>3$ and suppose the statement is true with k replaced by $k-1$. Compute in R_n, where we denote $\beta = \alpha_{k-1} + \alpha_k$:

$$\begin{aligned} \sigma_{k-1}&(x_{\alpha_1}, \dots, x_{\alpha_k}) \\ &= (x_{\alpha_k} + x_{\alpha_{k-1}})(x_{\alpha_1} \cdots x_{\alpha_{k-2}}) + x_{\alpha_k} x_{\alpha_{k-1}} \sigma_{k-3}(x_{\alpha_1}, \dots, x_{\alpha_{k-2}}) \\ &= (x_{\alpha_k} + x_{\alpha_{k-1}})(x_{\alpha_1} \cdots x_{\alpha_{k-2}}) + (x_{\alpha_k} + x_{\alpha_{k-1}}) x_\beta \sigma_{k-3}(x_{\alpha_1}, \dots, x_{\alpha_{k-2}}) \\ &= (x_{\alpha_k} + x_{\alpha_{k-1}}) \sigma_{k-2}(x_\beta, x_{\alpha_1}, \dots, x_{\alpha_{k-2}}). \end{aligned} \tag{20}$$

Now $\{\beta, \alpha_1, \dots, \alpha_{k-2}\}$ is also a minimal linearly dependent set (note that minimality is equivalent to the statement that $\alpha_1, \dots, \alpha_{k-1}$ are linearly independent and $\alpha_1 + \cdots + \alpha_k = 0$). Therefore, the right-hand side of (20) is 0 in R_n by the induction hypothesis. □

Proof of Theorem 2 We know that $\Phi^G_* H\mathbb{Z}/2$ is a ring, since $\Phi^G H\mathbb{Z}/2$ is an E_∞–ring spectrum. By (16), we know that the elements x_α represent elements of $\Phi^G_* H\mathbb{Z}/2$, and hence polynomials in the elements x_α do as well. Now it is important to note that (15) is not a spectral sequence of rings. However, there are maps arising from smashing n cubes (7) (over $H\mathbb{Z}/2$) for $n=1$, and from this, it is not difficult to deduce that for S linearly independent, a monomial of the form

$$\prod_{s \in S} x_s^{r_s} \quad \text{where } r_s \geq 1 \tag{21}$$

is represented in (15) by

$$S \cdot \prod_{s \in S} x_s^{r_s - 1}. \tag{22}$$

(Note that by Lemma 1, for S not linearly independent, (22) does not survive to E^2.)

By Lemma 1, we know that such elements generate the E^2–term as a $\mathbb{Z}/2$–module, so we have already proved that the spectral sequence associated with the cube (7) collapses to E^2.

Now counting basis elements in filtration degree 2 shows that $\Phi^G_* H\mathbb{Z}/2$ must have a quadratic relation among the elements x_α, x_β, x_γ when

$$\alpha+\beta+\gamma=0.$$

(It suffices to consider $n=2$.) The relation must be symmetric and homogeneous for reasons of dimensions, so the possible candidates for the relation are

$$x_\alpha x_\beta + x_\alpha x_\gamma + x_\beta x_\gamma = 0 \tag{23}$$

or

$$x_\alpha x_\beta + x_\alpha x_\gamma + x_\beta x_\gamma + x_\alpha^2 + x_\beta^2 + x_\gamma^2 = 0. \tag{24}$$

We will prove the theorem by finding a basis of the monomials (21) of the ring on the right-hand side of (18) and matching them, in the form (22), with the E^2–term (15).

Before determining which of the relations (23), (24) is correct, we observe (by induction) that the ring R_n given by relation (23) satisfies (with the identification $G^* \cong (\mathbb{Z}/2)^n$)

$$R_n = R_{n-1} \otimes \mathbb{Z}/2[x_{(0,\dots,0,1)}] + \sum_{\alpha\in((\mathbb{Z}/2)^{n-1}\smallsetminus\{0\})\times\{1\}} R_{n-1}\otimes x_\alpha \cdot \mathbb{Z}/2[x_\alpha] \tag{25}$$

and that the ring R'_n obtained from the relations (24) satisfies a completely analogous statement with R_i replaced by R'_i. By the identification between (21) and (22), we see that we obtain a $\mathbb{Z}/2$–module of the same rank as the E^2–term of the spectral sequence of (7) in each dimension if and only if the sum (25) for each n is a direct sum (and similarly for the case of R'_n). Since we already know that the spectral sequence collapses to E_2, we know that this direct sum must occur for whichever relation (23) or (24) is correct, and also that the "winning" relation (23) (resp. (24)), ranging over all applicable choices of α, β and γ generates all the relations in $\Phi^G_* H\mathbb{Z}/2$.

We will complete the proof by showing that (24) generates a spurious relation, and hence is eliminated. This cannot be done for $n=2$, as we actually have $R_2 \cong R'_2$ via the (nonfunctorial isomorphism) replacing the generators x_α, x_β and x_γ with $x_\alpha + x_\beta$, $x_\alpha + x_\gamma$ and $x_\beta + x_\gamma$.

We therefore must resort to $n=3$. Let $\alpha_1=(1,0,0)$, $\alpha_2=(0,1,0)$, $\alpha_3=(0,0,1)$, $\alpha_4=(1,1,1)$. Applying the computation (20) in the proof of Proposition 3 to compute $\sigma_3(x_{\alpha_1},x_{\alpha_2},x_{\alpha_3},x_{\alpha_4})$ in the ring R'_3, we obtain

$$\sigma_3(x_{\alpha_1},x_{\alpha_2},x_{\alpha_3},x_{\alpha_4}) = (x_{\alpha_1}+x_{\alpha_2})(x_{\alpha_3}^2+x_{\alpha_4}^2+x_\beta^2)+(x_{\alpha_3}+x_{\alpha_4})(x_{\alpha_1}^2+x_{\alpha_2}^2+x_\beta^2).$$

As this is clearly not symmetrical in x_{α_1}, x_{α_2}, x_{α_3}, x_{α_4}, by permuting (say, using a 4–cycle) and adding both relations, we obtain a spurious relation in dimension 3 and filtration degree 2, which shows that the analogue of (25) with R_i replaced by R_i' fails to be a direct sum for $n = 3$, thereby excluding the relation (24), and completing the proof. □

From the fact that (25) is a direct sum, we obtain the following:

Corollary 4 *The Poincaré series of the ring* R_n *is*

$$\frac{1}{(1-x)^n}\prod_{i=1}^{n}(1+(2^{i-1}-1)x). \qquad \square$$

4 The coefficients of $H\mathbb{Z}/2$ suspended by a G–representation

In this section, we will compute explicitly the coefficients of $H\mathbb{Z}/2$ suspended by

$$V = \sum_{\alpha\in G^*\smallsetminus\{0\}} m_\alpha \alpha \tag{26}$$

with $m_\alpha \geq 0$.

Theorem 5 (i) *For* $m_\alpha \geq 0$ *and* $G^* \cong (\mathbb{Z}/2)^n$, *recalling* (26), *the Poincaré series of*

$$\Sigma^V H\mathbb{Z}/2_*$$

is

$$\frac{1}{(1-x)^n}\Bigg(\sum_{(\mathbb{Z}/2)^k\cong H\subseteq G^*}(-1)^k\Bigg(\prod_{i=1}^{n-k}(1+(2^{i-1}-1)x)\Bigg)x^l\Bigg) \tag{27}$$

where

$$l = k + \sum_{\alpha\in H\smallsetminus\{0\}} m_\alpha.$$

(ii) *For* $m_\alpha \geq 0$, *the canonical map*

$$\Sigma^V H\mathbb{Z}/2 \to \widetilde{E\mathcal{F}[G]} \wedge H\mathbb{Z}/2$$

(given by the smash product of the inclusions $S^{m_\alpha\alpha} \to S^{\infty\alpha}$*) induces an injective map on* $\mathbb{Z}$*–graded homotopy groups.*

We need the following purely combinatorial result. Let

$$\begin{bmatrix} n \\ k \end{bmatrix} = \frac{(2^n-1)\cdot(2^{n-1}-1)\cdots(2^{n-k+1}-1)}{(2^k-1)\cdot(2^{k-1}-1)\cdots(2^1-1)}.$$

Note that this is the number of k–dimensional $\mathbb{Z}/2$–vector subspaces of $(\mathbb{Z}/2)^n$. The following statement amounts to part (i) of Theorem 5 for $m_\alpha = 0$.

Lemma 6 *We have*

$$\sum_{k=0}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} x^k \prod_{i=1}^{n-k} (1 + (2^{i-1} - 1)x) = (1-x)^n.$$

Proof Induction on n. We have

$$\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} n-1 \\ k \end{bmatrix} + 2^{n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix},$$

so by the induction hypothesis,

$$\sum_{k=0}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} x^k \prod_{i=1}^{n-k} (1 + (2^{i-1} - 1)x)$$
$$= \sum_{k=0}^{n} (-1)^k \left(\begin{bmatrix} n-1 \\ k \end{bmatrix} + 2^{n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} \right) x^k \prod_{i=1}^{n-k} (1 + (2^{i-1} - 1)x).$$

Splitting the right-hand side into two sums, we get

$$\sum_{k=0}^{n-1} (-1)^k \begin{bmatrix} n-1 \\ k \end{bmatrix} x^k \prod_{i=1}^{n-k} (1 + (2^{i-1} - 1)x)$$
$$+ \sum_{k=1}^{n} (-1)^k 2^{n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} x^k \prod_{i=1}^{n-k} (1 + (2^{i-1} - 1)x)$$
$$= (1-x)^n + \sum_{k=0}^{n-1} (-1)^k \begin{bmatrix} n-1 \\ k \end{bmatrix} x^k \prod_{i=1}^{n-k-1} (1 + (2^{i-1} - 1)x) 2^{n-k-1}$$
$$+ \sum_{k=1}^{n} (-1)^k 2^{n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} x^k \prod_{i=1}^{n-k} (1 + (2^{i-1} - 1)x)$$
$$= (1-x)^n. \qquad \square$$

Proof of Theorem 5 We will proceed by induction on n. Assume (i) and (ii) are true for lower values of n. Then, for the given n, we proceed by induction on

$$\ell = |\{\alpha \in G^* \mid m_\alpha > 0\}|.$$

For $\ell=0$, (i) follows from Lemma 6 and (ii) is obvious (by ring structure of $\Phi^G H\mathbb{Z}/2$). Suppose $\ell \geq 1$ and (i), (ii) are true for lower values of ℓ. Setting

(28) $$V_\ell = \sum_{i=1}^{\ell} m_{\alpha_i}\alpha_i,$$

we will study the effect on coefficients $(?)_*$ of the cofibration sequence

(29) $$S(m_\ell\alpha_\ell)_+ \wedge \Sigma^{V_{\ell-1}} H\mathbb{Z}/2 \longrightarrow \Sigma^{V_{\ell-1}} H\mathbb{Z}/2 \longrightarrow \Sigma^{V_\ell} H\mathbb{Z}/2.$$

First, we observed that the first map factors through the top row of the diagram:

(30) $$\begin{tikzcd} (EG/\mathrm{Ker}\,\alpha_\ell)_+ \wedge \Sigma^{V_{\ell-1}} H\mathbb{Z}/2 \arrow[r] \arrow[d] & \Sigma^{V_{\ell-1}} H\mathbb{Z}/2 \arrow[d] \\ (EG/\mathrm{Ker}\,\alpha_\ell)_+ \wedge \widetilde{E\mathcal{F}[G]} \wedge H\mathbb{Z}/2 \arrow[r] & \widetilde{E\mathcal{F}[G]} \wedge H\mathbb{Z}/2 \end{tikzcd}$$

Next, the right column of (30) is injective on $(?)_*$ by (ii) for $\ell-1$, and hence the top row, and hence also the first map (29), is 0 on $(?)_*$.

Now the Poincaré series of

(31) $$(S(m_\ell\alpha_\ell)_+ \wedge \Sigma^{V_{\ell-1}} H\mathbb{Z}/2)_*^G$$

is

$$\frac{1-x^{m_\ell}}{1-x}$$

times the Poincaré series of

(32) $$(\Sigma^{V_{\ell-1}} H\mathbb{Z}/2)_*^{\mathrm{Ker}\,\alpha_\ell},$$

which, when multiplied by x and added to the Poincaré series of

$$(\Sigma^{V_{\ell-1}} H\mathbb{Z}/2)_*^G,$$

is (27) by the induction hypothesis. This proves (i).

To prove (ii), we observe that the elements of (31) are generated by powers of x_{α_ℓ} multiplied by elements of (32), so again, we are done by the induction hypothesis. □

5 The complex calculating RO(G)–graded coefficients

To calculate the RO(G)–graded coefficients of $H\mathbb{Z}/2_G$ in dimensions given by virtual representations, we introduce another spectral sequence. In fact, we will again use the cofibration sequence (6), but we will rewrite it as

(33) $$S^0 \to S^{\infty\alpha} \to \Sigma EG/\mathrm{Ker}(\alpha)_+.$$

We will smash the second maps of (33) over all $\alpha \in G^* \smallsetminus \{0\}$, to obtain a cube

$$\bigwedge_{\alpha \in G^* \smallsetminus \{0\}} (S^{\infty\alpha} \to \Sigma EG/\mathrm{Ker}(\alpha)_+) \tag{34}$$

whose iterated fiber is S^0. Our method is to smash with $H\mathbb{Z}/2_G$ and take $\mathrm{RO}(G)$–graded coefficients to obtain

$$\Big(\bigwedge_{\alpha \in G^* \smallsetminus \{0\}} (S^{\infty\alpha} \to \Sigma EG/\mathrm{Ker}(\alpha)_+) \wedge H\mathbb{Z}/2 \Big)_\star, \tag{35}$$

thus yielding a spectral sequence calculating $H\mathbb{Z}/2_\star$.

However, there is a key point to notice which drastically simplifies this calculation. Namely, smashing (6) with $EG/\mathrm{Ker}(\alpha)_+$, the first morphism becomes an equivalence, thus showing that

$$EG/\mathrm{Ker}(\alpha)_+ \wedge S^{\infty\alpha} \simeq *. \tag{36}$$

Together with (8), this shows that the only vertices of the cube (34) which are nonzero are actually those of the form where all the elements α for which we take the term $S^{\infty\alpha}$ in (34) are those *not vanishing* on some subgroup $A \subseteq G$, while those elements α for which we take the term $\Sigma EG/\mathrm{Ker}(\alpha)_+$ are those nonzero elements of G^* which *do vanish* on A, ie nonzero elements of $(G/A)^*$. The corresponding vertex of (34) is then a suspension of

$$\mathrm{gr}_A(S^0) = EG/A_+ \wedge \widetilde{E\mathcal{F}[A]}. \tag{37}$$

We also put

$$\mathrm{gr}_A(H\mathbb{Z}/2) = \mathrm{gr}_A(S^0) \wedge H\mathbb{Z}/2.$$

Because of the high number of zero terms, the spectral sequence may be regraded by $\mathrm{rank}_{\mathbb{Z}/2}(A)$, thus having only n, instead of $2^n - 1$, filtration degrees. (Note that the cube (34) may be reinterpreted as a "filtration" of the spectrum S^0; from this point of view, we have simply observed that many of the filtered parts coincide.)

It is now important, however, to discuss the grading seriously. Since we index coefficients homologically, we will write the spectral sequence in homological indexing. Additionally, we want the term $\mathrm{gr}_G(S^0)$ be in filtration degree 0 (since that is where the unit is). Thus, the (homologically indexed) filtration degree of (37) is

$$p = \mathrm{rank}(A) - n$$

(a nonpositive number). Thus,

$$\pi_k\Big(\Sigma^{\sum_{\alpha \in G^* \smallsetminus \{0\}} m_\alpha \alpha} \mathrm{gr}_A(H\mathbb{Z}/2)\Big) \subseteq E^1_{\mathrm{rank}(A)-n,\,k+n-\mathrm{rank}(A)},$$

or, put differently, for a given choice of the elements m_α,

$$E^1_{p,q} = \bigoplus_{\mathrm{rank}(A)=n+p} \pi_{q+p-\sum m_\alpha \alpha}\, \mathrm{gr}_A(H\mathbb{Z}/2) \quad \text{for } p=-n,\dots,0. \tag{38}$$

We will next describe explicitly the differential

$$d^1\colon E^1_{p,q} \to E^1_{p-1,q}. \tag{39}$$

Let us first introduce some notation. To this end, we need to start out by describing the E^1–term more explicitly.

In effect, we can calculate $\mathrm{gr}_A(H\mathbb{Z}/2)_\star$ by taking first the A–fixed points using Theorem 2 with G replaced by A, and then applying the Borel homology spectral sequence for G/A. This spectral sequence collapses because there exists a splitting:

$$\begin{tikzcd} A \arrow[r, "\subseteq"] \arrow[dr, "="'] & G \arrow[d, "r"] \\ & A \end{tikzcd} \tag{40}$$

However, the splitting is not canonical, and this is reflected by the choice of generators we observe. More explicitly, the splitting determines for each representation

$$0 \neq \beta\colon A \to \mathbb{Z}/2$$

an extension

$$\widetilde{\beta}\colon G \to \mathbb{Z}/2.$$

One difficulty with describing Borel homology is that it does not naturally form a ring. Because of that, it is more convenient to describe first the coefficients of

$$\gamma_A(H\mathbb{Z}/2) := F(EG/A_+, \widetilde{E\mathcal{F}[A]}) \wedge H\mathbb{Z}/2. \tag{41}$$

This is an (E_∞–) ring spectrum, and its ring of coefficients is given by

$$\gamma_A(H\mathbb{Z}/2)_\star = M[(y_\alpha u_\alpha^{-1})^{\pm 1}][\![y_\alpha \mid \alpha \in (G/A)^* \smallsetminus \{0\}]\!]/(y_{\alpha+\alpha'} - y_\alpha - y_{\alpha'}) \tag{42}$$

where

$$M = \mathbb{Z}/2[x_{\widetilde{\beta}}, u_{\widetilde{\beta}}^{\pm 1}, u_{\widetilde{\beta}+\alpha}^{\pm 1} \mid \beta \in A^* \smallsetminus \{0\}, \alpha \in (G/A)^* \smallsetminus \{0\}]$$
$$/(x_{\widetilde{\alpha}} x_{\widetilde{\beta}} + x_{\widetilde{\alpha}} x_{\widetilde{\gamma}} + x_{\widetilde{\beta}} x_{\widetilde{\gamma}} \mid \alpha+\beta+\gamma = 0)$$

and the RO(G)–graded dimensions of the generators are

$$|u_\gamma| = -\gamma, \quad |x_\gamma| = 1 \quad \text{and} \quad |y_\gamma| = -1.$$

We may then describe $\mathrm{gr}_A(H\mathbb{Z}/2)_\star$ as the $(\dim_{\mathbb{Z}/2}(G/A))^{\mathrm{th}}$ (= only nontrivial) local cohomology module of the ring $\gamma_A(H\mathbb{Z}/2)$ with respect to the ideal generated by the elements y_α. Note that after taking A–fixed points first, this is the usual computation of G/A–Borel homology from the corresponding Borel cohomology. Recall that $H_I^*(R)$ for a finitely generated ideal I of a commutative ring R is obtained by choosing finitely many generators $y_1,\dots,y_\ell$ of I, tensoring, over R, the cochain complexes

$$R \to y_i^{-1}R$$

(with R in degree 0) and taking cohomology. It is, canonically, independent of the choice of generators. In the present case, we are simply dealing with the power series ring R in $\dim_{\mathbb{Z}/2}(G/A)$ generators over a $\mathbb{Z}/2$–algebra, and the augmentation ideal. Taking the defining generators of the power series ring, we see immediately that only the top local cohomology group survives.

We note that the basic philosophy of our notation is

$$\text{“}y_\alpha = x_\alpha^{-1}\text{”}. \tag{43}$$

As a first demonstration of this philosophy, let us investigate the effect of a change of the splitting (40). Writing metaphorically

$$x_{\widetilde{\beta}+\alpha}x_{\widetilde{\beta}} + x_{\widetilde{\beta}+\alpha}x_\alpha + x_{\widetilde{\beta}}x_\alpha = 0, \tag{44}$$

we get

$$x_{\widetilde{\beta}+\alpha}x_{\widetilde{\beta}}y_\alpha + x_{\widetilde{\beta}+\alpha} + x_{\widetilde{\beta}} = 0, \tag{45}$$

from which we calculate

$$x_{\widetilde{\beta}+\alpha} = x_{\widetilde{\beta}}(1+x_{\widetilde{\beta}}y_\alpha)^{-1} = \sum_{k=0}^{\infty} x_{\widetilde{\beta}}^{k+1}y_\alpha^k. \tag{46}$$

This formula is correct in $\gamma_A(H\mathbb{Z}/2)_\star$ and hence can also be used in the module $\mathrm{gr}_A(H\mathbb{Z}/2)_\star$.

Next, we will describe the differential d^1 of (38). These connecting maps will be the sums of maps of degree -1 of the form

$$d^{AB}\colon \mathrm{gr}_A(H\mathbb{Z}/2)_\star \to \mathrm{gr}_B(H\mathbb{Z}/2)_\star, \tag{47}$$

where $B \subset A$ is a subgroup with quotient isomorphic to $\mathbb{Z}/2$. Let $\beta\colon A \to \mathbb{Z}/2$ be the unique nontrivial representation which vanishes when restricted to A. The key point is to observe that the canonical map

$$EG/A_+ \wedge \widetilde{E\mathcal{F}[B]} \wedge S^{\infty\widetilde{\beta}} \xrightarrow{\sim} EG/A_+ \wedge \widetilde{E\mathcal{F}[A]} \tag{48}$$

is an equivalence, and hence (47) can be calculated by smashing with $H\mathbb{Z}/2$ the connecting map

$$EG/A_+ \wedge \widetilde{E\mathcal{F}[B]} \wedge S^{\infty\widetilde{\beta}} \to \Sigma EG/B_+ \wedge \widetilde{E\mathcal{F}[B]}. \tag{49}$$

Consequently, (47) is a homomorphism of $\gamma_A H\mathbb{Z}/2_\star$–modules, and is computed, just like in dimension 1, by making the replacement

$$x_{\widetilde{\beta}} \mapsto y_{\widetilde{\beta}}^{-1}$$

and multiplying by $y_{\widetilde{\beta}}$. (Note that independence of the splitting $\widetilde{\beta}$ at this point follows from topology; it is a nontrivial fact to verify purely algebraically.)

We have thereby finished describing the differential d^1 of the spectral sequence (38). The main result of the present section is the following:

Theorem 7 *The spectral sequence* (38) *collapses to* E^2.

We will first prove some auxiliary results.

Lemma 8 *The Borel homology spectral sequence of any cell* $H\mathbb{Z}/2_G$*–module with cells*

$$\Sigma^? G_+ \wedge H\mathbb{Z}/2 \tag{50}$$

collapses to E^2.

Proof Taking G–fixed points, we obtain a cell $H\mathbb{Z}/2$–module with one cell for each cell (50). Now the homotopy category of $H\mathbb{Z}/2$–modules is equivalent to the derived category of $\mathbb{F}_2$–vector spaces, and a chain complex of $\mathbb{F}_2$–modules is isomorphic to a sum of an acyclic module and suspensions of $\mathbb{F}_2$. □

Lemma 9 *Let G and H be finite groups, let X be an G–cell spectrum, and let Y be an H–cell spectrum (all indexed over the complete universe). Then*

$$(H\mathbb{Z}/2_{G\times H} \wedge i_\sharp X \wedge j_\sharp Y)^{G\times H} \simeq (H\mathbb{Z}/2_G \wedge X)^G \wedge_{H\mathbb{Z}/2} (H\mathbb{Z}/2_H \wedge Y)^H.$$

Here on the left-hand side, $i_\sharp$ is the functor introducing trivial H–action on a G–spectrum and pushing forward to the complete universe, while $j_\sharp$ is the functor introducing trivial G–action on an H–spectrum and pushing forward to the complete universe.

Proof First consider $Y = S^0$. Then we have the forgetful map

$$(i_\sharp X \wedge H\mathbb{Z}/2)^{G\times H} \to (X \wedge H\mathbb{Z}/2)^G$$

which is an equivalence because it is true on cells.

In general, we have a map

$$Z^\Gamma \wedge T^\Gamma \to (Z \wedge T)^\Gamma,$$

so take the composition

$$\begin{aligned}(X \wedge H\mathbb{Z}/2)^G \wedge (Y \wedge H\mathbb{Z}/2)^H &= (i_\sharp X \wedge H\mathbb{Z}/2)^{G\times H} \wedge (j_\sharp Y \wedge H\mathbb{Z}/2)^{G\times H} \\ &\to (i_\sharp H\mathbb{Z}/2 \wedge j_\sharp Y \wedge H\mathbb{Z}/2)^{G\times H} \\ &\to (i_\sharp X \wedge j_\sharp Y)^{G\times H}\end{aligned}$$

(the last map coming from the ring structure on $H\mathbb{Z}/2$). Then again this map is an equivalence on cells, and hence an equivalence. □

Lemma 10 *Recalling again the notation* (26)*, we have:*

(a) *The spectral sequence* (35) *for*

$$\pi_*\Sigma^V H\mathbb{Z}/2 \tag{51}$$

with all $m_\alpha \geq 0$ collapses to the E^2–term in filtration degree 0.

(b) *Let $m_\alpha \leq 0$ for all α and let*

$$S = \{\alpha \in G^* \smallsetminus \{0\} \mid m_\alpha \neq 0\}.$$

Suppose the subgroup of G^* spanned by S has rank m. Then the spectral sequence (35) *for* (51) *collapses to E^2 in filtration degree $-m$.*

Proof Recall the notation (28). Let $G^* \smallsetminus \{0\} = \{\alpha_1, \dots, \alpha_{2^n-1}\}$. When α_k is linearly independent of $\alpha_1, \dots, \alpha_{k-1}$, we have

$$\pi_*\Sigma^{V_k} H\mathbb{Z}/2 \cong \pi_*(\Sigma^{V_{k-1}} H\mathbb{Z}/2)^G \otimes \pi_*(\Sigma^{m_{\alpha_k}\gamma} H\mathbb{Z}/2)^{\mathbb{Z}/2}, \tag{52}$$

where γ is the sign representation of $\mathbb{Z}/2$ by Lemma 9. Note that in the case (b), we may, without loss of generality, assume $m = n$ (ie that S spans G^*) and that what we just said occurs for $k = 1, \dots, n$ and additionally that $m_{\alpha_i} < 0$ for $i = 1, \dots, n$.

When α_k is a linear combination of $\alpha_1, \dots, \alpha_{k-1}$, and $m_{\alpha_k} \neq 0$, we use the cofibration sequence

$$S(m_{\alpha_k}\alpha_k)_+ \wedge \Sigma^{V_{k-1}} H\mathbb{Z}/2 \to \Sigma^{V_{k-1}} H\mathbb{Z}/2 \to \Sigma^{V_k} H\mathbb{Z}/2 \tag{53}$$

in the case (a) and

(54) $$\Sigma^{V_k} H\mathbb{Z}/2 \to \Sigma^{V_{k-1}} H\mathbb{Z}/2 \to DS(-m_{\alpha_k}\alpha_k)_+ \wedge \Sigma^{V_{k-1}} H\mathbb{Z}/2$$

in the case (b). If we denote each of these cofibration sequences symbolically as

$$A \to B \to C,$$

then in the case (a), (53) gives a short exact sequence of the form

(55) $$0 \to E^1 A \to E^1 B \to E^1 C \to 0$$

of the spectral sequence of (35) where in the A–term, we replace G by $\mathrm{Ker}(\alpha_k)$ and $H\mathbb{Z}/2$ by $S(m_k\alpha_k)_+ \wedge H\mathbb{Z}/2$. By the induction hypothesis, however, the homology of $E^1 A$ is concentrated in the top filtration degree, which is -1 from the point of view of G, and the homology of $E^1 B$ is concentrated in filtration degree 0, so the long exact sequence in homology gives

(56) $$0 \to E^2 \to E^2 C \to \Sigma E^2 A \to 0,$$

which is all in filtration degree 0, so our statement follows.

In the case (b), by our assumptions, we have $k > n$. Additionally, (54) gives a short exact sequence

(57) $$0 \to \Sigma^{-1} E^1 C \to E^1 A \to E^1 B \to 0,$$

but by the induction hypothesis (using the fact that a set of generators of G^* projects to a set of generators of the factor group $\mathrm{Ker}(\alpha_k)^*$), the homology of the first and last term is concentrated in filtration degree $-n$, so (57) translates to the same short exact sequence with E^1 replaced by E^2, which is entirely in filtration degree $-n$, and the statement follows. □

To continue the proof of Theorem 7, let again

$$G^* \smallsetminus \{0\} = \{\alpha_1, \dots, \alpha_{2^n-1}\}.$$

Consider

(58) $$\Sigma^{V_{2^n-1}} H\mathbb{Z}/2,$$

and let, this time, without loss of generality,

$$m_{\alpha_1}, \dots, m_{\alpha_q} < 0 \quad \text{and} \quad m_{\alpha_{q+1}}, \dots, m_{\alpha_{2^n-1}} \geq 0.$$

Let $A = \mathrm{Ker}(\alpha_1) \cap \dots \cap \mathrm{Ker}(\alpha_q)$. We will consider the sequence of cofibrations (53) with $q \leq k < 2^n - 1$. Resolving this recursively, we may consider this as a cell object construction in the category of $H\mathbb{Z}/2_G$–modules, with "cells" of the form of suspensions (by an integer) of

$$G/(\mathrm{Ker}(\alpha_{j_1})\cap\cdots\cap\mathrm{Ker}(\alpha_{j_p}))_+ \wedge \Sigma^{V_q} H\mathbb{Z}/2 \quad \text{where } q<j_1<\cdots<j_p\leq 2^n-1. \tag{59}$$

By the *degree* of a cell c, we shall mean the number

$$\deg(c) = n - \mathrm{rank}(\mathrm{Ker}(\alpha_{j_1}) \cap \cdots \cap \mathrm{Ker}(\alpha_{j_p})),$$

and by the A*–relative degree* of c, we shall mean

$$\deg_A(c) = \mathrm{rank}(G/A) - \mathrm{rank}\big(\mathrm{Ker}(\alpha_{j_1})\cap\cdots\cap\mathrm{Ker}(\alpha_{j_p}) \,/\, \mathrm{Ker}(\alpha_{j_1})\cap\cdots\cap\mathrm{Ker}(\alpha_{j_p})\cap A\big).$$

We see easily from the construction that cells of a given degree are attached to cells of strictly lower degree, and that cells of a given A–relative degree are attached to cells of lesser or equal A–relative degree. (Roughly speaking, "more free" cells are attached to "less free" ones.)

Lemma 11 *The spectral sequence arising from the cube* (35) *with* $H\mathbb{Z}/2$ *replaced by the complex formed by our "cells" of* A*–relative degree* d *collapses to* E^2 *concentrated in filtration degree* $d - \mathrm{rank}(G/A)$.

Proof Within a given A–relative degree d, attaching cells of each consecutive degree results in a short exact sequence of the form (55) where the first two terms collapse to E^2 in filtration degree $d-\mathrm{rank}(G/A)-1$ and $d-\mathrm{rank}(G/A)$, respectively. Thus, there results a short exact sequence of the form (56) in filtration degree $d-\mathrm{rank}(G/A)$, as claimed. □

(The rest of) the proof of Theorem 7 Filtering cells of (58) by A–relative degree, we obtain a spectral sequence $\mathcal{E}$ converging to E^2 of the spectral sequence of the cube (35) for (58). By Lemma 11, all the terms will be of the same (35)-filtration degree $-\mathrm{rank}(G/A)$, which is the complementary degree of $\mathcal{E}$. (Note that in this discussion, we completely ignore the original topological degree.) Thus, being concentrated in one complementary degree, $\mathcal{E}$ collapses to E^2 in that complementary degree.

However, by precisely the same arguments, we can write a variant $\widetilde{\mathcal{E}}$ of the spectral sequence $\mathcal{E}$ in homotopy groups (rather than (35) E^1–terms) of the filtered pieces of (58) by A–relative degree. By Lemma 11, $\widetilde{\mathcal{E}}^1 \cong \mathcal{E}^1$, and $d^1_{\widetilde{\mathcal{E}}}$, $d^1_{\mathcal{E}}$ have the same rank (since they are computed by the same formula). It follows that $\widetilde{\mathcal{E}}^2 \cong \mathcal{E}^2$, both collapsing to a single complementary degree. Therefore, it follows that E^2 (of the spectral sequence associated with (35) for (58)) is isomorphic to the homotopy of (58), and hence the spectral sequence collapses to E^2 by a counting argument. □

6 Example: $n = 2$

In the case $n = 2$, there are only three sign representations α, β, γ which play a symmetrical role and satisfy

$$\alpha + \beta + \gamma = 0 \in G^*, \tag{60}$$

which means that the Poincaré series of the homotopy

$$\pi_*(\Sigma^{k\alpha+\ell\beta+m\gamma} H\mathbb{Z}/2) \tag{61}$$

can be written down explicitly.

First recall that by Theorem 5, for $k, \ell, m \geq 0$, the Poincaré series is

$$\frac{1}{(1-x)^2}(1 + x - x^{1+k} - x^{1+\ell} - x^{1+m} + x^{2+k+\ell+m}). \tag{62}$$

If $k, \ell < 0$ and $m \leq 0$, by the proof of Lemma 10, the formula (62) is still valid when multiplied by x^{-2} (since all the homotopy classes are in filtration degree -2).

If $k, \ell < 0$ and $m > 0$, in the proof of Theorem 7, $A = 0$, so the A–relative degree and the degree coincide. Further, by (60) and our formula for the differential d^1 of the spectral sequence of (35), the differential $d^1_{\mathcal{E}}$ has maximal possible rank (ie "everything that can cancel dimensionwise will"). We conclude that the E^2 is concentrated in filtration degrees -1 and -2. By the cancellation principle we just mentioned, the Poincaré series can still be recovered from the formula (62). If we write the expression (62) as

$$P_+(x) - P_-(x), \tag{63}$$

where $P_+(x)$ (resp. $-P_-(x)$) is the sum of monomial summands with a positive coefficient (resp. with a negative coefficient) then the correct Poincaré series in this case is

$$x^{-2} P_+(x) + x^{-1} P_-(x),$$

the two summands of which represent classes in filtration degree -2 and -1, respectively.

Similarly, one shows that if $k, \ell \geq 0$ and $m < 0$, the E^2 collapses to filtration degrees 0 and -1, and the Poincaré series in this case is

$$P_+(x) + x^{-1} P_-(x).$$

All other cases are related to these by a symmetry of $(\mathbb{Z}/2)^2$.

Remark It might seem natural to conjecture that the classes of different filtration degrees in E^2 may be of different dimensions, with a gap between them (evoking the "gap condition" which was proved for $\mathbb{Z}/2$ in [13], and made famous for the group $\mathbb{Z}/8$ by the Hill, Hopkins and Ravenel [12] work on the Kervaire invariant 1 problem). However, one easily sees that for $n \geq 3$, classes of different filtration degrees may occur in the same dimension. For example, by Lemma 9 and by what we just proved, such a situation always occurs for $\pi_* \Sigma^{4\alpha+4\beta-2\gamma+4\delta} H\mathbb{Z}/2$ where α, β, γ are the three sign representations of $\mathbb{Z}/2 \times \mathbb{Z}/2 \times \mathbb{Z}/2$ factoring through the projections to the first two copies of $\mathbb{Z}/2$, and δ is the sign representation which factors through the projection onto the last $\mathbb{Z}/2$.

References

[1] **W C Abram**, **I Kriz**, *The equivariant complex cobordism ring of a finite abelian group*, Math. Res. Lett. 22 (2015) 1573–1588 MR

[2] **S Araki**, **M Murayama**, *τ–cohomology theories*, Japan. J. Math. 4 (1978) 363–416 MR

[3] **M F Atiyah**, *K–theory and reality*, Quart. J. Math. Oxford Ser. 17 (1966) 367–386 MR

[4] **M F Atiyah**, **I M Singer**, *The index of elliptic operators, I*, Ann. of Math. 87 (1968) 484–530 MR

[5] **G E Bredon**, *Equivariant cohomology theories*, Lecture Notes in Mathematics 34, Springer (1967) MR

[6] **T tom Dieck**, *Bordism of G–manifolds and integrality theorems*, Topology 9 (1970) 345–358 MR

[7] **P Donovan**, **M Karoubi**, *Graded Brauer groups and K–theory with local coefficients*, Inst. Hautes Études Sci. Publ. Math. 38 (1970) 5–25 MR

[8] **A W M Dress**, *Notes on the theory of representations of finite groups, I: The Burnside ring of a finite group and some AGN-applications*, Universität Bielefeld, Fakultät für Mathematik (1971) MR

[9] **D S Freed**, **M J Hopkins**, **C Teleman**, *Loop groups and twisted K–theory, I*, J. Topol. 4 (2011) 737–798 MR

[10] **J P C Greenlees**, *Adams spectral sequences in equivariant topology*, PhD thesis, University of Cambridge (1985)

[11] **J P C Greenlees**, **J P May**, *Equivariant stable homotopy theory*, from "Handbook of algebraic topology" (I M James, editor), North-Holland, Amsterdam (1995) 277–323 MR

[12] **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

[13] **P Hu**, **I Kriz**, *Real-oriented homotopy theory and an analogue of the Adams–Novikov spectral sequence*, Topology 40 (2001) 317–399 MR

[14] **P Hu**, **I Kriz**, *Coefficients of the constant Mackey functor over cyclic* p*–groups*, preprint (2010)

[15] **P Hu**, **I Kriz**, *Topological Hermitian cobordism*, J. Homotopy Relat. Struct. 11 (2016) 173–197 MR

[16] **P Hu**, **I Kriz**, **K Ormsby**, *The homotopy limit problem for Hermitian* K*–theory, equivariant motivic homotopy theory and motivic Real cobordism*, Adv. Math. 228 (2011) 434–480 MR

[17] **I Kriz**, *The* $\mathbb{Z}/p$*–equivariant complex cobordism ring*, from "Homotopy invariant algebraic structures" (J-P Meyer, J Morava, W S Wilson, editors), Contemp. Math. 239, Amer. Math. Soc. (1999) 217–223 MR

[18] **S Kriz**, *Equivariant cohomology and the super reciprocal plane of a hyperplane arrangement*, preprint (2015)

[19] **P S Landweber**, *Conjugations on complex manifolds and equivariant homotopy of* MU, Bull. Amer. Math. Soc. 74 (1968) 271–274 MR

[20] **G Lewis**, **J P May**, **J McClure**, *Ordinary* $RO(G)$*–graded cohomology*, Bull. Amer. Math. Soc. 4 (1981) 208–212 MR

[21] **L G Lewis, Jr**, *The* $RO(G)$*–graded equivariant ordinary cohomology of complex projective spaces with linear* $\mathbb{Z}/p$ *actions*, from "Algebraic topology and transformation groups" (T tom Dieck, editor), Lecture Notes in Math. 1361, Springer (1988) 53–122 MR

[22] **L G Lewis, Jr**, **J P May**, **M Steinberger**, **J E McClure**, *Equivariant stable homotopy theory*, Lecture Notes in Mathematics 1213, Springer (1986) MR

[23] **N Proudfoot**, **D Speyer**, *A broken circuit ring*, Beiträge Algebra Geom. 47 (2006) 161–166 MR

[24] **H Terao**, *Algebras generated by reciprocals of linear forms*, J. Algebra 250 (2002) 549–558 MR

Department of Mathematics, University of Michigan
2074 East Hall, 530 Church Street, Ann Arbor, MI 48109-1043, United States

johnholl@umich.edu, ikriz@umich.edu

Received: 9 June 2015 Revised: 17 August 2016

Algebraic & Geometric Topology 17 (2017) 765–791

Homotopy theory of cocomplete quasicategories

KAROL SZUMIŁO

We prove that the homotopy theory of cocomplete quasicategories is equivalent to the homotopy theory of cofibration categories. This is achieved by presenting both theories as fibration categories and constructing an explicit exact equivalence between them.

55U35; 18G55

Introduction

There are a few notions that formalize the concept of a cocomplete homotopy theory, but it is not clear how they compare to each other. We consider two of them: cofibration categories and cocomplete quasicategories and prove that they are indeed equivalent. More precisely, our main result (Theorems 1.10, 2.14 and 4.9) is as follows.

Theorem *Both the category of cofibration categories and the category of cocomplete quasicategories carry structures of fibration categories and these two fibration categories are equivalent.*

These two models of cocomplete homotopy theories exemplify two different approaches to abstract homotopy theory: *homotopical algebra* and *higher category theory*. Homotopical algebra refers broadly to the theory of categories with equivalences and some further structure which provides tools for constructing derived functors. It was started by Quillen when he introduced model categories [17], but there are other notions of a similar flavor, eg *(co)fibration categories*, first defined by K Brown [6], which are crucial in the present paper. Higher category theory refers, in this context, to various models of $(\infty, 1)$–categories which provide the language to express homotopy coherent universal properties. Examples of such models include *quasicategories* introduced by Boardman and Vogt [5] and studied in detail by Joyal [14] and Lurie [16], *Segal categories* introduced by Dwyer, Kan and Smith [9] and developed by Hirschowitz and Simpson [12], and *complete Segal spaces* introduced by Rezk [18].

These (and other) notions of an $(\infty, 1)$–category are known to be equivalent to each other by the results of Bergner [4] and Joyal and Tierney [15]. An abstract axiomatization was also developed by Toën [25] and Barwick and Schommer-Pries [3]. Moreover,

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.765

Barwick and Kan [2; 1] established that these concepts are also equivalent to the notion of a *relative category*, ie a category equipped with a class of weak equivalences and no further structure.

Our main theorem can be seen as a structured version of the latter result that concerns cocomplete homotopy theories as opposed to arbitrary ones. In particular, the comparison between cofibration categories and cocomplete quasicategories includes a direct translation between homotopy colimits computed as derived functors of cofibration categories and colimits in quasicategories characterized by homotopy coherent universal properties. The result can be seen as an answer to a version of [13, Problem 8.2] which asks for a comparison between the theories of model categories and complete Segal spaces.

This paper is the last in the series of three that summarize the results of the author's thesis [21; 22] and relies heavily on the techniques of the previous two. The main result of the first one [24] was existence of a fibration category of cofibration categories. In the second one [23] we introduced the *quasicategory of frames* which is a new construction of the $(\infty, 1)$–category associated to a cofibration category. In the present paper we construct a fibration category of cocomplete quasicategories and prove that the quasicategory of frames functor is an equivalence of fibration categories.

Section 2 contains the basic theory of quasicategories, which is mostly cited from Joyal [14] and Dugger and Spivak [8]. In particular, we establish fibration categories of quasicategories and of cocomplete quasicategories. This section contains no new results, except possibly for the existence of the latter fibration category. (The completeness of the homotopy theory of cocomplete quasicategories is discussed in Lurie [16], but it is not stated in terms of fibration categories.)

In Section 4 we prove that N_f is a weak equivalence of fibration categories. To this end we associate with every cocomplete quasicategory $\mathcal{D}$ a cofibration category $\mathrm{Dg}\,\mathcal{D}$ called the *category of diagrams in* $\mathcal{D}$. This yields a functor Dg which is not exact but is an inverse to N_f up to weak equivalence. This suffices to conclude that N_f is an equivalence of homotopy theories.

The results are parametrized by a regular cardinal number κ and concern κ–cocomplete cofibration categories and κ–cocomplete quasicategories. In Section 4 the arguments split into two cases. First, we consider the easier case of $\kappa > \aleph_0$ and then point out the modifications necessary for the proof when $\kappa = \aleph_0$.

Acknowledgments This paper is based on a part of my thesis [21], which was written while I was a doctoral student in the Bonn International Graduate School in Mathematics and, more specifically, Graduiertenkolleg 1150 "Homotopy and Cohomology" and

International Max Planck Research School for Moduli Spaces. I want to thank everyone involved for creating an excellent working environment. Moreover, this material is partially based upon work supported by the National Science Foundation under Grant No. 0932078 000 while the author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2014 semester.

I want to thank Clark Barwick, Bill Dwyer, André Joyal, Chris Kapulkin, Lennart Meier, Thomas Nikolaus, Chris Schommer-Pries, Peter Teichner and Marek Zawadowski for conversations on various topics which were very beneficial to my research.

I am especially grateful to Viktoriya Ozornova and Irakli Patchkoria for reading an early draft of my thesis. Their feedback helped me make many improvements and avoid numerous errors.

Above all, I want to express my gratitude to my supervisor Stefan Schwede whose expertise was always invaluable and without whose support this thesis could not have been written.

1 Review of cofibration categories

Our results are based on the techniques of [24; 23] and we start by summarizing the contents of the first of these papers. The central notion is that of *cofibration categories* which are slightly modified duals of Brown's *categories of fibrant objects* [6].

Definition 1.1 [24, Definition 1.1] A *cofibration category* is a category $\mathcal{C}$ equipped with two subcategories: the subcategory of *weak equivalences* (denoted by $\xrightarrow{\sim}$) and the subcategory of *cofibrations* (denoted by $\rightarrowtail$) such that the following axioms are satisfied. (Here, an *acyclic cofibration* is a morphism that is both a weak equivalence and a cofibration.)

(C0) Weak equivalences satisfy the 2-out-of-6 property, ie if

$$W \xrightarrow{f} X \xrightarrow{g} Y \xrightarrow{h} Z$$

are morphisms of $\mathcal{C}$ such that both gf and hg are weak equivalences, then so are f, g and h (and thus also hgf).

(C1) Every isomorphism of $\mathcal{C}$ is an acyclic cofibration.

(C2) An initial object exists in $\mathcal{C}$.

(C3) Every object X of $\mathcal{C}$ is cofibrant, ie if 0 is the initial object of $\mathcal{C}$, then the unique morphism $0 \to X$ is a cofibration.

(C4) Cofibrations are stable under pushouts along arbitrary morphisms of $\mathcal{C}$ (in particular these pushouts exist in $\mathcal{C}$). Acyclic cofibrations are stable under pushouts along arbitrary morphisms of $\mathcal{C}$.

(C5) Every morphism of $\mathcal{C}$ factors as a composite of a cofibration followed by a weak equivalence.

(C6) Cofibrations are stable under sequential colimits, ie given a sequence of cofibrations

$$A_0 \rightarrowtail A_1 \rightarrowtail A_2 \rightarrowtail \cdots$$

its colimit A_∞ exists and the induced morphism $A_0 \to A_\infty$ is a cofibration. Acyclic cofibrations are stable under sequential colimits.

(C7-κ) Coproducts of κ–small families of objects exist. Cofibrations and acyclic cofibrations are stable under κ–small coproducts.

The last two axioms are optional. If we drop them, then cofibration categories can be considered as models of finitely cocomplete homotopy theories. If we include (C6) and (C7-κ) for a fixed regular cardinal $\kappa > \aleph_0$, we obtain models of κ–cocomplete homotopy theories; we call them *(homotopy) κ–cocomplete cofibration categories*. For $\kappa = \aleph_0$ the name *(homotopy) $\aleph_0$–cocomplete cofibration category* will refer to a cofibration category satisfying the axioms (C0)–(C5). The definition readily dualizes to yield *fibration categories* which are models of finitely complete homotopy theories or κ–complete homotopy theories depending on the choice of axioms.

First, we recall some classical results about cofibration categories, mostly following [20]. We fix a cofibration category $\mathcal{C}$.

Definition 1.2 (1) A *cylinder* of an object X is a factorization of the codiagonal morphism $X \amalg X \to X$ as $X \amalg X \rightarrowtail IX \xrightarrow{\sim} X$.

(2) A *left homotopy* between morphisms $f, g\colon X \to Y$ via a cylinder $X \amalg X \rightarrowtail IX \xrightarrow{\sim} X$ is a commutative square of the form

$$\begin{tikzcd} X \amalg X \arrow[r, "{[f,g]}"] \arrow[d, tail] & Y \arrow[d, tail, "\sim"] \\ IX \arrow[r] & Z \end{tikzcd}$$

(3) Morphisms $f, g\colon X \to Y$ are *left homotopic* (notation: $f \simeq_l g$) if there exists a left homotopy between them via some cylinder on X.

The definition of left homotopies differs from the standard definition as usually given in the context of model categories where the morphism $Y \overset{\sim}{\rightarrowtail} Z$ is required to be the identity. This modification is dictated by the lack of fibrant objects in cofibration categories and makes the definition well-behaved for arbitrary Y while the standard definition in a model category is only well-behaved for a fibrant Y.

We denote the homotopy category of $\mathcal{C}$ (ie its localization with respect to weak equivalences) by $\operatorname{Ho}\mathcal{C}$ and for a morphism f of $\mathcal{C}$ we write $[f]$ for its image under the localization functor $\mathcal{C} \to \operatorname{Ho}\mathcal{C}$. The homotopy category can be constructed in two steps: first dividing out left homotopies and then applying the calculus of fractions.

Proposition 1.3 *The relation of left homotopy is a congruence on $\mathcal{C}$. Moreover, every morphism of $\mathcal{C}$ that becomes an isomorphism in $\mathcal{C} / \simeq_l$ is a weak equivalence. Thus left homotopic morphisms become equal in $\operatorname{Ho}\mathcal{C}$ and $\mathcal{C} / \simeq_l$ comes equipped with a canonical functor $\mathcal{C} / \simeq_l \to \operatorname{Ho}\mathcal{C}$.*

Proof The first statement is [20, Theorem 6.3.3(1)]. The remaining ones follow by straightforward 2-out-of-3 arguments. □

The next theorem is a crucial tool in the theory of cofibration categories and can be used to verify many of their fundamental properties. It says that up to left homotopy all cofibration categories satisfy the left calculus of fractions in the sense of Gabriel and Zisman [10, Chapter I]. This fact was first proven by Brown [6, Proposition I.2]. In general, constructing $\operatorname{Ho}\mathcal{C}$ may involve using arbitrarily long zig-zags of morphisms in $\operatorname{Ho}\mathcal{C}$ and identifying them via arbitrarily long chains of relations. However, the previous proposition implies that $\mathcal{C} / \simeq_l \to \operatorname{Ho}\mathcal{C}$ is also a localization functor and in that case Theorem 1.4 says that it suffices to consider two-step zig-zags (called *left fractions*) up to a much simplified equivalence relation.

Theorem 1.4 *A cofibration category $\mathcal{C}$ satisfies the left calculus of fractions up to left homotopy, ie*

(1) *Every morphism $\varphi \in \operatorname{Ho}\mathcal{C}(X, Y)$ can be written as a left fraction $[s]^{-1}[f]$, where $f\colon X \to \tilde{Y}$ and $s\colon Y \overset{\sim}{\to} \tilde{Y}$ are morphisms of $\mathcal{C}$.*

(2) *Two fractions $[s]^{-1}[f]$ and $[t]^{-1}[g]$ are equal in $\operatorname{Ho}\mathcal{C}(X, Y)$ if and only if there exist weak equivalences u and v such that*

$$us \simeq_l vt \quad \text{and} \quad uf \simeq_l vg.$$

(3) *If $\varphi \in \operatorname{Ho}\mathcal{C}(X, Y)$ and $\psi \in \operatorname{Ho}\mathcal{C}(Y, Z)$ can be written as $[s]^{-1}[f]$ and $[t]^{-1}[g]$ respectively and a square*

$$\begin{tikzcd} Y \arrow[r, "g"] \arrow[d, "s"', "\sim"] & \widetilde{Z} \arrow[d, "u"', "\sim"] \\ \widetilde{Y} \arrow[r, "h"'] & \widehat{Z} \end{tikzcd}$$

commutes up to homotopy, then $\psi\varphi$ *can be written as* $[ut]^{-1}[hf]$.

Proof Parts (1) and (2) follow from [20, Theorem 6.4.4(1)], and (3) follows from the proof of [20, Theorem 6.4.1]. □

We will need the following technical lemma. Even though cofibrations in a cofibration category do not necessarily satisfy any lifting property, they can still be shown to have a version of the "homotopy extension property" with respect to left homotopies.

Lemma 1.5 *Let* $i\colon A \rightarrowtail B$ *be a cofibration in* $\mathcal{C}$. *Let* $f\colon A \to X$ *and* $g\colon B \to X$ *be morphisms such that* gi *is left homotopic to* f. *Then there exist a weak equivalence* $s\colon X \to \widehat{X}$ *and a morphism* $\widetilde{g}\colon B \to \widehat{X}$ *such that* $\widetilde{g}$ *is left homotopic to* sg *and* $\widetilde{g}i = sf$.

Proof Pick compatible cylinders on A and B, ie a diagram

$$\begin{tikzcd} A \amalg A \arrow[r, tail] \arrow[d, "i \amalg i"'] & IA \arrow[r, "\sim"] \arrow[d] & A \arrow[d, "i"] \\ B \amalg B \arrow[r, tail] & IB \arrow[r, "\sim"] & B \end{tikzcd}$$

such that the induced morphism $IA \amalg_{(A \amalg A)} (B \amalg B) \to IB$ is a cofibration. Let δ_0 and δ_1 denote the two structure morphisms $A \rightarrowtail IA$.

Pick a left homotopy

$$\begin{tikzcd} A \amalg A \arrow[r, "{[f, gi]}"] \arrow[d, tail, "{[\delta_0, \delta_1]}"'] & X \arrow[d, tail, "j"', "\sim"] \\ IA \arrow[r, "H"'] & \widetilde{X} \end{tikzcd}$$

between f and gi. Then we have in particular $jgi = H\delta_1$ and thus there is an induced morphism $[H, jg]\colon IA \amalg_A B \to \widetilde{X}$ so we can take a pushout:

$$\begin{tikzcd} IA \amalg_A B \arrow[r, "{[H, jg]}"] \arrow[d, tail, "\sim"'] & \widetilde{X} \arrow[d, tail, "\widetilde{j}"', "\sim"] \\ IB \arrow[r, "\widetilde{H}"'] & \widehat{X} \end{tikzcd}$$

Set $s = \widetilde{j}\, j$ and $\widetilde{g} = \widetilde{H}$. We have $sf = \widetilde{g}i$, and $\widetilde{H}$ and $\mathrm{id}_{\widehat{X}}$ constitute a left homotopy between $\widetilde{g}$ and sg. □

The main result of [24] establishes the homotopy theory of cofibration categories in the form of a fibration category. We recall the prerequisite definitions before stating the theorem.

Definition 1.6 A functor $F\colon \mathcal{C} \to \mathcal{D}$ between cofibration categories is *exact* if it preserves cofibrations, acyclic cofibrations, initial objects and pushouts along cofibrations.

If $\mathcal{C}$ and $\mathcal{D}$ are κ–cocomplete, then F is *κ–cocontinuous* if, in addition, it preserves colimits of sequences of cofibrations and κ–small coproducts.

The category of (small) κ–cocomplete cofibration categories and κ–cocontinuous functors will be denoted by CofCat_κ. It is equipped with classes of weak equivalences and fibrations as defined below.

Definition 1.7 An exact functor $F\colon \mathcal{C} \to \mathcal{D}$ is a *weak equivalence* if it induces an equivalence $\mathrm{Ho}\,\mathcal{C} \to \mathrm{Ho}\,\mathcal{D}$.

A typical way of proving that an exact functor is a weak equivalence is by using the *approximation properties* of the following proposition. They were originally introduced by Waldhausen [27, Section 1.6] in his work on algebraic K–theory and later adapted to the context of cofibration categories by Cisinski.

Proposition 1.8 [7, Théorème 3.19] *An exact functor $F\colon \mathcal{C} \to \mathcal{D}$ is a weak equivalence if and only if it satisfies the following properties:*

(App1) *F reflects weak equivalences.*

(App2) *Given a morphism $f\colon FA \to Y$ in $\mathcal{D}$, there exists a morphism $i\colon A \to B$ in $\mathcal{C}$ and a commutative diagram*

$$\begin{tikzcd} FA \arrow[r, "f"] \arrow[d, "Fi"'] & Y \arrow[d, "\sim"] \\ FB \arrow[r, "\sim"'] & Z \end{tikzcd}$$

in $\mathcal{D}$. □

Definition 1.9 [24, Definition 2.3] Let $P\colon \mathcal{E} \to \mathcal{D}$ be an exact functor of cofibration categories.

(1) P is an *isofibration* if for every object $A \in \mathcal{E}$ and an isomorphism $g\colon PA \to Y$ there is an isomorphism $f\colon A \to B$ such that $Pf = g$.

(2) It is said to satisfy the *lifting property for factorizations* if for any morphism $f\colon A \to B$ of $\mathcal{E}$ and a factorization

$$\begin{tikzcd} PA \arrow[rr, "Pf"] \arrow[dr, tail, "j"'] & & PB \\ & X \arrow[ur, "t"', "\sim"] & \end{tikzcd}$$

there exists a factorization

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, tail, "i"'] & & B \\ & C \arrow[ur, "s"', "\sim"] & \end{tikzcd}$$

such that $Pi = j$ and $Ps = t$ (in particular, $PC = X$).

(3) It has the *lifting property for pseudofactorizations* if for any morphism $f\colon A \to B$ of $\mathcal{E}$ and a diagram

$$\begin{tikzcd} PA \arrow[r, "Pf"] \arrow[d, tail, "j"'] & PB \arrow[d, tail, "v", "\sim"'] \\ X \arrow[r, "\sim", "t"'] & Y \end{tikzcd}$$

there exists a diagram

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, tail, "i"'] & B \arrow[d, tail, "u", "\sim"'] \\ C \arrow[r, "\sim", "s"'] & D \end{tikzcd}$$

such that $Pi = j$, $Ps = t$ and $Pu = v$ (in particular, $PC = X$ and $PD = Y$).

(4) We say that P is a *fibration* if it is an isofibration and satisfies the lifting properties for factorizations and pseudofactorizations.

Theorem 1.10 [24, Theorem 2.9] *The category* CofCat_κ *of small* κ*–cocomplete cofibration categories with weak equivalences and fibrations as above is a fibration category.* □

The goal of the paper is to prove that this fibration category is equivalent to the corresponding fibration category of κ–cocomplete quasicategories.

2 Cocomplete quasicategories

We will start with a concise summary of the theory of quasicategories. It is well covered in [14] and [16] so we do not go into much detail. Our main goal is to establish a fibration category of finitely cocomplete quasicategories in Theorem 2.14. We refer to [24] for background on fibration categories. We cite [14] for the proof that the fibration category of all quasicategories can be obtained without constructing the entire Joyal model structure (Theorem 2.4) which makes the proof rather elementary. (A more streamlined exposition of the same results can be found in the appendices to [8].) Then we briefly introduce colimits in quasicategories and state their basic properties used in the proof of Theorem 2.14.

We will denote the groupoid freely generated by an isomorphism $0 \to 1$ by $E(1)$ and its nerve by $E[1]$. Quasicategories are defined as certain special simplicial sets and are to be thought of as models of $(\infty, 1)$–categories where vertices are objects, edges are morphisms and higher simplices are higher morphisms (or higher homotopies). Functors between quasicategories are just simplicial maps. In particular, maps out of $E[1]$ are equivalences in quasicategories and $E[1]$–homotopies are natural equivalences between functors. The account of the homotopy theory of quasicategories below closely follows the classical approach to simplicial homotopy theory (see eg [11, Chapter I]) with Kan complexes replaced by quasicategories and usual simplicial homotopies replaced by $E[1]$–homotopies.

Definition 2.1 (1) Let $f, g\colon K \to L$ be simplicial maps. An $E[1]$–*homotopy* from f to g is a simplicial map $K \times E[1] \to L$ extending $[f, g]\colon K \times \partial\Delta[1] \to L$.

(2) Two simplicial maps $f, g\colon K \to L$ are $E[1]$–*homotopic* if there exists a zig-zag of $E[1]$–homotopies connecting f to g. (It suffices to consider sequences instead of zig-zags since $E[1]$ has an automorphism that exchanges the vertices.)

(3) A simplicial map $f\colon K \to L$ is an $E[1]$–*homotopy equivalence* if there is a simplicial map $g\colon L \to K$ such that fg is $E[1]$–homotopic to id_L and gf is $E[1]$–homotopic to id_K.

Definition 2.2 (1) A simplicial map is an *inner fibration* if it has the right lifting property with respect to the inner horn inclusions.

(2) A simplicial map is an *inner isofibration* if it is an inner fibration and has the right lifting property with respect to $\Delta[0] \hookrightarrow E[1]$.

(3) A simplicial map is an *acyclic Kan fibration* if it has the right lifting property with respect to $\partial\Delta[m] \hookrightarrow \Delta[m]$ for all m.

(4) A simplicial set $\mathcal{C}$ is a *quasicategory* if the unique map $\mathcal{C} \to \Delta[0]$ is an inner fibration.

We will refer to $E[1]$–equivalences between quasicategories as *categorical equivalences* and use them to introduce the homotopy theory of quasicategories. (It is also possible to extend this notion to maps of general simplicial sets, but we have no need to do it.) If K is any simplicial set and $\mathcal{C}$ is a quasicategory, then the relation of "being connected by a single $E[1]$–homotopy" is already an equivalence relation on the set of simplicial maps $K \to \mathcal{C}$ by [8, Proposition 2.3]. This simplifies the definition of categorical equivalences since it is always sufficient to consider one-step $E[1]$–homotopies. The following lemma provides a useful criterion for verifying that a functor between quasicategories is a categorical equivalence.

Lemma 2.3 [26] *A functor $F\colon \mathcal{C} \to \mathcal{D}$ between quasicategories is a categorical equivalence provided that for every commutative square of the form*

$$\begin{tikzcd} \partial\Delta[m] \arrow[r, "u"] \arrow[d, hook] & \mathcal{C} \arrow[d, "F"] \\ \Delta[m] \arrow[r, "v"'] & \mathcal{D} \end{tikzcd}$$

there exists a map $w\colon \Delta[m] \to \mathcal{C}$ such that $w|\partial\Delta[m] = u$ and Fw is $E[1]$–homotopic to v relative to $\partial\Delta[m]$. □

Theorem 2.4 *The category of small quasicategories with simplicial maps as morphisms, categorical equivalences as weak equivalences and inner isofibrations as fibrations is a fibration category.*

Proof Only two of the axioms require nontrivial proofs: stability of acyclic fibrations under pullbacks, which follows from the fact that acyclic (inner iso-) fibrations coincide with acyclic Kan fibrations by [14, Theorem 5.15], and the factorization axiom which is verified in [14, Proposition 5.16]. □

This fibration category is a part of the Joyal model structure on simplicial sets established in [14, Theorem 6.12]. Indeed, the theorem above is an intermediate step in the construction of this model category.

Quasicategories are models for homotopy theories and as such they have homotopy categories. Two morphisms $f, g\colon x \to y$ of a quasicategory $\mathcal{D}$ are *homotopic* if there exists a simplex $H\colon \Delta[2] \to \mathcal{D}$ such that $H\delta_0 = y\sigma_0$, $H\delta_1 = g$ and $H\delta_2 = f$. The *homotopy category* of $\mathcal{D}$ is the category $\operatorname{Ho}\mathcal{D}$ with the same objects as $\mathcal{D}$, homotopy classes of morphisms of $\mathcal{D}$ as morphisms and the composition induced by filling horns.

If f is a morphism of a quasicategory $\mathcal{C}$, then we say that f is an *equivalence* if the simplicial map $f\colon \Delta[1] \to \mathcal{C}$ extends to $E[1] \to \mathcal{C}$. (By [14, Proposition 4.22] a morphism is an equivalence if and only if it becomes an isomorphism in the homotopy category.) Two objects of $\mathcal{C}$ are *equivalent* if they are connected by an equivalence.

We proceed to the discussion of colimits in quasicategories. Such colimits are homotopy invariant by design and they serve as models for homotopy colimits. However, in quasicategories there is no corresponding notion of a "strict" colimit and thus it is customary to refer to "homotopy colimits" in quasicategories simply as colimits. The general theory of colimits is explored in depth in [16, Chapter 4]; here we only discuss its most basic aspects.

The quasicategorical notion of colimit is defined using the join construction for simplicial sets. As a functor $\star\colon \Delta \times \Delta \to \Delta$ it is defined by concatenation: $[m], [n] \mapsto [m+1+n]$. Then the general join is defined as the unique functor $\mathsf{sSet} \times \mathsf{sSet} \to \mathsf{sSet}$ which agrees with the above on the representable simplicial sets and such that for each K the resulting functor $K \star -\colon \mathsf{sSet} \to K \downarrow \mathsf{sSet}$ preserves colimits. As such, the functor $K \star -$ has a right adjoint which we will denote by $(X\colon K \to M) \mapsto X \setminus M$. ($X \setminus M$ is called the *slice* of M under X.)

Lemma 2.5 *Let $P\colon \mathcal{C} \twoheadrightarrow \mathcal{D}$ be a inner isofibration of quasicategories and $X\colon K \to \mathcal{C}$ a diagram. Then the induced map $X \setminus \mathcal{C} \to PX \setminus \mathcal{D}$ is an inner isofibration. In particular, $X \setminus \mathcal{C}$ is a quasicategory.*

Proof This follows from [14, Theorem 3.19(i) and Proposition 4.10]. □

For any simplicial set K we define the *under-cone* on K as $K^{\rhd} = K \star \Delta[0]$. We also fix a regular cardinal number κ.

Definition 2.6 Let $\mathcal{C}$ be a quasicategory and let $X\colon K \to \mathcal{C}$ be any simplicial map (which we consider as a K–indexed diagram in $\mathcal{C}$).

(1) A *cone under* X is a diagram $S\colon K^{\rhd} \to \mathcal{C}$ such that $S|K = X$.

(2) A cone S under X is *universal* or a *colimit of* X if for any $m > 0$ and any diagram of solid arrows

$$\begin{tikzcd} K \star \partial\Delta[m] \arrow[r, "U"] \arrow[d] & \mathcal{C} \\ K \star \Delta[m] \arrow[ur, dashed] & \end{tikzcd}$$

where $U|K^{\rhd} = S$, there exists a dashed arrow making the diagram commute.

(3) An *initial object* of $\mathcal{C}$ is a colimit of the unique empty diagram in $\mathcal{C}$.

(4) A simplicial map $f\colon K \to L$ is *cofinal* if for every quasicategory $\mathcal{C}$ and every universal cone $S\colon L^{\rhd} \to \mathcal{C}$ the induced cone $Sf^{\rhd}$ is also universal.

(5) The quasicategory $\mathcal{C}$ is *κ–cocomplete* if for every κ–small simplicial set K every diagram $K \to \mathcal{C}$ has a colimit.

(6) A functor $F\colon \mathcal{C} \to \mathcal{D}$ between finitely cocomplete quasicategories is said to be *κ–cocontinuous* if for every κ–small simplicial set K and every universal cone $S\colon K^{\rhd} \to \mathcal{C}$ the cone FS is also universal.

Lemma 2.7 *A cone S under X is universal if and only if it is an initial object of $X \setminus \mathcal{C}$.*

Proof This follows directly from the fact that the slice functor is a right adjoint of the join functor. □

We will now discuss the counterparts of a few classical statements of category theory saying that colimits are essentially unique and invariant under equivalences. For a quasicategory $\mathcal{C}$ and a diagram $X\colon K \to \mathcal{C}$ we let $(X \setminus \mathcal{C})^{\mathrm{univ}}$ denote the simplicial subset of $X \setminus \mathcal{C}$ consisting of those simplices whose vertices are all universal.

Lemma 2.8 *The simplicial set $(X \setminus \mathcal{C})^{\mathrm{univ}}$ is empty or a contractible Kan complex.*

Proof A simplicial set is empty or a contractible Kan complex if and only if it has the right lifting property with respect to the boundary inclusions $\partial\Delta[m] \hookrightarrow \Delta[m]$ for all $m > 0$. For $(X \setminus \mathcal{C})^{\mathrm{univ}}$ such lifting problems are equivalent to the lifting problems

$$\begin{tikzcd} K \star \partial\Delta[m] \arrow[r, "U"] \arrow[d] & \mathcal{C} \\ K \star \Delta[m] \arrow[ur, dashed] & \end{tikzcd}$$

with $U|(K \star \{i\})$ universal for each $i \in [m]$, which have solutions by the definition of universal cones. □

Corollary 2.9 *If $X\colon K \to \mathcal{C}$ is a diagram in a quasicategory and S and T are two universal cones under X, then they are equivalent under X, ie as objects of $X \setminus \mathcal{C}$.*

Proof The simplicial set $(X \setminus \mathcal{C})^{\mathrm{univ}}$ is nonempty and thus a contractible Kan complex by the previous lemma. Hence it has the right lifting property with respect to the inclusion $\partial\Delta[1] \hookrightarrow E[1]$, which translates to the lifting property

$$\begin{tikzcd} K \star \partial\Delta[1] \arrow[r, "{[S,T]}"] \arrow[d] & \mathcal{C} \\ K \star E[1] \arrow[ur, dashed] & \end{tikzcd}$$

which yields an equivalence of S and T. □

Lemma 2.10 *If $\mathcal{C}$ is a quasicategory and X and Y are equivalent objects of $\mathcal{C}$, then X is initial if and only if Y is.*

Proof Assume that X is initial and let $U\colon \partial\Delta[m] \to \mathcal{C}$ be such that $U|\Delta[0] = Y$. We can consider an equivalence from X to Y as a diagram $f\colon \Delta[0] \star \Delta[0] \to \mathcal{C}$. Then by the universal property of X there is a diagram $\Delta[0] \star \partial\Delta[m]$ extending both f and U. (We can iteratively choose extensions over $\Delta[0] \star \Delta[k]$ for all faces $\Delta[k] \hookrightarrow \partial\Delta[m]$.) This diagram is a special outer horn (under the isomorphism $\Delta[0] \star \partial\Delta[m] \cong \Lambda^0[m+1]$) and thus has a filler by [14, Theorem 3.14]. Therefore U extends over $\Delta[m]$ and hence Y is initial. □

Our goal is to compare cofibration categories to quasicategories, but we expect κ–cocomplete cofibration categories to correspond to κ–cocomplete quasicategories, not to arbitrary ones. In the remainder of this section we will restrict the fibration structure of Theorem 2.4 to the subcategory of κ–cocomplete quasicategories and κ–cocontinuous functors.

First, we need two lemmas about lifting colimits along inner isofibrations.

Lemma 2.11 *Consider a pullback square of quasicategories*

$$\begin{tikzcd} \mathcal{P} \arrow[r, "G"] \arrow[d, "Q"'] & \mathcal{E} \arrow[d, two heads, "P"] \\ \mathcal{C} \arrow[r, "F"'] & \mathcal{D} \end{tikzcd}$$

where P is an inner isofibration, and let $S\colon K^{\triangleright} \to \mathcal{P}$ be a cone. If all GS, QS and $PGS = FQS$ are universal, then so is S.

Proof Under these assumptions the square

$$\begin{tikzcd} X \setminus \mathcal{P} \arrow[r, "G"] \arrow[d, "Q"'] & GX \setminus \mathcal{E} \arrow[d, two heads, "P"] \\ QX \setminus \mathcal{C} \arrow[r, "F"'] & PGX \setminus \mathcal{D} \end{tikzcd}$$

(where $X = S|K$) is also a pullback along an inner isofibration by Lemma 2.5. Hence it suffices to verify the conclusion for initial objects.

Thus assume that $K = \varnothing$ and let $m > 0$ and $U\colon \partial\Delta[m] \to \mathcal{P}$ be such that $U|\Delta[0] = S$. Then we have

$$GU|\Delta[0] = GS \quad \text{and} \quad QU|\Delta[0] = QS$$

and since both GS and QS are initial we can find $V_{\mathcal{E}} \in \mathcal{E}_m$ and $V_{\mathcal{C}} \in \mathcal{C}_m$ such that $V_{\mathcal{E}}|\partial\Delta[m] = GU$ and $V_{\mathcal{C}}|\partial\Delta[m] = QU$. Next, define $\widetilde{V}\colon \partial\Delta[m+1] \to \mathcal{D}$ by replacing the 1st face of $PV_{\mathcal{E}}\sigma_1|\partial\Delta[m+1]$ with $FV_{\mathcal{C}}$ and $\widetilde{W}\colon \Lambda^1[m+1] \to \mathcal{E}$ by setting it to $V_{\mathcal{E}}\sigma_1|\Lambda^1[m+1]$.

By the assumption PGS is initial and $\widetilde{V}|\Delta[0] = PGS$ so $\widetilde{V}$ extends to $V \in \mathcal{D}_{m+1}$. Then we have a commutative square

$$\begin{tikzcd} \Lambda^1[m+1] \arrow[r, "\widetilde{W}"] \arrow[d] & \mathcal{E} \arrow[d, two heads, "P"] \\ \Delta[m+1] \arrow[r, "V"'] & \mathcal{D} \end{tikzcd}$$

which admits a lift W since P is an inner isofibration and $0 < 1 < m+1$. We have $FV_{\mathcal{C}} = PW\delta_1$ and thus $(V_{\mathcal{C}}, W\delta_1)$ is an m–simplex of $\mathcal{P}$ whose boundary is U. Hence S is initial. □

Lemma 2.12 *Let $P\colon \mathcal{C} \twoheadrightarrow \mathcal{D}$ be an inner isofibration, $X\colon K \to \mathcal{C}$ a diagram and $T\colon K^{\triangleright} \to \mathcal{D}$ a colimit of PX. If X has a colimit in $\mathcal{C}$ which is preserved by P, then there exists a colimit $S\colon K^{\triangleright} \to \mathcal{C}$ of X such that $PS = T$.*

Proof Let $\widetilde{S}\colon K^{\triangleright} \to \mathcal{C}$ be some colimit of X. Since both T and $P\widetilde{S}$ are universal, we have a simplicial map $U\colon K \star E[1] \to \mathcal{D}$ such that $U|(K \star \partial\Delta[1]) = [T, P\widetilde{S}]$ by Corollary 2.9. The conclusion now follows from Lemmas 2.5 and 2.10. □

The homotopical content of the next proposition is the same as that of [16, Lemma 5.4.5.5]. However, we need a stricter point-set level statement. See also [19, Sections 3 and 4] for a systematic approach to results of this type.

Proposition 2.13 *Let $F\colon \mathcal{C} \to \mathcal{D}$ and $P\colon \mathcal{E} \twoheadrightarrow \mathcal{D}$ be κ–cocontinuous functors between κ–cocomplete quasicategories with P an inner isofibration. Then a pullback of P along F exists in the category of κ–cocomplete quasicategories and κ–cocontinuous functors.*

Proof Form a pullback of P along F in the category of quasicategories:

$$\begin{tikzcd} \mathcal{P} \arrow[r, "G"] \arrow[d, "Q"'] & \mathcal{E} \arrow[d, two heads, "P"] \\ \mathcal{C} \arrow[r, "F"'] & \mathcal{D} \end{tikzcd}$$

We will check that this square is also a pullback in the category of κ–cocomplete quasicategories and κ–cocontinuous functors.

First, we verify that $\mathcal{P}$ has κ–small colimits. Let $X\colon K \to \mathcal{P}$ be a diagram with K κ–small. Let $S\colon K^{\triangleright} \to \mathcal{C}$ be a colimit of QX, then FS is a colimit of $FQX = PGX$ in $\mathcal{D}$. Lemma 2.12 implies that we can choose a colimit T of GX in $\mathcal{E}$ so that $PT = FS$. Then it follows by Lemma 2.11 that (S, T) is a colimit of $X = (QX, GX)$ in $\mathcal{P}$.

It remains to see that given a square

$$\begin{tikzcd} \mathcal{F} \arrow[r] \arrow[d] & \mathcal{E} \arrow[d, "P"] \\ \mathcal{C} \arrow[r, "F"'] & \mathcal{D} \end{tikzcd}$$

of κ–cocomplete quasicategories and κ–cocontinuous functors, the induced functor $\mathcal{F} \to \mathcal{P}$ preserves κ–small colimits. Indeed, this follows directly from Lemma 2.11. □

Theorem 2.14 *The category QCat_κ of small κ–cocomplete quasicategories with κ–cocontinuous functors as morphisms, categorical equivalences as weak equivalences and (κ–cocontinuous) inner isofibrations as fibrations is a fibration category.*

Proof By Theorem 2.4 it suffices to observe:

(1) A terminal quasicategory is also a terminal κ–cocomplete quasicategory (which is clear).

(2) A pullback (in the category of all quasicategories) of κ–cocomplete quasicategories and κ–cocontinuous functors one of which is an inner isofibration is also a pullback in the category of κ–cocomplete quasicategories, which follows by (the proof of) Proposition 2.13.

(3) For a κ–cocomplete quasicategory $\mathcal{C}$, the functor $\mathcal{C}^{E[1]} \to \mathcal{C} \times \mathcal{C}$ is a κ–cocontinuous functor between κ–cocomplete quasicategories. Indeed, $\mathcal{C}^{E[1]}$ is κ–cocomplete since it is categorically equivalent to $\mathcal{C}$ (by Lemmas 2.7 and 2.10) and $\mathcal{C} \times \mathcal{C}$ is κ–cocomplete by (2). Finally, $\mathcal{C}^{E[1]} \to \mathcal{C} \times \mathcal{C}$ preserves κ–small colimits by (2) since both projections $\mathcal{C}^{E[1]} \to \mathcal{C}$ do. □

3 The quasicategory of frames

In [23] we introduced a functor $\mathrm{N_f}\colon \mathsf{CofCat}_\kappa \to \mathsf{QCat}_\kappa$. Let us briefly recall the construction. For each m let $D[m]$ be the category of elements of $\Delta[m]$ with the face operators as morphisms. It comes equipped with a functor $p_{[m]}\colon D[m] \to [m]$ that evaluates a map $[k] \to [m]$ at m. We consider $D[m]$ as a homotopical category with weak equivalences created by $p_{[m]}$ (from the isomorphisms of $[m]$). Then for a cofibration category $\mathcal{C}$ we define a simplicial set $\mathrm{N_f}\mathcal{C}$ (called the *quasicategory of frames* in $\mathcal{C}$) whose m–simplices are homotopical, Reedy cofibrant diagrams $D[m] \to \mathcal{C}$. See [23, Section 2] for full details.

Theorem 3.1 *For a κ–cocomplete cofibration category $\mathcal{C}$, the simplicial set $\mathrm{N_f}\mathcal{C}$ is a κ–cocomplete quasicategory and $\mathrm{N_f}\colon \mathsf{CofCat}_\kappa \to \mathsf{QCat}_\kappa$ is an exact functor of fibration categories.*

Proof By [23, Theorem 2.3] $\mathrm{N_f}\mathcal{C}$ is a κ–cocomplete quasicategory. Moreover, [23, Propositions 3.5, 3.8 and 3.9] imply that $\mathrm{N_f}$ is indeed exact. □

The results of the last section heavily depend on the methods of [24; 23] which in turn involve a lot of notation useful in expressing properties of $\mathrm{N_f}\mathcal{C}$ in terms of various diagrams in $\mathcal{C}$. In this section, we recall some of that notation and prove a few auxiliary lemmas.

First of all, the categories $D[m]$ introduced above generalize to homotopical categories DK for all simplicial sets K. The underlying category of DK has all simplices of K as objects and face operators between them as morphisms. The weak equivalences in DK are induced from degenerate simplices of K in a manner described in [23, Section 2]. The following fact is a fundamental tool for translating between properties of $\mathcal{C}$ and $\mathrm{N_f}\mathcal{C}$.

Proposition 3.2 [23, Proposition 2.6] *Let $\mathcal{C}$ be a cofibration category and K a simplicial set. There is a natural bijection between*

- *the set of homotopical Reedy cofibrant diagrams $DK \to \mathcal{C}$, and*
- *the set of simplicial maps $K \to \mathrm{N_f}\mathcal{C}$.* □

Moreover, this construction admits useful variations most conveniently described in terms of marked simplicial complexes. A *marked simplicial complex* is a simplicial set K equipped with an embedding $K \hookrightarrow NP$, where P is a homotopical poset. In this case DK stands for the same category as above but with (possibly) richer homotopical structure, ie one created by the composite $DK \hookrightarrow DP \to P$. Here, DP stands for DNP and the latter functor evaluates an object $[k] \to P$ at k. $\operatorname{Sd} K$ stands for the homotopical poset defined as the full homotopical subcategory of DK spanned by the nondegenerate simplices of K. Diagrams over $\operatorname{Sd} K$ have the same homotopical content as diagrams over DK, as made precise by the following lemma.

Lemma 3.3 [23, Lemma 3.12] *Let $K \hookrightarrow L$ be an injective map of finite marked simplicial complexes (which means that it covers an injective homotopical map of the underlying homotopical posets). Then for every cofibration category $\mathcal{C}$ the inclusion $DK \cup \operatorname{Sd} L \hookrightarrow DL$ induces an acyclic fibration $\mathcal{C}_{\mathrm{R}}^{DL} \to \mathcal{C}_{\mathrm{R}}^{DK \cup \operatorname{Sd} L}$.* □

This lemma will be useful in various ways, for example in constructing $E[1]$–homotopies between maps into $\mathrm{N_f}\,\mathcal{C}$. An $E[1]$–homotopy $K \times E[1] \to \mathrm{N_f}\,\mathcal{C}$ corresponds to a homotopical Reedy cofibrant diagram $D(K \times E[1]) \to \mathcal{C}$. Moreover, [23, Corollary 3.7] says that in order to specify such a homotopy it is enough to give a diagram $D(K \times \widehat{[1]}) \to \mathcal{C}$. (Here, $\widehat{[1]}$ stands for the poset $[1]$ with all morphisms as weak equivalences.) These observations allow us to state and prove the following lemma.

Lemma 3.4 *Let $K \hookrightarrow L$ be an inclusion of marked simplicial complexes, X and Y homotopical Reedy cofibrant diagrams $DL \to \mathcal{C}$, and $f\colon X \,|\, \operatorname{Sd} L \to Y \,|\, \operatorname{Sd} L$ a natural weak equivalence such that $f \,|\, \operatorname{Sd} K$ is an identity transformation. Then X and Y are $E[1]$–homotopic relative to K as diagrams in $\mathrm{N_f}\,\mathcal{C}$.*

Proof By [23, Corollary 3.7] it suffices to construct a homotopical Reedy cofibrant diagram $D(L \times \widehat{[1]}) \to \mathcal{C}$ that restricts to $[X, Y]$ on $D(L \times \partial\Delta[1])$ and to the identity on $D(K \times \widehat{[1]})$, ie to a degenerate edge of $(\mathrm{N_f}\,\mathcal{C})^K$.

First, observe that we have a homotopical diagram $[f, \mathrm{id}]\colon (\operatorname{Sd} L \cup DK) \times \widehat{[1]} \to \mathcal{C}$ which is Reedy cofibrant when seen as a diagram $\operatorname{Sd} L \cup DK \to \mathcal{C}^{\widehat{[1]}}$. Hence by Lemma 3.3 it extends to a Reedy cofibrant diagram $DL \to \mathcal{C}^{\widehat{[1]}}$. We consider it as a diagram $DL \times \widehat{[1]} \to \mathcal{C}$ and pull it back to $D(L \times \widehat{[1]}) \to \mathcal{C}$. It restricts to $[X, Y]$ on $D(L \times \partial\Delta[1])$ and to the identity on $D(K \times \widehat{[1]})$. Thus it can be replaced Reedy cofibrantly relative to $D(L \times \partial\Delta[1] \cup K \times \widehat{[1]})$ by [23, Lemma 1.9], which finishes the proof. □

Another lemma that we will need says that up to equivalence all frames are Reedy cofibrant replacements of constant diagrams.

Lemma 3.5 *Any object of $X \in \mathrm{N_f}\,\mathcal{C}$ is equivalent to a Reedy cofibrant replacement of $p_{[0]}^* X_0$.*

Proof By [23, Lemma 3.2] there are homotopical functors $f\colon [0] \to D[0]$ and $s\colon D[0] \to D[0]$ such that $p_{[0]} f = \mathrm{id}_{[0]}$ and there are weak equivalences

$$\mathrm{id} \xrightarrow{\sim} s \xleftarrow{\sim} f p_{[0]}.$$

These equivalences evaluated at X form a diagram $D[0] \times \mathrm{Sd}\,[\widehat{1}] \to \mathcal{C}$ which we can pull back along $D[\widehat{1}] \to D[0] \times \mathrm{Sd}\,[\widehat{1}]$ and then replace Reedy cofibrantly to obtain a homotopical Reedy cofibrant diagram $Y\colon D[\widehat{1}] \to \mathcal{C}$ such that $Y\delta_1 = X$ by [23, Lemma 1.9]. By [23, Corollary 3.7] Y is an equivalence and by the construction $Y\delta_0$ is a Reedy cofibrant replacement of $p_{[0]}^* X_0$. □

Perhaps the most useful result of [23] characterizes universal cones $K^{\rhd} \to \mathrm{N_f}\,\mathcal{C}$ in terms of the corresponding diagram $D(K^{\rhd}) \to \mathcal{C}$. It comes in two versions depending on whether $\kappa > \aleph_0$ or $\kappa = \aleph_0$. First, we state it in the case of $\kappa > \aleph_0$.

Theorem 3.6 [23, Theorem 4.6] *Let $\mathcal{C}$ be a κ–cocomplete cofibration category, K a κ–small simplicial set and $S\colon K^{\rhd} \to \mathrm{N_f}\,\mathcal{C}$. Then S is universal as a cone under $S|K$ if and only if the induced morphism*

$$\operatorname{colim}_{DK} S \to \operatorname{colim}_{D(K^{\rhd})} S$$

is a weak equivalence (with S seen, by Proposition 3.2, as a homotopical Reedy cofibrant diagram $D(K^{\rhd}) \to \mathcal{C}$). □

Observe that the assumption $\kappa > \aleph_0$ is necessary for the colimits in the statement of the theorem to exist. If $\kappa = \aleph_0$, then K is a finite simplicial set, but DK is still infinite (unless K is empty). This problem makes both the statement and the proof more technical in the case of $\kappa = \aleph_0$.

We filter the category DK by finite subcategories

$$D^{(0)}K \hookrightarrow D^{(1)}K \hookrightarrow D^{(2)}K \hookrightarrow \cdots$$

as described in detail in [23, Section 5]. Then given a homotopical Reedy cofibrant diagram $X\colon DK \to \mathcal{C}$ the colimits of its restrictions to all $D^{(k)}K$ exist. The homotopy type of these colimits stabilizes for k sufficiently large and this stable value is the homotopy colimit of X. This allows us to state the remaining case of the theorem.

Theorem 3.7 [23, Theorem 5.12] *Let $\mathcal{C}$ be a cofibration category and K a finite simplicial set. A cone $S\colon K^{\triangleright} \to \mathrm{N_f}\,\mathcal{C}$ is universal if and only if the induced morphism*

$$\operatorname{colim}_{D^{(k)}K} S \to \operatorname{colim}_{D^{(k)}(K^{\triangleright})} S$$

is a weak equivalence for k sufficiently large (where S is seen as a homotopical Reedy cofibrant diagram $D(K^{\triangleright}) \to \mathcal{C}$ by Proposition 3.2). □

Both these theorems will be instrumental in the proof of our main result.

4 Cofibration categories of diagrams in quasicategories

In this section we will prove our main result, ie that $\mathrm{N_f}$ is a weak equivalence of fibration categories. This will be achieved by defining a functor Dg_κ from the category of κ–cocomplete quasicategories to the category of κ–cocomplete cofibration categories. The functor Dg_κ fails to be exact (eg it does not preserve the terminal object), but it will be verified to induce an inverse to $\mathrm{N_f}$ on the level of homotopy categories which is sufficient to complete the proof.

Definition 4.1 Let sSet_κ denote the category of κ–small simplicial sets. If $\mathcal{C}$ is a κ–cocomplete quasicategory we consider the slice category $\mathrm{sSet}_\kappa \downarrow \mathcal{C}$, we denote it by $\mathrm{Dg}_\kappa\, \mathcal{C}$ and call it the *category of κ–small diagrams in* $\mathcal{C}$. Then we define a morphism

$$\begin{tikzcd} K \arrow[rr, "f"] \arrow[dr, "X"'] & & L \arrow[dl, "Y"] \\ & \mathcal{C} & \end{tikzcd}$$

to be

- a *weak equivalence* if the induced morphism $\operatorname{colim}_K X \to \operatorname{colim}_L Y$ is an equivalence in $\mathcal{C}$ (more precisely, if for any universal cone $S\colon L^{\triangleright} \to \mathcal{C}$ under Y the induced cone $Sf^{\triangleright}$ is universal under X),
- a *cofibration* if f is injective.

In particular, a morphism of $\mathrm{Dg}_\kappa\, \mathcal{C}$ as above is a weak equivalence whenever f is cofinal, but there are of course many weak equivalences with f not cofinal. We will make use of the class of *right anodyne maps*, which is generated by the *right horn* inclusions $\Lambda^i[m] \hookrightarrow \Delta[m]$ (ie the ones with $0 < i \le m$) under coproducts, pushouts along arbitrary maps, sequential colimits and retracts.

Proposition 4.2 *With weak equivalences and cofibrations as defined above* $\mathrm{Dg}_\kappa\,\mathcal{C}$ *is a* κ*–cocomplete cofibration category.*

Proof (C0) Weak equivalences satisfy 2-out-of-6 since equivalences in $\mathcal{C}$ do.

(C1) Isomorphisms are weak equivalences since isomorphisms of simplicial sets are cofinal.

(C2)–(C3) The empty diagram is an initial object and hence every object is cofibrant.

(C4) Pushouts are created by the forgetful functor $\mathrm{Dg}_\kappa\,\mathcal{C} \to \mathsf{sSet}_\kappa$, thus pushouts along cofibrations exist and cofibrations are stable under pushouts. By [20, Lemma 1.4.3(1)] it suffices to verify that the gluing lemma holds, which follows by [16, Proposition 4.4.2.2].

(C5) It will suffice to verify that in the usual mapping cylinder factorization

$$K \to Mf \to L$$

the second map is cofinal. Indeed, we have a diagram

$$\begin{tikzcd}
K \times \Delta[0] \arrow[r, "f"] \arrow[d, "K \times \delta_0"'] & L \arrow[d, "j"'] \arrow[ddr, bend left, "\mathrm{id}_L"] & \\
K \times \Delta[1] \arrow[r] & Mf \arrow[dr] & \\
& & L
\end{tikzcd}$$

where the square is a pushout. The map $K \times \delta_0$ is right anodyne by [14, Theorem 2.17] and thus so is j. Hence it is cofinal by [16, Proposition 4.1.1.3(4)].

(C6)–(C7-κ) The proof is similar to that of (C4). (But there is no analogue of [16, Proposition 4.4.2.2] for sequential colimits explicitly stated in [16]. Instead, it follows from the more general [16, Proposition 4.2.3.10 and Remark 4.2.3.9].) □

Lemma 4.3 *A* κ*–cocontinuous functor* $F\colon \mathcal{C} \to \mathcal{D}$ *induces a* κ*–cocontinuous functor* $\mathrm{Dg}_\kappa\, F = \mathrm{Dg}_\kappa\,\mathcal{C} \to \mathrm{Dg}_\kappa\,\mathcal{D}$ *and thus we obtain a functor* $\mathrm{Dg}_\kappa\colon \mathsf{QCat}_\kappa \to \mathsf{CofCat}_\kappa$.

Proof Colimits in both $\mathrm{Dg}_\kappa\,\mathcal{C}$ and $\mathrm{Dg}_\kappa\,\mathcal{D}$ are created in sSet_κ and thus are preserved by $\mathrm{Dg}_\kappa\, F$. Cofibrations are clearly preserved and so are weak equivalences since F preserves κ–small colimits. □

For the moment, we focus on the case of $\kappa > \aleph_0$. The case of $\kappa = \aleph_0$ will be dealt with later.

Definition 4.4 For a κ–cocomplete cofibration category $\mathcal{C}$ we define a functor

$$\Phi_{\mathcal{C}}\colon \mathrm{Dg}_\kappa \mathrm{N_f}\, \mathcal{C} \to \mathcal{C}$$

by sending a diagram $X\colon K \to \mathrm{N_f}\, \mathcal{C}$ to $\operatorname{colim}_{DK} X$.

Observe that DK is κ–small since K is and $\kappa > \aleph_0$, so the colimit used in this definition exists in $\mathcal{C}$. It is clear that $\Phi_{\mathcal{C}}$ is a functor. While we may not be able to choose colimits so that $\Phi_{\mathcal{C}}$ is natural in $\mathcal{C}$, it is pseudonatural, ie natural up to coherent natural isomorphism.

Lemma 4.5 *The functor $\Phi_{\mathcal{C}}$ is κ–cocontinuous and a weak equivalence.*

Proof Preservation of cofibrations follows by [20, Theorem 9.4.1(1a)] since if $K \hookrightarrow L$ is an injective map of simplicial sets, then the induced functor $DK \hookrightarrow DL$ is a sieve. Colimits in $\mathcal{C}$ are compatible with colimits of indexing categories and thus $\Phi_{\mathcal{C}}$ is κ–cocontinuous. (Preservation of weak equivalences follows from the argument below.)

To see that it is a weak equivalence, it is enough to verify the approximation properties of Proposition 1.8. Lemma 4.1 of [23] and Theorem 3.6 imply that a morphism f in $\mathrm{Dg}_\kappa \mathrm{N_f}\, \mathcal{C}$ is a weak equivalence if and only if $\Phi_{\mathcal{C}} f$ is. Therefore $\Phi_{\mathcal{C}}$ preserves weak equivalences and satisfies (App1). It remains to check (App2), but it follows directly from [23, Lemma 4.2]. □

Next, we need a functor $\mathcal{D} \to \mathrm{N_f}\, \mathrm{Dg}_\kappa\, \mathcal{D}$ for every κ–cocomplete quasicategory $\mathcal{D}$. Let's start with unraveling the definition of $\mathrm{N_f}\, \mathrm{Dg}_\kappa\, \mathcal{D}$.

An m–simplex of $\mathrm{N_f}\, \mathrm{Dg}_\kappa\, \mathcal{D}$ consists of a Reedy cofibrant diagram $K\colon D[m] \to \mathsf{sSet}_\kappa$ and for each $\varphi \in D[m]$ a diagram $X_\varphi\colon K_\varphi \to \mathcal{D}$. These diagrams are compatible with each other in the sense that they form a cone under K with the vertex $\mathcal{D}$. Moreover, the entire structure is homotopical as a diagram in $\mathrm{Dg}_\kappa\, \mathcal{D}$, ie if $\varphi, \psi \in D[m]$ and $\chi\colon \varphi \to \psi$ is a weak equivalence, then the induced morphism $\operatorname{colim}_{K_\varphi} X_\varphi \to \operatorname{colim}_{K_\psi} X_\psi$ is an equivalence in $\mathcal{D}$.

If $\mu\colon [n] \to [m]$, then $(K, X)\mu = (K\mu, X\mu)$ is defined simply by $(K\mu)_\varphi = K_{\mu\varphi}$ and $(X\mu)_\varphi = X_{\mu\varphi}$.

We can now define a functor $\Psi_{\mathcal{D}}\colon \mathcal{D} \to \mathrm{N_f}\, \mathrm{Dg}_\kappa\, \mathcal{D}$ as follows.

Definition 4.6 For $x \in \mathcal{D}_m$ we set the underlying simplicial diagram of $\Psi_{\mathcal{D}} x$ to $\varphi \mapsto \Delta[k]$, where $\varphi\colon [k] \to [m]$, and the corresponding diagram in $\mathcal{D}$ to $x\varphi\colon \Delta[k] \to \mathcal{D}$. Then $\Psi_{\mathcal{D}} x$ is homotopical as a diagram $D[m] \to \mathrm{Dg}_\kappa\, \mathcal{D}$ since any weak equivalence in $D[m]$ induces a right anodyne (and hence cofinal by [16, Proposition 4.1.1.3(4)]) map of simplices.

Clearly, $\Psi_{\mathcal{D}}$ is a functor and is natural in $\mathcal{D}$. We check that it is also a categorical equivalence.

Proposition 4.7 *For every κ–cocomplete quasicategory $\mathcal{D}$ the functor $\Psi_{\mathcal{D}}$ is a categorical equivalence.*

Proof Consider a square as follows:

$$\begin{tikzcd} \partial\Delta[m] \arrow[r, "x"] \arrow[d] & \mathcal{D} \arrow[d, "\Psi_{\mathcal{D}}"] \\ \Delta[m] \arrow[r, "Y"'] & \mathrm{N_f}\,\mathrm{Dg}_\kappa\,\mathcal{D} \end{tikzcd}$$

By Lemma 2.3 it will be enough to extend x to a simplex $\hat{x}\colon \Delta[m] \to \mathcal{D}$ and construct an $E[1]$–homotopy from $\Psi_{\mathcal{D}}\hat{x}$ to Y relative to $\partial\Delta[m]$.

Let's start by finding $\hat{x}$. Consider $Y_{[m]}\colon A_{[m]} \to \mathcal{D}$. Since Y agrees with $\Psi_{\mathcal{D}}x$ over $\partial\Delta[m]$ the $[m]^{\mathrm{th}}$ latching object of Y is $x\colon \partial\Delta[m] \to \mathcal{D}$, ie we have an induced injective map $\partial\Delta[m] \hookrightarrow A_{[m]}$ and $Y_{[m]}|\partial\Delta[m] = x$. Choose a universal cone

$$\widetilde{Y}_{[m]}\colon A_{[m]}^{\rhd} \to \mathcal{D}$$

under $Y_{[m]}$ and consider $\widetilde{Y}_{[m]}|\partial\Delta[m]^{\rhd}$. We have

$$\partial\Delta[m]^{\rhd} \cong \Lambda^{m+1}[m+1]$$

which is an outer horn. However, $\widetilde{Y}_{[m]}|\partial\Delta[m]^{\rhd}$ is special since $\Psi_{\mathcal{D}}x$ is homotopical, and thus extends to $z\colon \Delta[m]^{\rhd} \to \mathcal{D}$ by [14, Theorem 4.13]. We set $\hat{x} = z|\Delta[m]$.

By Proposition 3.2, finding an $E[1]$–homotopy from $\Psi_{\mathcal{D}}\hat{x}$ to Y translates into constructing a homotopical Reedy cofibrant diagram $D([m] \times E(1)) \to \mathrm{Dg}_\kappa\,\mathcal{D}$ restricting to $[\Psi_{\mathcal{D}}\hat{x}, Y]$ on $D(\Delta[m] \times \partial\Delta[1])$. By [23, Corollary 3.7] it will be sufficient to construct such a diagram on $D([m] \times \widehat{[1]})$ and by Lemma 3.3 it will suffice to define it on $\mathrm{Sd}([m] \times \widehat{[1]})$.

We form a pushout on the left in $\mathrm{Dg}_\kappa\,\mathcal{D}$:

$$\begin{tikzcd} \widetilde{Y}|\partial\Delta[m]^{\rhd} \arrow[r, tail] \arrow[d, tail] & \widetilde{Y} \arrow[d, tail] \\ z \arrow[r, tail] & Z \end{tikzcd} \qquad \begin{tikzcd} \partial\Delta[m] \arrow[r, hook] \arrow[d, hook] & A_{[m]} \arrow[d, hook] \\ \Delta[m] \arrow[r, hook] & B \end{tikzcd}$$

Its underlying square of simplicial sets is $(-)^{\rhd}$ applied to the square on the right.

This yields the following sequence of morphisms of $\mathrm{Dg}_\kappa\,\mathcal{D}$ (with morphisms of the underlying simplicial sets displayed below):

$$\begin{array}{ccccccccc} \widehat{x} & \rightarrowtail & z & \rightarrowtail & Z & \leftarrowtail & \widetilde{Y}_{[m]} & \leftarrowtail & Y_{[m]} \\ \Delta[m] & \hookrightarrow & \Delta[m]^{\triangleright} & \hookrightarrow & B^{\triangleright} & \hookleftarrow & A^{\triangleright}_{[m]} & \hookleftarrow & A_{[m]} \end{array}$$

The first morphism is a weak equivalence since z is a filler of a special horn. So are the middle two since the underlying maps of simplicial sets preserve the cone points. The last one is also a weak equivalence since $\widetilde{Y}_{[m]}$ is universal. All these morphisms are maps of cones under $Y\,|\,\mathrm{Sd}\,\partial\Delta[m] = \Psi_{\mathcal{D}}x\,|\,\mathrm{Sd}\,\partial\Delta[m]$ and hence can be seen as transformations of diagrams over $\mathrm{Sd}[m]$ which restrict to identities over $\mathrm{Sd}\,\partial\Delta[m]$. The conclusion follows by Lemma 3.4. $\square$

Before we can prove the main theorem we need to know the following:

Lemma 4.8 *The functor* Dg_κ *is homotopical.*

Proof We begin by constructing a natural equivalence $\Theta_{\mathcal{C}}\colon \mathrm{Ho}\,\mathrm{N_f}\,\mathcal{C} \to \mathrm{Ho}\,\mathcal{C}$ for every cofibration category $\mathcal{C}$. We send an object $X\colon D[0] \to \mathcal{C}$ to X_0 and a morphism $Y\colon D[1] \to \mathcal{C}$ to the composite $[\upsilon_1]^{-1}[\upsilon_0]$, where υ_0 and υ_1 are the structure morphisms

$$Y_0 \xrightarrow{\upsilon_0} Y_{01} \underset{\sim}{\xleftarrow{\upsilon_1}} Y_1.$$

This assignment is well-defined and functorial since $\mathcal{C}$ has homotopy calculus of fractions, see Theorem 1.4.

We check that $\Theta_{\mathcal{C}}$ is an equivalence. It is surjective and full since both $\mathrm{Sd}[0] \hookrightarrow D[0]$ and $D\partial\Delta[1] \cup \mathrm{Sd}[1] \hookrightarrow D[1]$ have the Reedy left lifting property with respect to all cofibration categories by Lemma 3.3. For faithfulness, consider $X, \widetilde{X}\colon D[1] \to \mathcal{C}$ such that $X\,|\,D\partial\Delta[1] = \widetilde{X}\,|\,D\partial\Delta[1]$ and $\Theta_{\mathcal{C}}(X) = \Theta_{\mathcal{C}}(\widetilde{X})$. Since we have already verified that $\Theta_{\mathcal{C}}$ is essentially surjective, Lemma 3.5 allows us to assume that $X\delta_0$ is a Reedy cofibrant replacement of $p^*_{[0]}X_1$ so that the structure morphisms of X fit into a cylinder

$$X_1 \amalg X_1 \rightarrowtail X_{11} \xrightarrow{\sim} X_1.$$

By Theorem 1.4(2) we have a diagram

$$\begin{tikzcd} & X_{01} \arrow[d, "\varphi"', "\sim"] & \\ X_0 \arrow[ur, tail] \arrow[dr, tail] & Y & X_1 \arrow[ul, tail, "\sim"'] \arrow[l, phantom, "\nu"] \arrow[dl, tail, "\sim"] \\ & \widetilde{X}_{01} \arrow[u, "\widetilde{\varphi}", "\sim"'] & \end{tikzcd}$$

where both squares commute up to left homotopy. By Lemma 1.5 we can assume that the left square commutes strictly. Let

$$\begin{tikzcd}
X_1 \amalg X_1 \arrow[r, "{[\varphi\nu, \widetilde{\varphi}\widetilde{\nu}]}"] \arrow[d, tail, "{[\delta_0, \delta_1]}"'] & Y \arrow[d, tail, "\psi"', "\sim"] \\
X_{11} \arrow[r, "\chi"'] & \widehat{Y}
\end{tikzcd}$$

be a left homotopy. Then we can form a diagram

$$\begin{tikzcd}
 & & X_1 \arrow[dl, "\sim"', "\nu"] \arrow[dr, "\sim"] & & \\
 & X_{01} \arrow[dr, "\psi\varphi", "\sim"'] & & X_{11} \arrow[dl, "\sim"', "\chi"] & \\
 & & \widehat{Y} & & \\
X_0 \arrow[uur] \arrow[rr] & & \widetilde{X}_{01} \arrow[u, "\psi\widetilde{\varphi}", "\sim"'] & & X_1 \arrow[ll, "\widetilde{\nu}"', "\sim"] \arrow[uul, "\sim"']
\end{tikzcd}$$

which is a homotopical diagram on Sd[2] and Reedy cofibrant over $\operatorname{Sd} \partial\Delta[2]$. Thus it can be replaced Reedy cofibrantly without modifying it over $\operatorname{Sd} \partial\Delta[2]$ by [23, Lemma 1.9]. Then X, $\widetilde{X}$ and $X\delta_0\sigma_0$ provide an extension over $D\partial\Delta[2]$. We know that the inclusion $D\partial\Delta[2] \cup \operatorname{Sd}[2] \hookrightarrow D[2]$ has the Reedy left lifting property with respect to all cofibration categories by Lemma 3.3, so we can find an extension to $D[2]$ which is a homotopy between X and $\widetilde{X}$ in $\mathrm{N_f}\,\mathcal{C}$.

Since equivalences of quasicategories induce equivalences of homotopy categories, it follows that $\mathrm{N_f}$ reflects equivalences. Thus Dg_κ is homotopical by Proposition 4.7. □

Finally, we are ready to prove the main theorem.

Theorem 4.9 *The functor* $\mathrm{N_f}\colon \mathsf{CofCat}_\kappa \to \mathsf{QCat}_\kappa$ *is a weak equivalence of fibration categories.*

Proof (for $\kappa > \aleph_0$) The functor Dg_κ is homotopical by Lemma 4.8 and thus induces a functor on the homotopy categories. Since Ψ is a natural categorical equivalence by Proposition 4.7 the induced transformation $\operatorname{Ho}\Psi$ is a natural isomorphism $\mathrm{id} \to (\operatorname{Ho}\mathrm{N_f})(\operatorname{Ho}\mathrm{Dg}_\kappa)$. The transformation Φ is merely pseudonatural, but natural isomorphisms of exact functors induce right homotopies in CofCat_κ (by the construction of path objects in the proof of [24, Theorem 2.8]). Therefore $\operatorname{Ho}\Phi$ is a natural transformation and by Lemma 4.5 it is an isomorphism $(\operatorname{Ho}\operatorname{Dg}_\kappa)(\operatorname{Ho}\mathrm{N_f}) \to \mathrm{id}$. Hence $\operatorname{Ho}\mathrm{N_f}$ is an equivalence. □

The only part of the argument above that does not work for $\kappa = \aleph_0$ is the construction of a natural weak equivalence $\Phi_{\mathcal{C}}\colon \operatorname{Dg}_\kappa \mathrm{N_f}\,\mathcal{C} \to \mathcal{C}$ for every cofibration category $\mathcal{C}$. Indeed, $\Phi_{\mathcal{C}}$ was defined using colimits over categories DK which are infinite even for finite simplicial sets K. Instead, we will define a zig-zag of (pseudonatural) weak equivalences connecting $\operatorname{Dg}_{\aleph_0} \mathrm{N_f}\,\mathcal{C}$ to $\mathcal{C}$, namely,

$$\operatorname{Dg}_{\aleph_0} \mathrm{N_f}\,\mathcal{C} \xrightarrow{\Phi_{\mathcal{C}}^{(-)}} \mathcal{C}_{\mathrm{R}}^{\widetilde{\mathbb{N}}} \hookleftarrow \mathcal{C}_{\mathrm{R}}^{\widehat{\mathbb{N}}} \xrightarrow{\mathrm{ev}_0} \mathcal{C}.$$

Here, $\widehat{\mathbb{N}}$ is the homotopical poset of natural numbers with all morphisms as weak equivalences so that $\mathcal{C}_{\mathrm{R}}^{\widehat{\mathbb{N}}}$ is the cofibration category of Reedy cofibrant homotopically constant sequences. Similarly, $\mathcal{C}_{\mathrm{R}}^{\widetilde{\mathbb{N}}}$ stands for the cofibration category of Reedy cofibrant eventually homotopically constant sequences; see [23, Section 5] for details.

It was verified in [23, Lemma 5.9] that $\mathcal{C}_{\mathrm{R}}^{\widehat{\mathbb{N}}} \hookrightarrow \mathcal{C}_{\mathrm{R}}^{\widetilde{\mathbb{N}}}$ is a weak equivalence. Moreover, $\mathrm{ev}_0\colon \mathcal{C}_{\mathrm{R}}^{\widehat{\mathbb{N}}} \to \mathcal{C}$ is induced by a homotopy equivalence $[0] \to \widehat{\mathbb{N}}$ hence it is a weak equivalence, too.

It remains to define $\Phi_{\mathcal{C}}^{(-)}$ and prove that it is also a weak equivalence. For each k and an object $X\colon DK \to \mathrm{N_f}\,\mathcal{C}$ we set $\Phi_{\mathcal{C}}^{(k)} X = \operatorname{colim}_{D^{(k)}K} X$. This colimit exists since $D^{(k)}K$ is finite if K is finite.

Lemma 4.10 *For a cofibration category $\mathcal{C}$ the formula above defines an exact functor $\Phi_{\mathcal{C}}^{(-)}\colon \operatorname{Dg}_{\aleph_0} \mathrm{N_f}\,\mathcal{C} \to \mathcal{C}_{\mathrm{R}}^{\widetilde{\mathbb{N}}}$. Moreover, it is a weak equivalence.*

Proof First, we need to verify that $\Phi_{\mathcal{C}}^{(-)} X$ is an eventually constant sequence for all $(K, X) \in \operatorname{Dg}_{\aleph_0} \mathrm{N_f}\,\mathcal{C}$. Consider X as a diagram in $\mathrm{N_f}\,\mathcal{C}$ and choose a universal cone $S\colon K^{\rhd} \to \mathrm{N_f}\,\mathcal{C}$. Then [23, Lemma 4.8] implies that $\Phi_{\mathcal{C}}^{(-)} S$ is eventually constant and Theorem 3.7 implies that the induced morphism $\Phi_{\mathcal{C}}^{(-)} S \to \Phi_{\mathcal{C}}^{(-)} S$ is an eventual weak equivalence. Thus $\Phi_{\mathcal{C}}^{(-)} S$ is eventually constant.

Preservation of cofibrations follows by [20, Theorem 9.4.1(1a)] since if $K \hookrightarrow L$ is an injective map of simplicial sets, then the induced functors $D^{(k)}K \cup D^{(k-1)}L \to D^{(k)}L$ are sieves. Colimits in $\mathcal{C}$ are compatible with colimits of indexing categories and thus $\Phi_{\mathcal{C}}^{(-)}$ is exact. (Preservation of weak equivalences follows from the argument below.)

To see that it is a weak equivalence, it is enough to verify the approximation properties of Proposition 1.8. Theorem 3.7 and [23, Lemma 5.8] imply that a morphism f in $\operatorname{Dg}_{\aleph_0} \mathrm{N_f}\,\mathcal{C}$ is a weak equivalence if and only if $\Phi_{\mathcal{C}}^{(-)} f$ is an eventual weak equivalence. Therefore $\Phi_{\mathcal{C}}^{(-)}$ preserves weak equivalences and satisfies (App1). It remains to check (App2), but it follows directly from [23, Lemma 5.10]. □

This yields the proof of Theorem 4.9 in the remaining case of $\kappa = \aleph_0$ since the three weak equivalences described above induce a natural isomorphism

$$(\mathrm{Ho}\,\mathrm{Dg}_\kappa)(\mathrm{Ho}\,\mathrm{N_f}) \to \mathrm{id}$$

and the rest of the argument applies verbatim.

References

[1] **C Barwick**, **D M Kan**, *A characterization of simplicial localization functors and a discussion of DK equivalences*, Indag. Math. 23 (2012) 69–79 MR

[2] **C Barwick**, **D M Kan**, *Relative categories: another model for the homotopy theory of homotopy theories*, Indag. Math. 23 (2012) 42–68 MR

[3] **C Barwick**, **C Schommer-Pries**, *On the unicity of the homotopy theory of higher categories*, preprint (2013) arXiv

[4] **J E Bergner**, *Three models for the homotopy theory of homotopy theories*, Topology 46 (2007) 397–436 MR

[5] **J M Boardman**, **R M Vogt**, *Homotopy invariant algebraic structures on topological spaces*, Lecture Notes in Math. 347, Springer, Berlin (1973) MR

[6] **K S Brown**, *Abstract homotopy theory and generalized sheaf cohomology*, Trans. Amer. Math. Soc. 186 (1973) 419–458 MR

[7] **D-C Cisinski**, *Catégories dérivables*, Bull. Soc. Math. France 138 (2010) 317–393 MR

[8] **D Dugger**, **D I Spivak**, *Mapping spaces in quasi-categories*, Algebr. Geom. Topol. 11 (2011) 263–325 MR

[9] **W G Dwyer**, **D M Kan**, **J H Smith**, *Homotopy commutative diagrams and their realizations*, J. Pure Appl. Algebra 57 (1989) 5–24 MR

[10] **P Gabriel**, **M Zisman**, *Calculus of fractions and homotopy theory*, Ergeb. Math. Grenzgeb., Springer, New York (1967) MR

[11] **P G Goerss**, **J F Jardine**, *Simplicial homotopy theory*, Progress in Mathematics 174, Birkhäuser, Basel (1999) MR

[12] **A Hirschowitz**, **C Simpson**, *Descente pour les n–champs*, preprint (2001) arXiv

[13] **M Hovey**, *Model categories*, Mathematical Surveys and Monographs 63, Amer. Math. Soc., Providence, RI (1999) MR

[14] **A Joyal**, *The theory of quasi-categories and its applications*, from "Advanced course on simplicial methods in higher categories", Quaderns 45, CRM, Barcelona (2008) 149–496

[15] **A Joyal**, **M Tierney**, *Quasi-categories vs Segal spaces*, from "Categories in algebra, geometry and mathematical physics" (A Davydov, M Batanin, M Johnson, S Lack, A Neeman, editors), Contemp. Math. 431, Amer. Math. Soc., Providence, RI (2007) 277–326 MR

[16] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton University Press (2009) MR

[17] **D G Quillen**, *Homotopical algebra*, Lecture Notes in Math. 43, Springer, Berlin (1967) MR

[18] **C Rezk**, *A model for the homotopy theory of homotopy theory*, Trans. Amer. Math. Soc. 353 (2001) 973–1007 MR

[19] **E Riehl**, **D Verity**, *Completeness results for quasi-categories of algebras, homotopy limits, and related general constructions*, Homology Homotopy Appl. 17 (2015) 1–33 MR

[20] **A Rădulescu-Banu**, *Cofibrations in homotopy theory*, preprint (2009) arXiv

[21] **K Szumiło**, *Two models for the homotopy theory of cocomplete homotopy theories*, PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn (2014) Available at `http://hss.ulb.uni-bonn.de/2014/3692/3692.htm`

[22] **K Szumiło**, *Two models for the homotopy theory of cocomplete homotopy theories*, preprint (2014) arXiv

[23] **K Szumiło**, *Frames in cofibration categories*, J. Homotopy Relat. Struct. (2016) 1–40

[24] **K Szumiło**, *Homotopy theory of cofibration categories*, Homology Homotopy Appl. 18 (2016) 345–357 MR

[25] **B Toën**, *Vers une axiomatisation de la théorie des catégories supérieures*, K–Theory 34 (2005) 233–263 MR

[26] **R M Vogt**, *The HELP-lemma and its converse in Quillen model categories*, J. Homotopy Relat. Struct. 6 (2011) 115–118 MR

[27] **F Waldhausen**, *Algebraic K–theory of spaces*, from "Algebraic and geometric topology" (A Ranicki, N Levitt, F Quinn, editors), Lecture Notes in Math. 1126, Springer, Berlin (1985) 318–419 MR

Department of Mathematics, University of Western Ontario
1151 Richmond Street, London ON N6A 3K7, Canada

`kszumilo@uwo.ca`

`http://www.math.uwo.ca/~kszumilo`

Received: 29 June 2015 Revised: 30 April 2016

Algebraic & Geometric Topology 17 (2017) 793–830

Turaev genus and alternating decompositions

CODY W ARMOND
ADAM M LOWRANCE

We prove that the genus of the Turaev surface of a link diagram is determined by a graph whose vertices correspond to the boundary components of the maximal alternating regions of the link diagram. Furthermore, we use these graphs to classify link diagrams whose Turaev surface has genus one or two, and we prove that similar classification theorems exist for all genera.

57M25; 57M27

1 Introduction

The discovery of the Jones polynomial [17] led to the resolution of the famous Tait conjectures. In particular, Kauffman [18], Murasugi [26], and Thistlethwaite [31] use the Jones polynomial to prove that an alternating diagram of a link with no nugatory crossings has the fewest possible number of crossings. In Turaev's [32] alternate proof of this result, he associates a closed oriented surface to each link diagram D, now known as the Turaev surface of D. Let D be a diagram of a nonsplit link L with $c(D)$ crossings, let $V_L(t)$ be the Jones polynomial of L, and let $g_T(D)$ be the genus of the Turaev surface of D. Turaev shows that

$$\operatorname{span} V_L(t) + g_T(D) \leq c(D). \tag{1-1}$$

In recent years, the Turaev surface has been shown to have further connections to the Jones polynomial (see Dasbach et al [11; 12]), Khovanov homology (Champanerkar, Kofman and Stoltzfus [10], Dasbach and Lowrance [14]), and knot Floer homology (Lowrance [23], Dasbach and Lowrance [13]).

Thistlethwaite [31] uses a decomposition of a link diagram into maximal alternating pieces to compute a lower bound on crossing number similar to inequality (1-1). Consider a link diagram D as a 4–valent plane graph with over/under decorations at the vertices. An edge or face of D should be understood to refer to an edge or face of the 4–valent plane graph. An edge of D is called *nonalternating* if both of its endpoints are overstrands or both of its endpoints are understrands. An edge is called *alternating* if one of its endpoints is an overstrand and the other is an understrand. Mark each nonalternating edge of D with two distinct points, and in each face of D connect those

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.793

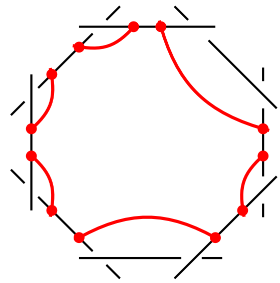


Figure 1: Each nonalternating edge is marked with two points. Inside of each face, draw arcs that connect marked points that are adjacent on the boundary but do not lie on the same edge of D.

marked points with arcs as depicted in Figure 1. This process results in a collection of pairwise disjoint simple closed curves $\{\gamma_1, \dots, \gamma_k\}$. The pair $(D, \{\gamma_1, \dots, \gamma_k\})$ is called the *alternating decomposition* of D.

Thistlethwaite associates to D a graph G, which we call the *alternating decomposition graph* of D, as follows. Suppose that D is a connected link diagram, ie when D is considered as a graph, it is a connected graph. If D is an alternating diagram, then G is a single vertex with no edges. Otherwise, the vertices of G are in one-to-one correspondence with the curves $\gamma_1, \dots, \gamma_k$ of the alternating decomposition of D. The edges of G are in one-to-one correspondence with the nonalternating edges of D. Let v_i and v_j be vertices of G corresponding to curves γ_i and γ_j respectively. An edge of G connects v_i to v_j if and only if the corresponding nonalternating edge of D intersects both γ_i and γ_j. If D is not a connected link diagram, then G is the disjoint union of the alternating decomposition graphs of its connected components.

The plane embedding of D induces an embedding of each component of G onto a sphere, as described in Section 3. Since each component of G can be embedded on a sphere, the graph G is planar. Whenever we refer to G with the sphere embeddings of its components induced by D, we use the notation $\mathbb{G}$ and call it the *sphere embedding induced by* D. We also consider $\mathbb{G}$ as an oriented ribbon graph of genus zero. See Section 3 for further discussion on oriented ribbon graphs. Each edge of G can be labeled as $+$ or $-$ according to whether it corresponds to an overstrand edge of D or an understrand edge of D respectively. Since the edges in each face of $\mathbb{G}$ rotate between $+$ and $-$ edges, it follows that every face has an even number of edges in its boundary. Therefore G is bipartite. Also, since every curve γ_i encloses a tangle, it follows that every vertex of G has even degree. Proposition 3.3 shows that a graph is an alternating decomposition graph if and only if it is planar, bipartite, and each vertex has even degree. See Section 3 for examples of alternating decompositions of link diagrams and their associated alternating decomposition graphs.

If D has alternating decomposition curves $\{\gamma_1, \dots, \gamma_k\}$, then an *alternating region* of D is a component of $S^2 - \{\gamma_1, \dots, \gamma_k\}$ that contains crossings of D. As the name suggests, if one follows a strand inside of an alternating region of D, then the crossings will alternate between over and under. Let $r_{\mathrm{alt}}(D)$ be the number of alternating regions in the alternating decomposition of D, and let $e(G)$ be the number of edges in G. Note that $e(G)$ is also the number of nonalternating edges in D. Thistlethwaite [31] proves that if D is a connected diagram of the link L, then

$$\operatorname{span} V_L(t) - r_{\mathrm{alt}}(D) + \tfrac{1}{2}e(G) + 1 \le c(D). \tag{1-2}$$

Bae and Morton [6] use Thistlethwaite's approach to study the extreme terms and the coefficients of the extreme terms in the Jones polynomial. Using combinatorial data from the planar dual of $\mathbb{G}$, a graph they call the *nonalternating spine* of D, they recover inequality (1-1) and show that it is a stronger bound than inequality (1-2).

In this paper, we use Thistlethwaite's alternating decompositions to study the Turaev surface of a link diagram. We show that the genus of the Turaev surface of a link diagram is determined by its alternating decomposition graph. If the Turaev surface is disconnected, then its genus refers to the sum of the genera of its connected components.

Theorem 1.1 *If D_1 and D_2 are link diagrams with isomorphic alternating decomposition graphs, then $g_T(D_1) = g_T(D_2)$.*

Champanerkar and Kofman [8] prove a version of Theorem 1.1 in the case where the two link diagrams are related by a rational tangle replacement. Lowrance [24] uses this special case to compute the Turaev genus of the $(3,q)$–torus links and of many other closed 3–braids; see also Abe and Kishimoto [2].

The *Turaev genus* of an alternating decomposition graph G, denoted $g_T(G)$, is defined to be $g_T(D)$, where D is a link diagram with alternating decomposition graph G. Corollary 3.9 gives a recursive algorithm to compute $g_T(G)$ without any reference to link diagrams. Theorem 1.1 coupled with our algorithm for computing $g_T(G)$ show that the genus of the Turaev surface is determined by how the various alternating regions of D are glued together along the nonalternating edges of D. The recursive algorithm is at the core of our classification theorems.

A *doubled path* of length k in G is a subgraph of G consisting of distinct vertices $v_0, \dots, v_k$ such that for each $i = 1, \dots, k$ there are two distinct edges $e_{i,1}$ and $e_{i,2}$ in G connecting vertices v_{i-1} and v_i, and such that $\deg v_i = 4$ for $i = 1, \dots, k-1$. If G is a graph with a doubled path consisting of vertices $v_0, \dots, v_k$, then let G' be $G/\{e_{i,1} \cup e_{i,2}\}$, the contraction of $e_{i,1}$ and $e_{i,2}$ from G for some i with $1 \le i \le k$. Then G' is called a *doubled path contraction* of G. The inverse operation of lengthening a doubled path inside of G is called a *doubled path extension* of G. Two alternating decomposition graphs G_1 and G_2 are called *doubled path equivalent* if there is a sequence of doubled path contractions and extensions transforming G_1 into G_2. Doubled path contraction/extension can make a graph nonbipartite (and hence not an alternating decomposition graph), but we do not require every graph in the sequence from G_1 to G_2 to be bipartite. Proposition 3.11 shows that if G_1 and G_2 are doubled path equivalent, then $g_T(G_1) = g_T(G_2)$.

A graph is *k–edge connected* for some positive integer k if the graph remains connected whenever fewer than k edges are removed. An alternating decomposition graph G is called *reduced* if G is a single vertex or every component of G is 3–edge connected. In Section 3, we study the behavior of alternating decomposition graphs under connected sum. We show that for any link L, there exists a diagram D of L with reduced alternating decomposition graph such that D minimizes Turaev genus. The classification theorems characterize all reduced alternating decomposition graphs of a fixed Turaev genus.

Our main theorems give classifications of all reduced alternating decomposition graphs of Turaev genus one and two. A *doubled cycle C_i^2* of length i is the graph obtained from the cycle C_i of length i by doubling every edge.

Theorem 1.2 *A reduced alternating decomposition graph G is of Turaev genus one if and only if G is doubled path equivalent to C_2^2, that is, if and only if G is a doubled cycle of even length.*

The previous theorem implies that every Turaev genus one link has a diagram D obtained by connecting an even number of alternating 2–tangles into a cycle, as in Figure 2. Dasbach and Lowrance [15] use Theorem 1.2 to compute the signature of all Turaev genus one knots and to show that either the leading or trailing coefficient of the Jones polynomial of a Turaev genus one link has absolute value one.

A link is *almost-alternating* if it is nonalternating and has a diagram D that can be transformed into an alternating diagram with a single crossing change; see Adams et al [4]. Abe and Kishimoto's work [2] implies that all almost-alternating links have Turaev genus one. It is unknown whether there is a link with Turaev genus one that is not almost-alternating; see Lowrance [25]. The following corollary shows another relationship between almost-alternating links and Turaev genus one links.

Corollary 1.3 *If L is a link of Turaev genus one, then there is an almost-alternating link L' such that L and L' are mutants of one another.*

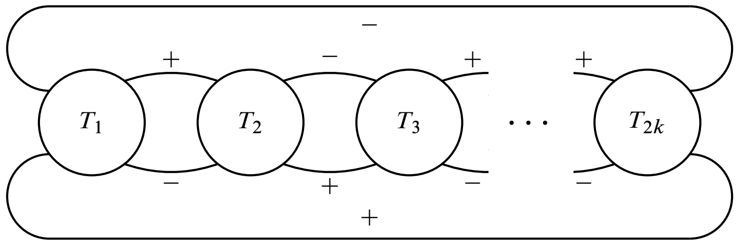


Figure 2: Every diagram D where $g_T(D) = 1$ and G is reduced has alternating decomposition as above. Each 2–tangle T_i is alternating. A $\pm$ sign on an edge indicates that it is a nonalternating edge of D with endpoints both over/under crossings respectively. The alternating decomposition graph G associated to such a diagram is a doubled cycle of length $2k$.

We present a similar classification theorem for reduced alternating decomposition graphs of Turaev genus two. However, instead of only one doubled path equivalence class, now there are five. Let G_1 and G_2 be two graphs. A *one-sum* $G_1 \oplus_1 G_2$ is the graph obtained by identifying a vertex of G_1 with a vertex of G_2. Let e_1 be an edge in G_1 connecting vertices v_1 and v_2, and let e_2 be an edge in G_2 connecting vertices u_1 and u_2. A *two-sum* $G_1 \oplus_2 G_2$ is the graph obtained by identifying the triple (v_1, v_2, e_1) with (u_1, u_2, e_2), and then deleting the edge corresponding to e_1 and e_2. For example the two-sum of two 3–cycles $C_3 \oplus_2 C_3$ is a four cycle C_4. Consider the following five classes of graphs, depicted in Figure 3:

(1) Let $C_i^2 \sqcup C_j^2$ denote the disjoint union of the doubled cycles C_i^2 and C_j^2.

(2) Let $C_i^2 \oplus_1 C_j^2$ be the graph obtained identifying a vertex of the doubled cycle C_i^2 with a vertex of C_j^2.

(3) Let $C_{i,j,k}$ be the graph obtained by identifying two paths of length k in the cycle C_{i+k} of length $i+k$ and the cycle C_{j+k} of length $j+k$. Furthermore, let $C_{i,j,k}^2$ be the graph $C_{i,j,k}$ with each edge doubled.

(4) Let $K_4(p,q)$ be the graph obtained by replacing two nonadjacent edges of the complete graph K_4 with doubled paths of lengths p and q respectively.

(5) Let $K_4(p)$ be the graph K_4 with one edge replaced by a doubled path of length p. Let $K_4(p) \oplus_2 K_4(q)$ be the two-sum of $K_4(p)$ and $K_4(q)$ taken along the unique edge in each summand that is not contained in or adjacent to the doubled path.

The graphs in the above families are not necessarily bipartite (depending on their parameters). Informally, the subsequent theorem states that a reduced alternating decomposition graph has Turaev genus two if and only if it is in one of the above five families and is bipartite. The precise statement uses doubled path equivalence.

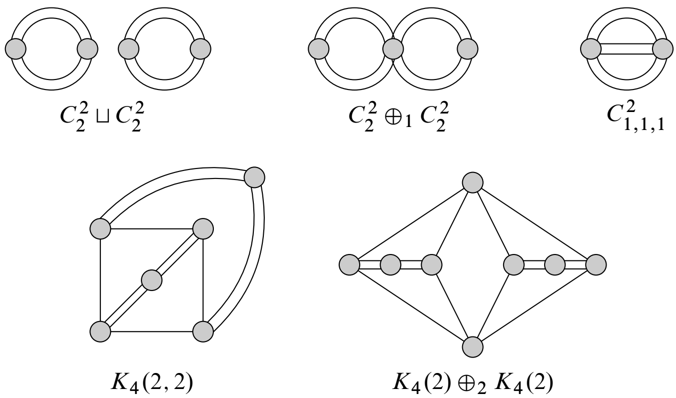


Figure 3: Representatives of the five doubled path equivalence classes of reduced alternating decompositions graphs of Turaev genus two. Informally, a Turaev genus two link diagram is obtained by inserting appropriate alternating tangles inside of the vertices of these graphs. In the case of $C_2^2 \sqcup C_2^2$ one should insert an annular alternating region bounded by two curves that correspond to vertices in distinct components. See Figure 9 for an example of a connected link diagram with disconnected alternating decomposition graph.

Theorem 1.4 *A reduced alternating decomposition graph G is of Turaev genus two if and only if G is doubled path equivalent to one of the following five graphs:*

$$C_2^2 \sqcup C_2^2, \quad C_2^2 \oplus_1 C_2^2, \quad C_{1,1,1}^2, \quad K_4(2,2), \quad \text{or} \quad K_4(2) \oplus_2 K_4(2).$$

Seungwon Kim [22] has independently proved versions of Theorem 1.2 and Theorem 1.4. The following theorem shows that for each nonnegative integer k, there exists a similar classification of reduced alternating decomposition graphs of Turaev genus k.

Theorem 1.5 *Let k be a nonnegative integer. There are a finite number of doubled path equivalence classes of reduced alternating decomposition graphs G with Turaev genus k.*

This paper is organized as follows. In Section 2, we review background material on the Turaev surface and discuss its connections to other areas of knot theory. In Section 3, we give the algorithm to compute $g_T(G)$ and prove Theorem 1.1. In Section 4, we classify alternating decomposition graphs of Turaev genus zero and show that all links have a Turaev genus minimizing diagram whose alternating decomposition graph is reduced. In Section 5, we prove the three main classification theorems (Theorems 1.2, 1.4, and 1.5).

The authors thank Sergei Chmutov, Oliver Dasbach, Nathan Druivenga, Charles Frohman, and Thomas Kindred for their helpful comments.

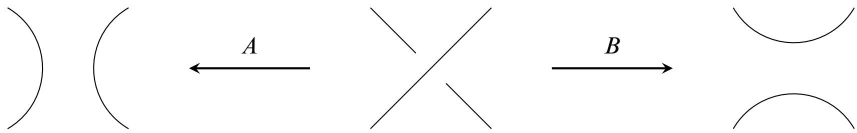


Figure 4: The A and B resolutions of a crossing

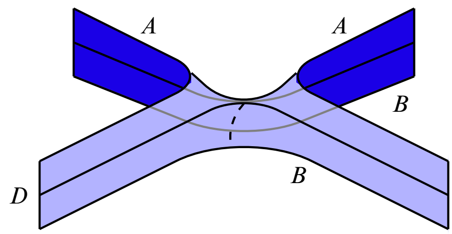


Figure 5: In a neighborhood of each crossing of D, a saddle surface transitions between the all-A and all-B states.

2 The Turaev surface

In this section, we give the construction of the Turaev surface of a link diagram D and discuss its connections to other link invariants. For a more in depth summary, see Champanerkar and Kofman's recent survey [9].

Each link diagram D has an associated Turaev surface $F(D)$, constructed as follows. Figure 4 shows the A and B resolutions of a crossing in D. The collection of simple closed curves obtained by performing either an A–resolution or a B–resolution for each crossing of D is a *state* of D. Performing an A–resolution for every crossing results in the *all-A state* of D. Similarly, performing a B–resolution for every crossings results in the *all-B state* of D. Let $s_A(D)$ and $s_B(D)$ denote the number of components in the all-A and all-B states of D respectively.

To construct the Turaev surface, we take a cobordism from the all-B state of D to the all-A state of D such that the cobordism consists of bands away from the crossings of D and saddles in neighborhoods of the crossing, as depicted in Figure 5. Finally, to obtain $F(D)$, we cap off the boundary components of the cobordism with disks. The Turaev surface $F(D)$ is oriented, and we denote the genus of the Turaev surface of D by $g_T(D)$. If the Turaev surface (or any oriented surface) is disconnected, then when we refer to its genus, we mean the sum of the genera of its connected components. Let $k(D)$ be the number of split components of the diagram D, ie the number of graph components of D when D is considered as a 4–valent graph whose vertices are the crossings. Also, let $c(D)$ be the number of crossings of D. It can be shown that

$$g_T(D) = \tfrac{1}{2}(2k(D) + c(D) - s_A(D) - s_B(D)). \tag{2-1}$$

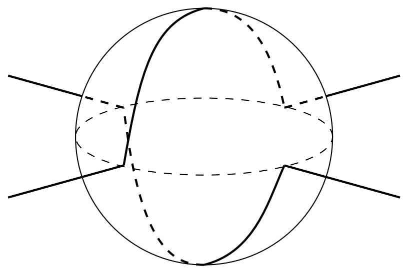


Figure 6: A crossing ball shows how L is embedded near a crossing of D.

The *Turaev genus* $g_T(L)$ of a link L is the minimum genus of the Turaev surface of D, where D is any diagram of L; ie

$$g_T(L) = \min\{g_T(D) \mid D \text{ is a diagram of } L\}.$$

Turaev [32] constructs his surface in a slightly different, but equivalent way. Turaev's construction allows us to see that a diagram D of the link L can be considered as a 4–valent graph simultaneously embedded on the sphere and the Turaev surface $F(D)$. First consider D as embedded on a sphere S. Then L can be embedded into S^3 by replacing crossings of D with suitably small balls where one strand passes over the other, as in Figure 6.

We construct the Turaev surface of D by first replacing each crossing of D with the disk that is the intersection of the associated crossing ball and S. Each alternating edge of D is replaced with an untwisted band that lies completely in the projection sphere S. Each nonalternating edge of D is replaced with a twisted band. One arc on the boundary of the twisted band will be an arc in a component of the all-A state of D, and one arc on the boundary of the twisted band will be an arc in a component of the all-B state of D. The band can be twisted so that the arc corresponding to the all-A state lies in the union of S and its exterior, while the arc corresponding to the all-B state lies in the union of S and its interior.

After replacing each crossing of D with a band, the boundary of the resulting surface is the union of the all-A state of D and the all-B state of D. Moreover, the boundary components corresponding to the all-A state lie in the union of S and its exterior, and the boundary components corresponding to the all-B state lie in the union of S and its interior. Therefore, the boundary components of this surface can be capped off with disks embedded in S^3, and the resulting surface is the Turaev surface $F(D)$. By projecting the link to S in the crossing balls, one can consider the diagram D to be embedded on both S and the Turaev surface $F(D)$; see Figure 7.

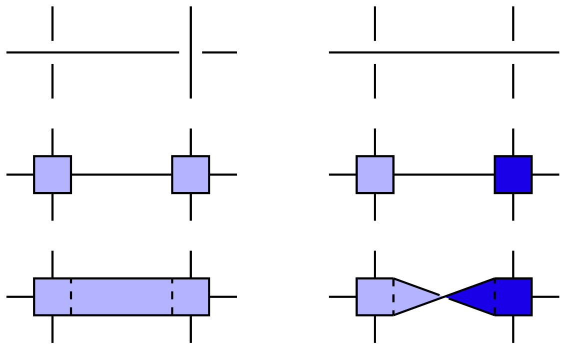


Figure 7: The disks and band associated to an alternating edge (left), and the disks and band associated to a nonalternating edge (right)

The Turaev surface of a link diagram and the Turaev genus of a link have the following properties; proofs of these facts can be found in [32; 11]:

(1) The Turaev surface $F(D)$ is a Heegaard surface in S^3, that is, $S^3 - F(D)$ is a union of two handlebodies.
(2) The diagram D is alternating on $F(D)$.
(3) The Turaev surface is a sphere if and only if D is a connected sum of alternating diagrams. Consequently, $g_T(L) = 0$ if and only if L is alternating.
(4) The complement $F(D) - D$ is a collection of disks.

The above conditions do not completely characterize Turaev surfaces. Let $g_{\text{alt}}(L)$ be the minimal genus of Heegaard surface F in S^3 on which the link L has an alternating projection such that the complement of that projection to F is a collection of disks. Adams [3] studies knots and links where $g_{\text{alt}}(L) = 1$, and Balm [7] studies the behavior of $g_{\text{alt}}(L)$ under connected sum. Lowrance [25] constructs a family of links where $g_{\text{alt}}(L) = 1$, but the Turaev genus is arbitrarily large. Armond, Druivenga, and Kindred [5] show how to determine whether a surface satisfying the above conditions is a Turaev surface using Heegaard diagrams. Indeed, the Heegaard diagrams corresponding to Turaev surfaces of genus one first inspired Theorem 1.2 and the subsequent work in this paper.

Like many link invariants defined as minimums over all diagrams, there is no algorithm to compute the Turaev genus of a link. Instead, our computations rely on various bounds of Turaev genus. The first bound follows immediately from inequality (1-2). We have

$$g_T(L) \leq c(L) - \operatorname{span} V_L(t),$$

where $c(L)$ is the minimum crossing number of L. Several other bounds on Turaev genus come from link homologies.

Khovanov [19] constructs a categorification $\mathrm{Kh}(L)$ of the Jones polynomial, now known as Khovanov homology. Khovanov homology is a bigraded $\mathbb{Z}$–module with homological grading i and quantum grading j, and one may write $\mathrm{Kh}(L)$ as a direct sum over its bigraded summands $\mathrm{Kh}(L) = \bigoplus_{i,j} \mathrm{Kh}^{i,j}(L)$. Define

$$\delta_{\min}(\mathrm{Kh}(L)) = \min\{j-2i \mid \mathrm{Kh}^{i,j}(L) \neq 0\},$$
$$\delta_{\max}(\mathrm{Kh}(L)) = \max\{j-2i \mid \mathrm{Kh}^{i,j}(L) \neq 0\}.$$

Champanerkar, Kofman, and Stoltzfus [10] show that

$$\delta_{\max}(\mathrm{Kh}(L)) - \delta_{\min}(\mathrm{Kh}(L)) - 2 \leq 2g_T(L). \tag{2-2}$$

A link diagram D is *adequate* if the number of components in the all-A (respectively all-B) state is strictly greater than the number of components in every state containing exactly one B–resolution (respectively exactly one A–resolution). A link is *adequate* if it has an adequate diagram. Khovanov [20] studies the Khovanov homology of adequate links, and Abe [1] proves that inequality (2-2) is tight when L is adequate.

Ozsváth and Szabó [28] and independently Rasmussen [29] construct a categorification $\widehat{\mathrm{HFK}}(K)$ of the Alexander polynomial of a knot K, called knot Floer homology. Knot Floer homology is also a bigraded $\mathbb{Z}$–module with homological (or Maslov) grading m and Alexander grading s, and one may write $\widehat{\mathrm{HFK}}(K)$ as a direct sum over its bigraded summands $\widehat{\mathrm{HFK}}(K) = \bigoplus_{m,s} \widehat{\mathrm{HFK}}_m(K,s)$. Define

$$\delta_{\min}(\widehat{\mathrm{HFK}}(K)) = \min\{s-m \mid \widehat{\mathrm{HFK}}_m(K,s) \neq 0\},$$
$$\delta_{\max}(\widehat{\mathrm{HFK}}(K)) = \max\{s-m \mid \widehat{\mathrm{HFK}}_m(K,s) \neq 0\}.$$

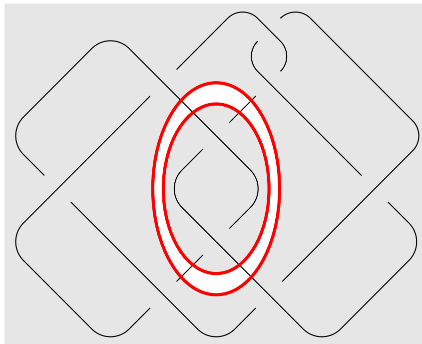
Lowrance [23] shows that

$$\delta_{\max}(\widehat{\mathrm{HFK}}(K)) - \delta_{\min}(\widehat{\mathrm{HFK}}(K)) \leq g_T(K). \tag{2-3}$$

Let $\sigma(K)$ be the signature of K, let $\tau(K)$ be the Ozsváth–Szabó τ–invariant [27], and let $s(K)$ be the Rasmussen s–invariant [30]. Dasbach and Lowrance [13] show that

$$\left|\tau(K) + \tfrac{1}{2}\sigma(K)\right| \leq g_T(K), \tag{2-4}$$
$$\left|\tfrac{1}{2}(s(K) + \sigma(K))\right| \leq g_T(K), \tag{2-5}$$
$$\left|\tau(K) - \tfrac{1}{2}s(K)\right| \leq g_T(K). \tag{2-6}$$

Essentially all known computations of the Turaev genus of a link rely on some inequality among (2-2)–(2-6). Finding a new method for computing the Turaev genus remains a challenging open question.



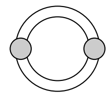


Figure 8: A diagram D of 9_{42} with its alternating regions shaded and its alternating decomposition graph $G = C_2^2$

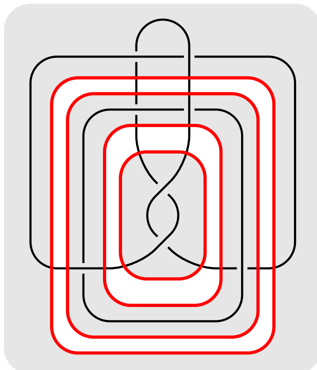
3 Alternating decomposition graphs

Throughout this section, we assume that D is a link diagram, G is the alternating decomposition graph of D, and $\mathbb{G}$ is the graph G with the sphere embedding induced by D. We begin the section with some examples.

Example 3.1 Figure 8 shows a diagram D of the knot 9_{42} from Rolfsen's table, along with its alternating decomposition curves $\{\gamma_1, \gamma_2\}$. Since the alternating decomposition of D has two curves that both intersect the same four nonalternating edges of D, it follows that the alternating decomposition graph of D is $G = C_2^2$, the graph with two vertices and four parallel edges between them. In this example, $g_T(D) = 1$ and since 9_{42} is nonalternating, it follows that $g_T(L) = 1$.

Example 3.2 Figure 9 shows a connected link diagram D with a disconnected alternating decomposition graph G. The alternating decomposition graph G is disconnected when D has an alternating region with more than one boundary component. In this case, the alternating decomposition graph G is $C_2^2 \sqcup C_2^2$, the disjoint union of two doubled 2–cycles. The disjoint union of two copies of the diagram from Figure 8 also has $C_2^2 \sqcup C_2^2$ as its alternating decomposition graph.

The embedding of D into the plane induces an embedding of each component of the alternating decomposition graph G onto a sphere. Each curve γ_i of the alternating decomposition of D is incident to two regions, precisely one of which contains crossings of D. In the examples of Figure 8 and Figure 9, the alternating regions with crossings are shaded, and the regions without crossings are unshaded. If γ_i and γ_j are different boundary curves of the same alternating region, then their associated vertices



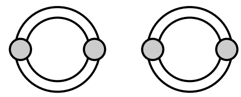


Figure 9: The alternating decomposition of D has an annular alternating region. Hence its alternating decomposition graph G is disconnected.

belong to different components of G. Let $\gamma_{i_1}, \dots, \gamma_{i_k}$ be the curves of the alternating decomposition graph associated to all of the vertices of a particular component of G. One may consider the diagram D as being embedded on the sphere S, and thus the curves $\gamma_{i_1}, \dots, \gamma_{i_k}$ are also embedded on S. The embedding of this component of G onto the sphere S is obtained by considering the vertex associated to γ_{i_j} to be the disk with boundary γ_{i_j} containing the alternating region incident to γ_{i_j}. This disk may contain other curves from the alternating decomposition of D, but these other curves are associated to a different component of G. The edges of this component are the segments of the nonalternating edges of D that go between two curves of the alternating decomposition of D. Thus each component of G has an induced embedding onto a sphere.

Thistlethwaite [31] proved that if G is an alternating decomposition graph of some link diagram, then G is planar, bipartite, and each vertex of G has even degree. Our first result of this section is the converse.

Proposition 3.3 *Let G be a planar, bipartite graph such that each vertex of G has even degree. Then G is the alternating decomposition graph of some link diagram D. Moreover, D may be chosen to be adequate.*

Proof Fix a planar embedding for G. For each vertex v_i in G, choose an alternating tangle T_i with $\deg v_i$ endpoints along the boundary. Each tangle T_i must contain at least one crossing, and each face of the tangle T_i can only meet the boundary circle in at most one arc. Assign to each endpoint the sign $+$ or $-$ based on whether the strand emanating from that point is the overstrand or the understrand, respectively, of the first crossing it meets. The signs $+$ and $-$ will alternate around the boundary of T_i.

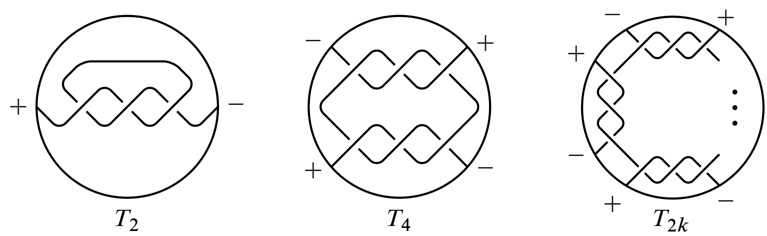


Figure 10: Inserting the tangles T_{2i} into an alternating decomposition graph G results in an adequate link diagram D whose alternating decomposition graph is G.

Since G is bipartite, the edges of G can also be assigned $+$ or $-$ in such a way that the signs alternate around each vertex in the planar embedding. Replace v_i with T_i in the planar embedding of G so that each endpoint of an arc in T_i and the edge of G which it gets connected to have the same sign. This produces a link diagram with the property that the nonalternating arcs exactly correspond to the edges of G.

To make the link diagram adequate, appropriate tangles must be chosen for the T_i. Choosing the tangles shown in Figure 10 will produce an adequate link diagram. This is because the circles in the all-A and all-B resolutions come in two types: those completely contained in one of the tangles, and those that pass through multiple tangles. Each crossing is either between two distinct circles of the first type, or between a circle of the first type and a circle of the second type. Specifically, each crossing is always between two distinct circles. Thus if one crossing is changed from the A–resolution to the B–resolution in the all-A state (or vice-versa in the all-B state), then the number of circles will decrease by one. □

Abe [1] proves that if D is adequate, D minimizes Turaev genus, ie $g_T(D) = g_T(L)$. Consequently, we have the following corollary.

Corollary 3.4 *Let G be a planar, bipartite graph such that each vertex has even degree. Then there is a link diagram D whose alternating decomposition graph is G such that $g_T(D) = g_T(L)$.*

An *oriented ribbon graph* is a graph G cellularly embedded in an oriented surface Σ. The genus of an oriented ribbon graph is the genus of Σ. We often visualize the vertices of an oriented ribbon graph as round disks and the edges of an oriented ribbon graph as rectangular bands attached on opposite ends to the round vertices. The sphere embedding $\mathbb{G}$ of an alternating decomposition graph G is a ribbon graph embedded on a disjoint union of spheres. From $\mathbb{G}$, we construct another ribbon graph $\widetilde{\mathbb{G}}$ such

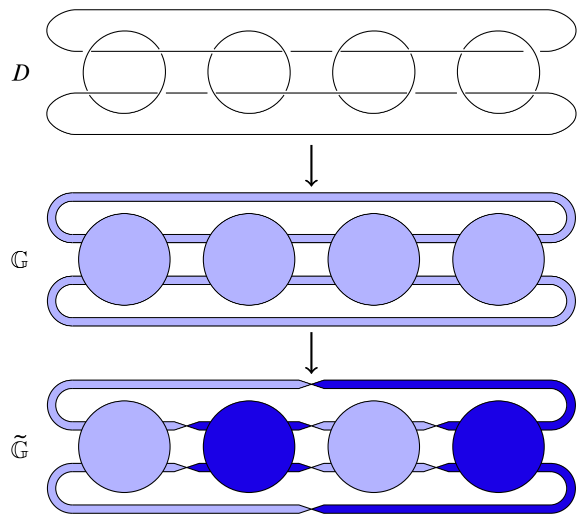


Figure 11: The link diagram D, the sphere embedding $\mathbb{G}$ of its alternating decomposition graph G and the twisted embedding $\widetilde{\mathbb{G}}$ of G

that the genus of $\widetilde{\mathbb{G}}$ is equal to $g_T(D)$. The ribbon graph $\widetilde{\mathbb{G}}$ has the same vertices and edges as $\mathbb{G}$. To obtain $\widetilde{\mathbb{G}}$ from $\mathbb{G}$ a half-twist is applied to each edge band of $\mathbb{G}$. We say that $\widetilde{\mathbb{G}}$ is the *twisted embedding of the alternating decomposition graph* G; see Figure 11. The operation of twisting some edges in a ribbon graph has been recently studied by Ellis-Monaghan and Moffatt under the name *partial petrials* [16].

Proposition 3.5 *Let $\widetilde{\mathbb{G}}$ be the twisted embedding of the alternating decomposition graph of a link diagram D. The genus of $\widetilde{\mathbb{G}}$ is $g_T(D)$.*

Proof Each vertex in $\widetilde{\mathbb{G}}$ corresponds to a curve in the alternating decomposition of D. Suppose a collection of curves $\{\gamma_{i_1}, \dots, \gamma_{i_j}\}$ bound an alternating region R in the alternating decomposition of D, and let $v_{i_1}, \dots, v_{i_j}$ be their corresponding vertices in $\widetilde{\mathbb{G}}$. The region R is a surface of genus zero with j boundary components. The vertices $v_{i_1}, \dots, v_{i_j}$ all lie in different components $\widetilde{\mathbb{G}}_{i_1}, \dots, \widetilde{\mathbb{G}}_{i_j}$ of $\widetilde{\mathbb{G}}$. Consider the vertices $v_{i_1}, \dots v_{i_j}$ as disks. Form the connected sum $\widetilde{\mathbb{G}}_{i_1} \# \cdots \# \widetilde{\mathbb{G}}_{i_j}$ by identifying disks inside of vertices $v_{i_1}, \dots, v_{i_j}$. What was a collection of j disks is now a single planar surface with j boundary components, just like R. Repeat this process for each collection of curves that bound an alternating region to form the surface Σ.

We partially construct the Turaev surface $F(D)$ as follows. Consider D as embedded on a sphere S sitting inside of S^3. Replace crossings of D with round disks, and replace all edges of D with either flat or twisted bands according to whether the edge is alternating or nonalternating. The boundary components of the resulting surface correspond to the union of the all-A and all-B states of D. If one such boundary component lies completely in S (ie each arc in the component contained in an edge band is contained in a flat edge band), then cap that boundary component off with a disk as follows. If the boundary component corresponds to a component of the all-B state, the interior of the disk should be contained inside S, and if the boundary component corresponds to a component of the all-A state, the interior of the disk should be contained outside S. The resulting surface is Σ, and so $g(\widetilde{\mathbb{G}}) = g(\Sigma) = g_T(D)$. □

Proposition 3.5 implies that the genus of the Turaev surface of D is determined by the sphere embedding $\mathbb{G}$ of its alternating decomposition graph G. Hence we define $g_T(\mathbb{G})$ to be $g_T(D)$ for any diagram D with sphere embedding $\mathbb{G}$ of its alternating decomposition graph G. We give a recursive algorithm to compute $g_T(\mathbb{G})$ without referring to the link diagram D. Our recurrence depends on the following lemma.

Lemma 3.6 *Let $\mathbb{G}$ be a sphere embedding of a connected, alternating decomposition graph G, and suppose the number of edges in G is nonzero.*

(1) *Either $\mathbb{G}$ contains a face bounded by exactly two edges or $\mathbb{G}$ contains at least four vertices of degree two.*

(2) *Either G contains a pair of parallel edges or G contains at least four vertices of degree two.*

Proof The degree of a face is defined to be the number of edges in its boundary. Suppose that $\mathbb{G}$ has no face of degree two and three or fewer vertices of degree two. Since every vertex in $\mathbb{G}$ has even degree, it follows that the other vertices of $\mathbb{G}$ have degree at least four. Let $v(\mathbb{G})$, $e(\mathbb{G})$ and $f(\mathbb{G})$ denote the number of vertices, edges and faces of $\mathbb{G}$ respectively. Also, let $\mathcal{V}(\mathbb{G})$ and $\mathcal{F}(\mathbb{G})$ be the vertex and face sets of $\mathbb{G}$. The handshaking lemma implies

$$4(v(\mathbb{G})-3)+6 = 4v(\mathbb{G})-6 \le \sum_{v\in\mathcal{V}(\mathbb{G})} \deg v = 2e(\mathbb{G}).$$

Thus $v(\mathbb{G}) \le \frac{1}{2}e(\mathbb{G}) + \frac{3}{2}$. Since $\mathbb{G}$ is bipartite, all of its faces have even degree, and since $\mathbb{G}$ has no face of degree two, the handshaking lemma applied to the planar dual of $\mathbb{G}$ implies

$$4f(\mathbb{G}) \le \sum_{f\in\mathcal{F}(\mathbb{G})} \deg f = 2e(\mathbb{G}).$$

Thus $f(\mathbb{G}) \leq \frac{1}{2}e(\mathbb{G})$. Now since $\mathbb{G}$ is connected and planar, its Euler characteristic is two. Therefore, we have

$$2 = v(\mathbb{G}) - e(\mathbb{G}) + f(\mathbb{G}) \leq \tfrac{1}{2}e(\mathbb{G}) + \tfrac{3}{2} - e(\mathbb{G}) + \tfrac{1}{2}e(\mathbb{G}) = \tfrac{3}{2},$$

which is a contradiction. Therefore $\mathbb{G}$ must have at least four vertices of degree two. The second statement follows immediately from the first. □

For any graph Γ (or oriented ribbon graph), let $k(\Gamma)$ denote the number of connected components in Γ. If e is an edge in Γ incident to vertices v_1 and v_2, then the contraction of e, denoted Γ/e is the graph obtained by identifying the vertices v_1 and v_2 and deleting the edge e. Any graph that can be obtained from Γ via a sequence of edge contractions and edge or vertex deletions is called a *minor* of Γ. The sphere embedding of a graph induces a sphere embedding on any of its minors. If Γ is bipartite, then $\Gamma - e$ is also bipartite. If Γ is bipartite and $k(\Gamma) = k(\Gamma - e) - 1$, then Γ/e is also bipartite. In the following proposition, whenever a set of edges is deleted or contracted, the induced sphere embedding on the subgraph is assumed. Proposition 3.7 gives a recursive algorithm to compute $g_T(\mathbb{G})$.

Proposition 3.7 *Let $\mathbb{G}$ be a sphere embedding of an alternating decomposition graph G.*

(1) *If $\mathbb{G}$ is a collection of isolated vertices, then $g_T(\mathbb{G}) = 0$.*

(2) *Suppose that $\mathbb{G}$ contains a face bounded by exactly two edges e_1 and e_2. Let $\mathbb{G}' = \mathbb{G} - \{e_1, e_2\}$, and let $\mathbb{G}'' = \mathbb{G}/\{e_1, e_2\}$. If $k(\mathbb{G}') = k(\mathbb{G})$, then $\mathbb{G}'$ is a sphere embedding of an alternating decomposition graph and $g_T(\mathbb{G}') = g_T(\mathbb{G}) - 1$. If $k(\mathbb{G}') = k(\mathbb{G}) + 1$, then both $\mathbb{G}'$ and $\mathbb{G}''$ are sphere embeddings of alternating decomposition graphs and $g_T(\mathbb{G}') = g_T(\mathbb{G}'') = g_T(\mathbb{G})$.*

(3) *Suppose that $\mathbb{G}$ contains a vertex v of degree two, incident to edges e_1 and e_2. Let $\mathbb{G}' = \mathbb{G}/\{e_1, e_2\}$. Then $\mathbb{G}'$ is a sphere embedding of an alternating decomposition graph, and $g_T(\mathbb{G}') = g_T(\mathbb{G})$.*

Proof (1) Let D be the disjoint union of m alternating diagrams. Then $g_T(D) = 0$ and $\mathbb{G}$ is m isolated vertices. Thus $g_T(\mathbb{G}) = 0$.

(2) Deleting or contracting two edges from a graph embedded on a disjoint union of spheres results in a graph embedded on a disjoint union of spheres. Moreover, since e_1 and e_2 bound a face, they are incident to the same two vertices. Hence all vertices of $\mathbb{G}'$ and $\mathbb{G}''$ have even degree. As $\mathbb{G}'$ is obtained from $\mathbb{G}$ by deleting two edges, it follows that $\mathbb{G}'$ is bipartite. Also, since e_1 and e_2 are parallel, it follows that if the deletion of e_1 and e_2 increases the number of components in $\mathbb{G}$, then $\mathbb{G}''$ is bipartite. Thus $\mathbb{G}'$ is a sphere embedding of an alternating decomposition graph, and if $k(\mathbb{G}') = k(\mathbb{G}) + 1$, then $\mathbb{G}''$ is a sphere embedding of an alternating decomposition graph.

Let $\widetilde{\mathbb{G}}$, $\widetilde{\mathbb{G}}'$ and $\widetilde{\mathbb{G}}''$ be the twisted embeddings of $\mathbb{G}$, $\mathbb{G}'$ and $\mathbb{G}''$ respectively. Define $f(\widetilde{\mathbb{G}})$ to be the number of components of $\Sigma - \widetilde{\mathbb{G}}$, where Σ is the surface on which $\mathbb{G}$ is embedded. Note that $f(\widetilde{\mathbb{G}})$ is also the number of boundary components of $\widetilde{\mathbb{G}}$. Similarly define $f(\widetilde{\mathbb{G}}')$ and $f(\widetilde{\mathbb{G}}'')$.

We have $v(\widetilde{\mathbb{G}}') = v(\widetilde{\mathbb{G}})$, $e(\widetilde{\mathbb{G}}') = e(\widetilde{\mathbb{G}}) - 2$ and $f(\widetilde{\mathbb{G}}') = f(\widetilde{\mathbb{G}})$. If $\mathbb{H}$ is an oriented ribbon graph, then its genus is

$$g(\mathbb{H}) = \tfrac{1}{2}(2k(\mathbb{H}) - v(\mathbb{H}) + e(\mathbb{H}) - f(\mathbb{H})).$$

Both $\mathbb{G}$ and $\widetilde{\mathbb{G}}$ have the same underlying graph G, and so they have the same number of components. A similar statement holds for $\mathbb{G}'$ and $\widetilde{\mathbb{G}}'$. If $k(\mathbb{G}') = k(\mathbb{G}) + 1$, then

$$g_T(\mathbb{G}') = g(\widetilde{\mathbb{G}}') = g(\widetilde{\mathbb{G}}) = g_T(\mathbb{G}),$$

and if $k(\mathbb{G}') = k(\mathbb{G})$, then

$$g_T(\mathbb{G}') = g(\widetilde{\mathbb{G}}') = g(\widetilde{\mathbb{G}}) - 1 = g_T(\mathbb{G}) - 1.$$

Also, if $k(\mathbb{G}') = k(\mathbb{G}) + 1$, then $\widetilde{\mathbb{G}}''$ can be obtained from $\widetilde{\mathbb{G}}'$ by taking a connected sum along the two vertices incident with e_1 and e_2 in $\widetilde{\mathbb{G}}$. Hence $g_T(\mathbb{G}'') = g_T(\mathbb{G}')$.

(3) As in the previous case, contracting two edges from a graph embedded on a disjoint union of spheres leads to a graph embedded on a disjoint union of spheres. Let v_1 and v_2 be the two vertices adjacent to v, and let v_{12} be the vertex in $\mathbb{G}'$ corresponding to vertices v_1 and v_2 in $\mathbb{G}$. If $v_1 \neq v_2$, then the degree of v_{12} is $\deg v_1 + \deg v_2 - 2$, which is even. If $v_1 = v_2$, then $\deg v_{12} = \deg v_1 - 2$, which is also even. All other vertices in $\mathbb{G}'$ have the same degree as their corresponding vertices in $\mathbb{G}$. Also, the bipartition of the vertices of $\mathbb{G}$ induces a bipartition of the vertices of $\mathbb{G}'$. Thus $\mathbb{G}'$ is a sphere embedding of an alternating decomposition graph.

Let $\widetilde{\mathbb{G}}$ and $\widetilde{\mathbb{G}}'$ be the twisted embeddings associated to $\mathbb{G}$ and $\mathbb{G}'$, respectively. Then $k(\widetilde{\mathbb{G}}') = k(\widetilde{\mathbb{G}})$ and $e(\widetilde{\mathbb{G}}') = e(\widetilde{\mathbb{G}}) - 2$. If $v_1 \neq v_2$, then $v(\widetilde{\mathbb{G}}') = v(\widetilde{\mathbb{G}}) - 2$ and $f(\widetilde{\mathbb{G}}') = f(\widetilde{\mathbb{G}})$, and if $v_1 = v_2$, then $v(\widetilde{\mathbb{G}}') = v(\widetilde{\mathbb{G}}) - 1$ and $f(\widetilde{\mathbb{G}}') = f(\widetilde{\mathbb{G}}) - 1$. Hence $g_T(\mathbb{G}') = g_T(\mathbb{G})$. □

As the following theorem shows, the Turaev genus of the sphere embedding $\mathbb{G}$ of the alternating decomposition graph G does not depend on its embedding at all.

Theorem 3.8 *Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be sphere embeddings of the same alternating decomposition graph G. Then $g_T(\mathbb{G}_1) = g_T(\mathbb{G}_2)$.*

Proof We proceed by induction on the number of edges in G. If G has no edges, then both $\mathbb{G}_1$ and $\mathbb{G}_2$ are embeddings of a disjoint union of vertices. Hence $g_T(\mathbb{G}_1) = g_T(\mathbb{G}_2) = 0$.

Suppose that G has n edges and that any two embeddings of an alternating decomposition graph with fewer than n edges have the same Turaev genus. Suppose that $\mathbb{G}_1$ has a vertex v of degree two incident to edges e_1 and e_2. Since $\mathbb{G}_2$ has the same underlying graph G as $\mathbb{G}_1$, the same statement holds for $\mathbb{G}_2$, that is, the vertex v in $\mathbb{G}_2$ has degree two and is incident to edges e_1 and e_2. Set $\mathbb{G}_1' = \mathbb{G}_1/\{e_1, e_2\}$, $\mathbb{G}_2' = \mathbb{G}_2'/\{e_1, e_2\}$ and $G' = G/\{e_1, e_2\}$. By Proposition 3.7, we have that $g_T(\mathbb{G}_1') = g_T(\mathbb{G}_1)$ and $g_T(\mathbb{G}_2') = g_T(\mathbb{G}_2)$. Since $\mathbb{G}_1'$ and $\mathbb{G}_2'$ are sphere embeddings of the same graph G', the inductive hypothesis implies that $g_T(\mathbb{G}_1') = g_T(\mathbb{G}_2')$. Therefore $g_T(\mathbb{G}_1) = g_T(\mathbb{G}_2)$.

Now suppose that $\mathbb{G}_1$ does not have a vertex of degree two. By Lemma 3.6, $\mathbb{G}_1$ has a face bounded by exactly two edges, say e_1 and e_2. Let $\mathbb{G}_1' = \mathbb{G}_1 - \{e_1, e_2\}$. Then Proposition 3.7 implies that if $k(\mathbb{G}_1') = k(\mathbb{G}_1)$, then $g_T(\mathbb{G}_1) = g_T(\mathbb{G}_1') + 1$, and if $k(\mathbb{G}_1') = k(\mathbb{G}_1) + 1$, then $g_T(\mathbb{G}_1) = g_T(\mathbb{G}_1')$. Since $\mathbb{G}_1$ and $\mathbb{G}_2$ have the same underlying graph G, the edges e_1 and e_2 are parallel in $\mathbb{G}_2$, but do not necessarily bound a face of degree two. Let $\mathbb{G}_2' = \mathbb{G}_2 - \{e_1, e_2\}$.

The twisted embedding $\widetilde{\mathbb{G}}_2$ is obtained from $\widetilde{\mathbb{G}}_2'$ by adding the two twisted edges corresponding to e_1 and e_2. The twisted edges e_1 and e_2 contain four boundary arcs that are pieces of boundary components of $\widetilde{\mathbb{G}}_2$. Fix one of the boundary arcs and fix an endpoint of that boundary arc. As one travels along the boundary of $\widetilde{\mathbb{G}}_2$ starting from the fixed endpoint, one of the other seven endpoints of boundary arcs of e_1 and e_2 must be encountered first. The planarity of $\mathbb{G}_2$ lets us rule out four of those endpoints. Furthermore, each edge in $\mathbb{G}_2$ corresponds to a nonalternating edge in some link diagram D. The two boundary arcs of that edge correspond to a segment in a component of the all-A state of D and a segment in a component of the all-B state of D. In particular, two boundary arcs of the same edge must belong to different components of the boundary of the twisted embedding of the associated alternating decomposition graph. This rules out one more of the endpoints as being the next endpoint encountered. There are two remaining cases, each depicted in Figure 12.

The four boundary arcs of e_1 and e_2 lie in exactly two components of the boundary of $\widetilde{\mathbb{G}}_2$. Moreover, if the twisted edges e_1 and e_2 are removed, then the two boundary components containing boundary arcs of e_1 and e_2 are transformed into two boundary components of the twisted embedding $\widetilde{\mathbb{G}}_2'$. Since no other boundary components of $\widetilde{\mathbb{G}}_2$ are changed by deleting e_1 and e_2, it follows that $f(\widetilde{\mathbb{G}}_2') = f(\widetilde{\mathbb{G}}_2)$. Since $v(\widetilde{\mathbb{G}}_2') = v(\widetilde{\mathbb{G}}_2)$ and $e(\widetilde{\mathbb{G}}_2') = e(\widetilde{\mathbb{G}}_2) - 2$, it follows that if $k(\mathbb{G}_2') = k(\mathbb{G}_2)$, then $g_T(\mathbb{G}_2) = g_T(\mathbb{G}_2') + 1$, and if $k(\mathbb{G}_2') = k(\mathbb{G}_2) + 1$, then $g_T(\mathbb{G}_2) = g_T(\mathbb{G}_2')$. The embedded graphs $\mathbb{G}_1'$ and $\mathbb{G}_2'$ have the same underlying graph, and hence the inductive hypothesis implies that $g_T(\mathbb{G}_1') = g_T(\mathbb{G}_2')$. Deleting e_1 and e_2 from $\mathbb{G}_1$ increases the number of components if and only if deleting e_1 and e_2 from $\mathbb{G}_2$ increases the number of components. Therefore $g_T(\mathbb{G}_1) = g_T(\mathbb{G}_2)$, and the desired result is proven. $\square$

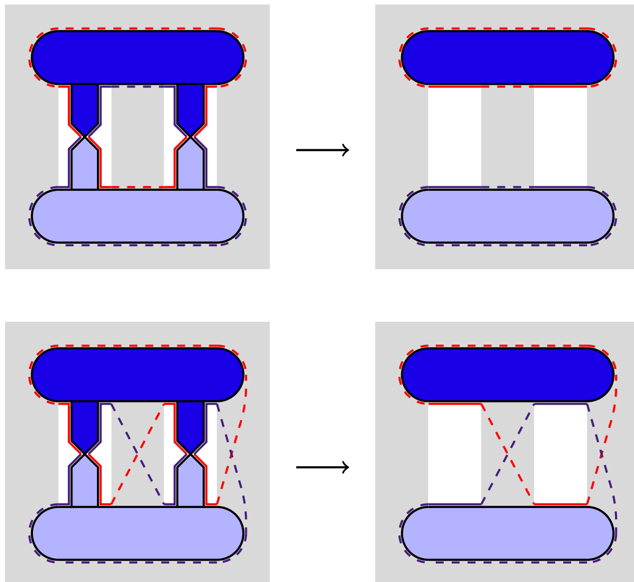


Figure 12: The two figures on the left show the boundary components of $\widetilde{\mathbb{G}}_2$ that contain the boundary arcs of e_1 and e_2, and the two figures on the right show the corresponding boundary components of $\widetilde{\mathbb{G}}'_2$. Other vertices and edges of the graph lie inside the two shaded areas.

Proof of Theorem 1.1 Let D_1 and D_2 be two link diagrams with the same alternating decomposition graph G. Let $\mathbb{G}_1$ be the sphere embedding of G induced by D_1, and let $\mathbb{G}_2$ be the sphere embedding of G induced by D_2. Theorem 3.8 implies that $g_T(D_1) = g_T(\mathbb{G}_1) = g_T(\mathbb{G}_2) = g_T(D_2)$, as desired. □

Since the Turaev genus of an alternating decomposition graph G does not depend on the sphere embedding of G, we can define $g_T(G)$ to be $g_T(D)$, where D is any link diagram with alternating decomposition graph G. The recursive algorithm in Proposition 3.7 can be restated without reference to embedding.

Corollary 3.9 *Let G be an alternating decomposition graph.*

1. *If G is a collection of isolated vertices, then $g_T(G) = 0$.*
2. *Suppose that G contains a set of parallel edges $\{e_1, e_2\}$. Let $G' = G - \{e_1, e_2\}$ and let $G'' = G/\{e_1, e_2\}$. If $k(G) = k(G')$, then $g_T(G') = g_T(G) - 1$, and if $k(G') = k(G) + 1$, then $g_T(G') = g_T(G'') = g_T(G)$.*
3. *Suppose that G contains a vertex v of degree two, incident to edges e_1 and e_2. Let $G' = G/\{e_1, e_2\}$. Then $g_T(G') = g_T(G)$.*

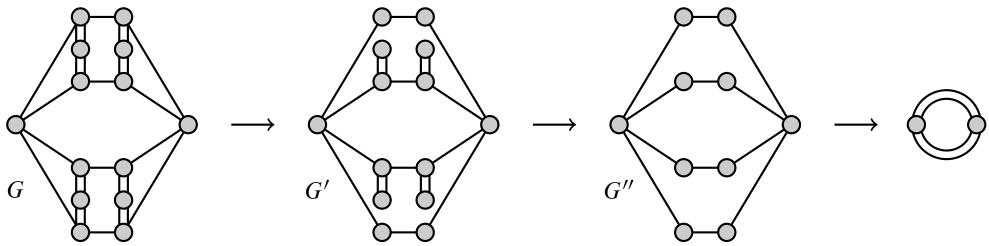


Figure 13: The graph G is transformed into C_2^2 via the algorithm of Corollary 3.9. The first step decreases Turaev genus by four, while the second and third steps do not change Turaev genus. Since $g_T(C_2^2) = 1$, it follows that $g_T(G) = 5$.

Example 3.10 Let G be the alternating decomposition graph on the top left of Figure 13. One can apply the algorithm of Corollary 3.9 to G as follows. First, delete four pairs of parallel edges as shown to obtain the graph G'. Since $k(G) = k(G')$, it follows that $g_T(G) = g_T(G') + 4$. Second, contract the remaining four pairs of parallel edges to obtain G'', and note that $g_T(G'') = g_T(G')$. Finally, apply operation (3) of Corollary 3.9 to four degree-two vertices of G'' to obtain C_2^2. Since $g_T(C_2^2) = 1$, it follows that $g_T(G) = 5$. This example shows that it is not always possible to find $g_T(G)$ pairs of parallel edges in G whose deletion does not increase the number of components.

Proposition 3.11 *Suppose that G_1 and G_2 are doubled path equivalent alternating decomposition graphs. Then $g_T(G_1) = g_T(G_2)$.*

Proof Let G be an alternating decomposition graph with sphere embedding $\mathbb{G}$ and twisted embedding $\widetilde{\mathbb{G}}$. A doubled path extension adds one vertex, two edges and one face to $\widetilde{\mathbb{G}}$, and a doubled path contraction removes one vertex, two edges and one face from $\widetilde{\mathbb{G}}$. Therefore the Euler characteristic of $\widetilde{\mathbb{G}}$ is unchanged by either doubled path extensions or doubled path contractions. If G_1 and G_2 are doubled path equivalent alternating decomposition graphs with twisted embeddings $\widetilde{\mathbb{G}}_1$ and $\widetilde{\mathbb{G}}_2$, then the Euler characteristics (and hence genera) of $\widetilde{\mathbb{G}}_1$ and $\widetilde{\mathbb{G}}_2$ agree. Thus $g_T(G_1) = g_T(G_2)$. □

We remind the reader that doubled path extensions and contractions can transform an alternating decomposition graph into a nonbipartite graph whose associated twisted embedding is nonorientable. However, the Euler characteristic argument in the proof of Proposition 3.11 applies in both the orientable or nonorientable cases. We also warn the reader that doubled path extensions and contractions only change the length of existing doubled paths. Creating new doubled paths or entirely destroying doubled paths will change the Turaev genus of the graph.

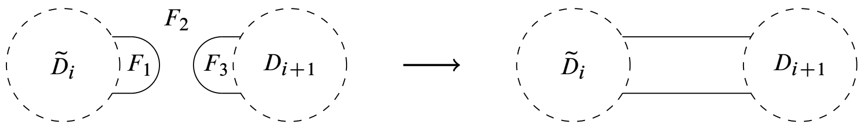


Figure 14: On the left is the disjoint union of $\widetilde{D}_i$ and D_{i+1}, and on the right is a connected sum of $\widetilde{D}_i$ and D_{i+1}. The diagram D_{i+1} is alternating. For $k = 1$, 2 and 3, let F_k denote the indicated face of $\widetilde{D}_k \sqcup D_{k+1}$.

4 Alternating decomposition graphs of Turaev genus zero

Turaev [32] showed that the genus of the Turaev surface of a link diagram D is zero if and only if D is a connected sum of alternating diagrams. In this section, we use Turaev's result to give a classification of alternating decomposition graphs of Turaev genus zero. In order to accomplish this, we will study the behavior of the alternating decomposition graph under certain types of connected sums.

Suppose that D is a link diagram with $g_T(D) = 0$. Hence $D = D_1 \# \cdots \# D_k$ is a connected sum of alternating diagrams $D_1, \dots, D_k$. Let $\widetilde{D}_i = D_1 \# \cdots \# D_i$ for $i = 1, \dots, k$. Then $D = \widetilde{D}_k$ and $\widetilde{D}_{i+1} = \widetilde{D}_i \# D_{i+1}$. Thus to classify connected sums of alternating diagrams, it suffices to examine the connected sum of a (possibly nonalternating) diagram $\widetilde{D}_i$ and an alternating diagram D_{i+1}; see Figure 14.

Let $\widetilde{G}_i$ be the alternating decomposition graph of $\widetilde{D}_i$, for each $i = 1, \dots, k$. Since D_{i+1} is alternating, its alternating decomposition graph is a single vertex. We examine how $\widetilde{G}_{i+1}$ is obtained from $\widetilde{G}_i$. A face of a link diagram is said to be *alternating* if every edge in the boundary of that face is alternating. Otherwise, the face is said to be *nonalternating*. Let e_i be the edge of $\widetilde{D}_i$ and let e_{i+1} be the edge of D_{i+1} along which we are taking the connected sum. The edge e_{i+1} is necessarily alternating, but e_i can be either alternating or nonalternating. Figure 15 shows the alternating decomposition curves in the seven relevant cases, which we describe in detail below.

Case 1 Suppose that e_i is nonalternating. Figure 15 shows the endpoints of e_i passing under the crossing, but the case where the endpoints pass over the crossing is exactly the same. Taking the connected sum merges the curve in the alternating decomposition of D_{i+1} with one of the curves in the alternating decomposition of $\widetilde{D}_i$. Therefore $\widetilde{G}_{i+1} = \widetilde{G}_i$.

Case 2 Suppose that e_i is alternating and the connected sum is taken as in Figure 15. Also, suppose that both F_1 and F_2 are alternating faces of $\widetilde{D}_i$. Then there are no alternating decomposition curves of $\widetilde{D}_i$ in either F_1 or F_2. Hence $\widetilde{G}_{i+1} = \widetilde{G}_i \sqcup C_2$, where C_2 is a 2–cycle.

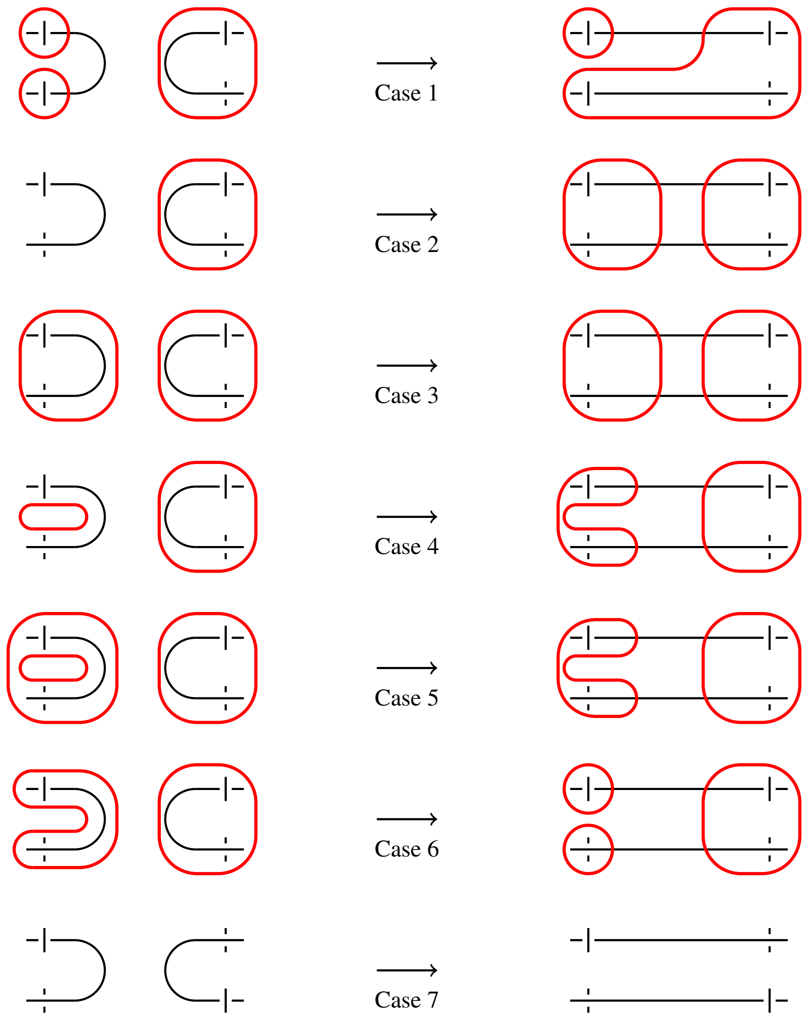


Figure 15: Taking the connected sum of $\widetilde{D}_i = D_1 \# \cdots \# D_i$ and the alternating diagram D_{i+1}

Case 3 Suppose that e_i is alternating and the connected sum is taken as in Figure 15. Also, suppose that F_1 is an alternating face of $\widetilde{D}_i$, while F_2 is a nonalternating face of $\widetilde{D}_i$. Let γ be the alternating decomposition curve in F_2 that runs along e_i. After performing the connected sum, the curve γ transforms into a curve that runs along the same portion of the boundary of F_2 and also along all of F_1. Thus the connected sum attaches the alternating decomposition curve of D_{i+1} to γ by two edges. Hence $\widetilde{G}_{i+1} = \widetilde{G}_i \oplus_1 C_2$. The transformation $G \mapsto G \oplus_1 C_2$ is called a *doubled pendant* move and is depicted in Figure 16.

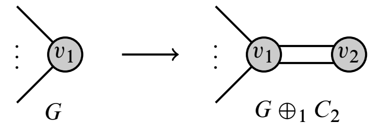


Figure 16: A doubled pendant move on G results in the graph $G \oplus_1 C_2$.

Case 4 Suppose that e_i is alternating and the connected sum is taken as in Figure 15. Also, suppose that F_1 is a nonalternating face of $\widetilde{D}_i$, while F_2 is an alternating face of $\widetilde{D}_i$. Let γ be the alternating decomposition curve in F_1 that runs along e_i. After performing the connected sum, the curve γ transforms into a curve that runs along the same portion of the boundary of F_1 and also along all of F_2. Thus the connected sum attaches the alternating decomposition curve of D_{i+1} to γ by two edges. Hence $\widetilde{G}_{i+1} = \widetilde{G}_i \oplus_1 C_2$.

Case 5 Suppose that e_i is alternating and the connected sum is taken as in Figure 15. Also, suppose that both F_1 and F_2 are nonalternating faces of $\widetilde{D}_i$ and that the alternating decomposition curves γ_1 and γ_2 that run along e_i are distinct curves. Since the region bounded by γ_1 and γ_2 contains crossings of $\widetilde{D}_i$, it follows that the vertices of $\widetilde{G}_i$ corresponding to γ_1 and γ_2 lie in different components of $\widetilde{G}_i$. Performing the connected sum operation merges γ_1 and γ_2, and connects the alternating decomposition curve of D_{i+1} to the newly merged γ_1 and γ_2 with two edges. Therefore, $\widetilde{G}_{i+1}$ is obtained from $\widetilde{G}_i$ by taking a one-sum along two vertices in separate components of $\widetilde{G}_i$ and then an additional one-sum with C_2.

Case 6 Suppose that e_i is alternating and the connected sum is taken as in Figure 15. Also, suppose that both F_1 and F_2 are nonalternating faces of $\widetilde{D}_i$ and that there is a single alternating decomposition curve that runs along e_i in both F_1 and F_2. Performing a connected sum operation splits this alternating decomposition curve into two curves, each of which has a single edge attached to the alternating decomposition curve of D_{i+1}. Thus the graph $\widetilde{G}_{i+1}$ is obtained from $\widetilde{G}_i$ by

(1) picking a vertex v of $\widetilde{G}_i$,
(2) partitioning the edges incident to v into two sets A and B each of odd order,
(3) splitting the vertex v into two new vertices v_1 and v_2 where, the edge set A is incident to v_1 and the edge set B is incident to v_2, and
(4) creating a new vertex v_3 of degree two adjacent to both v_1 and v_2.

See Figure 17 for a depiction of this operation, which we call a *two-path extension*.

Case 7 Suppose that e_i is alternating and the connected sum is taken as in Figure 15. Note that this connected sum is different than Cases 2–6. In this case, it does not matter whether either, neither, or both of F_1 and F_2 are alternating or nonalternating. In each case, we have $\widetilde{G}_{i+1} = \widetilde{G}_i$.

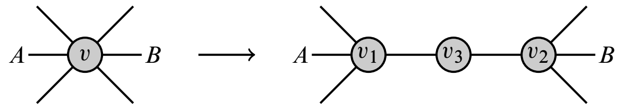


Figure 17: A two-path extension. The edge sets A and B must each be of odd order.

Theorem 4.1 *Let G be an alternating decomposition graph with $g_T(G) = 0$. Then G can be obtained from a collection of isolated vertices via a sequence of doubled pendant moves, two-path extensions and one-sums along vertices in different components.*

Proof Suppose D is a link diagram with alternating decomposition graph G. Then $g_T(D) = g_T(G) = 0$, and hence $D = D_1 \# \cdots \# D_k$ is a connected sum of alternating diagrams $D_1, \dots, D_k$. Let $\widetilde{D}_i = D_1 \# \cdots \# D_i$, and let $\widetilde{G}_i$ be the alternating decomposition graph of $\widetilde{G}_i$. Our analysis above shows that there is a sequence $\widetilde{G}_1, \widetilde{G}_2, \dots, \widetilde{G}_k = G$ of alternating decomposition graphs such that $\widetilde{G}_1$ is a collection of isolated vertices and $\widetilde{G}_{i+1}$ can be obtained from $\widetilde{G}_i$ by either doing nothing, a doubled pendant move, a two-path extension, a disjoint union with C_2, or the multistep operation of Case 5 (which stipulated that we glue together two components of $\widetilde{G}_i$ along a vertex, and then perform a doubled pendant move to the same vertex).

We modify the sequence $\widetilde{G}_1, \dots, \widetilde{G}_k = G$ so that it still begins in a collection of isolated vertices, still ends in G, and each graph can be obtained from the previous one via a doubled pendant move, a two-path extension, or by identifying two vertices in different components. For each i where $\widetilde{G}_{i+1}$ is obtained from $\widetilde{G}_i$ via a disjoint union with C_2, we modify $\widetilde{G}_j$ for $j \leq i$ by adding an isolated vertex v. Since $G \sqcup C_2 = G \sqcup \{v\} \oplus_1 C_2$, we have changed adding a disjoint union of C_2 into doubled pendant move.

For each i where $\widetilde{G}_{i+1}$ is obtained from $\widetilde{G}_i$ via the operation in Case 5, we note that $\widetilde{G}_{i+1}$ is obtained from $\widetilde{G}_i$ by taking a one-sum of vertices in different components and then performing a doubled pendant move. In order to satisfy the conditions in the theorem, these two operations must be completed in separate steps. Thus we modify the sequence by increasing the index of each $\widetilde{G}_j$ by one, with $j \geq i+1$. Then we set $\widetilde{G}_{i+1}$ to be the graph obtained from $\widetilde{G}_i$ by taking the prescribed one-sum of vertices in different components, and then $\widetilde{G}_{i+2}$ can be obtained from $\widetilde{G}_{i+1}$ by a doubled pendant move. □

Recall that an alternating decomposition graph G is reduced if it is a single vertex or if each component of G is 3–edge connected. In the following proposition, we prove that there exists a Turaev genus minimizing diagram of every nonsplit link with reduced alternating decomposition graph.

Proposition 4.2 *Every nonsplit link L has a diagram D with alternating decomposition graph G such that G is reduced and $g_T(G) = g_T(L)$.*

Proof Equation (2-1) implies that for any choice of edge along which to take a connected sum of D_1 and D_2, we have $g_T(D_1 \# D_2) = g_T(D_1) + g_T(D_2)$. Let D' be a diagram of L that minimizes Turaev genus, ie such that $g_T(D') = g_T(L)$. Suppose that D' can be written as a connected sum $D_1 \# \cdots \# D_k$ where each D_i cannot be realized as a connected sum. Let G_i be the alternating decomposition graph of D_i.

Since each D_i cannot be realized as a connected sum, there is no circle in the plane that intersects D_i exactly twice such that the two 1–tangles formed are nontrivial. Therefore, there is no circle in the plane that intersects the alternating decomposition graph of D_i exactly twice in two distinct edges. Hence the alternating decomposition graph G_i is reduced.

However, the alternating decomposition graph G' of D' is not necessarily reduced. We construct another diagram D of L such that $g_T(D) = g_T(D') = g_T(L)$, and such that the alternating decomposition graph G of D is reduced. Suppose the connected sum of two diagrams $\widetilde{D}_1$ and $\widetilde{D}_2$ is formed in the same manner as Case 7 of Figure 15. Let e_1 and e_2 be the edges along which the connected sum is being taken, and let F_1, F_2 and F_3 be the three faces with e_1 and e_2 in their boundary, as in Figure 14. If at least two of F_1, F_2 and F_3 are alternating faces, then the alternating decomposition graph of $\widetilde{D}_1 \# \widetilde{D}_2$ is either the one-sum or disjoint union of the alternating decomposition graphs of $\widetilde{D}_1$ and $\widetilde{D}_2$. Therefore, if the alternating decomposition graphs of $\widetilde{D}_1$ and $\widetilde{D}_2$ are reduced, then the alternating decomposition graph of $\widetilde{D}_1 \# \widetilde{D}_2$ is reduced.

For each summand $D_1, \dots, D_k$ in $D = D_1 \# \cdots \# D_k$, insert a small twist into the edge on which a connected sum occurs, as in Figure 18. Inserting the twist does not change the alternating decomposition graph of each D_i, and thus does not change the genus of the associated Turaev surface. Each new twisted edge is an alternating edge, and the face bounded by that single alternating edge is an alternating face. Therefore, if all of the connected sums are taken along these twisted edges, then the alternating decomposition graph G of the resulting diagram D will be reduced. Moreover, since adding the twists does not change the genus of the Turaev surface, $g_T(D) = g_T(D') = g_T(L)$. □

5 Turaev genus classification results

In this section, we classify all reduced alternating decomposition graphs of Turaev genus one and two. We also show that for any nonnegative integer k, there are a finite number doubled path equivalence classes of alternating decomposition graphs of Turaev genus k. Hence there exists a classification of all reduced alternating decomposition graphs of Turaev genus k for any nonnegative integer k.

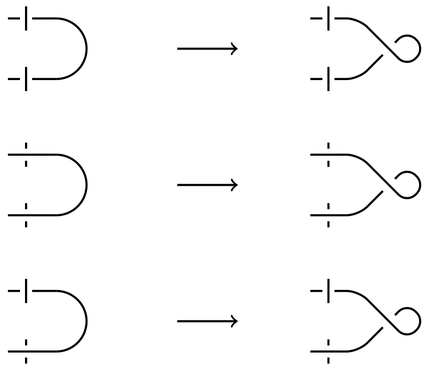


Figure 18: Inserting twists into edges where a connected sum is taken makes the resulting diagram have reduced alternating decomposition graph.

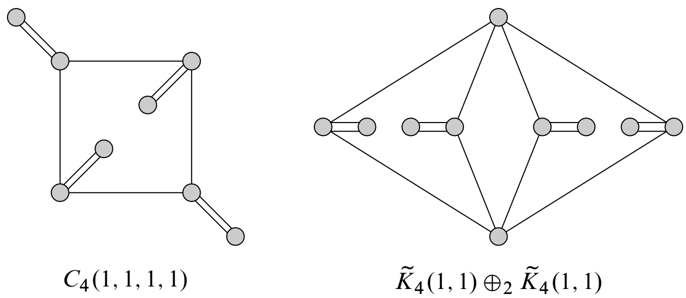


Figure 19: The graphs $C_4(1,1,1,1)$ and $\widetilde{K}_4(1,1) \oplus_2 \widetilde{K}_4(1,1)$

A graph G is called a *doubled forest* if it is obtained from a forest by doubling every edge. A *doubled tree* is a doubled forest with one component. Let $C_4(p,q,r,s)$ be the graph obtained by attaching doubled paths of lengths p, q, r and s to the vertices of a four cycle. Also, let $\widetilde{K}_4(p,q)$ be the graph obtained by removing an edge of the complete graph on four vertices K_4 and then attaching doubled paths of lengths p and q to the vertices incident to the removed edge. Let $\widetilde{K}_4(p,q) \oplus_2 \widetilde{K}_4(r,s)$ be the two-sum of $\widetilde{K}_4(p,q)$ and $\widetilde{K}_4(r,s)$ taken along the unique edge in each summand that is not contained in nor adjacent to a doubled path; see Figure 19.

Lemma 5.1 *Let H be an alternating decomposition graph without isolated vertices such that $g_T(H) = 0$ and H has at most four vertices of degree two. Then H is either*

(1) *a disjoint union of two doubled paths,*

(2) *a doubled tree with two, three or four leaves,*

(3) *$C_4(p,q,r,s)$ for nonnegative integers p, q, r and s, or*

(4) *$\widetilde{K}_4(p,q) \oplus_2 \widetilde{K}_4(r,s)$ for nonnegative integers p, q, r and s.*

Proof Each of the above graphs clearly has four or fewer vertices of degree two, and the algorithm of Corollary 3.9 implies that each of the above graphs is indeed Turaev genus zero. It remains to show that the above list is exhaustive.

Theorem 4.1 states that every Turaev genus zero alternating decomposition graph can be obtained from a collection of isolated vertices via a sequence of doubled pendant moves, two-path extensions and one-sums of vertices in distinct components. If H is obtained from a collection of isolated vertices via a sequence of doubled pendant moves and one-sums from distinct components, then H is a doubled forest. Since H has four or fewer vertices of degree two and no isolated vertices, H is either a disjoint union of two doubled paths or a doubled tree with two, three, or four leaves.

If a doubled tree H has a vertex of degree $2d$ for some positive integer d, then H contains at least d vertices of degree two. A two-path extension always increases the number of degree-two vertices in the graph. Therefore, we can only apply a two-path extension to a vertex of degree two, four, or six. Let H' be obtained from the doubled tree H via a two-path extension applied at a vertex v where the set of edges incident to v is partitioned into sets A and B of odd order, as in Figure 17. Without loss of generality, assume $|A| \geq |B|$.

If the degree of v is two, then $|A| = |B| = 1$. Therefore, a two-path extension will add two new vertices of degree two. Hence H must be a doubled path, and H' is $C_4(p,0,0,0)$ for some p. If the degree of v is four, then $|A| = 3$ and $|B| = 1$. A two-path extension will again add two vertices of degree two, and hence H must be a doubled path. Thus H' is $C_4(p,q,0,0)$ for some p and q.

If the degree of v is six, then H already has at least three vertices of degree two. If $|A| = 5$ and $|B| = 1$, then a two-path extension would create two new vertices of degree two, resulting in at least five vertices of degree two. Therefore $|A| = 3$ and $|B| = 3$, and H is a doubled tree with three degree-two vertices. Let $\mathcal{N}(A)$ (respectively $\mathcal{N}(B)$) be the set of vertices adjacent to v and incident to an edge in A (respectively B). There are two cases: either $|\mathcal{N}(A)| = |\mathcal{N}(B)| = 2$ or $|\mathcal{N}(A)| = |\mathcal{N}(B)| = 3$. In the former case, $H' = C_4(p,0,r,0)$ for some p and r. In the latter, $H' = \widetilde{K}_4(p,0) \oplus_2 \widetilde{K}_4(r,s)$ for some p, r and s.

In each of the above instances, H' already has four vertices of degree two. Thus the only allowable operation is a doubled pendant move applied to a vertex that is already of degree two. Alternately, one could take a one-sum between H' and a doubled path that identifies two degree-two vertices. However, this is the same as a doubled pendant move applied to a vertex of degree two. The only effect this has is changing the parameters in $C_4(p,q,r,s)$ or $\widetilde{K}_4(p,q) \oplus_2 \widetilde{K}_4(r,s)$, and hence the result holds. □

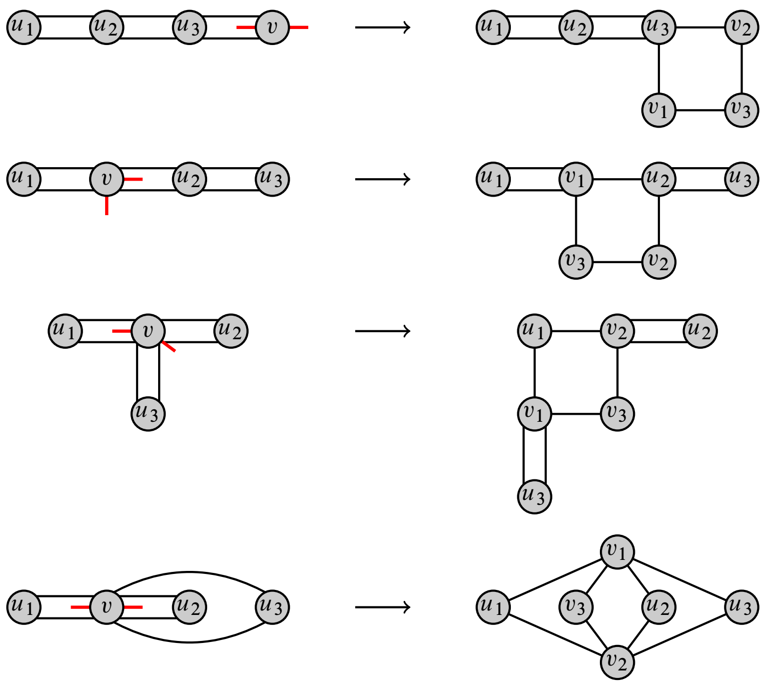


Figure 20: Applying two-path extensions to doubled trees. The short red lines denote the partition of the edges incident to v into the sets A and B.

Figure 20 shows examples of a two-path extension being applied to a doubled tree with two or three vertices of degree two.

The previous classification of alternating decomposition graphs of Turaev genus zero with at most four vertices of degree two leads directly to the classification reduced alternating decomposition graphs of Turaev genus one and two.

Proof of Theorem 1.2 If G is a doubled cycle of even length, then it is reduced and Corollary 3.9 implies that $g_T(G)=1$.

Let G be a reduced alternating decomposition graph with $g_T(G)=1$. Lemma 3.6 implies G contains a pair of parallel edges $\{e_1,e_2\}$. Let $G'=G-\{e_1,e_2\}$. Since G is reduced $k(G')=k(G)$ and thus $g_T(G')=0$. Because G has no vertices of degree two, it follows that G' has at most two vertices of degree two. Lemma 5.1 implies that G' is a doubled path. Therefore G is a doubled cycle of even length. □

Suppose L is a link containing a 2–tangle T inside the ball B. A *mutation* of L is a link L' obtained by removing the ball B, rotating it $180°$ about any of its principle axes, and gluing B back into the link. Two links that are related by a sequence of mutations are said to be mutants of one another.

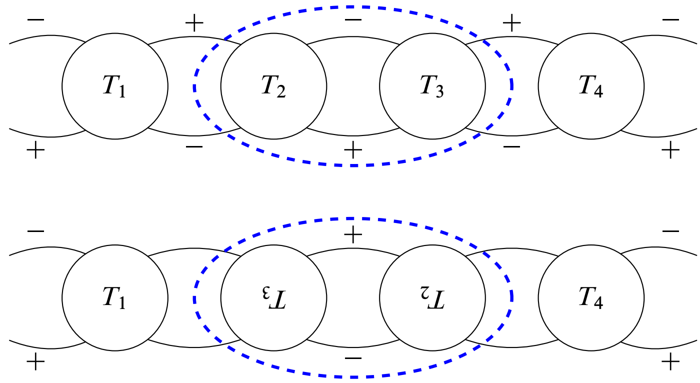


Figure 21: The 2–tangle in the upper diagram is rotated 180° to obtain the lower diagram. In the lower diagram, the 2–tangle containing T_1 and T_3 and the 2–tangle containing T_2 and T_4 are alternating.

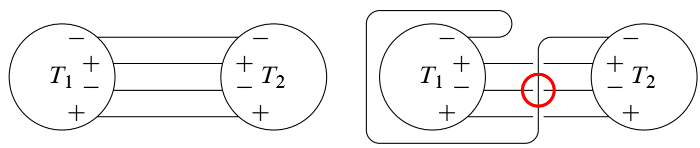


Figure 22: A diagram with alternating decomposition graph C_2^2 is transformed into an almost-alternating diagram by pulling one of the nonalternating edges over one of the tangles. If the circled crossing is changed, then the diagram will be alternating.

Proof of Corollary 1.3 Since L is Turaev genus one, it has a diagram D as in Figure 2. The alternating decomposition graph of this diagram is C_{2k}^2, a doubled cycle of length $2k$. Let T be a the tangle consisting of T_i and T_{i+1}. Rotating the tangle T by 180° in the plane of the diagram results in a new diagram whose alternating decomposition graph is C_{2k-2}^2, a doubled cycle of length $2k-2$; see Figure 21. Therefore, through a sequence of mutations, the diagram D can be transformed into a diagram whose alternating decomposition graph is C_2^2.

It remains to show that any diagram D' with alternating decomposition graph C_2^2 is an almost-alternating link. We may assume that D' consists of two alternating 2–tangles T_1 and T_2 connected together by four nonalternating edges. If one of those nonalternating edges is pulled over the tangle T_1 as in Figure 22, then the resulting diagram is almost-alternating. □

Many Turaev genus one links are known to be almost-alternating. Kim and Lee [21] show that nonalternating, three-stranded pretzel links are almost-alternating. If each tangle T_i in Figure 2 is a rational tangle, then the link L is called a *Montesinos link*.

In the appendix to [2], Jong shows that nonalternating Montesinos links are almost-alternating. Non-alternating Montesinos links include nonalternating pretzel links on arbitrarily many strands. The manipulation of Figure 22 is a key step in Jong's work. All almost-alternating links are Turaev genus one, but it remains open whether all Turaev genus one links are almost-alternating.

Proof of Theorem 1.4 Suppose that

$$G \in \{C_2^2 \sqcup C_2^2, C_2^2 \oplus_1 C_2^2, C_{1,1,1}^2, K_4(2,2), K_4(2) \oplus_2 K_4(2)\}.$$

Corollary 3.9 implies that $g_T(G) = 2$. Proposition 3.11 implies that any alternating decomposition graph that is doubled path equivalent to G also has Turaev genus two.

Let G be a reduced alternating decomposition graph with $g_T(G) = 2$. Since G is reduced and $g_T(G) = 2$, it follows that G contains a pair of parallel edges $\{e_1, e_2\}$ such that $g_T(G') = 1$, where $G' = G - \{e_1, e_2\}$. The graph G' has at most two vertices of degree two. Lemma 3.6 implies that G' contains at least one pair of parallel edges. If the deletion of every pair of parallel edges in G' increased the number of components of G', then every pair could be contracted to obtain the graph $\widetilde{G}'$. Then $g_T(\widetilde{G}') = g_T(G') = 1$, and the graph $\widetilde{G}'$ has at most two vertices of degree two and no pairs of parallel edges. Hence Lemma 3.6 implies $\widetilde{G}'$ has no edges, which contradicts $g_T(\widetilde{G}') = 1$. Thus G' contains a pair of parallel edges $\{e_3, e_4\}$ such that their deletion results in a graph with no more components.

Let $G'' = G - \{e_1, e_2, e_3, e_4\}$. Since G'' is an alternating decomposition graph of Turaev genus zero with at most four vertices of degree two, it is one of the graphs in Lemma 5.1. It remains to show that if G'' is one of the graphs in Lemma 5.1, G can be obtained from G'' by adding two pairs of parallel edges, and G is a reduced alternating decomposition graph of Turaev genus two, then G is doubled path equivalent to one of the five graphs in the statement of the theorem.

Suppose that G'' is a disjoint union of two doubled paths. Then G'' has four vertices v_1, v_2, v_3 and v_4 of degree two, and thus each pair of parallel edges added to G'' must connect two of the degree-two vertices. There are two ways to add these parallel edges, one that results in a disjoint union of two doubled cycles and the other that results in a single doubled cycle. However, a doubled cycle only has Turaev genus one, and so G must be a disjoint union of two doubled cycles, ie G is doubled path equivalent to $C_2^2 \sqcup C_2^2$; see Figure 23.

Suppose that G'' is a doubled path where v_1 and v_2 are its degree-two vertices. If one adds a pair of parallel edges connecting v_1 and v_2, then adds a pair of parallel edges anywhere else to obtain G, then G is doubled path equivalent to $C_{1,1,1}^2$. If one

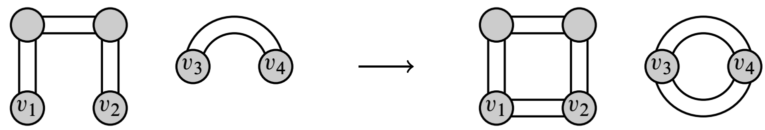


Figure 23: If G'' is a disjoint union of two doubled paths, then G is a disjoint union of two doubled cycles of even length.

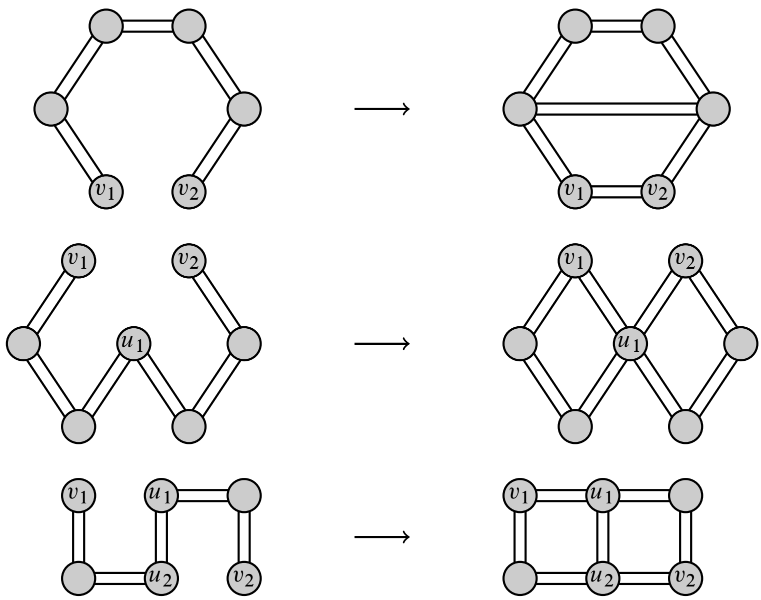


Figure 24: If G'' is a doubled path, then G is doubled path equivalent to either $C_2^2 \oplus_1 C_2^2$ or $C_{1,1,1}^2$.

adds a pair of parallel edges connecting v_1 and some other vertex u_1 and a pair of parallel edges connecting v_2 and some other vertex u_2 to obtain G, then there are three possibilities for G. If u_1 is between v_1 and u_2, then G is not reduced. If $u_1 = u_2$, then G is doubled path equivalent to $C_2^2 \oplus_1 C_2^2$. If u_2 is between v_1 and u_1, then G is doubled path equivalent to $C_{1,1,1}^2$; see Figure 24.

Suppose that G'' is a doubled tree with three vertices v_1, v_2 and v_3 of degree two. Let v be the unique vertex in G'' of degree six. Since G'' contains three vertices of degree two, it follows that two of those vertices must be connected by a pair of parallel edges in G. Without loss of generality, assume we add a pair of parallel edges connecting v_1 and v_2. Also, suppose that we add the other pair of parallel edges connecting v_3 and some other vertex u. If v is between u and v_3, then G is doubled path equivalent to $C_{1,1,1}^2$. If $u = v$, then G is doubled path equivalent to $C_2^2 \oplus_2 C_2^2$. If u is between v and v_3, then G is not reduced; see Figure 25.

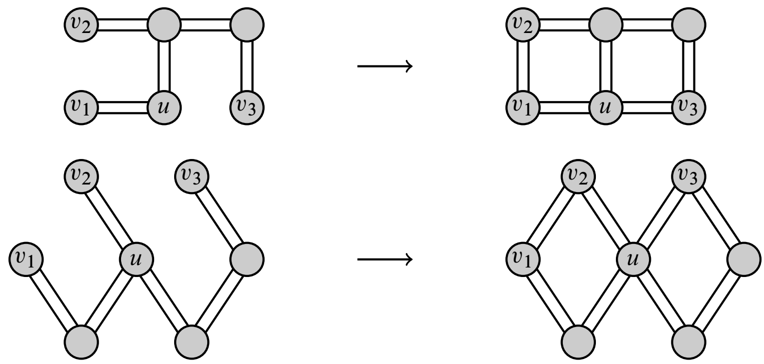


Figure 25: If G'' is a doubled tree with three vertices of degree two, then G is doubled path equivalent to either $C^2_{1,1,1}$ or $C^2_2 \oplus_1 C^2_2$.

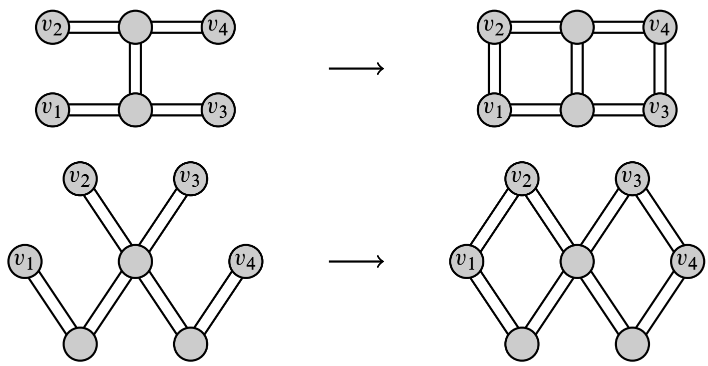


Figure 26: If G'' is a doubled tree with four vertices of degree two, then G is doubled path equivalent to either $C^2_{1,1,1}$ or $C^2_2 \oplus_1 C^2_2$.

Suppose that G'' is a doubled tree with four vertices of degree two. Then one must add one pair of parallel edges connecting two of the degree-two vertices and another pair of parallel edges connecting the other two degree-two vertices. Furthermore G'' either contains two vertices of degree six or one vertex of degree eight. If G'' contains two vertices of degree six, then G is either not reduced or doubled path equivalent to $C^2_{1,1,1}$. If G'' contains a vertex of degree eight, then G is doubled path equivalent to $C^2_2 \oplus_1 C^2_2$; see Figure 26.

Suppose that $G'' = C_4(p,q,r,s)$ for some nonnegative integers p, q, r and s. Since G'' has four vertices of degree two, each pair of parallel edges added to G'' must connect two of the degree-two vertices. The resulting graph is $K_4(\tilde{p},\tilde{q})$ for some values of $\tilde{p}$ and $\tilde{q}$. Thus G is doubled path equivalent to $K_4(2,2)$.

Suppose that $G'' = \widetilde{K}_4(p,q) \oplus_2 \widetilde{K}_4(r,s)$ for some nonnegative integers p, q, r and s. Since G'' has four vertices of degree two, each pair of parallel edges added to G'' must connect two of the degree-two vertices. The resulting graph is $K_4(\tilde{p}) \oplus_2 K_4(\tilde{q})$ for some values of $\tilde{p}$ and $\tilde{q}$. Thus G is doubled path equivalent to $K_4(2) \oplus_2 K_4(2)$.

Hence if G is a reduced alternating decomposition graph with $g_T(G) = 2$, then G is doubled path equivalent to one of $C_2^2 \sqcup C_2^2$, $C_2^2 \oplus_1 C_2^2$, $C_{1,1,1}^2$, $K_4(2,2)$, or $K_4(2) \oplus_2 K_4(2)$. □

Suppose G has $v(G)$ vertices, $e(G)$ edges, and $k(G)$ components. The *nullity* $n(G)$ of G is defined as

$$n(G) = e(G) - v(G) + k(G).$$

One can equivalently define the nullity of G to be the nullity of the incidence matrix of G or to be the number of edges not in a maximal spanning forest of G. The simplification $\mathrm{si}(G)$ of the graph G is the graph obtained from G by deleting loops and replacing each set of multiple edges connecting two distinct vertices v_1 and v_2 with a single edge connecting v_1 and v_2. As long as an alternating decomposition graph G does not have any vertices of degree two, its Turaev genus is bounded below by the nullity of the simplification of G in the following manner.

Proposition 5.2 *Let G be an alternating decomposition graph, and let $\mathrm{si}(G)$ be the simplification of G. If G contains no vertices of degree two, then $3g_T(G) \geq n(\mathrm{si}(G))$.*

Proof Since G is assumed to have no vertices of degree two, the base case is $G = C_2^2$, a doubled cycle of length two, ie G contains two vertices with four parallel edges between them. In this case $g_T(G) = 1$ and $n(\mathrm{si}(G)) = 0$, and so the result holds.

Now suppose that the desired inequality holds for all alternating decomposition graphs with no vertices of degree two that have fewer edges than G. Since G does not contain any vertices of degree two, Lemma 3.6 implies that G contains a pair of parallel edges e_1 and e_2. Let $G' = G - \{e_1, e_2\}$, and let e_{12} be the edge in $\mathrm{si}(G)$ corresponding to e_1 and e_2.

Suppose that $k(G') = k(G) + 1$. Then $g_T(G') = g_T(G)$. The edge e_{12} is a bridge in $\mathrm{si}(G)$, and thus $n(\mathrm{si}(G')) = n(\mathrm{si}(G))$. By induction, $3g_T(G') \geq n(\mathrm{si}(G'))$, and hence $3g_T(G) \geq n(\mathrm{si}(G))$.

Suppose that $k(G') = k(G)$. Then $g_T(G) = g_T(G') + 1$ and $n(\mathrm{si}(G)) \leq n(\mathrm{si}(G')) + 1$. Let v_1 and v_2 be the two vertices incident to e_1 and e_2 in G. For $i = 1$ or 2, we consider three cases:

(1) The degree of v_i is greater than two.

(2) The vertex v_i has degree two and two distinct neighbors.

(3) The vertex v_i has degree two and only one distinct neighbor.

In order to apply our inductive hypothesis, we eliminate all vertices of degree two in G' as follows. If $\deg v_i > 2$, then nothing needs to be done. If $\deg v_i = 2$ and v_i has two distinct neighbors, then perform a two-path contraction at v_i. A two-path contraction does not change the Turaev genus of the graph but could decrease the nullity of the simplification of the graph by one. Suppose that $\deg v_i = 2$ and v_i has only one neighbor. Let P_i be the maximal doubled path embedded in G' with endpoints v_i and u_i such that every interior vertex of P_i has exactly two neighbors. If every edge in P_i is contracted, then both the Turaev genus and the nullity of the simplification of the resulting graph remain unchanged.

Let G'' be the graph obtained from G' by performing the above operations on v_1 and v_2. Then G'' has no vertices of degree two. We have

$$g_T(G'') = g_T(G') \quad \text{and} \quad n(\mathrm{si}(G'')) + 2 \geq n(\mathrm{si}(G')).$$

Therefore

$$g_T(G) = g_T(G'') + 1 \quad \text{and} \quad n(\mathrm{si}(G)) \leq n(\mathrm{si}(G'')) + 3.$$

By the inductive hypothesis, we have $n(\mathrm{si}(G'')) \leq 3g_T(G'')$. Therefore

$$n(\mathrm{si}(G)) \leq n(\mathrm{si}(G'')) + 3 \leq 3g_T(G'') + 3 = 3g_T(G). \qquad \square$$

We use the next lemma in the proof of Theorem 1.5, which will conclude the paper.

Lemma 5.3 *Let n_1 and n_2 be nonnegative integers. There are a finite number of graphs G such that $n(G) = n_1$ and such that G contains n_2 vertices of degree two.*

Proof Because nullity is additive with respect to disjoint union, it suffices to show the above statement for connected graphs. Let T be a tree, and let $d_{12}(T)$ be the number of degree-one or degree-two vertices in T. Suppose that T is a spanning tree of a graph G with $n(G) = n_1$ where G contains n_2 vertices of degree two. Hence G is obtained from T by adding n_1 edges. Each of the n_1 edges added to T can make at most two of the vertices of degree one or two in T have degree larger than two in G. Also, every degree-two vertex in G is either a degree one or a degree-two vertex in T. Therefore $d_{12}(T) \leq 2n_1 + n_2$.

Every tree can be obtained from a single vertex by repeatedly adding pendant edges. Each pendant edge addition increases $d_{12}(T)$, and for a given tree, there are only finitely many ways to add a pendant edge. Thus the number of trees T with $d_{12}(T) \leq 2n_1 + n_2$ is finite. There are only a finite number of ways to add n_1 edges to such a tree, and hence there exists a finite number of graphs G with nullity n_1 that contain n_2 vertices of degree two. $\square$

We end the paper with the proof of Theorem 1.5.

Proof of Theorem 1.5 For each doubled path equivalence class c of reduced alternating decomposition graphs G with $g_T(G) = k$, let G_c be a representative such that no other representative of c can be obtained from G_c via a sequence of doubled path contractions. Let V' be the set of vertices v in G such that $\deg v = 4$, each v has exactly two distinct neighbors u and w, there are two edges incident to both u and v, and there are two edges incident to both w and v.

For each vertex $v \in V'$, there are two pairs of parallel edges incident to v, say parallel edges $e_{v,1}$ and $e_{v,2}$ and parallel edges $e_{v,3}$ and $e_{v,4}$. Let E' be a set of edges containing exactly one pair of these parallel edges for each $v \in V'$, that is, $E' = \{e_{v,1}, e_{v,2} \mid v \in V'\}$. We claim that the graph $G_c - E'$, ie the graph obtained by deleting the edges set E' from G_c, has the same number of components as G_c.

By way of contradiction, suppose that $G_c - E'$ has more components than G_c. Then there exists a minimal subset E'' of E' such that $G_c - E''$ has one more component than G_c, but $G_c - S$ has the same number of components as G_c for any proper subset S of E''. Note that if an edge $e_{v,1}$ is in E'', then its parallel edge $e_{v,2}$ is also in E''. Therefore if $G_c'' = G_c/E''$, ie the contraction of the edges in E'' from G_c, then G_c'' is obtained from G_c via a sequence of doubled path contractions.

Let C'' be a cycle in G_c''. Then there is a cycle C in G_c such that $C'' = C/(C \cap E'')$. Since G_c is bipartite, it follows that C has an even number of edges. Since adding any single edge of E'' to $G_c - E''$ connects two components of G_c, it follows that $C \cap E''$ has an even number of edges. Therefore, C'' has an even number of edges. Because each cycle of G_c'' has an even number of edges, the graph G_c'' is bipartite. Thus G_c'' is an alternating decomposition graph, which contradicts that no other representative of c can be obtained from G_c via a sequence of doubled path contractions.

Therefore $G_c - E'$ has the same number of components as G_c. Hence deleting each pair of parallel edges in E' from G_c decreases the Turaev genus by one, which implies that $|E'| \le 2k$ and $|V'| \le k$. Each vertex $v \in V'$ has degree two in the simplification $\mathrm{si}(G_c)$.

Any other vertex of degree two in $\mathrm{si}(G_c)$ arises from a vertex v in G_c with two distinct neighbors v_1 and v_2 such that there are r edges between v and v_1 and s edges between v and v_2, where $r+s$ is even and $\max\{r,s\} > 2$. For each such vertex, there are two parallel edges whose removal decreases Turaev genus by one and does not change the simplification $\mathrm{si}(G_c)$. Because pairs of such vertices could be adjacent, there are at most $2k$ in G_c.

Therefore $\mathrm{si}(G_c)$ has at most $3k$ vertices of degree two. Moreover $3k = 3g_T(G_c) \ge n(\mathrm{si}(G_c))$. Because the nullity and the number of degree-two vertices are bounded, Lemma 5.3 implies that there are only a finite number of candidates for the graph $\mathrm{si}(G_c)$. Because adding arbitrarily many parallel edges to an alternating decomposition graph increases its Turaev genus without bound, there are only a finite number of alternating decomposition graphs of a fixed Turaev genus whose simplification is a given graph. Therefore, there are only finitely many doubled path equivalence classes of alternating decomposition graphs of Turaev genus k. □

References

[1] **T Abe**, *The Turaev genus of an adequate knot*, Topology Appl. 156 (2009) 2704–2712 MR

[2] **T Abe**, **K Kishimoto**, *The dealternating number and the alternation number of a closed 3–braid*, J. Knot Theory Ramifications 19 (2010) 1157–1181 MR

[3] **C C Adams**, *Toroidally alternating knots and links*, Topology 33 (1994) 353–369 MR

[4] **C C Adams**, **J F Brock**, **J Bugbee**, **T D Comar**, **K A Faigin**, **A M Huston**, **A M Joseph**, **D Pesikoff**, *Almost alternating links*, Topology Appl. 46 (1992) 151–165 MR

[5] **C Armond**, **N Druivenga**, **T Kindred**, *Heegaard diagrams corresponding to Turaev surfaces*, J. Knot Theory Ramifications 24 (2015) art. id. 1550026, 1–14 MR

[6] **Y Bae**, **H R Morton**, *The spread and extreme terms of Jones polynomials*, J. Knot Theory Ramifications 12 (2003) 359–373 MR

[7] **C L J Balm**, *Topics in knot theory: on generalized crossing changes and the additivity of the Turaev genus*, PhD thesis, Michigan State University (2013) MR Available at `http://search.proquest.com/docview/1356063624`

[8] **A Champanerkar**, **I Kofman**, *Twisting quasi-alternating links*, Proc. Amer. Math. Soc. 137 (2009) 2451–2458 MR

[9] **A Champanerkar**, **I Kofman**, *A survey on the Turaev genus of knots*, Acta Math. Vietnam. 39 (2014) 497–514 MR

[10] **A Champanerkar**, **I Kofman**, **N Stoltzfus**, *Graphs on surfaces and Khovanov homology*, Algebr. Geom. Topol. 7 (2007) 1531–1540 MR

[11] **O T Dasbach**, **D Futer**, **E Kalfagianni**, **X-S Lin**, **N W Stoltzfus**, *The Jones polynomial and graphs on surfaces*, J. Combin. Theory Ser. B 98 (2008) 384–399 MR

[12] **O T Dasbach**, **D Futer**, **E Kalfagianni**, **X-S Lin**, **N W Stoltzfus**, *Alternating sum formulae for the determinant and other link invariants*, J. Knot Theory Ramifications 19 (2010) 765–782 MR

[13] **O T Dasbach**, **A M Lowrance**, *Turaev genus, knot signature, and the knot homology concordance invariants*, Proc. Amer. Math. Soc. 139 (2011) 2631–2645 MR

[14] **O T Dasbach**, **A M Lowrance**, *A Turaev surface approach to Khovanov homology*, Quantum Topol. 5 (2014) 425–486 MR

[15] **O T Dasbach**, **A M Lowrance**, *Invariants for Turaev genus one links*, preprint (2016) arXiv

[16] **J A Ellis-Monaghan**, **I Moffatt**, *Graphs on surfaces: dualities, polynomials, and knots*, Springer Briefs in Mathematics, Springer (2013) MR

[17] **V F R Jones**, *A polynomial invariant for knots via von Neumann algebras*, Bull. Amer. Math. Soc. 12 (1985) 103–111 MR

[18] **L H Kauffman**, *State models and the Jones polynomial*, Topology 26 (1987) 395–407 MR

[19] **M Khovanov**, *A categorification of the Jones polynomial*, Duke Math. J. 101 (2000) 359–426 MR

[20] **M Khovanov**, *Patterns in knot cohomology, I*, Experiment. Math. 12 (2003) 365–374 MR

[21] **D Kim**, **J Lee**, *Some invariants of pretzel links*, Bull. Austral. Math. Soc. 75 (2007) 253–271 MR

[22] **S Kim**, *Link diagrams with low Turaev genus*, preprint (2015) arXiv

[23] **A M Lowrance**, *On knot Floer width and Turaev genus*, Algebr. Geom. Topol. 8 (2008) 1141–1162 MR

[24] **A Lowrance**, *The Khovanov width of twisted links and closed* 3*–braids*, Comment. Math. Helv. 86 (2011) 675–706 MR

[25] **A M Lowrance**, *Alternating distances of knots and links*, Topology Appl. 182 (2015) 53–70 MR

[26] **K Murasugi**, *Jones polynomials and classical conjectures in knot theory*, Topology 26 (1987) 187–194 MR

[27] **P Ozsváth**, **Z Szabó**, *Knot Floer homology and the four-ball genus*, Geom. Topol. 7 (2003) 615–639 MR

[28] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and knot invariants*, Adv. Math. 186 (2004) 58–116 MR

[29] **J A Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard University (2003) MR Available at `http://search.proquest.com/docview/305332635`

[30] **J Rasmussen**, *Khovanov homology and the slice genus*, Invent. Math. 182 (2010) 419–447 MR

[31] **M B Thistlethwaite**, *An upper bound for the breadth of the Jones polynomial*, Math. Proc. Cambridge Philos. Soc. 103 (1988) 451–456 MR

[32] **V G Turaev**, *A simple proof of the Murasugi and Kauffman theorems on alternating links*, Enseign. Math. 33 (1987) 203–225 MR

Department of Mathematics, Ohio State University at Mansfield
1760 University Drive, Mansfield, OH 44906, United States

Department of Mathematics, Vassar College
124 Raymond Ave Box 257, Poughkeepsie, NY 12604, United States

armond.2@osu.edu, adlowrance@vassar.edu

Received: 9 July 2015 Revised: 15 June 2016

Algebraic & Geometric Topology 17 (2017) 831–846

Constructing geometrically equivalent hyperbolic orbifolds

David McReynolds
Jeffrey S Meyer
Matthew Stover

We construct families of nonisometric hyperbolic orbifolds that contain the same isometry classes of nonflat totally geodesic subspaces. The main tool is a variant of the well-known Sunada method for constructing length-isospectral Riemannian manifolds that handles totally geodesic submanifolds of multiple codimensions simultaneously.

51M10, 58J53; 11F06

1 Introduction

Classical spectra like the eigenvalue spectrum of the Laplace–Beltrami operator or the primitive geodesic length spectrum have played an important role in dynamics, geometry, and representation theory. We continue the investigation of higher-dimensional spectra that encode the geometry of the nonflat totally geodesic submanifolds of a fixed complete, finite-volume, Riemannian manifold M. We will refer to the set of such submanifolds, counted with multiplicity, as the *geometric spectrum*.

To construct our examples, we restrict ourselves to closed arithmetic locally symmetric orbifolds, where recent work shows that the geometric spectrum, when nonempty, carries much information. In McReynolds and Reid [3] it was shown that if M_1, M_2 are arithmetic hyperbolic 3–manifolds with the same geometric spectrum, provided the geometric spectrum is nonempty, then M_1 and M_2 are commensurable. For higher dimensions, Meyer [4, Theorem C] proved that if M_1 and M_2 are standard arithmetic hyperbolic m–manifolds (see Section 2) with the same geometric spectrum, then M_1 and M_2 are commensurable. It is well-known that the geometric spectrum of a standard arithmetic hyperbolic m–manifold is nonempty with representatives in every possible proper codimension.

For any finite-volume, hyperbolic 3–manifold M, there exist infinitely many pairs of nonisometric finite covers (M_j, N_j) of M such that M_j and N_j have the same totally geodesic surfaces [3]. This has two parts. First, there are infinitely many pairs of finite covers (M_j', N_j') with the same geometric spectrum. It is a feature of this

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.831

construction that $\mathrm{Vol}(M_j') = \mathrm{Vol}(N_j')$, though we know no general reason why that must hold. Secondly, there exist infinitely many pairs $\{M_j, N_j\}$ with the same set of totally geodesic surfaces (ie without multiplicity) such that $\mathrm{Vol}(M_j)/\,\mathrm{Vol}(N_j)$ is unbounded.

The main result of this article is the generalization of these covering constructions to higher-dimensional hyperbolic manifolds. We use a variant of the well-known Sunada method for producing length-isospectral Riemannian manifolds [11] that allows one to handle totally geodesic submanifolds of varying codimensions. The case of totally geodesic subsurfaces of a hyperbolic 3–manifold is handled by [3], and the challenge we overcome is to address all codimensions simultaneously.

Define the *totally geodesic spectrum* of a locally symmetric Riemannian orbifold M to be the set

$$\mathcal{TG}(M) = \left\{ \begin{array}{c} \text{isometry classes of orientable nonflat finite-} \\ \text{volume totally geodesic subspaces } X \subset M \\ \text{with multiplicity } m_X \end{array} \right\} = \{(X_j, m_{X_j})\}. \tag{1}$$

We say that M_1 and M_2 are *geometrically isospectral* if $\mathcal{TG}(M_1) = \mathcal{TG}(M_2)$. The *totally geodesic set* of a locally symmetric, Riemannian orbifold is

$$\mathrm{TG}(M) = \left\{ \begin{array}{c} \text{isometry classes of orientable nonflat finite-} \\ \text{volume totally geodesic subspaces } X \subset M \end{array} \right\} = \{(X_j)\}. \tag{2}$$

We say that M_1, M_2 are *geometrically equivalent* if $\mathrm{TG}(M_1) = \mathrm{TG}(M_2)$.

Theorem 1.1 *For every commensurability class $\mathcal{C}$ of closed arithmetic hyperbolic m–orbifolds with $m \geq 3$, we have the following:*

(a) *For each $M \in \mathcal{C}$, there exist nonisometric finite covers M' and N' of M such that $\mathcal{TG}(M') = \mathcal{TG}(N')$.*

(b) *For each $M \in \mathcal{C}$, there exist infinitely many pairs of nonisometric, finite covers (M_j, N_j) of M such that*
   - (i) $\mathrm{TG}(M_j) = \mathrm{TG}(N_j)$ *for all j;*
   - (ii) *the ratio $\mathrm{Vol}(M_j)/\,\mathrm{Vol}(N_j)$ is unbounded.*

The orientability condition in (2) is a matter of taste, as a small modification of our methods allows for nonorientable geodesic subspaces. Our methods can produce examples modeled on other symmetric spaces of noncompact type, but the technicalities would obscure the basic ideas behind our construction, which is general enough to highlight the basic procedure (see Theorem 5.3 for a generalization of Theorem 1.1).

Acknowledgements The authors acknowledge support from the US National Science Foundation grants DMS-1107452, 1107263, 1107367 “RNMS: GEometric structures

And Representation varieties" (the GEAR Network). McReynolds was supported by NSF grants DMS-1105710 and DMS-1408458. Stover was supported by NSF grant DMS-1361000.

2 Notation and overview

In this section, we outline the construction of the covers required to prove our main results. Before providing this outline, we briefly set some notation and terminology that will be used throughout the article.

2.1 Preliminaries

A finite-volume hyperbolic m–manifold M is *arithmetic* if its fundamental group Γ has a commensurator $\mathrm{Comm}(\Gamma) = \{g \in \mathrm{Isom}^+(\mathbb{H}^m) \mid \Gamma, g\Gamma g^{-1} \text{ are commensurable}\}$ that is dense in $\mathrm{Isom}^+(\mathbb{H}^m)$ (see [5, (16.3.3)]). The subclass that exhibits the richest collections of totally geodesic submanifolds is the subclass of so-called standard arithmetic manifolds, which we now describe.

Throughout this paper, k denotes a number field, $\mathcal{O}_k$ its ring of integers, and q a nondegenerate quadratic form over k. For a prime ideal $\mathfrak{p}$ of $\mathcal{O}_k$, let $k_{\mathfrak{p}}$ denote the localization of k at $\mathfrak{p}$ and $\mathcal{O}_{\mathfrak{p}}$ is its ring of integers. Call (k, q) an *admissible hyperbolic pair* when k is totally real and q is positive definite at all but one real place of k, at which it has signature $(m, 1)$. Set $\mathbf{G} = \mathbf{SO}(q)$, fix a k–rational embedding $\iota\colon \mathbf{G} \to \mathbf{GL}_d$, and define $\mathbf{G}(\mathcal{O}_k) = \iota^{-1}(\iota(\mathbf{G}(k)) \cap \mathbf{GL}_d(\mathcal{O}_k))$. Since the k–isomorphism class of $\mathbf{G}$ is independent of the similarity class of q, we can assume that the matrix representative $\iota(q)$ for q lies in $\mathbf{GL}_d(\mathcal{O}_k)$.

An admissible hyperbolic pair gives rise to a commensurability class of m–dimensional hyperbolic orbifolds as follows. Restriction of scalars followed by the appropriate projection induces a map $\pi\colon \mathbf{G}(k) \to \mathbf{PSO}_0(m, 1)$ with finite kernel, and we call the image $\Gamma_q = \pi(\mathbf{G}(\mathcal{O}_k))$ a *principle arithmetic lattice* in $\mathbf{PSO}_0(m, 1)$. As $\mathbf{PSO}_0(m, 1) = \mathrm{Isom}^+(\mathbb{H}^m)$, the lattice Γ_q is also the orbifold fundamental group of the orientable hyperbolic orbifold $M_{\Gamma_q} = \Gamma_q \backslash \mathbb{H}^m$.

We call hyperbolic manifolds commensurable with M_{Γ_q} *standard arithmetic manifolds*, and emphasize that every even-dimensional arithmetic hyperbolic manifold is standard. However, when m is odd, there are infinitely many commensurability classes of non-standard arithmetic lattices. See [4] for more details on parametrizing commensurability classes of arithmetic hyperbolic orbifolds.

For any lattice Γ in $\mathbf{PSO}_0(m, 1)$, let $\widetilde{\Gamma}$ be the lift of Γ to $\mathbf{SO}_0(m, 1)$. When m is even, the groups $\mathbf{PSO}_0(m, 1)$, $\mathbf{SO}_0(m, 1)$ are isomorphic and so $\widetilde{\Gamma} \cong \Gamma$. When m is odd, $\mathbf{SO}_0(m,1)$ is a two-fold covering of $\mathbf{PSO}_0(m,1)$, and hence we have a central exact sequence

$$1 \longrightarrow \mu_2 \longrightarrow \widetilde{\Gamma} \longrightarrow \Gamma \longrightarrow 1,$$

where μ_2, the group of 2^{nd} roots of unity, is the center of $\mathbf{SO}_0(m,1)$. If this exact sequence does not split, there is an index-two subgroup of Γ for which the associated sequence does split. In other words, possibly passing to an index-two subgroup when m is odd, we can assume that Γ embeds as a lattice in $\mathbf{SO}_0(m,1)$.

Associated with any totally geodesic embedding $f\colon \mathbb{H}^n \hookrightarrow \mathbb{H}^m$ is an injection

$$f_*\colon \mathbf{PS}_0(\mathbf{O}(n,1)\times\mathbf{O}(m-n)) \hookrightarrow \mathbf{PSO}_0(m,1),$$

and we will denote the image by H_f. Given a torsion-free lattice Γ in $\mathbf{PO}_0(m,1)$, proper, totally geodesic, finite-volume submanifolds of $M_\Gamma = \Gamma\backslash\mathbb{H}^m$ are then associated with embeddings f as above such that $\Gamma\cap H_f$ is a lattice in H_f. Notice that, while M_Γ is an orientable manifold, a geodesic submanifold can be nonorientable. Moreover, the submanifold is oriented if and only if $(\Gamma\cap H_f) \subset f_*\big(\mathbf{P}_0(\mathbf{SO}(n,1)\times\mathbf{SO}(m-n))\big)$.

We now relate $\Gamma\cap H_f$ to the fundamental group of the geodesic submanifold. Let $N_\Lambda = \Lambda\backslash\mathbb{H}^n$ be an oriented totally geodesic submanifold of M_Γ of dimension n. Then we have an injective homomorphism $\Lambda\to\Gamma$. Choosing a lifting of $N_\Lambda \to M_\Gamma$ to an embedding $f\colon \mathbb{H}^n \hookrightarrow \mathbb{H}^m$ of universal coverings, we see that Λ is a subgroup of $\Gamma\cap H_f$. Assuming that Γ lifts to $\mathbf{SO}_0(m,1)$, we obtain an injective homomorphism $f_\star\colon \Lambda\to\mathbf{SO}(n,1)\times\mathbf{SO}(m-n)$. The real Zariski closure of $f_\star(\Lambda)$ is then of the form $\mathbf{SO}_0(n,1)\times H_\Lambda$ for some closed subgroup H_Λ of $\mathbf{SO}(m-n)$.

As is well-known, an orientable finite-volume totally geodesic subspace N_Λ of M_Γ is also arithmetic [4, Section 3]. Associated with N_Λ is an $(n+1)$–dimensional quadratic subform r of q with orthogonal complement t (ie q is k–isometric to $r\oplus t$) such that the k–groups $\mathbf{H}_r = \mathbf{SO}(r)$, $\mathbf{H}_t = \mathbf{SO}(t)$ and $\mathbf{H} = \mathbf{H}_r\times\mathbf{H}_t$ satisfy

$$\mathbf{H}_r(\mathbb{R}) = \mathbf{SO}(n,1), \quad \mathbf{H}_t(\mathbb{R}) = \mathbf{SO}(m-n) \quad \text{and} \quad \widetilde{\Lambda} = f_\star(\Lambda)\subset\mathbf{H}(k).$$

The semisimple k–group $\mathbf{H}$ is naturally a k–subgroup of $\mathbf{G}$. We call Λ *a totally geodesic subgroup* of either Γ or the lift $\widetilde{\Gamma}$ of Γ to $\mathbf{G}(\mathcal{O}_k)$; recall from above that Λ is isomorphic to a subgroup of both Γ and $\widetilde{\Gamma}$.

2.2 Strategy of proof for geometric equivalence

We will find a finite group G, a surjective homomorphism $\rho\colon \Gamma\to G$, and two subgroups $C_1, C_2\subset G$ such that

$$\rho(\Lambda)\cap C_1 = \rho(\Lambda)\cap C_2 \tag{3}$$

for all totally geodesic $\Lambda \subset \Gamma$. It then follows from covering space theory that the finite covers M_1, M_2 associated with $\Gamma_1 = \rho^{-1}(C_1)$, $\Gamma_2 = \rho^{-1}(C_2)$ contain exactly the same totally geodesic submanifolds (see [3, Lemma 4.1]). Thus, it suffices to find a map $\rho\colon \mathbf{G}(\mathcal{O}_k) \to G$ such that $\gcd([\mathbf{G}(\mathcal{O}_K) : \ker\pi \cap \mathbf{G}(\mathcal{O}_k)], |C_i|) = 1$ and such that (3) holds. Let S_0 denote the set of nondyadic primes of $\mathcal{O}_k$ not lying over a prime dividing the index $[\mathbf{G}(\mathcal{O}_K) : \ker\pi \cap \mathbf{G}(\mathcal{O}_k)]$. The candidates for G and ρ are the natural reduction maps $\rho_{\mathfrak{p}}\colon \mathbf{G}(\mathcal{O}_k) \to \mathbf{G}(\mathcal{O}_k/\mathfrak{p})$, where $\mathfrak{p}$ is a prime ideal of $\mathcal{O}_k$. Set $\mathbb{F}_{p^r} = \mathcal{O}_k/\mathfrak{p}$, where $|\mathcal{O}_k/\mathfrak{p}| = p^r$. For a totally geodesic subgroup Λ, set $H_{\mathfrak{p}} = \rho_{\mathfrak{p}}(\widetilde{\Lambda})$, which sits inside of $\rho_{\mathfrak{p}}(\mathbf{G}(\mathcal{O}_k))$. For our examples, C_1 will be the trivial subgroup and C_ℓ will be a cyclic group of prime order ℓ such that ℓ does not divide the order of $H_{\mathfrak{p}}$ for any totally geodesic subgroup. In that case, (3) will be satisfied and the manifolds M_1 and M_ℓ associated with the pullbacks of C_1 and C_ℓ will be geometrically equivalent. Furthermore, notice that, since our covering has odd degree, nonorientable manifolds only lift to nonorientable manifolds, so $\mathrm{TG}(M_1)$, which only contains oriented submanifolds, indeed equals $\mathrm{TG}(M_\ell)$.

Finding the desired prime ℓ requires two main steps:

(a) Compute $|\rho_{\mathfrak{p}}(\mathbf{G}(\mathcal{O}_k))|$. This step uses structure theory of algebraic groups, basic Galois cohomology, and strong approximation. We obtain the diagram

$$\begin{tikzcd} & & \widetilde{\mathbf{G}}(\mathcal{O}_k) \arrow[r] \arrow[d, two heads, "\rho_p"'] & \mathbf{G}(\mathcal{O}_k) \arrow[d, "\rho_p"'] & & \\ 1 \arrow[r] & \mathbf{F}(\mathbb{F}_{p^r}) \arrow[r] & \widetilde{\mathbf{G}}(\mathbb{F}_{p^r}) \arrow[r] & \mathbf{G}(\mathbb{F}_{p^r}) \arrow[r] & H^1(\mathbb{F}_{p^r}, \mathbf{F}) \arrow[r] & 1 \end{tikzcd} \tag{4}$$

where $\widetilde{\mathbf{G}}$ is the simply connected cover of $\mathbf{G}$ and $\mathbf{F}$ is a finite $\mathbb{F}_{p^r}$–group.

(b) Determine all possible divisors of $|H_{\mathfrak{p}}|$. This step uses Bruhat–Tits-theoretic computations associated with the diagram

$$\begin{tikzcd} \widetilde{\Lambda} \arrow[r, hook] \arrow[d, "\rho_p"'] & \mathbf{H}(k_{\mathfrak{p}}) \cap \mathbf{G}(\mathcal{O}_{\mathfrak{p}}) \arrow[d, two heads] \\ H_{\mathfrak{p}} \arrow[r, hook] & \overline{\mathbf{H}}(\mathbb{F}_{p^r}) \end{tikzcd} \tag{5}$$

where $\overline{\mathbf{H}}$ is a certain algebraic $\mathbb{F}_{p^r}$–group associated with $\mathbf{H}$. We know the right vertical map is surjective, and hence we can realize $H_{\mathfrak{p}}$ as a subgroup of $\overline{\mathbf{H}}(\mathbb{F}_{p^r})$. Recall that $k_{\mathfrak{p}}$ denotes the localization of k at $\mathfrak{p}$ and $\mathcal{O}_{\mathfrak{p}}$ is its ring of integers.

Using the calculations for the orders of the groups $\rho_{\mathfrak{p}}(\mathbf{G}(\mathcal{O}_k))$ and the subgroups $H_{\mathfrak{p}}$, we find the prime ℓ using Zsigmondy's theorem [15].

2.3 Strategy of proof for geometric isospectrality

Following [3], to produce geometrically isospectral manifolds we require two good primes $\mathfrak{p}_1, \mathfrak{p}_2$ where we can use the same prime ℓ for both $\mathfrak{p}_1$ and $\mathfrak{p}_2$ in the above construction. The key observation in using the two primes $\mathfrak{p}_1, \mathfrak{p}_2$ is that, since M_1 is a cyclic cover of degree ℓ to which every geodesic submanifold of M_ℓ has exactly ℓ distinct lifts, the geometric spectrum of the orbifolds satisfies

$$\mathcal{TG}(M_1) = \{(X, m_{X,1})\} = \{(X, \ell m_{X,\ell})\}, \tag{6}$$

where $\mathcal{TG}(M_\ell) = \{(X, m_{X,\ell})\}$. The validity of (6) follows from the argument used in [3, page 178] to establish this for totally geodesic subsurfaces of a hyperbolic 3–manifold. That there exists a prime ℓ that satisfies the necessary properties for both $\mathfrak{p}_1$ and $\mathfrak{p}_2$ is a straightforward application of the Chebotarev density theorem. In particular, there is a positive-density set of primes $\mathfrak{p}$ for which our methods apply.

3 Step (a) Computing $|\rho_{\mathfrak{p}}(\mathbf{G}(\mathcal{O}_k))|$

For each $\mathfrak{p} \in S_0$, let $q_{\mathfrak{p}}$ denote the reduction of q to $\mathcal{O}_k/\mathfrak{q} = \mathbb{F}_{p^r}$. We will say q has a *good reduction* at $\mathfrak{p}$ if $q_{\mathfrak{p}}$ is nondegenerate; note that the subset $S_1 \subset S_0$ where q has good reduction is cofinite. For $\mathfrak{p} \in S_1$, set $G_{\mathfrak{p}} = \mathrm{SO}(m+1; p^r)$ to be the $\mathbb{F}_{p^r}$–points of $\mathbf{SO}(q_{\mathfrak{p}})$. Over a finite field, orthogonal groups are always quasisplit, and hence come in one of three types (see [7, Table 1] for the orders of these groups):

- $B_{n,n}$, the only form of B_n, arises when $\dim q = 2n+1$. It has order

$$|\mathrm{SO}(2n+1; p^r)| = p^{rn^2} \prod_{j=1}^{n} (p^{2rj} - 1). \tag{7}$$

- $D_{n,n}$, the split form of D_n, arises when $\dim q = 2n$ and $\operatorname{disc} q$ is a square in $\mathbb{F}_{p^r}$. It has order

$$|\mathrm{SO}^+(2n; p^r)| = p^{rn(n-1)}(p^{rn} - 1) \prod_{j=1}^{n-1} (p^{2rj} - 1). \tag{8}$$

- $D_{n,n-1}$, the nonsplit quasisplit form of D_n, arises when $\dim q = 2n$ and $\operatorname{disc} q$ is not square in $\mathbb{F}_{p^r}$. It has order

$$|\mathrm{SO}^-(2n; p^r)| = p^{rn(n-1)}(p^{rn} + 1) \prod_{j=1}^{n-1} (p^{2rj} - 1). \tag{9}$$

We have the exact sequence of algebraic k–groups (see [8, Section 2.3])

$$1 \longrightarrow \mu_2 \longrightarrow \mathbf{Spin}(q) \longrightarrow \mathbf{SO}(q) \longrightarrow 1,$$

where μ_2 is the cyclic group of order two. This sequence yields the exact sequence for $\mathbb{F}_{p^r}$–points

$$1 \longrightarrow \mu_2 \longrightarrow \mathbf{Spin}(q)(\mathbb{F}_{p^r}) \longrightarrow \mathbf{SO}(q)(\mathbb{F}_{p^r}) \longrightarrow \mathbb{F}_{p^r}^{\times}/(\mathbb{F}_{p^r}^{\times})^2 \longrightarrow 1.$$

Strong approximation (see Lemma 1.1 and Theorem 2.3 in [10]) gives us that

$$\rho_{\mathfrak{p}} \colon \mathbf{Spin}(q)(\mathcal{O}_k) \to \mathbf{Spin}(q)(\mathbb{F}_{p^r})$$

is surjective, and we obtain the following commutative diagram:

$$\begin{tikzcd}
 & & \mathbf{Spin}(q)(\mathcal{O}_k) \arrow[r] \arrow[d, two heads, "\rho_{\mathfrak{p}}"'] & \mathbf{SO}(q)(\mathcal{O}_k) \arrow[d, "\rho_{\mathfrak{p}}"] & & \\
1 \arrow[r] & \mu_2 \arrow[r] & \mathbf{Spin}(q)(\mathbb{F}_{p^r}) \arrow[r] & \mathbf{SO}(q)(\mathbb{F}_{p^r}) \arrow[r] & \mathbb{F}_{p^r}^{\times}/(\mathbb{F}_{p^r}^{\times})^2 \arrow[r] & 1
\end{tikzcd}$$

Using this commutative diagram and noting that $|\mathbb{F}_{p^r}^{\times}/(\mathbb{F}_{p^r}^{\times})^2| = 2$, we obtain:

Proposition 3.1 *The index* $[G_{\mathfrak{p}} : \rho_{\mathfrak{p}}(\mathbf{G}(\mathcal{O}_k))]$ *is either one or two.*

This result and the above list of group orders completes our calculation of $|\rho_{\mathfrak{p}}(\mathbf{G}(\mathcal{O}_k))|$.

4 Step (b) Computing $|H_{\mathfrak{p}}|$ for a totally geodesic $\widetilde{\Lambda}$

Our goal of this section is the computations of $|H_{\mathfrak{p}}|$ for a generic totally geodesic $\widetilde{\Lambda} \subset \mathbf{G}(\mathcal{O}_k)$. We use the notation established in Section 2. Let $\mathfrak{p} \in S_1$ and $\mathcal{G}_{\mathfrak{p}} = \mathbf{G}(\mathcal{O}_{\mathfrak{p}})$ denote the parahoric of $\mathbf{G}(k_{\mathfrak{p}})$ with pro–p unipotent radical $\mathcal{G}_{\mathfrak{p}}^{+}$. It follows that $\mathcal{H}_{\mathfrak{p}} = \mathbf{H}(k_{\mathfrak{p}}) \cap \mathcal{G}_{\mathfrak{p}}$ is a parahoric of $\mathbf{H}(k_{\mathfrak{p}})$ containing $\widetilde{\Lambda}$, and $\mathcal{H}_{\mathfrak{p}}^{+} = \mathcal{G}_{\mathfrak{p}}^{+} \cap \mathcal{H}_{\mathfrak{p}}$ is the pro–p unipotent radical of $\mathcal{H}_{\mathfrak{p}}$. Set $\overline{\mathbf{H}}$ to be the $\mathbb{F}_{p^r}$–group whose $\mathbb{F}_{p^r}$–points are $\mathcal{H}_{\mathfrak{p}}/\mathcal{H}_{\mathfrak{p}}^{+}$. We have the following commutative diagram where we know the right two vertical arrows are surjections by [12, 3.4.4]:

$$\begin{tikzcd}
\widetilde{\Lambda} \arrow[r, hook] \arrow[d, "\rho_{\mathfrak{p}}"'] & \mathcal{H}_{\mathfrak{p}} \arrow[r, hook] \arrow[d, two heads] & \mathcal{G}_{\mathfrak{p}} \arrow[d, two heads] \\
H_{\mathfrak{p}} \arrow[r, hook] & \overline{\mathbf{H}}(\mathbb{F}_{p^r}) \arrow[r, hook] & \mathrm{SO}(m+1, p^r)
\end{tikzcd}$$

Hence $H_{\mathfrak{p}}$ is a subgroup of $\overline{\mathbf{H}}(\mathbb{F}_{p^r})$, which is in turn a subgroup of $\mathrm{SO}(m+1, p^r)$.

4.1 A simplification

The group $\overline{\mathbf{H}}(\mathbb{F}_{p^r})$ fits into the exact sequence

$$1 \longrightarrow \mathcal{R}_u(\overline{\mathbf{H}})(\mathbb{F}_{p^r}) \longrightarrow \overline{\mathbf{H}}(\mathbb{F}_{p^r}) \longrightarrow \overline{\mathbf{H}}^{\mathrm{red}}(\mathbb{F}_{p^r}) \longrightarrow 1, \tag{10}$$

where $\overline{\mathbf{H}}^{\mathrm{red}}$ is a reductive group whose Dynkin diagram can be read off of local Dynkin diagrams. From (10) we obtain

$$|\overline{\mathbf{H}}(\mathbb{F}_{p^r})| = |\mathcal{R}_u(\overline{\mathbf{H}})(\mathbb{F}_{p^r})| \cdot |\overline{\mathbf{H}}^{\mathrm{red}}(\mathbb{F}_{p^r})|. \tag{11}$$

Therefore, computing $|\overline{\mathbf{H}}(\mathbb{F}_{p^r})|$ reduces to computing the size of unipotent $\mathbb{F}_{p^r}$–groups and the size of $\overline{\mathbf{H}}^{\mathrm{red}}(\mathbb{F}_{p^r})$. We compute the former with the following proposition.

Proposition 4.1 *If $\mathbf{U}$ is a unipotent group over a finite field $\mathbb{F}_{p^r}$, then $|\mathbf{U}(\mathbb{F}_{p^r})| = p^s$ for some $s \in \mathbb{Z}_{\geq 0}$.*

Proof Since $\mathbb{F}_{p^r}$ is perfect, $\mathbf{U}$ splits [1, Corollary 15.5(ii)]. Therefore $\mathbf{U}$ admits a composition series

$$\mathbf{U} = \mathbf{U}_0 \supset \mathbf{U}_1 \supset \mathbf{U}_2 \supset \cdots \supset \mathbf{U}_s = \{1\}$$

of connected $\mathbb{F}_{p^r}$–groups such that $\mathbf{U}_i/\mathbf{U}_{i+1}$ is $\mathbb{F}_{p^r}$–isomorphic to $\mathbf{G}_a$. Since each $\mathbf{U}_{i+1}$ is connected, $H^1(\mathbb{F}_{p^r}, \mathbf{U}_{i+1})$ is trivial by Lang's theorem [8, Theorem 6.1], and hence

$$1 \longrightarrow \mathbf{U}_{i+1}(\mathbb{F}_{p^r}) \longrightarrow \mathbf{U}_i(\mathbb{F}_{p^r}) \longrightarrow \mathbf{G}_a(\mathbb{F}_{p^r}) \longrightarrow 1$$

is exact. We proceed by induction on the length of the composition series. If the series has length 0, then $\mathbf{U} \cong \mathbf{G}_a$, and hence $|\mathbf{U}(\mathbb{F}_{p^r})| = p^r$. If the statement is true for series of length j, then the above exact sequence implies it follows for series of length $j+1$, and the result follows. □

4.2 Computing $|\overline{\mathbf{H}}^{\mathrm{red}}(\mathbb{F}_{p^r})|$

We are now left computing the orders of $\overline{\mathbf{H}}^{\mathrm{red}}(\mathbb{F}_{p^r})$. To do so, we use the classification of local indices [12]. A $\mathfrak{p}$–adic group $\mathbf{H}$ is called *residually split* if $\mathrm{rank}_{k_\mathfrak{p}}(\mathbf{H}) = \mathrm{rank}_{k_\mathfrak{p}^{\mathrm{un}}}(\mathbf{H})$, where $k_\mathfrak{p}^{\mathrm{un}}$ is the maximal unramified extension of $k_\mathfrak{p}$. The classification of local Dynkin diagrams of simple $k_\mathfrak{p}$–groups falls into two classes: residually split and not residually split. As we explain later, we can restrict ourselves to computing these orders for totally geodesic groups of maximal dimension for both $\overline{\mathbf{H}}_r^{\mathrm{red}}(\mathbb{F}_{p^r})$ and $\overline{\mathbf{H}}_t^{\mathrm{red}}(\mathbb{F}_{p^r})$.

Proposition 4.2 *Continuing the notation of the earlier sections, suppose* $\mathbf{H}_0 = \mathbf{SO}(q_0)$ *for some quadratic subform* $q_0 \subset q$ *of odd dimension* $2n-1 \geq 4$ *and let* $\mathfrak{p} \subset S_1$. *Then* $|\overline{\mathbf{H}}_0^{\mathrm{red}}(\mathbb{F}_{p^r})|$ *divides* $p^X Y$, *where* $X \in \mathbb{Z}_{\geq 0}$ *and* Y *is one of the following:*

(T1) $\prod_{j=1}^{n-1}(p^{2rj}-1)$

(T2) $(p^{2r}-1)^2 \prod_{j=1}^{n-3}(p^{2rj}-1)$

(T3) $(p^{r(k-1)} \pm 1)\big(\prod_{j=1}^{k-2}(p^{2rj}-1)\big)\big(\prod_{j=1}^{n-k}(p^{2rj}-1)\big)$ *for* $3 \leq k \leq n-3$

(T4) $(p^{2r}-1)(p^{r(n-2)} \pm 1) \prod_{j=1}^{n-3}(p^{2rj}-1)$

(T5) $(p^{r(n-1)} \pm 1) \prod_{j=1}^{n-2}(p^{2rj}-1)$

(T6) $\prod_{j=1}^{n-2}(p^{2rj}-1)$

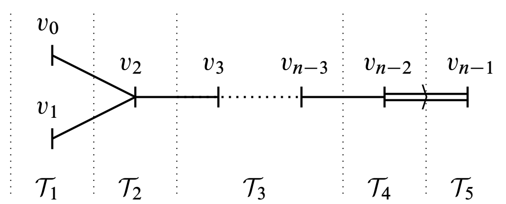
(T7) $(p^{2r}-1) \prod_{j=1}^{n-3}(p^{2rj}-1)$

(T8) $\big(\prod_{j=1}^{k-1}(p^{2rj}-1)\big)\big(\prod_{j=1}^{n-k-1}(p^{2rj}-1)\big)$ *for* $3 \leq k \leq n-3$

Proof Since every parahoric lies in a maximal one it suffices to compute the orders of all possible reductions of maximal parahorics. We analyze all possible local indices of $\mathbf{H}$ and remove one vertex to obtain the Dynkin diagram of $\overline{\mathbf{H}}^{\mathrm{red}}$ [12]. We then use the orders of Section 3, [7], and Proposition 4.1 to compute the size of each possible quotient. For each case below, we give the local diagram, where we have distinguished the nodes associated with similar reductions. We follow the diagram with a table listing the Killing–Cartan type and order of the reduction group associated with each class of node.

Case 1 $\mathbf{H}$ is residually split of type B_{n-1}:

B_{n-1}

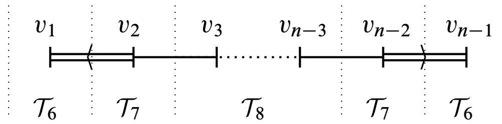


The Killing–Cartan types and orders of the reduction groups are given by:

	type of $\overline{\mathbf{H}}^{\mathrm{red}}$	order of $\overline{\mathbf{H}}^{\mathrm{red}}$
$\mathcal{T}_1$	B_{n-1}	$p^{r(n-1)^2} \prod_{j=1}^{n-1}(p^{2rj}-1)$
$\mathcal{T}_2$	$A_1 \times A_1 \times B_{n-3}$	$(p^r(p^{2r}-1))^2 \big(p^{r(n-3)^2} \prod_{j=1}^{n-3}(p^{2rj}-1)\big)$
$\mathcal{T}_3$	$D_k \times B_{n-k-1}$ $(3 \leq k \leq n-3)$	$\big(p^{rk(k-1)}(p^{rk} \pm 1) \prod_{j=1}^{k-1}(p^{2rj}-1)\big) \times \big(p^{r(n-k-1)^2} \prod_{j=1}^{n-k-1}(p^{2rj}-1)\big)$
$\mathcal{T}_4$	$D_{n-2} \times A_1$	$\big(p^{r(n-2)(n-3)}(p^{r(n-2)} \pm 1) \prod_{j=1}^{n-3}(p^{2rj}-1)\big)(p^r(p^{2r}-1))$
$\mathcal{T}_5$	D_{n-1}	$p^{r(n-1)(n-2)}(p^{r(n-1)} \pm 1) \prod_{j=1}^{n-2}(p^{2rj}-1)$

Case 2 $\mathbf{H}$ is not residually split of type B_{n-1}:

${}^{(2)}B_{n-1}$



The Killing–Cartan types and orders of the reduction groups are given by:

	type of $\overline{\mathbf{H}}^{\mathrm{red}}$	order of $\overline{\mathbf{H}}^{\mathrm{red}}$
$\mathcal{T}_6$	B_{n-2}	$p^{r(n-2)^2}\prod_{j=1}^{n-2}(p^{2rj}-1)$
$\mathcal{T}_7$	$A_1 \times B_{n-3}$	$(p^r(p^{2r}-1))\big(p^{r(n-3)^2}\prod_{j=1}^{n-3}(p^{2rj}-1)\big)$
$\mathcal{T}_8$	$B_{k-1}\times B_{n-k-1}$ $(3 \le k \le n-3)$	$\big(p^{r(k-1)^2}\prod_{j=1}^{k-1}(p^{2rj}-1)\big) \times \big(p^{r(n-k-1)^2}\prod_{j=1}^{n-k-1}(p^{2rj}-1)\big)$

This concludes the proof. □

Proposition 4.3 *Continuing the notation of the earlier sections, suppose* $\mathbf{H}_0 = \mathbf{SO}(q_0)$ *for some quadratic subform* $q_0 \subset q$ *of even dimension* $2n \ge 4$ *and let* $\mathfrak{p} \subset S_1$*. Then* $|\overline{\mathbf{H}}_0^{\mathrm{red}}(\mathbb{F}_{p^r})|$ *divides* $p^X Y$*, where* $X \in \mathbb{Z}_{\ge 0}$ *and* Y *is one of the following:*

(S1) $(p^{rn} \pm 1)\prod_{j=1}^{n-1}(p^{2rj}-1)$

(S2) $(p^{2r}-1)^2(p^{r(n-2)} \pm 1)\prod_{j=1}^{n-3}(p^{2rj}-1)$

(S3) $(p^{rk} \pm 1)(p^{r(n-k)} \pm 1)\big(\prod_{j=1}^{k-1}(p^{2rj}-1)\big)\big(\prod_{j=1}^{n-k-1}(p^{2rj}-1)\big)$ *for* $3 \le k \le n-3$

(S4) $\prod_{j=1}^{n-1}(p^{2rj}-1)$

(S5) $(p^{2r}-1)\prod_{j=1}^{n-2}(p^{2rj}-1)$

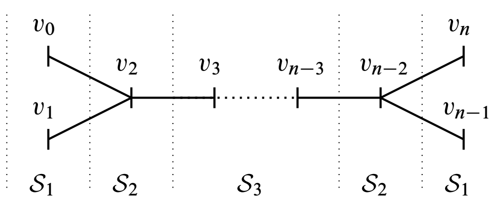
(S6) $\big(\prod_{j=1}^{k-1}(p^{2rj}-1)\big)\big(\prod_{j=1}^{n-k}(p^{2rj}-1)\big)$ *for* $3 \le k \le n-2$

or any of (T1)–(T8) *listed in the previous proposition.*

Proof The idea and presentation of this proof is the same as for Proposition 4.2.

Case 1 $\mathbf{H}$ is residually split of type D_n and in fact $\mathbf{H}$ splits over $k_{\mathfrak{p}}$:

${}^{1}D_{n,n}^{(1)}$

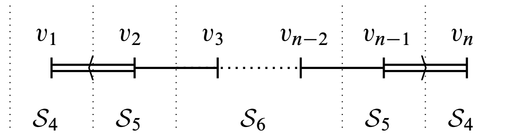


The Killing–Cartan types and orders of the reduction groups are given by:

	type of $\overline{\mathbf{H}}^{\mathrm{red}}$	order of $\overline{\mathbf{H}}^{\mathrm{red}}$
$\mathcal{S}_1$	D_n	$p^{rn(n-1)}(p^{rn} \pm 1) \prod_{j=1}^{n-1}(p^{2rj} - 1)$
$\mathcal{S}_2$	$A_1 \times A_1 \times D_{n-2}$	$(p^r(p^{2r}-1))^2 \big(p^{r(n-2)(n-3)}(p^{r(n-2)} \pm 1) \prod_{j=1}^{n-3}(p^{2rj}-1)\big)$
$\mathcal{S}_3$	$D_k \times D_{n-k}$ $(3 \le k \le n-3)$	$\big(p^{rk(k-1)}(p^{rk} \pm 1) \prod_{j=1}^{k-1}(p^{2rj}-1)\big) \times \big(p^{r(n-k)(n-k-1)}(p^{r(n-k)} \pm 1) \prod_{j=1}^{n-k-1}(p^{2rj}-1)\big)$

Case 2 $\mathbf{H}$ is residually split of type D_n and $\mathbf{H}$ is nonsplit quasisplit over both $k_{\mathfrak{p}}$ and $k_{\mathfrak{p}}^{\mathrm{un}}$:

$${}^2D^{(1)}_{n,n-1}$$

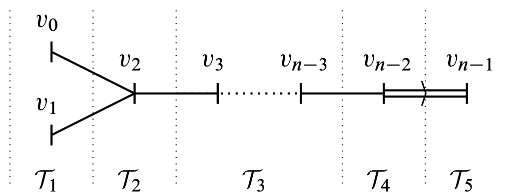


The Killing–Cartan types and orders of the reduction groups are given by:

	type of $\overline{\mathbf{H}}^{\mathrm{red}}$	order of $\overline{\mathbf{H}}^{\mathrm{red}}$
$\mathcal{S}_4$	B_{n-1}	$p^{r(n-1)^2} \prod_{j=1}^{n-1}(p^{2rj}-1)$
$\mathcal{S}_5$	$A_1 \times B_{n-2}$	$(p^r(p^{2r}-1)) \big(p^{r(n-2)^2} \prod_{j=1}^{n-2}(p^{2rj}-1)\big)$
$\mathcal{S}_6$	$B_{k-1} \times B_{n-k}$ $(3 \le k \le n-2)$	$\big(p^{r(k-1)^2} \prod_{j=1}^{k-1}(p^{2rj}-1)\big)\big(p^{r(n-k)^2} \prod_{j=1}^{n-k}(p^{2rj}-1)\big)$

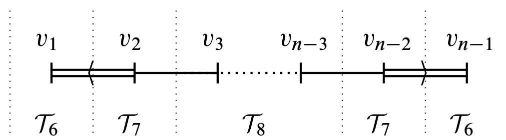
Case 3 $\mathbf{H}$ is not residually split of type D_n and $\mathbf{H}$ is nonsplit quasisplit over $k_{\mathfrak{p}}$ but splits over $k_{\mathfrak{p}}^{\mathrm{un}}$:

$${}^2D^{(1)}_{n,n-1}$$



Case 4 $\mathbf{H}$ is not residually split of type D_n and $\mathbf{H}$ is not quasisplit over $k_{\mathfrak{p}}$, but splits over $k_{\mathfrak{p}}^{\mathrm{un}}$:

$${}^1D^{(1)}_{n,n-2}$$



These last two diagrams are precisely the same as the diagrams analyzed in the previous proof, and hence the corresponding Killing–Cartan types and orders are the same. □

5 Proof of Theorem 1.1

Recall that $G_{\mathfrak{p}} = \mathbf{G}(\mathcal{O}_k/\mathfrak{p}) = \mathrm{SO}(m+1, p^r)$, and in the previous two sections, we analyzed the orders of its subgroups $\rho_{\mathfrak{p}}(\mathbf{G}(\mathcal{O}_k))$ and $H_{\mathfrak{p}}$. To prove Theorem 1.1, we need the following result of Zsigmondy:

Theorem 5.1 [15] *Let p be an odd prime and d be an integer greater than one. There exists a prime divisor of p^d+1 that does not divide p^j+1 for all $0<j<d$ and does not divide p^j-1 for all $0<j<2d$.*

Lemma 5.2 *Let (k,q) be an admissible hyperbolic pair and S_1 the set of nondyadic primes in $\mathcal{O}_k$ where q has good reduction. Then for each $\mathfrak{p} \in S_1$, there exists a subgroup $C_{\mathfrak{p}} < G_{\mathfrak{p}}$ such that $C_{\mathfrak{p}} \cap H_{\mathfrak{p}} = \{1\}$ for any $H_{\mathfrak{p}}$.*

Proof When $\dim(q) = 2n+1$, we know that $p^{nr}+1$ divides $|G_{\mathfrak{p}}|$ for any prime $\mathfrak{p} \in S_1$ by (7). For the groups $H_{\mathfrak{p}}$, we know that $|H_{\mathfrak{p}}|$ divides $p^{\alpha} \prod_j (p^j-1) \prod_{j'} (p^{j'}+1)$, where $j \leq 2r(n-1)$ and $j' \leq r(n-1)$. Consequently, $p^{nr}+1$ is not a divisor of $|H_{\mathfrak{p}}|$ for any totally geodesic subgroup. By Theorem 5.1, there exists a prime divisor $\ell_{\mathfrak{p}}$ of $p^{nr}+1$ that does not divide p^j+1 for $0<j<nr$ or $p^{2jr}-1$ for $0<j<n$. It follows that $\ell_{\mathfrak{p}}$ divides $|G_{\mathfrak{p}}|$ but not $|H_{\mathfrak{p}}|$ for any totally geodesic subgroup. By Cauchy's theorem, there exists $g \in G_{\mathfrak{p}}$ of order $\ell_{\mathfrak{p}}$ and it follows for $C_{\mathfrak{p}} = \langle g \rangle$ that $C_{\mathfrak{p}} \cap H_{\mathfrak{p}} = \{1\}$ for any totally geodesic subgroup.

When $\dim(q) = 2n$ and $\mathfrak{p} \in S_1$, we must modify the argument above. If $\det(q)$ is not a square modulo $\mathfrak{p}$, then we can proceed as above since $p^{nr}+1$ divides $|G_{\mathfrak{p}}|$. When $\det(q)$ is a square modulo $\mathfrak{p}$, we have $G_{\mathfrak{p}} = \mathrm{SO}^+(2n; p^r)$. In this case, there exists $g \in \mathrm{SO}^+(2n; p^r)$ such that g has $n/2$ eigenvalues λ_{p^r} and $n/2$ eigenvalues $\lambda_{p^r}^{-1}$, where $\lambda_{p^r} \in \mathbb{F}_{p^r}^{\times}$ is a generator for the group of units; we can take a generator for the diagonal subgroup of $(\mathrm{SO}^+(2, p^r))^n$. Taking ℓ to be an odd prime divisor of p^r-1, which exists by Theorem 5.1, and setting $a = (p^r-1)/\ell$, we assert that $C_{\mathfrak{p}} = \langle g^a \rangle$ is the desired subgroup. To see this, note that if $\gamma \in \mathbf{PSO}_0(2n-2, 1)$, then γ has an eigenvalue of ± 1 since $2n-2$ is even. As every totally geodesic m'–suborbifold with $m' \geq 2$ in a standard arithmetic orbifold is contained in a codimension-one totally geodesic suborbifold (see [4]), it follows that $\rho_{\mathfrak{p}}(\gamma)$ has ± 1 as an eigenvalue. As no nontrivial element of $C_{\mathfrak{p}}$ has this property, $C_{\mathfrak{p}} \cap H_{\mathfrak{p}} = \{1\}$. □

Proof of Theorem 1.1 for standard arithmetic orbifolds As Theorem 1.1 for $m=3$ was proven in [3], we will assume $m \geq 4$ and so $\dim(q) \geq 5$. We first prove (b). By definition, $\Gamma = \pi_1(M)$ is commensurable with $\mathbf{G}(\mathcal{O}_k)$ associated with some admissible hyperbolic pair (k,q). Strong approximation implies that $\rho_{\mathfrak{p}}(\Gamma) = \rho_{\mathfrak{p}}(\mathbf{G}(\mathcal{O}_k))$ for all but finitely many $\mathfrak{p}$, hence by Proposition 3.1 there is an infinite subset S_2 of S_1

such that $[G_{\mathfrak{p}} : \rho_{\mathfrak{p}}(\Gamma)] = 1$ or 2 for each $\mathfrak{p} \in S_2$. By Lemma 5.2, there exists a subgroup $C_{\mathfrak{p}} < G_{\mathfrak{p}}$ such that $C_{\mathfrak{p}} \cap H_{\mathfrak{p}} = \{1\}$. Since $C_{\mathfrak{p}}$ is cyclic and of odd prime order, it follows that $C_{\mathfrak{p}} < \rho_{\mathfrak{p}}(\Gamma)$. The subgroups $C_{\mathfrak{p}}, \{1\}$ satisfy (3) and so the covers $M_1, M_{C_{\mathfrak{p}}}$ corresponding to the finite-index subgroups $\ker \rho_{\mathfrak{p}}$, $\rho_{\mathfrak{p}}^{-1}(C_{\mathfrak{p}})$ are geometrically equivalent.

To produce geometrically equivalent covers with unbounded volume ratio, for each odd prime ℓ, we set S_ℓ to be the subset of primes $\mathfrak{p} \in S_2$ such that $C_{\mathfrak{p}}$ has order ℓ. We first assume that S_ℓ is infinite for some ℓ. In that case, for each $j \in \mathbb{N}$ and for any $\mathfrak{p}_1, \dots, \mathfrak{p}_j \in S_\ell$, the image of $\pi_1(M)$ under reduction modulo $\prod_i \mathfrak{p}_i$ has index 2^{s_j} in $\prod_i G_{\mathfrak{p}_i}$ for some $s_j \in \mathbb{N}$. By our choice of ℓ, the subgroup $C_{j,\ell} = \prod_i C_{\mathfrak{p}_i}$ of $\prod_i G_{\mathfrak{p}_i}$ has trivial intersection with the image of any totally geodesic subgroup, and visibly this property holds for any subgroup of $C_{j,\ell}$. Setting M_j and N_j to be the finite covers of M corresponding to the finite-index subgroups $\rho^{-1}_{\mathfrak{p}_1 \dots \mathfrak{p}_j}(1)$ and $\rho^{-1}_{\mathfrak{p}_1 \dots \mathfrak{p}_j}(C_{j,\ell})$ of Γ, respectively, we obtain a pair of geometric equivalent finite covers of M with volume ratio $\mathrm{Vol}(M_j)/\mathrm{Vol}(N_j) = \ell^j$.

We now assume that $|S_\ell|$ is finite for all odd primes ℓ. Since S_2 is infinite and each prime $\mathfrak{p} \in S_2$ is in S_ℓ for some odd prime ℓ, there must be infinitely many odd primes ℓ with $S_\ell \neq \varnothing$. Fixing an infinite sequence $\{\ell_j\}$ of distinct odd primes with $S_{\ell_j} \neq \varnothing$, for any j and any $\mathfrak{p}_j \in S_{\ell_j}$, we again have $[G_{\mathfrak{p}_j} : \rho_{\mathfrak{p}_j}(\Gamma)] = 1$ or 2. By our choice of $\mathfrak{p}_j$, we have a subgroup $C_{\mathfrak{p}_j} < G_{\mathfrak{p}_j}$ that intersects the image of every totally geodesic subgroup trivially. Setting the manifolds M_j and N_j to be the finite covers of M corresponding to the finite-index subgroups $\rho_{\mathfrak{p}_j}^{-1}(1)$ and $\rho_{\mathfrak{p}_j}^{-1}(C_{\mathfrak{p}_j})$ of Γ, respectively, we obtain geometrically equivalent finite covers with volume ratio ℓ_j.

We now prove (a). As M is compact and $\dim(q) \geq 5$, we see that $k \neq \mathbb{Q}$ by Godement's compactness criterion (see [5, Corollary 5.3.2]) and Meyer's theorem (see [5, Proposition 6.4.1]). Since $k \neq \mathbb{Q}$, by the Chebotarev density theorem there is a prime p with two overlying primes $\mathfrak{p}_1, \mathfrak{p}_2 \in S_2$ such that $\mathcal{O}_k/\mathfrak{p}_1 \cong \mathcal{O}_k/\mathfrak{p}_2$. For a pair of such primes $\mathfrak{p}_1, \mathfrak{p}_2$ we have $G_{\mathfrak{p}_1} \cong G_{\mathfrak{p}_2}$, and can apply Lemma 5.2 to both. We obtain finite-index subgroups $\rho^{-1}_{\mathfrak{p}_1\mathfrak{p}_2}(C_{\mathfrak{p}_1} \times \{1\})$ and $\rho^{-1}_{\mathfrak{p}_1\mathfrak{p}_2}(\{1\} \times C_{\mathfrak{p}_2})$ of Γ. The associated finite covers $M_{\ell,1}$ and $M_{1,\ell}$ of M have the same geometric spectra. To see that $\mathcal{TG}(M_{1,\ell}) = \mathcal{TG}(M_{\ell,1})$, we first note that the finite cover $M_{\ell,\ell}$ associated with the finite-index subgroup $\rho^{-1}_{\mathfrak{p}_1\mathfrak{p}_2}(C_{\mathfrak{p}_1} \times C_{\mathfrak{p}_2})$ in $\pi_1(M)$ is geometrically equivalent to both $M_{\ell,1}$ and $M_{1,\ell}$ and so $\mathrm{TG}(M_{\ell,1}) = \mathrm{TG}(M_{1,\ell})$. To see that the multiplicities are equal simply note that both manifolds are cyclic covers of $M_{\ell,\ell}$ of degree ℓ and thus separately satisfy (6) with $M_{\ell,\ell}$. That the manifolds are nonisometric follows from an argument similar to one used in [3, page 179]. Briefly, each element $\gamma \in \pi_1(M_{1,\ell})$ is trivial under reduction modulo $\mathfrak{p}_1$ while there are infinitely many elements in $\pi_1(M_{\ell,1})$ with image that generates $C_{\mathfrak{p}_1}$. Consequently, these elements in

$\pi_1(M_{\ell,1})$ with order ℓ image under modulo $\mathfrak{p}_1$ cannot be conjugate to any element in $\pi_1(M_{1,\ell})$ in $\mathrm{Isom}(\mathbb{H}^m)$. However, if $M_{1,\ell}$ and $M_{\ell,1}$ are isometric, by Mostow rigidity, $\pi_1(M_{1,\ell})$ and $\pi_1(M_{\ell,1})$ are conjugate in $\mathrm{Isom}(\mathbb{H}^m)$, and so $M_{1,\ell}$ and $M_{\ell,1}$ are nonisometric. □

The proof for a nonstandard arithmetic hyperbolic orbifold $M = \Gamma\backslash\mathbb{H}^m$ is similar. As in the standard arithmetic setting, there is an associated number field k and an algebraic k–group $\mathbf{G}$ for which Γ is commensurable with the group $\mathbf{G}(\mathcal{O}_k)$. There is an infinite set of primes S_0' of $\mathcal{O}_k$ such that for each $\mathfrak{p} \in S_0'$, the local group $\mathbf{G}(k_{\mathfrak{p}})$ is isomorphic to $\mathbf{SO}(V_{\mathfrak{p}}, q_{\mathfrak{p}})$, where $(V_{\mathfrak{p}}, q_{\mathfrak{p}})$ is a quadratic space over $k_{\mathfrak{p}}$. Restricting to primes in S_0', the proof then follows as in the standard arithmetic case. For (a), we note that when M is a closed arithmetic hyperbolic m–orbifold with $m \geq 4$, the field of definition of M is not $\mathbb{Q}$ (see [5, Section 6.4]).

This method can be implemented for any finite-volume, complete, hyperbolic m–orbifold when $m \geq 4$.

Theorem 5.3 *If M is a complete, orientable, finite-volume hyperbolic m–orbifold with $m \geq 4$, then the following holds:*

(a) *If the field of definition of M is not $\mathbb{Q}$, then there exist finite, nonisometric covers M' and N' that are geometrically isospectral.*

(b) *There exists a sequence (M_j, N_j) of pairs of nonisometric finite covers of M such that M_j and N_j are geometrically equivalent and $\mathrm{Vol}(M_j)/\mathrm{Vol}(N_j)$ is unbounded as a function of j.*

Proof Given M with $\Gamma = \pi_1(M)$, there exists an injective homomorphism $\rho\colon \Gamma \to \mathbf{PSO}_0(m,1)$ such that the field generated by the matrix coefficients is a number field k (see [13] or [2, Section 4.1]); this field is the so-called field of definition. If R is the $\mathcal{O}_k$–submodule of k generated by the entries of $\rho(\Gamma)$, there is a cofinite subset of the set of prime ideals $\mathcal{P}$ of $\mathcal{O}_k$ such that $R/\mathfrak{P} \cong \mathcal{O}_k/\mathfrak{p} = \mathbb{F}_{p^r}$ for each $\mathfrak{p} \in \mathcal{P}$, where $\mathfrak{P} = R\mathfrak{p}$. Since $\rho(\Gamma) < \mathbf{PSO}_0(m,1)$ is Zariski dense, we can apply Nori–Weisfeiler strong approximation [6; 14]. When $m+1$ is odd (resp. even), there exists an infinite set of nondyadic primes $S_2 \subset \mathcal{P}$ such that the image of $\rho_{\mathfrak{p}}(\Gamma)$ contains the commutator subgroup $\Omega(m+1; p^r)$ (resp. $\Omega^{\pm}(m+1; p^r)$) of $\mathrm{SO}(m+1; p^r)$ (resp. $\mathrm{SO}^{\pm}(m+1; p^r)$) for each $\mathfrak{P} \in S_2$ (see [2, Theorem 5.3]). The argument now follows as in the previous case of standard arithmetic hyperbolic m–orbifolds. □

Remark Our use of Zsigmondy's theorem was inspired by [2], where Long and Reid proved that any lattice $\Gamma < \mathbf{SO}(n,1)$ contains hyperbolic elements with infinite-order holonomy. In [3], the use of Zsigmondy's theorem was replaced by a direct argument. Prasad and Rapinchuk [9] have general results on the existence of semisimple elements whose Zariski closure is dense in a maximal torus. It is possible to replace our elementary counting argument with an argument based on [9], though one must still determine the possible images of subgroups associated with totally geodesic submanifolds as in Section 4.

References

[1] **A Borel**, *Linear algebraic groups*, 2nd edition, Graduate Texts in Mathematics 126, Springer, New York (1991) MR

[2] **D D Long**, **A W Reid**, *Eigenvalues of hyperbolic elements in Kleinian groups*, from "In the tradition of Ahlfors–Bers, V" (M Bonk, J Gilman, H Masur, Y Minsky, M Wolf, editors), Contemp. Math. 510, Amer. Math. Soc., Providence, RI (2010) 197–208 MR

[3] **D B McReynolds**, **A W Reid**, *The genus spectrum of a hyperbolic 3–manifold*, Math. Res. Lett. 21 (2014) 169–185 MR

[4] **J S Meyer**, *Totally geodesic spectra of arithmetic hyperbolic spaces*, preprint (2015) arXiv To appear in Trans. Amer. Math. Soc.

[5] **D W Morris**, *Introduction to arithmetic groups*, Deductive Press (2015) MR

[6] **M V Nori**, *On subgroups of $\mathrm{GL}_n(\mathbf{F}_p)$*, Invent. Math. 88 (1987) 257–275 MR

[7] **T Ono**, *On algebraic groups and discontinuous groups*, Nagoya Math. J. 27 (1966) 279–322 MR

[8] **V Platonov**, **A Rapinchuk**, *Algebraic groups and number theory*, Pure and Applied Mathematics 139, Academic Press, Boston (1994) MR

[9] **G Prasad**, **A S Rapinchuk**, *Existence of irreducible $\mathbb{R}$–regular elements in Zariski-dense subgroups*, Math. Res. Lett. 10 (2003) 21–32 MR

[10] **A S Rapinchuk**, *Strong approximation for algebraic groups*, from "Thin groups and superstrong approximation" (E Breuillard, H Oh, editors), Math. Sci. Res. Inst. Publ. 61, Cambridge Univ. Press (2014) 269–298 MR

[11] **T Sunada**, *Riemannian coverings and isospectral manifolds*, Ann. of Math. 121 (1985) 169–186 MR

[12] **J Tits**, *Reductive groups over local fields*, from "Automorphic forms, representations and L–functions, Part 1" (A Borel, W Casselman, editors), Proc. Sympos. Pure Math. 33.1, Amer. Math. Soc., Providence, RI (1979) 29–69 MR

[13] **E B Vinberg**, *Rings of definition of dense subgroups of semisimple linear groups*, Izv. Akad. Nauk SSSR Ser. Mat. 35 (1971) 45–55 MR In Russian; translated in Math. USSR-Izv. 5 (1971) 45–55

[14] **B Weisfeiler**, *Strong approximation for Zariski-dense subgroups of semisimple algebraic groups*, Ann. of Math. 120 (1984) 271–315 MR

[15] **K Zsigmondy**, *Zur Theorie der Potenzreste*, Monatsh. Math. Phys. 3 (1892) 265–284 MR

Department of Mathematics, Purdue University
150 N. University St., West Lafayette, IN 47907, United States

Department of Mathematics, California State University, San Bernardino
5500 University Parkway, San Bernardino, CA 92407-2318, United States

Department of Mathematics, Temple University
1805 N Broad St., Philadelphia, PA 19122, United States

dmcreyno@purdue.edu, jeffrey.meyer@csusb.edu, mstover@temple.edu

Received: 24 July 2015 Revised: 28 August 2016

Algebraic & Geometric Topology 17 (2017) 847–867

On phantom maps into co-H–spaces

JAMES SCHWASS

We study the existence of essential phantom maps into co-H–spaces, motivated by Iriye's observation that every suspension space Y of finite type with $H_i(Y;\mathbb{Q}) \neq 0$ for some $i > 1$ is the target of essential phantom maps. We show that Iriye's observation can be extended to the collection of nilpotent, finite-type co-H–spaces. This work hinges on an enhanced understanding of the connections between homotopy decompositions of looped co-H–spaces and coalgebra decompositions of tensor algebras due to Grbić, Theriault and Wu.

55P45, 55S37

1 Introduction

We will work in the category **Top** of spaces having the homotopy type of a pointed CW complex and pointed maps between them. We will restrict our attention throughout to simply connected spaces, or their loop spaces. A map $X \to Y$ is called a *phantom map* if for every n the composite

$$X_n \to X \to Y$$

is nullhomotopic, where $X_n \to X$ is an n–skeleton for some CW structure of X. We offer an alternative characterization of this concept to illustrate that the choice of a CW structure X is insignificant; according to Bousfield and Kan [4], $X \to Y$ is phantom if and only if $X \to Y \to Y^{(n)}$ is nullhomotopic for every n, where $Y^{(n)}$ denotes the n^{th} Postnikov approximation of Y.

From the definition and characterization given above, it is clear that a phantom map must induce the zero map on homotopy groups, and on any homology theory, and so these maps appear trivial upon passage to such common algebraic models for topological spaces. On the other hand, phantom maps can be of genuine topological interest. The theory of phantom maps has been used by Harper and Roitberg [12] and Gray [9], among many others, to produce and study examples of distinct homotopy classes of spaces X and Y which have the same n–type, ie $X^{(n)} \simeq Y^{(n)}$ for all n. Roitberg [20] has also used the theory of phantom maps to compute the homotopy automorphism groups of particular spaces; in general the computation of homotopy automorphism

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.847

groups is intractable. These examples serve to illustrate that phantom maps play a significant role in **Top**. But, since these maps vanish under many of our favorite functors, they prove difficult to study, or even to locate. The purpose of this work is to locate new examples of phantom maps; the analysis of particular invariants of these phantom maps and the structure of the collection of phantom maps will take place elsewhere.

The constant map is an obvious example of a phantom map. Of more interest are essential (ie homotopically nontrivial) phantom maps, which abound in **Top**. We offer, as evidence of this fact, the following theorems of Iriye, and McGibbon and Møller. We will say a space X is of finite type (over $\mathbb{Z}$) if each $H_n(X;\mathbb{Z})$ and $\pi_n(X)$ is a finitely generated group. We write $\mathrm{Ph}(X,Y)$ for the subset of $[X,Y]$ consisting of homotopy classes of phantom maps.

Theorem 1.1 [13] *Suppose $Y \simeq \Sigma X$ is a nilpotent suspension space of finite type. If $H_i(Y;\mathbb{Q}) \neq 0$ for some $i > 1$ then Y is the target of essential phantom maps from finite-type domains.*

Theorem 1.2 [17] *If X and Y are of finite type and $\mathrm{Ph}(X,Y)$ is not the one point set, then $\mathrm{Ph}(X,Y)$ is uncountably large.*

In many senses, the concept of a co-H–space is a mild generalization of that of a suspension space. As such, many statements that hold true for the collection of suspension spaces are also true for the collection of co-H–spaces. We wondered if one could replace the suspension space Y in Theorem 1.1 with any nilpotent co-H–space of finite type. Our main result is a positive answer to this question.

Theorem 1.3 *Suppose Y is a nilpotent co-H–space with $H_i(Y;\mathbb{Q}) \neq 0$ for some $i > 1$. Then Y is the target of essential phantom maps from finite-type domains.*

The proof of Theorem 1.3 is comprised of several pieces. For a co-H–space whose rational homology is "large" we develop decomposition methods in phantom map theory and appeal to recently developed highly structured decompositions of the loop space of a co-H–space due to Selick, Grbič, Theriault and Wu. For a co-H–space with "small" rational homology we exploit strong connections between phantom map theory and rational homotopy theory discovered by McGibbon and Roitberg.

Through the theory of Lusternik–Schnirelmann category, this work can be viewed as providing a solution to the case $n = 1$ of the following question. Our exposition of Lusternik–Schnirelmann category here will be limited to the following three observations: $\mathrm{cat}(X)$ is a nonnegative integer, assigned to a space X, which we think of as a measure of the complexity of X; $\mathrm{cat}(X)=0$ if and only if X is contractible; the spaces of Lusternik–Schnirelmann category one are precisely the noncontractible co-H–spaces.

Question 1.4 Suppose Y has finite type, and $\mathrm{cat}(Y)=n<\infty$. If $H_i(Y;\mathbb{Q})\neq 0$ for some $i>1$, is Y the target of essential phantom maps from finite-type domains?

In Section 2.1 we lay out the preliminaries on phantom map theory. In Section 2.2 we describe recently developed connections between coalgebra decompositions of tensor algebras and homotopy decompositions of looped co-H–spaces. In Section 3 we develop techniques to bridge the gap between the decompositions of Section 2.2 and the theory of phantom maps. Section 4 contains the proof of Theorem 1.3. Examples and applications are given in Section 5.

Acknowledgements We would like to thank Jeff Strom, the author's dissertation advisor, for many helpful conversations regarding the content and preparation of this paper. This work grew out of the author's doctoral dissertation at Western Michigan University. We are indebted to Kouyemon Iriye for supplying a critical insight into the proof of Proposition 4.5, which helped this work reach its maturity. We would also like to thank the referee for suggesting the inclusion of additional examples.

2 Preliminaries

Localization will play a central role in what is to follow. We assume familiarity with the rudiments of localization; a detailed reference is [15]. Since a rationally nontrivial p–local space is not of finite type over $\mathbb{Z}$, we will have a need for a p–local analog of the notion of a finite-type space; a space X is of *finite type over* $\mathbb{Z}_{(p)}$ if each $H_n(X;\mathbb{Z})$ and $\pi_n(X)$ is a finitely generated $\mathbb{Z}_{(p)}$–module. We should note that a space of finite type over $\mathbb{Z}_{(p)}$ is necessarily p–local. Though we will be primarily interested in phantom maps between finite-type spaces, we will have occasion to examine phantom maps from finite-type domains into targets having finite type over $\mathbb{Z}_{(p)}$.

2.1 Background on phantom maps

In Section 2.1.1 we describe a critical identification of $\mathrm{Ph}(X,Y)$ with a particular functor which factors through the category of towers of groups. In Section 2.1.2 we describe connections between phantom map theory and rational homotopy theory that are indispensable in discovering new examples of phantom maps from old, among other things. Most of the material in this section can be found in the wonderful survey article [16] of McGibbon.

2.1.1 The tower perspective By a tower $\{G_n\}$ of groups we mean a diagram

$$\cdots \xrightarrow{p_{n+1}} G_n \xrightarrow{p_n} \cdots \xrightarrow{p_3} G_2 \xrightarrow{p_2} G_1 \tag{1}$$

in the category of groups. We mean something similar by a tower of abelian groups, or a tower of sets, or really a tower of any sort of gadget — these are $\mathbb{N}^{\mathrm{op}}$–shaped diagrams in various categories. A morphism of towers is a natural transformation of $\mathbb{N}^{\mathrm{op}}$ shaped diagrams. By $\lim G_n$ we mean the limit of the diagram (1) in the appropriate category.

We now set about describing the functor $\lim^1$. On the category of towers of abelian groups, by $\lim^1$ we mean the first derived functor of lim; more concretely, if $\{G_n\}$ is a tower of abelian groups, then $\lim G_n$ is the kernel and $\lim^1 G_n$ is the cokernel of the map

$$\prod G_n \xrightarrow{\mathrm{id}-(p_n)} \prod G_n$$

given by

$$(a_1, a_2, \dots) \mapsto (a_1 - p_2(a_2), a_2 - p_3(a_3), \dots).$$

Bousfield and Kan [4, pages 254–255] extend the definition of $\lim^1$ to the category of towers of arbitrary groups as follows: Given a tower $\{G_n\}$ of groups let $\prod G_n$ act on $\prod G_n$ by

$$(g_n) \cdot (x_n) = (g_n x_n (p_{n+1}(g_{n+1})^{-1})),$$

where $G_{n+1} \xrightarrow{p_{n+1}} G_n$ is the structure map in the tower $\{G_n\}$. Then $\lim^1 G_n$ is the orbit space of this action. This is important to us because we will have occasion to refer to $\lim^1 G_n$ where $\{G_n\}$ is a tower of not necessarily abelian groups.

In particular, if X and Y have the homotopy type of CW complexes, then a CW structure for X gives rise to a tower $\{[\Sigma X_n, Y]\}$ of (generally nonabelian) groups; dually the Postnikov tower for Y gives rise to a tower $\{[X, \Omega Y^{(n)}]\}$ of (generally nonabelian) groups. We now arrive at a fundamental identification in phantom map theory.

Corollary 2.1 [4] *For spaces X and Y there are bijections of pointed sets*

$$\lim{}^1[\Sigma X_n, Y] \cong \mathrm{Ph}(X, Y) \cong \lim{}^1[X, \Omega Y^{(n)}].$$

The identification made in Corollary 2.1 allows for the introduction of algebraic methods for characterizing the condition $\mathrm{Ph}(X, Y) = *$. Given a tower of gadgets (groups, sets, etc) $\{G_n\}$ let $G_k^{(n)}$ be the image in G_k of the composite of the structure maps

$$G_n \to G_{n-1} \to \cdots \to G_k$$

when $n \geq k$ and for $n < k$ set $G_k^{(n)} = 1$. This defines, for each $k \geq 1$ a subtower $\{G_k^{(n)}\}$, indexed by n, of the tower $\{G_n\}$. Notice that for fixed k the sequence of images $G_k^{(n)}$

are nested; we say the tower $\{G_n\}$ *satisfies the Mittag-Leffler condition* if all of the nested sequences $G_k^{(n)}$ satisfy a descending chain condition: explicitly, for each k there is some N such that for all $n \geq N$ one has $G_k^{(n)} = G_k^{(N)}$.

It is well known that if a tower $\{G_n\}$ satisfies the Mittag-Leffler condition, then $\lim^1 G_n = *$. When the tower $\{G_n\}$ is comprised of countable groups, the converse of this statement is also true:

Theorem 2.2 [17] *Suppose G_n is a tower of countable groups. Then $\lim^1 G_n = *$ if and only if the tower G_n satisfies the Mittag-Leffler condition. Moreover, if $\lim^1 G_n \neq *$, then $\lim^1 G_n$ is uncountably large.*

It is worthwhile to note that, when X and Y are of finite type over $\mathbb{Z}$ or $\mathbb{Z}_{(p)}$ for some prime p, for each n the groups

$$[\Sigma X_n, Y] \quad \text{and} \quad [X, \Omega Y^{(n)}]$$

are countable. Theorem 2.2 will be used to develop decomposition methods in phantom map theory in Section 3.

2.1.2 Phantom maps and rational equivalences McGibbon and Roitberg have characterized the finite-type spaces that are not the targets of essential phantom maps from finite-type domains in terms of the existence of particular rational equivalences.

Theorem 2.3 [18] *For a nilpotent, finite-type space Y, the following are equivalent:*

(i) $\mathrm{Ph}(X, Y) = *$ *for all finite-type domains X.*
(ii) $\mathrm{Ph}(K(\mathbb{Z}, m), Y) = *$ *for all m.*
(iii) *There is a rational equivalence $\prod_\alpha K(\mathbb{Z}, m_\alpha) \to \Omega Y$.*

We should note that the direction of the rational equivalence in Theorem 2.3(iii) is significant; for any space Y there is a rational equivalence $\Omega Y \to \prod K(\mathbb{Z}, m_\beta)$.

We will need a p–local version of the implication (i) $\Longrightarrow$ (iii) of Theorem 2.3, which we record as Proposition 2.4. This will be used to establish a lemma in Section 3 required to develop decomposition methods in phantom map theory.

We have previously observed that if X and Y are of finite type over $\mathbb{Z}$ or $\mathbb{Z}_{(p)}$, then the groups

$$[\Sigma X_n, Y] \quad \text{and} \quad [X, \Omega Y^{(n)}]$$

are countable for all n. As such, Theorem 2.2 can be used to characterize the condition $\mathrm{Ph}(X, Y) = *$ in terms of the Mittag-Leffler condition. This is the main point required to complete the construction of the rational equivalence $\prod K(\mathbb{Z}, m_\beta) \to \Omega Y$ as given by McGibbon and Roitberg, given the hypothesis $\mathrm{Ph}(X, Y) = *$ for all finite-type domains X, and so we arrive at the following partial refinement of Theorem 2.3.

Proposition 2.4 *Suppose Y is nilpotent and has finite type over $\mathbb{Z}_{(p)}$. If $\mathrm{Ph}(X, Y) = *$ for all finite-type domains X, then there is a rational equivalence*

$$\prod K(\mathbb{Z}, m_\beta) \to \Omega Y.$$

The converse of this statement could feasibly hold, but we have not yet had occasion to check this. Indeed, if conjugacy classes in $[X, \Omega Y^{(n)}]$ are of finite cardinality for every n, then the converse of Proposition 2.4 can be established using the proof of Theorem 2.3 given by McGibbon and Roitberg [18].

Theorem 2.3 only begins to hint at the connections between phantom map theory and rational homotopy theory. The next result is another glimpse of these strong connections. We should note that the result stated here is slightly stronger than in [18], though the authors' argument establishes the result in light of the observation that $[X, \Omega Y^{(n)}]$ is a countable group when X and Y are of finite type over $\mathbb{Z}$ or $\mathbb{Z}_{(p)}$. Before stating the result, we remark that $\mathrm{Ph}(X, Y)$ is a contravariant functor in X and a covariant functor in Y.

Theorem 2.5 [18] *Suppose Y and Y' are of finite type over $\mathbb{Z}$ or $\mathbb{Z}_{(p)}$. If $Y \to Y'$ induces a surjection on $\pi_* \otimes \mathbb{Q}$, then for every finite-type domain X the induced map*

$$\mathrm{Ph}(X, Y) \to \mathrm{Ph}(X, Y')$$

is surjective.

Note that for each prime p and each nilpotent space Y the p–localization $Y \to Y_{(p)}$ is a rational equivalence, hence induces surjections on $\pi_* \otimes \mathbb{Q}$, and so we arrive at a corollary which has been well-known in the phantom map literature, and will be one of our primary tools for detecting essential phantom maps.

Corollary 2.6 *Suppose Y is a nilpotent, finite-type space. If $Y_{(p)}$ is the target of essential phantom maps from finite-type domains, then so is Y.*

2.2 Homotopy decompositions of looped co-H–spaces

Our jumping off point is the generalized Bott-Samelson theorem, due to Berstein.

Bott–Samelson theorem [2] *If Y is a simply connected co-H–space, then there is a natural algebra isomorphism*

$$H_*(\Omega Y) \cong T(\Sigma^{-1}\widetilde{H}_*(Y)),$$

where $H_*(\Omega Y)$ is equipped with the Pontryagin product. (Here homology has coefficients in a PID k and $\widetilde{H}_*(Y)$ is a free k–module.)

For the rest of this section we fix a prime p; the ground ring for all algebraic objects will be $\mathbb{F}_p$, the field with p elements. All homology in this section has $\mathbb{F}_p$ coefficients. Many of the results of this section remain true if we replace $\mathbb{F}_p$ with an arbitrary field, though we will have no need for such generality. We write T for the free graded tensor algebra functor taking the category of vector spaces to the category of graded algebras.

In the 1980s, F Cohen, Moore and Neisendorfer developed a technique fueled by the Bott–Samelson theorem which they use to determine the homotopy exponents of odd-dimensional spheres; the difficulty of drawing concrete conclusions regarding homotopy groups of spheres is well documented, and illustrates the power of this technique. We now loosely outline one component of this program. Cohen, Moore and Neisendorfer sought out algebraic decompositions of $T(\Sigma^{-1}\widetilde{H}_*(Y))$, and showed that these algebraic decompositions have geometric realizations in the form of homotopy decompositions of ΩY for $Y = S^{2n+1}$, among a few other specific spaces.

In [22], Selick and Wu begin developing functorial analogs of the ad hoc decomposition methods of Cohen, Moore and Neisendorfer, apparently motivated by the power of these methods, along with a conjecture of Cohen. The functorial decomposition methods reach maturity in [11], after contributions by Grbić, Theriault, Selick and Wu spanning the course of about a decade. Before describing these functorial analogs, we lay out some nomenclature and conventions.

Of course as vector spaces $T(V) = \bigoplus_{n\geq 0} V^{\otimes n}$, where $V^{\otimes 0} = \mathbb{F}_p$. This identifies V as a submodule of $T(V)$. The algebra $T(V)$ is equipped with a unit $\mathbb{F}_p \to T(V)$ and augmentation $T(V) \to \mathbb{F}_p$ defined by inclusion of and projection onto $\mathbb{F}_p = V^{\otimes 0}$, respectively. The tensor algebra $T(V)$ is naturally endowed with the structure of a Hopf algebra by declaring the elements of V to be primitive. More explicitly, since $T(V)$ is the free algebra on V, the linear map $V \to T(V) \otimes T(V)$ given by $v \mapsto 1 \otimes v + v \otimes 1$ extends uniquely to a map of algebras $\Delta\colon T(V) \to T(V) \otimes T(V)$, giving a comultiplication on $T(V)$. One can check that the unit and augmentation are morphisms of coalgebras and algebras, respectively, and so we have given $T(V)$ the structure of a Hopf algebra. This discussion serves to illustrate that we can think of the tensor algebra functor T as taking its values in the categories of algebras, coalgebras or Hopf algebras. We will specify which category we mean to take for the target of the functor T if there is potential for confusion.

A *natural coalgebra retract* of T is a functor A from vector spaces to coalgebras equipped with natural transformations $A \xrightarrow{I} T$ and $T \xrightarrow{R} A$ such that RI is the identity natural transformation on A. A *natural coalgebra decomposition* of T is a pair of functors A, B from vector spaces to coalgebras equipped with natural coalgebra isomorphisms $T \cong A \otimes B$. Since $\otimes$ is the categorical product in the category of coalgebras, which happens to be a pointed category, it follows that if $T \cong A \otimes B$ is a natural coalgebra decomposition, then both A and B are natural coalgebra retracts of T. A *natural sub-Hopf algebra* of T is a subfunctor B from vector spaces to Hopf algebras. A natural sub-Hopf algebra B of T is *coalgebra split* if B is a natural coalgebra retract of T when regarded as a functor into the category of coalgebras.

We will write $\mathbf{CoH}_{(p)}$ for the category of p–local co-H–spaces and co-H–maps between them. A *natural homotopy retract* of $\Omega\colon \mathbf{CoH}_{(p)} \to \mathbf{Top}$ is a functor $\bar{A}\colon \mathbf{CoH}_{(p)} \to \mathbf{Top}$ equipped with natural transformations $\bar{A} \xrightarrow{I} \Omega$ and $\Omega \xrightarrow{R} \bar{A}$ such that RI is naturally homotopic to the identity natural transformation on $\bar{A}$. Such a functor $\bar{A}$ is a *geometric realization over* $\mathbf{CoH}_{(p)}$ of a natural coalgebra retract A of T if there is a natural isomorphism of functors from $\mathbf{Top}$ to the category of coalgebras

$$H_* \circ \bar{A} \cong A \circ \Sigma^{-1} \tilde{H}_*.$$

A *natural homotopy decomposition* of $\Omega\colon \mathbf{CoH}_{(p)} \to \mathbf{Top}$ is a pair of functors $\bar{A}$ and $\bar{B}$ from $\mathbf{CoH}_{(p)} \to \mathbf{Top}$ equipped with natural homotopy equivalences $\Omega \simeq \bar{A} \times \bar{B}$. A natural homotopy decomposition $\Omega \simeq \bar{A} \times \bar{B}$ is a *geometric realization over* $\mathbf{CoH}_{(p)}$ of the natural coalgebra decomposition $T \cong A \otimes B$ if $\bar{A}$ and $\bar{B}$ are geometric realizations of A and B, respectively.

We are now equipped to describe the functorial analogs of the decomposition methods of Cohen, Moore and Neisendorfer. These results give a wonderful algebraic source of homotopy decompositions of looped co-H–spaces.

Theorem 2.7 [21] *Every natural coalgebra retract of T has a geometric realization over $\mathbf{CoH}_{(p)}$.*

Corollary 2.8 [21] *Every natural coalgebra decomposition of T has a geometric realization over $\mathbf{CoH}_{(p)}$.*

We will be interested in a particular natural coalgebra decomposition of the tensor algebra functor known as the minimal decomposition, which we now set about describing. Beginning with Cohen, there was an interest in studying the minimal functorial coalgebra retract $A^{\min}$ of T for which $V \subseteq A^{\min}(V)$ for every vector space V; we should note that constructions of $A^{\min}$ are theoretical, and concrete information regarding this functor can be difficult to come by [23]. Cohen conjectured that the primitives of $T(V)$, considered as a Hopf algebra, having tensor length not a power of p must lie in the coalgebra complement of $A^{\min}(V)$ in $T(V)$. This was confirmed by Selick and Wu, who discovered the minimal decomposition and began studying its structural properties in [22].

Theorem 2.9 [22] *There is a natural coalgebra-split sub-Hopf algebra $B^{\max}$ of T and a natural coalgebra decomposition*

$$T \cong A^{\min} \otimes B^{\max}. \tag{2}$$

Moreover, $L_n(V) \subseteq B^{\max}(V)$ if n is not a power of p. Here $L_n(V)$ denotes the submodule of homogeneous Lie elements of tensor length n in $T(V)$. The natural coalgebra decomposition (2) *is known as the* minimal decomposition*.*

By Corollary 2.8, the minimal decomposition has a geometric realization as $\Omega \simeq \bar{A}^{\min} \times \bar{B}^{\max}$ over $\mathbf{CoH}_{(p)}$. We can find more structure in this homotopy decomposition of Ω by making use of the observation that $B^{\max}$ is a natural sub-Hopf algebra of T. For a Hopf algebra M, write IM for the augmentation ideal of M, and write $QM = IM/(IM)^2$ for the module of indecomposables of M. Suppose B is any natural coalgebra-split sub-Hopf algebra B of T. Since $B(V)$ is a sub-Hopf algebra of $T(V)$ for each vector space V, it follows that $B(V)$ is also a tensor algebra. That is, there is a natural isomorphism of algebras

$$B(V) \cong T\Big(\bigoplus_{n\geq 1} Q_n B(V)\Big),$$

where $Q_n B(V)$ is the image of the submodule

$$B_n(V) = IB(V) \cap V^{\otimes n} \subseteq T(V)$$

of $B(V)$ consisting of elements of tensor length n in $T(V)$ lying in the augmentation ideal of $B(V)$ under the natural map $B(V) \to QB(V)$. The construction of each $Q_n B(V)$ is natural, so we obtain natural isomorphisms

$$B \cong T \circ \bigoplus_{n\geq 1} Q_n B.$$

Ideally one can geometrically realize this additional structure as well; this is the content of the following theorem of Grbiċ, Theriault and Wu:

Theorem 2.10 [11] *Suppose B is a natural coalgebra-split sub-Hopf algebra of T. There exist functors $\bar{Q}_n B\colon \mathbf{CoH}_{(p)} \to \mathbf{Top}$ with*

(1) $\Sigma^{-1}\widetilde{H}_*(\bar{Q}_n B(Y)) \cong Q_n B(\Sigma^{-1}\widetilde{H}_*(Y))$,

(2) $\bar{Q}_n B(Y)$ *is naturally a retract of an* $(n-1)$*–fold desuspension of* $Y^{\wedge n}$, *the* n^{th} *smash power of* Y,

(3) $\bar{B}(Y) \simeq \Omega\big(\bigvee_{n\geq 1} \bar{Q}_n B(Y)\big)$.

The statement (2) requires some justification. Theriault [24] has shown that if X and Y are coassociative co-H–spaces then $X \wedge Y \simeq \Sigma Z$ for some co-H–space Z. Gray [10] showed that the coassociativity requirement could be relaxed — we need only require that one of the factors in the smash product be simply connected or a suspension space. Inductively, it follows that an n–fold smash product of simply connected co-H–spaces is an $(n-1)$–fold suspension of a co-H–space; symbolically, for simply connected co-H–spaces X_i, $i = 1, \ldots, n$,

$$\bigwedge_{i=1}^{n} X_i \simeq \Sigma^{n-1} Z \tag{3}$$

for some co-H–space Z. Of course there may be many choices for the space Z. For example, the well-known decomposition

$$\Sigma(X \times Y) \simeq \Sigma X \vee \Sigma Y \vee \Sigma(X \wedge Y)$$

and the failure of the identity

$$X \times Y \simeq X \vee Y \vee (X \wedge Y)$$

witnesses the failure of a cancellation property for Σ. This ambiguity need not worry us, since we will only have a need to describe the homology of a space Z fitting in $\Sigma^{n-1} Z \simeq Y^{\wedge n}$. That the space Z can be chosen to admit a co-H–structure also illustrates that $\bar{Q}_n B(Y)$ can be endowed with the structure of a co-H–space, which will be of importance in the proof of Theorem 1.3.

3 Decomposition methods in phantom map theory

In this section we develop tools which will be used to bridge the gap between the decompositions of Section 2.2 and phantom map theory. The loop- and wedge-splitting theorems (and their duals) have many applications outside our present scope, due to the existence of a vast library of decompositions in the literature to which these theorems can be applied. To substantiate this claim, we provide an application of the loop-splitting theorem to special cases of Question 1.4 in Example 5.3.

Theorem 3.1 (loop-splitting theorem) *Suppose Y has finite type over $\mathbb{Z}$ or $\mathbb{Z}_{(p)}$ for some prime p, and $\Omega Y \simeq A \times \Omega B$. If B is the target of essential phantom maps from finite-type domains, then so is Y.*

Proof Take X to be an arbitrary finite-type domain and write

$$G_n = [X, \Omega Y^{(n)}] \quad \text{and} \quad H_n = [X, \Omega B^{(n)}].$$

We make use of the identification

$$\mathrm{Ph}(X, Y) \cong \lim{}^1 G_n \quad \text{and} \quad \mathrm{Ph}(X, B) \cong \lim{}^1 H_n.$$

By Theorem 2.2 if $\mathrm{Ph}(X, Y) = *$ then $\{G_n\}$ is Mittag-Leffler. Since $\Omega Y \simeq A \times \Omega B$ we have a natural projection $f\colon \Omega Y \to \Omega B$ inducing surjections $f_n\colon G_n \to H_n$ of pointed sets.

If we knew each f_n was a homomorphism of groups, we could conclude $\mathrm{Ph}(X, B) \cong \lim^1 H_n = *$ by noting $\lim^1 f\colon \lim^1 G_n \to \lim^1 H_n$ is surjective and $\lim^1 G_n = *$. In general, however, we cannot expect the functions f_n to be homomorphisms, and so we must work marginally harder.

Fortunately, the Mittag-Leffler condition makes no reference to the group structure of the individual stages of a tower, and is more a property of the underlying tower of sets. In light of Theorem 2.2, to show $\lim^1 H_n = *$ it suffices to show the Mittag-Leffler condition is preserved under epimorphisms of towers of pointed sets. This is the content of the following lemma:

Lemma 3.2 *If $f\colon \{G_n\} \to \{H_n\}$ is an epimorphism of towers of pointed sets, and $\{G_n\}$ satisfies the Mittag-Leffler condition, then so does $\{H_n\}$.*

Proof Since $\{G_n\}$ is Mittag-Leffler then for each k there is some $N \in \mathbb{N}$ so that for $n \geq N$ one has

$$G_k^{(N)} = G_k^{(n)}.$$

A quick diagram chase shows that the surjections $f_k\colon G_k \to H_k$ induce surjections $f_k^{(n)}\colon G_k^{(n)} \to H_k^{(n)}$. In other words,

$$H_k^{(n)} = \{f(x) \mid x \in G_k^{(n)}\}.$$

But, for $n \geq N$ we have $G_k^{(n)} = G_k^{(N)}$ and so this shows $H_k^{(n)} = H_k^{(N)}$. So, the tower $\{H_n\}$ is Mittag-Leffler, which completes the proof of the lemma, and hence the proof of the loop-splitting theorem. □

Theorem 3.3 (wedge-splitting theorem) *Suppose Y is simply connected and has finite type over $\mathbb{Z}$ or $\mathbb{Z}_{(p)}$ and $Y \simeq A \vee B$. If both A and B are rationally nontrivial, then Y is the target of essential phantom maps from finite-type domains.*

For the proof we will need the following variation of Iriye's Corollary 1.5 from [13]. The proof is a simple modification of the argument there, replacing Theorem 2.1 with our Proposition 2.4.

Lemma 3.4 *Suppose Y has finite type over $\mathbb{Z}_{(p)}$. If either*

(1) *there is some $\alpha \in \pi_{2n+1}(Y)$ of infinite order whose image under the Hurewicz map is also of infinite order, or*

(2) *there is some $v \in H^{2n}(Y;\mathbb{Z})$ of infinite order whose square v^2 is also of infinite order,*

then ΣY is the target of essential phantom maps from finite-type domains.

Proof of the wedge-splitting theorem We note that since Y is simply connected, so too are A and B. In the long fiber sequence induced by the inclusion $i\colon A \vee B \to A \times B$,

$$\cdots \to \Omega F \xrightarrow{\Omega f} \Omega(A \vee B) \xrightarrow{\Omega i} \Omega A \times \Omega B \xrightarrow{\partial} F \xrightarrow{f} A \vee B \xrightarrow{i} A \times B$$

we can identify $F \simeq (\Omega A) * (\Omega B)$, where $X * Y$ denotes the join of topological spaces X and Y, and we find that $\partial \simeq *$. It follows that Ωi has a section, and Ωf has a retraction, which gives a natural homotopy equivalence

$$\Omega(A \vee B) \simeq \Omega A \times \Omega B \times \Omega((\Omega A) * (\Omega B)). \tag{4}$$

For a more complete account of this discussion we refer the reader to the work of Porter [19]. We now proceed by cases.

Case I Suppose Y has finite type over $\mathbb{Z}$. Then so do A and B. Now, if both A and B are rationally nontrivial, then $(\Omega A) * (\Omega B)$ is a simply connected, rationally nontrivial suspension space, hence is the target of essential phantom maps from finite-type domains by Theorem 1.1. Applying the loop-splitting theorem to the splitting, (4) then implies $A \vee B$ is the target of essential phantom maps from finite-type domains.

Case II In case Y has finite type over $\mathbb{Z}_{(p)}$ our goal will be, as above, to show that $\Omega A * \Omega B$ is the target of essential phantom maps from finite-type domains and appeal to the loop-splitting theorem. But, since $\Omega A * \Omega B$ is not of finite type over $\mathbb{Z}$ we must make use of Lemma 3.4. To do so we need to discover more about $\Omega A \wedge \Omega B$. Suppose $\mathrm{conn}_{\mathbb{Q}}(A) = n$ and $\mathrm{conn}_{\mathbb{Q}}(B) = m$, where by $\mathrm{conn}_{\mathbb{Q}}(X) = k-1$ we mean $\pi_i(X) \otimes \mathbb{Q} = 0$ for $i < k$ and $\pi_k(X) \otimes \mathbb{Q} \neq 0$. Choose $a \in H^n(\Omega A;\mathbb{Z})$ and $b \in H^m(\Omega B;\mathbb{Z})$ of infinite order. We proceed by cases.

Case A If n and m are both even, then a^2 and b^2 can be seen to be of infinite order, since $H^*(\Omega A;\mathbb{Q})$ contains $\mathbb{Q}[\overline{a}]$ as a subalgebra, where $\overline{a}$ is the image of a under rationalization, and similarly $\mathbb{Q}[\overline{b}]$ is a subalgebra of $H^*(\Omega B;\mathbb{Q})$. Then $(a\otimes b)^2$ has infinite order in $H^*(\Omega A\wedge\Omega B;\mathbb{Z})$, since $(\overline{a}\otimes\overline{b})^2$ is nonzero in $H^*(\Omega A\wedge\Omega B;\mathbb{Q})$ and Lemma 3.4(2) applies. Here we use the Künneth theorem to embed $H^*(\Omega A;\mathbb{Z})\otimes H^*(\Omega B;\mathbb{Z})$ in $H^*(\Omega A\wedge\Omega B;\mathbb{Z})$ as a submodule.

Case B If n is even and m is odd, then $\mathrm{conn}_{\mathbb{Q}}(\Omega A\wedge\Omega B)=n+m-1$ and, by the Hurewicz theorem, $\pi_{n+m}(\Omega A\wedge\Omega B)\to H_{n+m}(\Omega A\wedge\Omega B)$ is an isomorphism, with $n+m$ odd, so Lemma 3.4(1) applies.

Case C Suppose n and m are both odd, and without loss of generality assume $n\le m$. Since $\mathrm{conn}_{\mathbb{Q}}(\Omega A\wedge\Omega B)=n+m-1$ the rational Hurewicz homomorphism $\pi_{2n+m}\otimes\mathbb{Q}\to H_{2n+m}(-;\mathbb{Q})$ is an isomorphism by the rational Hurewicz theorem. Since n and m are odd, $2n+m$ is odd, while $\pi_{2n+m}(\Omega A\wedge\Omega B)\otimes\mathbb{Q}\neq 0$, and so Lemma 3.4(1) applies. $\square$

4 Proof of Theorem 1.3

We begin by showing it suffices to prove Theorem 1.3 when the nilpotent co-H–space Y in question is simply connected, so that we may appeal to the decompositions of looped co-H–spaces described in Section 2.2. To this end, assume Y is a co-H–space with $H_i(Y;\mathbb{Q})\neq 0$ for some $i>1$. By Fox [6], $\widetilde{Y}$ is a co-H–space, and as a consequence of the work of Iwase, Saito and Toshio [14] on homology of universal covers of co-H–spaces we see that if $H_i(Y;\mathbb{Q})\neq 0$ then $H_i(\widetilde{Y};\mathbb{Q})\neq 0$. In light of these facts and the upcoming Lemma 4.1 we replace Y with its universal cover for the proof of Theorem 1.3.

Lemma 4.1 *Suppose Y is a nilpotent co-H–space and let $c\colon \widetilde{Y}\to Y$ be the universal cover. If $\widetilde{Y}$ is the target of essential phantom maps from finite-type domains, then so too is Y.*

Proof By Theorem 2.3 if $\widetilde{Y}$ is the target of essential phantom maps from finite-type domains, then $\mathrm{Ph}(K(\mathbb{Z},n),\widetilde{Y})\neq *$ for some $n\ge 2$. We argue that c induces a weak injection $\mathrm{Ph}(K(\mathbb{Z},n),\widetilde{Y})\to\mathrm{Ph}(K(\mathbb{Z},n),Y)$.

Suppose $\varphi\colon K(\mathbb{Z},n)\to\widetilde{Y}$ is an essential phantom map. The map c is the fiber of the classifying map $Y\to B\pi_1(Y)$. Since Y is a co-H–space $\pi_1(Y)$ is a free group, and since Y is nilpotent $\pi_1(Y)$ is either trivial or congruent to $\mathbb{Z}$. Since the result is trivial in case $\pi_1(Y)=1$ we assume $\pi_1(Y)\cong\mathbb{Z}$. So $B\pi_1(Y)\simeq S^1$ and we have a fiber sequence

$$\Omega S^1 \xrightarrow{\delta} \widetilde{Y} \xrightarrow{c} Y.$$

We proceed by contradiction. Suppose $c\varphi\simeq *$. Then there is a lift $\lambda\colon K(\mathbb{Z},n)\to\Omega S^1$ of φ through δ. But $\Omega S^1\simeq\mathbb{Z}$ is discrete and $K(\mathbb{Z},n)$ is connected so $\lambda\simeq *$ and $\varphi\simeq\delta\lambda$ is trivial, a contradiction. Hence $c\varphi\colon K(\mathbb{Z},n)\to Y$ is essential. □

We now derive Theorem 1.3 as a consequence of the following three propositions. We begin with the case $\dim_{\mathbb{Q}}\widetilde{H}_*(Y;\mathbb{Q})\geq 2$. This condition ensures the decompositions of Section 2.2 are algebraically rich enough to detect essential phantom maps into Y via techniques developed in Section 3.

Proposition 4.2 *Suppose Y is a simply connected co-H–space with $\dim_{\mathbb{Q}}\widetilde{H}_*(Y;\mathbb{Q})$ at least 2. Then Y is the target of essential phantom maps from finite-type domains.*

Proof Choose a homogeneous basis of integral classes $\{x_1,x_2,\dots\}$ for $\widetilde{H}_*(Y;\mathbb{Q})$ with $|x_i|\leq|x_{i+1}|$ for each i, where $|x|$ denotes the homogeneous degree of x in $\widetilde{H}_*(Y;\mathbb{Q})$. Write

$$a=\Sigma^{-1}x_1\in\Sigma^{-1}\widetilde{H}_{m+1}(Y;\mathbb{Q})\quad\text{and}\quad b=\Sigma^{-1}x_2\in\Sigma^{-1}\widetilde{H}_{n+1}(Y;\mathbb{Q}).$$

Choose a prime $p\geq 5$ such that

$$H_{\leq m+n+2}(Y^{\wedge 2};\mathbb{Z})\quad\text{and}\quad H_{\leq 2m+n+3}(Y^{\wedge 3};\mathbb{Z})$$

have no p–torsion. We identify a and b as elements of $H_m(\Omega Y;\mathbb{Q})$ and $H_n(\Omega Y;\mathbb{Q})$, respectively, via that Bott–Samelson theorem. We will also write a, $b\in H_*(\Omega Y;\mathbb{Z})$ for lifts of a and b, and we will use the same notation for the mod p reductions of these elements in $H_*(\Omega Y;\mathbb{F}_p)$, making the context clear by indicating coefficient rings. We replace Y with its p–localization to avoid cumbersome notation; that is, we write Y for $Y_{(p)}$.

To show Y is the target of essential phantom maps from finite-type domains, we consider the geometric realization

$$\Omega Y\simeq\Omega\bar{A}^{\min}(Y)\times\Omega\Big(\bigvee_{n\geq 2}\bar{Q}_n B^{\max}(Y)\Big)$$

of the minimal decomposition from Section 2.2. We justify the indexing $n\geq 2$ by noting that $Q_1B^{\max}=0$, since $V\subseteq A^{\min}(V)$ for all vector spaces V. By the loop-splitting theorem, it suffices to show that $\bigvee_{n\geq 2}\bar{Q}_nB^{\max}(Y)$ is the target of essential phantom maps from finite-type domains. By the wedge-splitting theorem, this will follow if $\bar{Q}_iB^{\max}(Y)$ is rationally nontrivial for at least two i. We will show this is the case.

Write $V=\Sigma^{-1}\widetilde{H}_*(Y;\mathbb{F}_p)$ and identify

$$H_*(\Omega Y;\mathbb{F}_p)\cong T(V)$$

through the Bott–Samelson theorem. By Theorem 2.9, when i is not a power of p one has $L_i(V)\subseteq B^{\max}(V)$. So, since $p\geq 5$ we see that $[a,b]$, $[[b,a],a]\in B^{\max}(V)$. Moreover, $[a,b]$ is indecomposable in $B^{\max}(V)$, since the tensor length of $[a,b]$ in $T(V)$ is two and $B^{\max}(V)$ contains no elements of tensor length one in $T(V)$ (again, since $V\subseteq A^{\min}(V)$). Similarly, $[[b,a],a]$ is indecomposable, and we have $[a,b]\in Q_2B^{\max}(V)$ and $[[b,a],a]\in Q_3B^{\max}(V)$.

It follows that $[a,b]$ is in the image of

$$H_{n+m}(\Omega\bar{Q}_2B^{\max}(Y);\mathbb{F}_p)\to H_{n+m}(\Omega Y;\mathbb{F}_p),$$

so $H_{n+m}(\Omega\bar{Q}_2B^{\max}(Y);\mathbb{F}_p)\neq 0$. Finally, we note $\bar{Q}_2B^{\max}(Y)$ is a co-H–space by Theorem 2.10 and so, by the Bott–Samelson theorem,

$$H_*(\Omega\bar{Q}_2B^{\max}(Y);\mathbb{F}_p)\cong T(\Sigma^{-1}\widetilde{H}_*(\bar{Q}_2B^{\max}(Y);\mathbb{F}_p)).$$

Hence, we infer

$$\widetilde{H}_{\leq m+n+1}(\bar{Q}_2B^{\max}(Y);\mathbb{F}_p)\neq 0. \tag{5}$$

Similarly,

$$\widetilde{H}_{\leq 2m+n+1}(\bar{Q}_3B^{\max}(Y);\mathbb{F}_p)\neq 0. \tag{6}$$

Now, according to Theorem 2.10 for each i the space $\bar{Q}_iB^{\max}(Y)$ is a retract of an $(i-1)$–fold desuspension of $Y^{\wedge i}$. In particular, $H_k(\bar{Q}_iB^{\max}(Y);\mathbb{Z})$ is a retract of $H_{k+i-1}(Y^{\wedge i};\mathbb{Z})$. So, if $H_{\leq m+n+1}(\bar{Q}_2B^{\max}(Y);\mathbb{Z})$ has p–torsion, then so does $H_{\leq m+n+2}(Y^{\wedge 2};\mathbb{Z})$. Similarly, if $H_{\leq 2m+n+1}(\bar{Q}_3B^{\max}(Y);\mathbb{Z})$ has p–torsion, so does $H_{\leq 2m+n+3}(Y^{\wedge 3};\mathbb{Z})$. So, since

$$H_{\leq m+n+2}(Y^{\wedge 2};\mathbb{Z})\quad\text{and}\quad H_{\leq 2m+n+3}(Y^{\wedge 3};\mathbb{Z})$$

have no p–torsion we find

$$\widetilde{H}_{\leq m+n+2}(\bar{Q}_2B^{\max}(Y);\mathbb{Q})\quad\text{and}\quad \widetilde{H}_{\leq 2m+n+3}(\bar{Q}_3B^{\max}(Y);\mathbb{Q})$$

are nonzero. □

In case Y is a simply connected, finite-type co-H–space with $\dim_{\mathbb{Q}}\widetilde{H}_*(Y;\mathbb{Q})=1$ we are unable to use the method of the proof of Proposition 4.2 to witness the existence of essential phantom maps into Y from finite-type domains; we cannot expect to produce rationally nontrivial commutators in $H_*(\Omega Y;\mathbb{Z})$, which ultimately were the driving force behind that argument. In this case, Y is rationally equivalent to a sphere. We proceed by cases on the parity of the dimension of this sphere.

Proposition 4.3 *Suppose Y is a nilpotent co-H–space with $H^{2n}(Y;\mathbb{Q}) \neq 0$ for some $n \geq 1$. Then Y is the target of essential phantom maps from finite-type domains.*

Corollary 4.4 *Suppose Y is a nilpotent co-H–space with $Y \sim_{\mathbb{Q}} S^{2n}$ for some $n \geq 1$. Then Y is the target of essential phantom maps from finite-type domains.*

Proof of Proposition 4.3 Let $Y \xrightarrow{g} K(\mathbb{Z}, 2n)$ represent an element of $H^{2n}(Y;\mathbb{Z})$ of infinite order. According to Ganea [7], since Y is a co-H–space there is a lift λ in the diagram

$$\begin{tikzcd} & \Sigma K(\mathbb{Z}, 2n-1) \arrow[d, "p"] \\ Y \arrow[ur, bend left, "\lambda"] \arrow[r, "g"'] & K(\mathbb{Z}, 2n) \end{tikzcd}$$

where $p\colon \Sigma K(\mathbb{Z}, 2n-1) \simeq \Sigma\Omega K(\mathbb{Z}, 2n) \to K(\mathbb{Z}, 2n)$ is the evaluation map. Since g induces a surjection on $\pi_{2n} \otimes \mathbb{Q}$ and p induces an isomorphism on π_{2n} we can be sure $\pi_{2n}(\lambda) \otimes \mathbb{Q}$ is surjective. Since $\Sigma K(\mathbb{Z}, 2n-1)$ is rationally equivalent to S^{2n} we have an isomorphism of vector spaces

$$\pi_*(\Sigma K(\mathbb{Z}, 2n-1)) \otimes \mathbb{Q} \cong \mathbb{Q} \cdot \alpha \oplus \mathbb{Q} \cdot [\alpha, \alpha], \tag{7}$$

where $\alpha \in \pi_{2n}(\Sigma K(\mathbb{Z}, 2n-1)) \otimes \mathbb{Q}$ is a nonzero element and $[-,-]$ denotes the Whitehead product. Since α is in the image of $\pi_{2n}(\lambda)$, it follows from the naturality of the Whitehead product that $\pi_*(\lambda) \otimes \mathbb{Q}$ is surjective.

Finally, note that by Theorem 2.5 the map $\lambda\colon Y \to \Sigma K(\mathbb{Z}, 2n-1)$ induces surjections

$$\mathrm{Ph}(X, Y) \to \mathrm{Ph}(X, \Sigma K(\mathbb{Z}, 2n-1))$$

for all finite-type spaces X. By Theorem 1.1 there is a finite-type space X for which $\mathrm{Ph}(X, \Sigma K(\mathbb{Z}, 2n-1)) \neq *$, so $\mathrm{Ph}(X, Y) \neq *$. □

Proposition 4.5 *If Y is a nilpotent co-H–space with $Y \sim_{\mathbb{Q}} S^{2n+1}$, $n \geq 1$, then Y is the target of essential phantom maps from finite-type domains.*

Proof We first reduce to the case where Y is $2n$–connected. According to Golasiński and Klein [8], if Y is a co-H–space, then one can choose compatible co-H–structures Y and on each skeleton Y_k so that the inclusion maps $Y_k \hookrightarrow Y$ are co-H–maps. Berstein and Hilton have shown the cofiber of a co-H–map is a co-H–space [3, Theorem 3.4], so Y/Y_k is a co-H–space. Finally, $Y \to Y/Y_{2n}$ is a rational equivalence, so by Theorem 2.5 this map induces a surjection $\mathrm{Ph}(X,Y) \to \mathrm{Ph}(X,Y/Y_{2n})$ for all finite-type domains X. Hence, if Y/Y_{2n} is the target of essential phantom maps from finite-type domains, then so too is Y.

Henceforth we assume the space Y to be $2n$–connected. We proceed by contradiction. Suppose Y is not the target of essential phantom maps from finite-type domains. For brevity, write $K = K(\mathbb{Z}, n)$. Then by Theorem 2.3 there is a rational equivalence $f\colon K \to \Omega Y$. Let $u\colon \Omega Y \to K$ represent a cohomology class of infinite order, and write F for the homotopy fiber of u. Since f and u are rational equivalences we can localize at a large enough prime p and find that $f_{(p)}$ and $u_{(p)}$ induce isomorphisms on π_{2n}. For the rest of this section all spaces and maps will be localized at this large prime p, though the notation will not be burdened with this assumption; we write Y for $Y_{(p)}$.

Now uf is a self-equivalence of K by the Whitehead theorem, so K is a retract of ΩY. Thus $\Omega Y \simeq K \times F$, which gives rise to a homotopy equivalence

$$\Sigma\Omega Y \simeq \Sigma K \vee \Sigma F \vee \Sigma K \wedge F.$$

Choose a section $s\colon Y \to \Sigma\Omega Y$ of the evaluation map, ensured to exist since Y is a co-H–space. Let $i\colon Y \to K$ be the composite

$$Y \xrightarrow{s} \Sigma\Omega Y \simeq \Sigma K \vee \Sigma F \vee \Sigma K \wedge F \to \Sigma K$$

and let q be the map

$$\Sigma K \hookrightarrow \Sigma K \vee \Sigma F \vee \Sigma K \wedge F \simeq \Sigma\Omega Y \to Y,$$

where the last map is the evaluation map. Then qi induces an isomorphism on $\pi_{2n+1}(Y)$. Since Y is $(2n)$–connected and of finite type, it follows from the Hurewicz theorem that

$$q^*\colon H^{2n+1}(Y;\mathbb{Z}) \to H^{2n+1}(\Sigma K;\mathbb{Z})$$

is an isomorphism.

We take a generator $v \in H^{2n+1}(\Sigma K;\mathbb{Z}/p)$ and let $w = (q^*)^{-1}(v) \in H^{2n+1}(Y;\mathbb{Z}/p)$. Then $v = \Sigma\tilde{v}$ for $\tilde{v}$ a generator of $H^{2n}(K;\mathbb{Z}/p)$, where

$$\Sigma\colon H^{2n}(K) \to H^{2n+1}(\Sigma K;\mathbb{Z}/p)$$

is the suspension isomorphism. We then consider the morphism of Bockstein spectral sequences $q^*\colon E^*(Y) \to E^*(\Sigma K)$. Write $\mathcal{P}^n$ for the n^{th} reduced p^{th} power map. Then $\tilde{v}^p = \mathcal{P}^n(\tilde{v})$ survives to $E^\infty_{2np}(K)$, so $\mathcal{P}^n(v)$ survives to $E^\infty_{2np+1}(\Sigma K)$. Since $\mathcal{P}^n(v) = \mathcal{P}^n(q^*(w)) = q^*\mathcal{P}^n(w)$, we infer $\mathcal{P}^n(w)$ survives to $E^\infty_{2np+1}(Y)$. It follows that $H_{2np+1}(Y;\mathbb{Q}) \neq 0$, contradicting the assumption $Y \sim_{\mathbb{Q}} S^{2n+1}$. $\square$

5 Examples

In Examples 5.1 and 5.2 we describe co-H–spaces that satisfy the hypotheses of Theorem 1.3, but not Theorem 1.1. More specifically, we construct nonsuspension co-H–spaces whose rational homology is nontrivial. We prefer to present infinite-dimensional examples, since Zabrodsky obtained much stronger results than we have herein on phantom maps into finite complexes in [25].

Example 5.1 For each prime $p \geq 3$ write $\alpha_p\colon S^{2p} \to S^3$ for a representative of an element of order p in $\pi_{2p}(S^3)$. The homotopy cofibers C_{α_p} of these maps are classical examples, due to Berstein and Hilton [3, page 444], of co-H–spaces that do not have the homotopy type of suspension spaces. One key to establishing these examples is to prove, via Berstein–Hilton–Hopf invariant techniques, that each map α_p is a co-H–map. By [3, Theorem 3.4], the cofiber of a co-H–map is a co-H–space.

Write $\alpha\colon \bigvee_{p\geq 3} S^{2p} \to S^3$, where the wedge is taken over all odd primes, for the map whose restriction to each summand S^{2p} is α_p. Since each α_p is a co-H–map, so is α. It follows that the homotopy cofiber C_α of α is a co-H–space. Evidently $\dim_{\mathbb{Q}} \widetilde{H}_*(C_\alpha;\mathbb{Q}) = \infty$.

We now argue that C_α is not a suspension space. Assume to the contrary that $C_\alpha \simeq \Sigma Z$. Then, by the proof of [3, Lemma 3.6], we can choose Z to be 1–connected, so that Z has a homology decomposition, ie there is a diagram

$$\begin{tikzcd}
M_1 \arrow[d, "k_1"'] & M_2 \arrow[d, "k_2"'] & \cdots & M_n \arrow[d] & M_{n+1} \arrow[d] & \cdots \\
Z_1 \arrow[r, "i_1"'] & Z_2 \arrow[r, "i_2"'] & \cdots \arrow[r] & Z_n \arrow[r, "i_n"'] & Z_{n+1} \arrow[r] & \cdots
\end{tikzcd}$$

in which $M_i = M(H_{i+1}(Z), i)$ for each i, $M_i \to Z_i \to Z_{i+1}$ is a cofiber sequence and Z is the homotopy colimit of the tower along the bottom of this diagram. The space Z_i is called the i^{th} stage of the homology decomposition. It follows that ΣZ has a homology decomposition in which each stage is a suspension.

Suppose $h\colon C_\alpha \to \Sigma Z$ is a homotopy equivalence. Write $(C_\alpha)_k$ for the k^{th} stage of the homology decomposition for C_α. According to Arkowitz [1, Proposition 3.4], since $\mathrm{Ext}(H_n(C_\alpha;\mathbb{Z}); H_{n+1}(\Sigma Z;\mathbb{Z})) = 0$ for all n and ΣZ is 2–connected, h induces homotopy equivalences $h_n\colon (C_\alpha)_n \to (\Sigma Z)_n \simeq \Sigma(Z_n)$. But then $(C_\alpha)_6 \simeq C_{\alpha_3}$ must be a suspension space, a contradiction.

Example 5.2 By modifying the construction from Example 5.1 we can obtain an infinite-dimensional, nonsuspension co-H–space Y with $Y \sim_{\mathbb{Q}} S^3$. Replace each map $\alpha_p\colon S^{2p} \to S^3$ with a map $\beta_p\colon M(\mathbb{Z}/p, 2p) \to S^3$ representing an element of $\pi_{2p}(S^3; \mathbb{Z}/p)$ of order p. The argument of Berstein and Hilton [3] shows that the cofiber C_{β_p} of each β_p is a co-H–space which is not a suspension space, and so the argument in Example 5.1 shows that C_β is a co-H–space which is not a suspension space.

Finally we present an application of the loop-splitting theorem to spaces that are not necessarily co-H–spaces. For a space Y write $G_m(Y)$ for the m^{th} space of Ganea over Y (see [5]; the reader may more readily recognize this space as $G_m(Y) = B_m\Omega Y$, where B_m is the m^{th} stage of Milnor's classifying space construction). The spaces $G_m(Y)$ can be thought of as prototypes for spaces of Lusternik–Schnirelmann category at most m. We view this example as a test case for Question 1.4.

Example 5.3 We show that if $H_i(G_m(Y); \mathbb{Q}) \neq 0$ for some $i > 1$ then $G_m(Y)$ is the target of essential phantom maps from finite-type domains.

There is a well-known homotopy decomposition

$$\Omega G_m(Y) \simeq \Omega Y \times \Omega((\Omega Y)^{*m+1}),$$

where X^{*k} denotes the k–fold join of X. Since $H_i(G_m(Y); \mathbb{Q}) \neq 0$ we must have $H_j(Y; \mathbb{Q}) \neq 0$ for some $j > 1$ and similarly $H_*((\Omega Y)^{*m+1}; \mathbb{Q})$ is similarly nontrivial, so, by Theorem 1.1, $(\Omega Y)^{*m+1}$ is the target of essential phantom maps from finite-type domains. The loop-splitting theorem then implies $G_m(Y)$ is the target of essential phantom maps.

References

[1] **M Arkowitz**, *Induced mappings of homology decompositions*, from "Homotopy and geometry" (J Oprea, A Tralle, editors), Banach Center Publ. 45, Polish Acad. Sci., Warsaw (1998) 225–233 MR

[2] **I Berstein**, *On co-groups in the category of graded algebras*, Trans. Amer. Math. Soc. 115 (1965) 257–269 MR

[3] **I Berstein**, **P J Hilton**, *Category and generalized Hopf invariants*, Illinois J. Math. 4 (1960) 437–451 MR

[4] **A K Bousfield**, **D M Kan**, *Homotopy limits, completions and localizations*, Lecture Notes in Mathematics 304, Springer, Berlin (1972) MR

[5] **O Cornea**, **G Lupton**, **J Oprea**, **D Tanré**, *Lusternik–Schnirelmann category*, Mathematical Surveys and Monographs 103, Amer. Math. Soc., Providence, RI (2003) MR

[6] **R H Fox**, *On the Lusternik–Schnirelmann category*, Ann. of Math. 42 (1941) 333–370 MR

[7] **T Ganea**, *Lusternik–Schnirelmann category and strong category*, Illinois J. Math. 11 (1967) 417–427 MR

[8] **M Golasiński**, **J R Klein**, *On maps into a co-* H*–space*, Hiroshima Math. J. 28 (1998) 321–327 MR

[9] **B I Gray**, *Spaces of the same* n*–type, for all* n, Topology 5 (1966) 241–243 MR

[10] **B Gray**, *On generalized Whitehead products*, Trans. Amer. Math. Soc. 363 (2011) 6143–6158 MR

[11] **J Grbić**, **S Theriault**, **J Wu**, *Decompositions of looped co-* H*–spaces*, Proc. Amer. Math. Soc. 141 (2013) 1451–1464 MR

[12] **J R Harper**, **J Roitberg**, *Phantom maps and spaces of the same* n*–type for all* n, J. Pure Appl. Algebra 80 (1992) 123–137 MR

[13] **K Iriye**, *On phantom maps into suspension spaces*, J. Math. Kyoto Univ. 43 (2003) 661–669 MR

[14] **N Iwase**, **S Saito**, **T Sumi**, *Homology of the universal covering of a co-H–space*, Trans. Amer. Math. Soc. 351 (1999) 4837–4846 MR

[15] **J P May**, **K Ponto**, *More concise algebraic topology: localization, completion, and model categories*, University of Chicago Press (2012) MR

[16] **C A McGibbon**, *Phantom maps*, from "Handbook of algebraic topology" (I M James, editor), North-Holland, Amsterdam (1995) 1209–1257 MR

[17] **C A McGibbon**, **J M Møller**, *On spaces with the same* n*–type for all* n, Topology 31 (1992) 177–201 MR

[18] **C A McGibbon**, **J Roitberg**, *Phantom maps and rational equivalences*, Amer. J. Math. 116 (1994) 1365–1379 MR

[19] **G J Porter**, *The homotopy groups of wedges of suspensions*, Amer. J. Math. 88 (1966) 655–663 MR

[20] **J Roitberg**, *Note on phantom phenomena and groups of self-homotopy equivalences*, Comment. Math. Helv. 66 (1991) 448–457 MR

[21] **P Selick**, **S Theriault**, **J Wu**, *Functorial homotopy decompositions of looped co-* H *spaces*, Math. Z. 267 (2011) 139–153 MR

[22] **P Selick**, **J Wu**, *On natural coalgebra decompositions of tensor algebras and loop suspensions*, Mem. Amer. Math. Soc. 701, Amer. Math. Soc., Providence, RI (2000) MR

[23] **P Selick**, **J Wu**, *The functor* $A^{\min}$ *on* p*–local spaces*, Math. Z. 253 (2006) 435–451 MR

[24] **S D Theriault**, *Homotopy decompositions involving the loops of coassociative co- H spaces*, Canad. J. Math. 55 (2003) 181–203 MR

[25] **A Zabrodsky**, *On phantom maps and a theorem of H Miller*, Israel J. Math. 58 (1987) 129–143 MR

Grand Rapids, MI 49503
United States

schwassj@gmail.com

Received: 17 August 2015 Revised: 9 March 2016

Algebraic & Geometric Topology 17 (2017) 869–893

Infinite loop spaces and nilpotent K–theory

ALEJANDRO ADEM
JOSÉ MANUEL GÓMEZ
JOHN A LIND
ULRIKE TILLMANN

Using a construction derived from the descending central series of the free groups, we produce filtrations by infinite loop spaces of the classical infinite loop spaces $B\mathrm{SU}$, BU, $B\mathrm{SO}$, BO, $B\mathrm{Sp}$, $B\mathrm{GL}_\infty(R)^+$ and $Q_0(\mathbb{S}^0)$. We show that these infinite loop spaces are the zero spaces of nonunital E_∞–ring spectra. We introduce the notion of q–nilpotent K–theory of a CW–complex X for any $q \geq 2$, which extends the notion of commutative K–theory defined by Adem and Gómez, and show that it is represented by $\mathbb{Z} \times B(q, U)$, where $B(q, U)$ is the q^{th} term of the aforementioned filtration of BU.

For the proof we introduce an alternative way of associating an infinite loop space to a commutative $\mathbb{I}$–monoid and give criteria for when it can be identified with the plus construction on the associated limit space. Furthermore, we introduce the notion of a commutative $\mathbb{I}$–rig and show that they give rise to nonunital E_∞–ring spectra.

55N15, 55R35

1 Introduction

Let G denote a locally compact, Hausdorff topological group such that $1_G \in G$ is a nondegenerate base point. It is well known that we can obtain a model for the classifying space BG as the geometric realization of the classical bar construction B_*G. Now fix an integer $q \geq 2$ and let Γ_n^q be the q^{th} stage of the descending central series of the free group on n letters F_n, with the convention $\Gamma_n^1 = F_n$. Consider the set of homomorphisms $B_n(q, G) := \mathrm{Hom}(F_n / \Gamma_n^q, G)$. If $e_1, \dots, e_n$ are generators of F_n, then evaluation on the classes corresponding to $e_1, \dots, e_n$ provides a natural inclusion $B_n(q, G) \subset G^n$. Using this inclusion we can give $B_n(q, G)$ the subspace topology. Therefore $B_n(q, G)$ is precisely the space of ordered n–tuples in G generating a subgroup of G with nilpotence class less than q. For each fixed $q \geq 2$ the collection $\{B_n(q, G)\}_{n \geq 0}$ forms a simplicial space with face and degeneracy maps induced by those in the bar construction. The geometric realization of this simplicial space is

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.869

denoted by $B(q,G)$. These spaces were first introduced by Adem, Cohen and Torres Giese [1], where many of their basic properties were established. They give rise to a natural filtration of the classifying space

$$B(2,G) \subset B(3,G) \subset \cdots \subset B(q,G) \subset B(q+1,G) \subset \cdots \subset BG.$$

For $q=2$ we obtain $B_{\mathrm{com}}G := B(2,G)$, which is constructed by assembling the different spaces of ordered commuting n–tuples in the group G. Adem and Gómez [2] showed that for Lie groups this space plays the role of a classifying space for commutativity. More generally $B(q,G)$ is a classifying space for G–bundles of transitional nilpotency class less than q.

For the infinite unitary group $U = \mathrm{colim}_{n\to\infty} U(n)$, it is well known that BU is the infinite loop space underlying a nonunital E_∞–ring spectrum, namely the homotopy fiber of the Postnikov section $ku \to H\mathbb{Z}$. In other words, BU is a so-called nonunital E_∞–ring space. A basic question is whether the above gives rise to a filtration of BU by nonunital E_∞–ring spaces. The main purpose of this paper is to show that indeed this is the case, not only for U but also for other linear groups.

Theorem 1.1 *The spaces $B(q,\mathrm{SU})$, $B(q,U)$, $B(q,\mathrm{SO})$, $B(q,O)$ and $B(q,\mathrm{Sp})$ provide a filtration by nonunital E_∞–ring spaces of the classical infinite loop spaces $B\mathrm{SU}$, BU, $B\mathrm{SO}$, BO and $B\mathrm{Sp}$, respectively.*

The q–nilpotent K–theory of a space X is defined using isomorphism classes of bundles on X whose transition functions generate subgroups of nilpotence class less than q. We show that $K_{q\text{–nil}}(X) \cong [X, \mathbb{Z} \times B(q,U)]$, from which we obtain:

Corollary 1.2 *$K_{q\text{–nil}}(-)$ is the zeroth term of a generalized multiplicative cohomology theory.*

In particular we obtain a sequence of multiplicative cohomology theories

$$K_{\mathrm{com}}(X) = K_{2\text{–nil}}(X) \to K_{3\text{–nil}}(X) \to \cdots \to K_{q\text{–nil}}(X) \to \cdots \to K(X).$$

We also show that $B(q,U) \to BU$ splits as a map of infinite loop spaces, whence we see that topological K–theory is a direct summand in $K_{q\text{–nil}}$.

The infinite loop space structure on $B(q,G)$ for $G = U$, SU, SO, O, Sp is obtained by using the machinery of commutative $\mathbb{I}$–monoids first introduced by Bökstedt and developed by Schlichtkrull [19], Sagave and Schlichtkrull [18] and Lind [9]. Here $\mathbb{I}$ is the category of finite sets and injections. In addition to the usual construction, we associate an infinite loop space to a commutative $\mathbb{I}$–monoid by restricting the usual homotopy colimit construction to the subcategory $\mathbb{P}$ of finite sets and isomorphisms. This allows us to identify the homotopy type of the homotopy colimit under certain conditions. Another addition to infinite loop space theory is the introduction of the notion of a commutative $\mathbb{I}$–rig, which we show to give rise to a bipermutative category and hence an E_∞–ring spectrum.

Our main examples above all arise from commutative $\mathbb{I}$–rigs where we can identify the infinite loop space as the plus construction of the associated limit space. A more complicated situation arises for $Q_0(\mathbb{S}^0) \simeq B\Sigma_\infty^+$ and $B\mathrm{GL}_\infty(R)^+$. Our methods give rise to natural sequences of E_∞–ring spaces but the terms are not easy to describe.

The outline of this article is as follows. In Section 2 we use the machinery of commutative $\mathbb{I}$–monoids to produce two associated infinite loop spaces, one of which is a nonunital E_∞–ring space when the $\mathbb{I}$–monoid is an $\mathbb{I}$–rig. In Section 3 we show that these are homotopy equivalent and identify them under suitable assumptions. Then in Section 4 we apply these results to prove Theorem 1.1 and show that the spaces $B(q,U)$ for $q \geq 2$ are infinite loop spaces and that BU splits off. Finally, in Section 5 we introduce the notion of q–nilpotent K–theory and show that it is represented by the infinite loop spaces $\mathbb{Z} \times B(q,U)$, answering the question raised for commutative K–theory in [2].

We would like to thank Christian Schlichtkrull for helpful conversations about commutative $\mathbb{I}$–monoids, Simon Gritschacher for drawing our attention to Fiedorowicz and Ogle [6] and the referee for providing very useful comments.

Acknowledgements Adem was supported by NSERC. Gómez acknowledges the financial support of COLCIENCIAS through grant number 121565840569 of the Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Inovación, Fondo Francisco José de Caldas.

2 Commutative $\mathbb{I}$–monoids and infinite loop spaces

The standard construction of the infinite loop space structure on BU from the permutative category of complex vector spaces and their isomorphisms does not restrict to give an infinite loop space structure on $B(q,U)$. Instead we are going to use certain constructions on commutative $\mathbb{I}$–monoids. More precisely, we will give two constructions of permutative categories from commutative $\mathbb{I}$–monoids. For the case of interest the permutative categories are actually bipermutative and hence give rise to E_∞–ring spectra. We start by setting up some notations and basic definitions following [19; 18; 9]. We will use [5] as a reference for bipermutative categories and the associated multiplicative infinite loop space machinery.

2.1 The category $\mathbb{I}$ and its subcategories $\mathbb{P}$ and $\mathbb{N}$

These three categories are skeletons of the category of finite sets and injections, the category of finite sets and isomorphisms, and the translation category associated to the monoid of natural numbers. We will use the following notation.

For every integer $n \geq 0$, let $\boldsymbol{n}$ denote the set $\{1, 2, \dots, n\}$. When $n = 0$ we use the convention $\mathbf{0} := \varnothing$. Let $\mathbb{I}$ denote the category whose objects are the elements of the form $\boldsymbol{n}$ for all integers $n \geq 0$ with morphisms given by all injective maps. Note that in particular $\mathbf{0}$ is an initial object in the category $\mathbb{I}$ and $\mathbb{I}$ is a symmetric monoidal category under the concatenation $\boldsymbol{m} \sqcup \boldsymbol{n} := \{1, 2, \dots, m+n\}$ with the symmetry morphism given by the (m, n)–shuffle map

$$\tau_{\boldsymbol{m},\boldsymbol{n}} \colon \boldsymbol{m} \sqcup \boldsymbol{n} \to \boldsymbol{n} \sqcup \boldsymbol{m}.$$

It is also symmetric monoidal under the Cartesian product

$$\boldsymbol{m} \times \boldsymbol{n} := \{1 = (1,1),\ 2 = (1,2), \dots,\ n+1 = (2,1), \dots,\ mn = (m,n)\}$$

given by lexicographic ordering. By definition, $\mathbf{0} \times \boldsymbol{n} = \mathbf{0} = \boldsymbol{n} \times \mathbf{0}$. The associated symmetry morphism is given by a permutation

$$\tau^{\times}_{\boldsymbol{m}\boldsymbol{n}} \colon \boldsymbol{m} \times \boldsymbol{n} \to \boldsymbol{n} \times \boldsymbol{m}.$$

The latter monoidal product is distributive over the former. More precisely, left distributivity

$$\delta^{l}_{\boldsymbol{m},\boldsymbol{n},\boldsymbol{k}} \colon \boldsymbol{m} \times \boldsymbol{k} \sqcup \boldsymbol{n} \times \boldsymbol{k} \to (\boldsymbol{m} \sqcup \boldsymbol{n}) \times \boldsymbol{k}$$

is given by the identity and right distributivity is given by a permutation

$$\delta^{r}_{\boldsymbol{m},\boldsymbol{n},\boldsymbol{k}} \colon \boldsymbol{m} \times \boldsymbol{n} \sqcup \boldsymbol{m} \times \boldsymbol{k} \to \boldsymbol{m} \times (\boldsymbol{n} \sqcup \boldsymbol{k}).$$

These two structures make $\mathbb{I}$ into a bipermutative category, as in [5, Definition 3.6].

The category $\mathbb{I}$ has two natural subcategories. Let $\mathbb{P}$ be the totally disconnected subcategory containing all objects and all isomorphisms $\sigma \colon \boldsymbol{n} \to \boldsymbol{n}$ but no other morphisms, and let $\mathbb{N}$ denote the connected subcategory containing all objects, their identities and only the canonical inclusions $j \colon \boldsymbol{n} \to \boldsymbol{m}$. While $\mathbb{P}$ is a bipermutative subcategory, $\mathbb{N}$ does not inherit any monoidal structure from $\mathbb{I}$.

2.2 Definitions of commutative $\mathbb{I}$–monoids and $\mathbb{I}$–rigs

An $\mathbb{I}$–space is a functor $X \colon \mathbb{I} \to \mathrm{Top}$. Every morphism in $\mathbb{I}$ can be factored as a composition of a canonical inclusion $j \colon \boldsymbol{n} \hookrightarrow \boldsymbol{m}$ and a permutation $\sigma \colon \boldsymbol{m} \to \boldsymbol{m}$. Therefore an $\mathbb{I}$–space $X \colon \mathbb{I} \to \mathrm{Top}$ determines a sequence of spaces $X(\boldsymbol{n})$ together with an induced action of the symmetric group Σ_n for $n \geq 0$, and structural maps $j_n\colon X(\boldsymbol{n}) \to X(\boldsymbol{n}+\boldsymbol{1})$ that are equivariant in the sense that $j_n(\sigma \cdot x) = \sigma \cdot j_n(x)$ for every $\sigma \in \Sigma_n$ and $x \in X(\boldsymbol{n})$. On the right-hand side we see σ as element in Σ_{n+1} via the canonical inclusion $\Sigma_n \hookrightarrow \Sigma_{n+1}$. Vice versa, given such a sequence of Σ_n–spaces $X(\boldsymbol{n})$ and compatible structure maps j_n, they give rise to an $\mathbb{I}$–space if and only if for $m \geq n$ and any two elements $\sigma, \sigma' \in \Sigma_m$ which restrict to the same permutation of $\boldsymbol{n}$ we have $\sigma(x) = \sigma'(x)$ for all $x \in j(X(\boldsymbol{n}))$. We note that this condition is not satisfied by the sequence $X(\boldsymbol{n}) = \Sigma_n$ with the left or right multiplication action, but *is* satisfied by the sequence $X(\boldsymbol{n}) = \boldsymbol{n}$ with the natural permutation action since $\boldsymbol{n} \cong \mathbb{I}(\boldsymbol{1}, \boldsymbol{n})$.

We say that an $\mathbb{I}$–space is an $\mathbb{I}$–monoid if it comes equipped with a natural transformation

$$\mu_{\boldsymbol{m},\boldsymbol{n}}\colon X(\boldsymbol{m}) \times X(\boldsymbol{n}) \to X(\boldsymbol{m} \sqcup \boldsymbol{n})$$

of functors defined on $\mathbb{I} \times \mathbb{I}$ and a natural transformation

$$\eta_{\boldsymbol{n}}\colon * \to X(\boldsymbol{n})$$

from the constant $\mathbb{I}$–space $*(\boldsymbol{n}) = *$ to X satisfying associativity and unit axioms for $* \in X(\boldsymbol{0})$. We say that X is a commutative $\mathbb{I}$–monoid if μ is commutative, meaning that the diagram

$$\begin{tikzcd}
X(\boldsymbol{m}) \times X(\boldsymbol{n}) \arrow[r, "\mu_{\boldsymbol{m},\boldsymbol{n}}"] \arrow[d, "\tau"'] & X(\boldsymbol{m} \sqcup \boldsymbol{n}) \arrow[d, "\tau_{\boldsymbol{m},\boldsymbol{n}}"'] \\
X(\boldsymbol{n}) \times X(\boldsymbol{m}) \arrow[r, "\mu_{\boldsymbol{n},\boldsymbol{m}}"] & X(\boldsymbol{n} \sqcup \boldsymbol{m})
\end{tikzcd}$$

commutes, where $\tau(x, y) = (y, x)$.

An $\mathbb{I}$–rig is a commutative $\mathbb{I}$–monoid equipped with a natural transformation

$$\pi_{\boldsymbol{m},\boldsymbol{n}}\colon X(\boldsymbol{m}) \times X(\boldsymbol{n}) \to X(\boldsymbol{m} \times \boldsymbol{n})$$

of functors defined on $\mathbb{P} \times \mathbb{P}$ and an element $1 \in X(1)$ satisfying associativity and unit axioms, as well as left distributivity, ie that the diagram

$$\begin{tikzcd}
(X(\boldsymbol{m}) \times X(\boldsymbol{n})) \times X(\boldsymbol{k}) \arrow[r, "\pi_{\boldsymbol{m}\sqcup\boldsymbol{n},\boldsymbol{k}} \circ (\mu_{\boldsymbol{m},\boldsymbol{n}} \times 1)"] \arrow[d, "(1\times\tau\times1)\circ(1\times1\times\Delta)"'] & X((\boldsymbol{m} \sqcup \boldsymbol{n}) \times \boldsymbol{k}) \\
X(\boldsymbol{m}) \times X(\boldsymbol{k}) \times X(\boldsymbol{n}) \times X(\boldsymbol{k}) \arrow[r, "\mu_{\boldsymbol{m}\times\boldsymbol{k},\boldsymbol{n}\times\boldsymbol{k}} \circ (\pi_{\boldsymbol{m},\boldsymbol{k}} \times \pi_{\boldsymbol{n},\boldsymbol{k}})"] & X(\boldsymbol{m} \times \boldsymbol{k} \sqcup \boldsymbol{n} \times \boldsymbol{k}) \arrow[u, "\delta^l_{\boldsymbol{m},\boldsymbol{n},\boldsymbol{k}}"']
\end{tikzcd}$$

commutes, and right distributivity, which is given by an analogous commutative diagram. Here $\triangle$ is the diagonal map. We emphasize that π is only required to be natural on the subcategory $\mathbb{P}\times\mathbb{P}$ of $\mathbb{I}\times\mathbb{I}$.[1]

A commutative $\mathbb{I}$–rig is an $\mathbb{I}$–rig in which π is commutative in the sense that the diagram

$$\begin{tikzcd} X(\boldsymbol{m})\times X(\boldsymbol{n}) \arrow[r, "\pi_{\boldsymbol{m},\boldsymbol{n}}"] \arrow[d, "\tau"'] & X(\boldsymbol{m}\times\boldsymbol{n}) \arrow[d, "\tau^{\times}_{\boldsymbol{m},\boldsymbol{n}}"'] \\ X(\boldsymbol{n})\times X(\boldsymbol{m}) \arrow[r, "\pi_{\boldsymbol{n},\boldsymbol{m}}"] & X(\boldsymbol{n}\times\boldsymbol{m}) \end{tikzcd}$$

commutes. A natural transformation T between two $\mathbb{I}$–spaces X and Y defines a map of commutative $\mathbb{I}$–monoids ($\mathbb{I}$–rigs) if it commutes with μ (and π) in the sense that $T\circ\mu_{\boldsymbol{m},\boldsymbol{n}}=\mu_{\boldsymbol{m},\boldsymbol{n}}\circ T\times T$ (and $T\circ\pi_{\boldsymbol{m},\boldsymbol{n}}=\pi_{\boldsymbol{m},\boldsymbol{n}}\circ T\times T$). We have thus defined a category of $\mathbb{I}$–spaces, a category of commutative $\mathbb{I}$–monoids and a category of $\mathbb{I}$–rigs.

2.3 Associated (bi)permutative translation categories

We will use the following notation for translation categories. If $Y\colon\mathcal{C}\to\mathrm{Top}$ is a functor from a category $\mathcal{C}$ to the category of topological spaces, we let $\mathcal{C}\ltimes Y$ denote the translation category on Y. The translation category, also known as the Grothendieck construction, is a topological category whose objects are pairs (c,x) consisting of an object c of $\mathcal{C}$ and a point $x\in Y(c)$. A morphism in $\mathcal{C}\ltimes Y$ from (c,x) to (c',x') is a morphism $\alpha\colon c\to c'$ in $\mathcal{C}$ satisfying the equation $Y(\alpha)(x)=x'$. For example, if $\mathcal{C}=G$ is a group, thought of as a one object category, then the translation category $G\ltimes Y$ is the action groupoid for the G–space Y and its classifying space is the homotopy orbit space $B(G\ltimes Y)=EG\times_G Y$. In general, the classifying space $B(\mathcal{C}\ltimes Y)$ is homeomorphic to the homotopy colimit $\mathrm{hocolim}_{\mathcal{C}}\,Y$ of Y over $\mathcal{C}$ defined using the bar construction.

Suppose now that X is a commutative $\mathbb{I}$–monoid. Then the translation category $\mathbb{I}\ltimes X$ is a permutative category, as we now explain. The monoidal structure $\oplus$ is defined on objects $(\boldsymbol{m},x)$ and $(\boldsymbol{n},y)$ by

$$(\boldsymbol{m},x)\oplus(\boldsymbol{n},y)=(\boldsymbol{m}\sqcup\boldsymbol{n},\mu_{\boldsymbol{m},\boldsymbol{n}}(x,y))$$

[1] In fact, we do not know of any nontrivial examples where π may be extended to a natural transformation of functors defined on $\mathbb{I}\times\mathbb{I}$. The examples of $\mathbb{I}$–rigs that we discuss in Section 2.5 do not satisfy this additional naturality condition. Indeed, as we will see in the following sections, an $\mathbb{I}$–rig that does satisfy this condition and has each level $X(\boldsymbol{n})$ a connected space would give rise to a connected E_∞–ring space $\mathrm{hocolim}_{\mathbb{I}}\,X$. An E_∞–ring space whose multiplicative unit and additive unit lie in the same path component is contractible, so such examples would only give rise to trivial E_∞–ring spectra.

and on morphisms $\alpha\colon (\boldsymbol{m}, x) \to (\boldsymbol{m}', x')$ and $\beta\colon (\boldsymbol{n}, y) \to (\boldsymbol{n}', y')$ by letting

$$\alpha \oplus \beta\colon (\boldsymbol{m}, x) \oplus (\boldsymbol{n}, y) \to (\boldsymbol{m}', x') \oplus (\boldsymbol{n}', y')$$

be determined by the morphism

$$\alpha \sqcup \beta\colon \boldsymbol{m} \sqcup \boldsymbol{n} \to \boldsymbol{m}' \sqcup \boldsymbol{n}'$$

in the category $\mathbb{I}$. Notice that $X(\alpha \sqcup \beta)(\mu_{\boldsymbol{m},\boldsymbol{n}}(x, y)) = \mu_{\boldsymbol{m}',\boldsymbol{n}'}(x', y')$ by the naturality of μ, so that this is well-defined. The associativity and unit conditions for X imply that $\mathbb{I} \ltimes X$ is a strict monoidal category with strict unit object $(\mathbf{0}, *)$ determined by the unit η of the $\mathbb{I}$–monoid X. The commutativity of X implies that $\mathbb{I} \ltimes X$ is a permutative category, see for example [5, Definition 3.1]. Note that the permutative structure on $\mathbb{I} \ltimes X$ restricts to the subcategory $\mathbb{P} \ltimes X$.

Suppose now that X is a commutative $\mathbb{I}$–rig. Then by the same reasoning as above, there is another permutative category structure on $\mathbb{P} \ltimes X$ with product $\otimes$ induced by π and strict unit object $(\mathbf{1}, 1)$. The distributivity axioms for X translate to distributivity axioms for bipermutative categories [5, Definition 3.6].

Furthermore, a natural transformation T between two $\mathbb{I}$–spaces X and Y induces a functor $\mathbb{I} \ltimes X \to \mathbb{I} \ltimes Y$. If X and Y are commutative $\mathbb{I}$–monoids ($\mathbb{I}$–rigs) and T is a morphism of such then the induced functor of translation categories is a functor of (bi)permutative categories.

We have thus proved the following result:

Proposition 2.1 *The assignment $X \mapsto \mathbb{I} \ltimes X$ defines a functor from the category of commutative $\mathbb{I}$–monoids to the category of permutative categories, and the assignment $X \mapsto \mathbb{P} \ltimes X$ defines a functor from the category of commutative $\mathbb{I}$–monoids ($\mathbb{I}$–rigs) to the category of (bi)permutative categories.*

2.4 Construction of two infinite loop spaces

Let X be a commutative $\mathbb{I}$–monoid. As explained in [12], the classifying space of a permutative category is an E_∞–space structured by an action of the Barratt–Eccles operad. We have proved the next theorem.

Theorem 2.2 *Suppose that $X\colon \mathbb{I} \to \mathrm{Top}$ is a commutative $\mathbb{I}$–monoid. Then the homotopy colimit*

$$\operatorname{hocolim}_{\mathbb{I}} X = B(\mathbb{I} \ltimes X)$$

is an E_∞–space.

Without further assumptions on X, this E_∞–space need not be grouplike (ie the monoid $\pi_0(\mathrm{hocolim}_{\mathbb{I}} X)$ need not be a group). However, we can always form the group completion $\Omega B(\mathrm{hocolim}_{\mathbb{I}} X)$ to get the associated infinite loop space. Note that an algebra over the Barratt–Eccles operad has an underlying monoid structure that is always strictly associative (and homotopy commutative) so that the usual functorial construction of the classifying space for monoids built using the bar construction can be applied. We will always use this model for B in defining the group completion functor $\Omega B(-)$. The consistency results in [12] guarantee that the group completion $\Omega B(\mathrm{hocolim}_{\mathbb{I}} X)$ defines an infinite loop space weakly equivalent to that obtained using any other delooping machine.

Schlichtkrull [19] defined a different infinite loop space associated to X, using the language of Γ–spaces. Schlichtkrull's construction is the same as May's construction [14] of a Γ–space applied to the permutative category $\mathbb{I} \ltimes X$. By the uniqueness result of [14], the infinite loop space $\Omega B(\mathrm{hocolim}_{\mathbb{I}} X)$ is equivalent to that defined by Schlichtkrull.

We now give a different construction of an infinite loop space associated to X. To start note the decomposition of categories

$$\mathbb{P} \ltimes X = \coprod_{n \geq 0} \Sigma_n \ltimes X(\boldsymbol{n}),$$

where Σ_n is seen as a category with one object. Thus $\mathbb{P} \ltimes X$ is a topological category with classifying space

$$M := \mathrm{hocolim}_{\mathbb{P}} X = B(\mathbb{P} \ltimes X) \simeq \coprod_{n \geq 0} E\Sigma_n \times_{\Sigma_n} X(\boldsymbol{n}).$$

As $\mathbb{P} \ltimes X$ is a permutative category, $M = B(\mathbb{P} \ltimes X)$ is an E_∞–space and thus its group completion, ΩBM, is an infinite loop space. The reduction maps $X(\boldsymbol{n}) \to *$ define a map of permutative categories $\mathbb{P} \ltimes X \to \mathbb{P} \ltimes *$ and hence a map of infinite loop spaces

$$\rho^X \colon \Omega B(\mathrm{hocolim}_{\mathbb{P}} X) \to \Omega B(\mathrm{hocolim}_{\mathbb{P}} *).$$

In particular, the homotopy fiber $\mathrm{hofib}\, \rho^X$ is naturally an infinite loop space.

When X is a commutative $\mathbb{I}$–rig, we process the associated bipermutative category $\mathbb{P} \ltimes X$ using the machinery of Elmendorf and Mandell. To a bipermutative category $\mathcal{C}$, they functorially associate a commutative symmetric ring spectrum [5, Corollary 3.9 and Theorem 9.3.8]. By [5, Theorem 4.6] and the original work of Segal [22], its underlying infinite loop space is weak homotopy equivalent to $\Omega BB\mathcal{C}$. By a theorem due to Schwede [21] and later refined by Mandell and May [10, Section 1], the homotopy category of commutative symmetric ring spectra is equivalent to that of E_∞–ring spectra. We write $K\mathcal{C}$ for the E_∞–ring spectrum associated to $\mathcal{C}$ under this equivalence of homotopy categories. The underlying infinite loop space of an E_∞–ring spectrum is an E_∞–ring space, as defined in [13, Chapter VI], so we may functorially associate to each bipermutative category an E_∞–ring space $\Omega^\infty K\mathcal{C}$. Moreover, by [9, Theorem 1.2], the space $\Omega^\infty K\mathcal{C}$ is weak homotopy equivalent to the group completion $\Omega BB\mathcal{C}$.

We now apply this machinery to the morphism $\mathbb{P} \ltimes X \to \mathbb{P} \ltimes *$ of bipermutative categories. We obtain a map of E_∞–ring spectra

$$K(\mathbb{P} \ltimes X) \to K(\mathbb{P} \ltimes *)$$

which is equivalent to ρ^X after applying Ω^∞. The homotopy fiber of a map of E_∞–ring spectra is a nonunital E_∞–ring spectrum. By a nonunital E_∞–ring space, we mean the underlying infinite loop space of a nonunital E_∞–ring spectrum. Since Ω^∞ preserves homotopy fiber sequences, this means that the homotopy fiber of a map of E_∞–ring spaces is a nonunital E_∞–ring space. We have proved the next theorem.

Theorem 2.3 *For any commutative $\mathbb{I}$–monoid X the homotopy fiber* hofib ρ^X *of*

$$\rho^X\colon \Omega B(\mathrm{hocolim}_{\mathbb{P}}\, X) \to \Omega B(\mathrm{hocolim}_{\mathbb{P}}\, *).$$

is an infinite loop space. If furthermore X is a commutative $\mathbb{I}$–rig, then hofib ρ^X *is a nonunital E_∞–ring space.*

2.5 The main example

For any group G, conjugation by G or action by any other automorphism of G induces a well-defined action on $B_n(q,G) = \mathrm{Hom}(F_n/\,\Gamma_n^q, G)$ by postcomposition. The action is also compatible with the simplicial face and degeneracy maps in the bar construction and hence induces an action on $B(q,G)$.

For every $q \geq 2$ we define an $\mathbb{I}$–space $B(q,U(-))$ by setting $\boldsymbol{n} \mapsto B(q,U(n))$ with morphisms induced by the natural inclusions and the action of Σ_n on $B(q,U(n))$ given by conjugation through permutation matrices. Being induced by the natural action of Σ_n on $\boldsymbol{n}$, it can be checked that this compatible sequence defines indeed an $\mathbb{I}$–space.

We give $B(q,U(-))$ the structure of an $\mathbb{I}$–monoid by defining the unit map $\eta_{\boldsymbol{n}}\colon * \to B(q,U(n))$ to be the inclusion of the base-point and defining the monoid structure map

$$\mu_{\boldsymbol{n},\boldsymbol{m}}\colon B(q,U(n)) \times B(q,U(m)) \to B(q,U(n+m))$$

to be induced by the block sum of matrices. To see that $\mu_{\boldsymbol{n},\boldsymbol{m}}$ is well-defined note that block sum defines a group homomorphism $U(n)\times U(m)\to U(n+m)$. When taking elements of the symmetric groups to permutation matrices, the disjoint union of sets corresponds to block sum of matrices. Thus μ defines a natural transformation of functors defined on $\mathbb{I}\times\mathbb{I}$. One checks compatibility with τ and hence $B(q,U(-))$ is a commutative $\mathbb{I}$–monoid.

Next we note that tensor product of matrices induces a well-defined map

$$\pi_{\boldsymbol{n},\boldsymbol{m}}\colon B(q,U(n))\times B(q,U(m))\to B(q,U(nm)).$$

To see this note that tensor product commutes with matrix multiplication and hence induces a homomorphism $U(n)\times U(m)\to U(nm)$. The map is equivariant for the symmetric group actions because the permutation matrix associated to the product of two permutations is the same as the tensor product of the corresponding permutation matrices. Hence π is a natural transformation of functors defined on the category $\mathbb{P}\times\mathbb{P}$. Note, however, that π is not natural for proper injections. The map π is compatible with τ and the distributivity of block sum and tensor product of matrices induces distributivity maps for μ and π. We have shown:

Theorem 2.4 *$B(q,U(-))$ is a commutative $\mathbb{I}$–rig.*

As a consequence, we may apply Theorems 2.2 and 2.3 to get a pair of infinite loop spaces, the latter of which carries a nonunital E_∞–ring structure. In the next section, we will show that these two infinite loop spaces are equivalent.

3 Identifying and comparing the infinite loop spaces

Let X be a commutative $\mathbb{I}$–monoid. We will first identify $\operatorname{hofib}\rho^X$ under certain assumptions and then show it is homotopy equivalent as an infinite loop space to $\operatorname{hocolim}_{\mathbb{I}} X$.

Consider the space

$$X_\infty := \operatorname{hocolim}_{n\in\mathbb{N}} X(\boldsymbol{n}).$$

Note that $X_\infty\simeq\operatorname{colim}_{n\in\mathbb{N}} X(\boldsymbol{n})$ if the structural maps $j_n\colon X(\boldsymbol{n})\to X(\boldsymbol{n}+\mathbf{1})$ are cofibrations. In our applications this will always be the case. Let X_∞^+ denote Quillen's plus construction applied with respect to the maximal perfect subgroup of $\pi_1(X_\infty)$ (which we take to be understood to be done in each component separately, if X_∞ is not connected). Also recall that a space Z is abelian if $\pi_1(Z)$ is abelian and acts trivially on homotopy groups $\pi_*(Z)$. It is well known that H–spaces are abelian.

Theorem 3.1 *Let* $X\colon \mathbb{I} \to \mathrm{Top}$ *be a commutative* $\mathbb{I}$*–monoid. Assume that*

- *the action of* Σ_∞ *on* $H_*(X_\infty)$ *is trivial;*
- *the inclusions induce natural isomorphisms* $\pi_0(X(\boldsymbol{n})) \simeq \pi_0(X_\infty)$ *of finitely generated abelian groups with multiplication compatible with the Pontrjagin product and in the center of the homology Pontrjagin ring;*
- *the commutator subgroup of* $\pi_1(X_\infty)$ *is perfect (for each component) and* X_∞^+ *is abelian.*

Then $\operatorname{hofib}\rho^X \simeq X_\infty^+$ *and, in particular,* X_∞^+ *is an infinite loop space.*

Proof Let $M = \operatorname{hocolim}_{\mathbb{P}} X = B(\mathbb{P} \ltimes X)$ and m be the point corresponding to the base point in $X(\mathbf{1})$ (in the identity component of $\pi_0(X(\mathbf{1}))$). Then

$$\mathrm{Tel}(M \xrightarrow{\cdot m} M \xrightarrow{\cdot m} M \xrightarrow{\cdot m} \cdots) \simeq \mathbb{Z} \times (E\Sigma_\infty \times_{\Sigma_\infty} X_\infty).$$

As $\mathbb{P} \ltimes X$ is a symmetric monoidal category, its classifying space M is a homotopy commutative topological monoid. The hypotheses imply that $\pi_0(M)$ is in the center of $H_*(M)$. Hence $H_*(M)[\pi_0(M)^{-1}]$ can be constructed by right fractions, so that we may apply the group completion theorem [15; 17]. Therefore there is a map

$$f\colon \mathbb{Z} \times (E\Sigma_\infty \times_{\Sigma_\infty} X_\infty) \to \Omega BM$$

which induces an isomorphism on homology with all systems of local coefficients on ΩBM. Furthermore, the fundamental group (of each component) of $E\Sigma_\infty \times_{\Sigma_\infty} X_\infty$ has a perfect commutator subgroup by [17], and f extends to a homology equivalence between abelian spaces

$$f^+\colon \mathbb{Z} \times (E\Sigma_\infty \times_{\Sigma_\infty} X_\infty)^+ \to \Omega BM,$$

which is thus a homotopy equivalence. This shows, in particular, that the space $\mathbb{Z} \times (E\Sigma_\infty \times_{\Sigma_\infty} X_\infty)^+$ is an infinite loop space as ΩBM is the group completion of an E_∞–space.

Consider now the fibration sequence

$$X_\infty \to E\Sigma_\infty \times_{\Sigma_\infty} X_\infty \xrightarrow{p} B\Sigma_\infty \tag{1}$$

and the associated map of plus constructions

$$p^+\colon \mathbb{Z} \times (E\Sigma_\infty \times_{\Sigma_\infty} X_\infty)^+ \to \mathbb{Z} \times B\Sigma_\infty^+.$$

Since f^+ is a homotopy equivalence and $\Omega B(\operatorname{hocolim}_{\mathbb{P}} *) \simeq \mathbb{Z} \times B\Sigma_\infty^+$, we can identify the homotopy fiber of p^+ with $\operatorname{hofib}\rho^X$. By assumption the action of Σ_∞ on X_∞ is homologically trivial. We are also assuming that X_∞^+ is abelian and in particular nilpotent. Under these conditions the fiber sequence (1) remains a fiber sequence after passing to plus constructions; see [4, Theorem 1.1]. Thus we have a homotopy fibration

$$X_\infty^+ \to \mathbb{Z} \times (E\Sigma_\infty \times_{\Sigma_\infty} X_\infty)^+ \to \mathbb{Z} \times B\Sigma_\infty^+.$$

This shows that the homotopy fiber of p^+ is X_∞^+ and so $X_\infty^+ \simeq \operatorname{hofib} \rho^X$. □

Remark 3.2 For any commutative $\mathbb{I}$–monoid X, the multiplication on $M_X := \coprod_{n \geq 0} X(\boldsymbol{n})$ is commutative up to the action of the shuffle maps $\tau_{\boldsymbol{m},\boldsymbol{n}}$. These are induced by the action of the symmetric group. So, assuming that these actions are trivial in homology, it follows that the Pontrjagin product is commutative on the level of homology. In particular $\pi_0(M_X)$ is in the center of the Pontrjagin ring $H_*(M_X)$. Thus by the group completion theorem [15], the map

$$\mathbb{Z} \times X_\infty \to \Omega B(M_X)$$

is a homology isomorphism. In recent work, Gritschacher [7] has shown that without any further assumption, the commutator subgroup of $\pi_1(X_\infty)$ is always perfect and that X_∞^+ is always an abelian space. In other words, the assumptions in Theorem 3.1 on $\pi_1(X_\infty)$ and X_∞^+ are actually consequences.[2]

In contrast, the condition that the symmetric groups act homologically trivially is necessary. To see this consider the commutative $\mathbb{I}$–space X with $X(\boldsymbol{n}) := Z^n$ for some pointed connected space Z. Then, by the parametrized version of the Barratt–Priddy–Quillen theorem (see for example [12; 22]),

$$\Omega B(\operatorname{hocolim}_{\mathbb{P}} X) \simeq Q(Z_+)$$

and thus $\operatorname{hofib} \rho^X \simeq \operatorname{hofib} p^+ \simeq Q(Z)$ while $X_\infty \simeq \operatorname{hocolim}_n Z^n$. Here $Q = \Omega^\infty \Sigma^\infty$ and Z_+ denotes the space Z with an additional base point.

We now turn to the question of comparing the infinite loop spaces $\operatorname{hofib} \rho^X$ and $\operatorname{hocolim}_{\mathbb{I}} X$. Suppose that X is a commutative $\mathbb{I}$–monoid. Consider the following commutative diagram of strict functors between permutative categories:

$$\begin{tikzcd}
\mathbb{P} \ltimes X \arrow[r, "\alpha_X"] \arrow[d, "\rho^X"'] & \mathbb{I} \ltimes X \arrow[d, "\rho_1^X"] \\
\mathbb{P} \ltimes * \arrow[r, "\alpha_*"] & \mathbb{I} \ltimes *
\end{tikzcd}$$

The horizontal maps are induced by the inclusion $\mathbb{P} \to \mathbb{I}$. In the above diagram $*$ is the terminal commutative $\mathbb{I}$–monoid and the vertical maps ρ^X and ρ_1^X are induced by

[2] As we do not know whether M_X is homotopy commutative, the results of [17] cannot be applied directly to conclude that the induced map $\mathbb{Z} \times X_\infty^+ \to \Omega B(M_X)$ is a homotopy equivalence.

the projection maps to a point. Passing to the level of classifying spaces and applying group completion we obtain a commutative diagram of infinite loop spaces:

$$\begin{tikzcd} \Omega B(\mathrm{hocolim}_{\mathbb{P}} X) \arrow[r, "\alpha_X"] \arrow[d, "\rho^X"'] & \Omega B(\mathrm{hocolim}_{\mathbb{I}} X) \arrow[d, "\rho_1^X"] \\ \Omega B(\mathrm{hocolim}_{\mathbb{P}} *) \arrow[r, "\alpha_*"] & \Omega B(\mathrm{hocolim}_{\mathbb{I}} *) \simeq * \end{tikzcd} \tag{2}$$

Note that the empty set is an initial object for $\mathbb{I}$ and hence $\mathrm{hocolim}_{\mathbb{I}} * = B\mathbb{I} \simeq *$.

The above diagram induces an infinite loop map between the homotopy fibers of the maps ρ^X and ρ_1^X. By definition the homotopy fiber on the left is the space $\mathrm{hofib}\,\rho^X$. Also, since $\mathrm{hocolim}_{\mathbb{I}} *$ is contractible, the homotopy fiber on the right can be identified with $\Omega B(\mathrm{hocolim}_{\mathbb{I}} X)$. This shows that we have a map of infinite loop spaces

$$\mathrm{hofib}\,\rho^X \xrightarrow{g} \Omega B(\mathrm{hocolim}_{\mathbb{I}} X).$$

Note that ρ^X has a canonical splitting of permutative categories induced by the unit $* \to X$ of the $\mathbb{I}$–monoid X. Thus it follows from the following theorem that g is a homotopy equivalence whenever the stated conditions on X are satisfied.

Theorem 3.3 *Let X be a commutative $\mathbb{I}$–monoid such that all maps $j\colon X(\boldsymbol{n}) \to X(\boldsymbol{m})$ induced by injections $j\colon \boldsymbol{n} \to \boldsymbol{m}$ are monomorphisms. Furthermore, assume that, for all $x \in X(\boldsymbol{n})$ and $y \in X(\boldsymbol{m})$, the sum $\mu_{\boldsymbol{n},\boldsymbol{m}}(x, y)$ is in the image of a map induced by a nonidentity order preserving injection if and only if x or y is. Then*

$$\alpha_X \times \rho^X : \Omega B(\mathrm{hocolim}_{\mathbb{P}} X) \to \Omega B(\mathrm{hocolim}_{\mathbb{I}} X) \times \Omega B(\mathrm{hocolim}_{\mathbb{P}} *)$$

is a weak homotopy equivalence of infinite loop spaces which is natural for commutative $\mathbb{I}$–monoids.

Notice that, when X is a commutative $\mathbb{I}$–rig, we may use the theorem to transfer the nonunital E_∞–ring space structure on $\mathrm{hofib}\,\rho^X$ along g to obtain a nonunital E_∞–ring space structure on the group completion of $\mathrm{hocolim}_{\mathbb{I}} X$.

A version of the theorem was proved by Fiedorowicz and Ogle [6] in the setting of simplicial sets. This was revisited in Gritschacher [7, Section 4]. For convenience of the reader we sketch a streamlined argument following [7].

Proof Given $x \in X(\boldsymbol{n})$ we can write it as $x = j_x(\bar{x})$, where $\bar{x} \in X(\bar{\boldsymbol{n}})$, $j_x\colon \bar{\boldsymbol{n}} \to \boldsymbol{n}$ is an order-preserving injection and $\bar{n}$ is minimal. We call x reduced if $x = \bar{x}$. Note that $\bar{x}$ and j_x are uniquely determined. Denote by $\overline{X}(\boldsymbol{n})$ the set of reduced elements in $X(\boldsymbol{n})$. The assignment $\boldsymbol{n} \mapsto \overline{X}(\boldsymbol{n})$ defines a $\mathbb{P}$–diagram. By the assumption on μ the commutative $\mathbb{I}$–monoid structure of X induces the structure of a permutative category on $\mathbb{P} \ltimes \overline{X}$.

Assume now that X is discrete. Then the assignment $(\boldsymbol{n}, x) \mapsto (\overline{\boldsymbol{n}}, \overline{x})$ on objects extends to define a functor

$$R_X \colon \mathbb{I} \ltimes X \to \mathbb{P} \ltimes \overline{X}.$$

It has a right inverse given by the inclusion $\iota_X \colon \mathbb{P} \ltimes \overline{X} \to \mathbb{I} \ltimes X$. Furthermore, the maps j_x define a natural transformation from $\iota_X \circ R_X$ to the identity on $\mathbb{I} \ltimes X$. Hence, R_X defines a homotopy deformation retract on classifying spaces. We also note that by our assumption on μ, the functor R_X is a strict symmetric monoidal functor.

The inclusions $\mathbb{P} \ltimes \overline{X} \to \mathbb{P} \ltimes X$ and $\mathbb{P} \to \mathbb{P} \ltimes X$ combine via the monoidal product functor to a functor

$$T_X \colon (\mathbb{P} \ltimes \overline{X}) \times \mathbb{P} \to \mathbb{P} \ltimes X$$

that maps the object $((\overline{\boldsymbol{n}}, \overline{x}), \boldsymbol{n})$ to $(\overline{\boldsymbol{n}} + \boldsymbol{n}, j(\overline{x}))$, where j is the canonical inclusion $\overline{\boldsymbol{n}} \hookrightarrow \overline{\boldsymbol{n}} + \boldsymbol{n}$. We claim this is a homotopy equivalence on classifying spaces. Indeed, an analysis of the effect of permutations on reduced points shows that the functor is bijective on automorphism groups of objects. As both source and target categories are groupoids and every isomorphism class of the target category has a representative in the image, this is an equivalence of categories. We note that T_X is not a strict monoidal functor (only up to conjugation by a block permutation). However, the left inverse functor $(\boldsymbol{n}, x) \mapsto ((\overline{\boldsymbol{n}}, \overline{x}), \boldsymbol{n} - \overline{\boldsymbol{n}})$ does commute strictly with the monoidal structure. Hence, this defines a homotopy equivalence of monoids on classifying spaces, and induces a homotopy equivalence of group completions. Compare [6, Lemma 1.7].

Consider now the map of permutative categories

$$\alpha_X \times \rho^X \colon \mathbb{P} \ltimes X \to (\mathbb{I} \ltimes X) \times \mathbb{P}$$

and take the group completion of their classifying spaces

$$\alpha_X \times \rho^X \colon \Omega B(B(\mathbb{P} \ltimes X)) \to \Omega B(B(\mathbb{I} \ltimes X)) \times \Omega B(B\mathbb{P}). \tag{3}$$

We claim that this is a weak homotopy equivalence which is natural in commutative $\mathbb{I}$–monoids. To see this precompose with the map of group completed classifying spaces induced by T_X and postcompose with the map induced by $R_X \times \mathrm{Id}$. The resulting composite is homotopic to the endofunctor of $(\mathbb{P} \ltimes \overline{X}) \times \mathbb{P}$ given by

$$((\overline{\boldsymbol{n}}, \overline{x}), \boldsymbol{m}) \mapsto ((\overline{\boldsymbol{n}}, \overline{x}), \overline{\boldsymbol{n}} + \boldsymbol{m}).$$

This map is the identity on the first component and an equivalence on the second component because we are working with group-complete monoids.

Using the naturality of the weak homotopy equivalence in (3) and applying it to boundary and face maps allows us to extend it to $\mathbb{I}$–diagrams in simplicial sets. More precisely, for any commutative $\mathbb{I}$–monoid X in simplicial sets that satisfies levelwise the condition on μ, we have a map of simplicial permutative categories which is a weak homotopy equivalence on applying $\Omega B(B(-))$ to each simplicial level, and hence a weak homotopy equivalence on total spaces:

$$\alpha_X \times \rho^X : |\boldsymbol{n} \mapsto \Omega B(B(\mathbb{P} \ltimes X(\boldsymbol{n})))| \simeq |\boldsymbol{n} \mapsto \Omega B(B(\mathbb{I} \ltimes X(\boldsymbol{n}))) \times \Omega B(B\mathbb{P})|.$$

As Ω commutes with Cartesian product, and as $|\boldsymbol{n} \mapsto \Omega Z(\boldsymbol{n})| \simeq \Omega|\boldsymbol{n} \mapsto Z(\boldsymbol{n})|$ whenever each $Z(\boldsymbol{n})$ is connected (see [11, Theorem 12.3]), we also have

$$\alpha_X \times \rho^X : \Omega|\boldsymbol{n} \mapsto B(B(\mathbb{P} \ltimes X(\boldsymbol{n})))| \simeq \Omega|\boldsymbol{n} \mapsto B(B(\mathbb{I} \ltimes X(\boldsymbol{n}))) \times B(B\mathbb{P})|.$$

Furthermore, as realizations of multisimplicial sets can be taken in any order, we deduce that

$$\alpha_X \times \rho^X \colon \Omega B\big(B(\mathbb{P} \ltimes |\boldsymbol{n} \mapsto X(\boldsymbol{n})|)\big) \simeq \Omega B\big(B(\mathbb{I} \ltimes |\boldsymbol{n} \mapsto X(\boldsymbol{n})|)\big) \times \Omega B(B\mathbb{P}).$$

Compare [6, Lemma 1.8]. Finally, by replacing every space by its singular simplicial set, any $\mathbb{I}$–diagram X in topological spaces gives rise to an $\mathbb{I}$–diagram in simplicial sets, taking commutative $\mathbb{I}$–monoids to simplicial ones. Note that the conditions on μ are pointwise conditions and are automatically satisfied by the singular p–simplices for each p. As a space is weakly homotopy equivalent to the realization of its singular simplicial set, the theorem follows. □

Example 3.4 Consider the commutative $\mathbb{I}$–space X with $X(\boldsymbol{n}) := Z^n$, where Z is a well-pointed connected space. Note that in this case Σ_n does not act trivially on $H_*(Z^n)$ and hence Theorem 3.1 does not apply. As before, by the parametrized version of the Barratt–Priddy–Quillen theorem,

$$\Omega B(\operatorname{hocolim}_{\mathbb{P}} X) \simeq Q(Z_+) \simeq Q(\mathbb{S}^0) \times Q(Z)$$

and hence $\operatorname{hofib} \rho^X \simeq Q(Z)$. Thus, by Theorem 3.3 we also have $\operatorname{hocolim}_{\mathbb{I}} X \simeq Q(Z)$, which is in agreement with a result of Schlichtkrull [20].

4 Constructing filtrations by infinite loop spaces

In this section we use the results obtained in the previous sections to produce filtrations of classical infinite loop spaces by sequences of infinite loop spaces arising from the descending central series of the free groups.

Theorem 4.1 *The spaces $B(q,U)$, $B(q,\mathrm{SU})$, $B(q,\mathrm{SO})$ $B(q,O)$ and $B(q,\mathrm{Sp})$ provide a filtration by nonunital E_∞–ring spaces of the classical nonunital E_∞–ring spaces BU, $B\mathrm{SU}$, $B\mathrm{SO}$, BO and $B\mathrm{Sp}$, respectively.*

Proof Consider first the case of BU. Recall that the spaces $B(q,U)$ provide a filtration of the space BU

$$B(2,U) \subset B(3,U) \subset \cdots \subset B(q,U) \subset B(q+1,U) \subset \cdots \subset BU.$$

We will show that this filtration is a filtration by nonunital E_∞–ring spaces. For this notice that by the main example in Section 2, each $\boldsymbol{n} \mapsto B(q,U(n))$ for $q \geq 2$ is a commutative $\mathbb{I}$–rig. In what follows we are going to show that the conditions of Theorem 3.1 are satisfied, and hence $B(q,U) \simeq \operatorname{hofib} \rho^{B(q,U(-))}$ is a nonunital E_∞–ring space by Theorem 2.3.

The conjugation action of Σ_n on $B(q,U(n))$ is homologically trivial because this action factors through the conjugation action of $U(n)$. The conjugation action by any element in $U(n)$ is trivial, up to homotopy, since the action of the identity matrix is trivial and $U(n)$ is path-connected. This implies in particular that the action of Σ_∞ on $B(q,U)$ is homologically trivial.

Note that $B(q,U(n))$ and hence $B(q,U)$ is path connected. Next, we argue that the space $B(q,U)$ is an H–space under direct sum multiplication. To be more precise, consider the injection $\mathbb{N} \sqcup \mathbb{N} \to \mathbb{N}$ defined by $(1,2,3,4,\dots) \cup (1',2',3',4',\dots) \mapsto (1,2,1',2',3,4,3',4',\dots)$. It defines a map of vector spaces $\mathbb{C}^\infty \times \mathbb{C}^\infty \to \mathbb{C}^\infty$ and hence a continuous homomorphisms $U \times U \to U$. The image of $U(n)$ in U under right or left multiplication by the identity matrix I differs from the image under the standard inclusion by conjugation of an even permutation. As such a permutation is in the path-component of the identity matrix, we see that the multiplication is unital up to homotopy.

H–spaces have abelian fundamental group and hence Theorem 3.1 applies. We conclude that $B(q,U) \simeq \operatorname{hofib} \rho^{B(q,U(-))}$ for every $q \geq 2$ and is a nonunital E_∞–ring space by Theorem 2.3. The very same arguments can be used to prove analogous statements for the commutative $\mathbb{I}$–rig $\boldsymbol{n} \mapsto B(q,\mathrm{SU}(n))$, and $\boldsymbol{n} \mapsto B(q,\mathrm{Sp}(n))$ for any $q \geq 2$.

In case of the commutative $\mathbb{I}$–rig $\boldsymbol{n} \mapsto B(q,\mathrm{SO}(n))$ we note that Σ_n is not a subgroup of $\mathrm{SO}(n)$. Nevertheless, the alternating group A_n is contained in $\mathrm{SO}(n)$ and by the same argument as above acts therefore trivially on the homology of $B(q,\mathrm{SO}(n))$. Furthermore, when n is odd, any odd permutation is represented by a matrix with determinant equal to -1. Hence it can be path-connected to the diagonal matrix $-I$ with constant entry -1. As $-I$ is in the center of $O(n)$ it acts trivially by conjugation on $B(q,\mathrm{SO}(n))$ and hence also on its homology. But then so does any odd permutation. This proves that when n is odd the action of Σ_n on $B(q,\mathrm{SO}(n))$ is homologically trivial. This in turn implies that the action of Σ_∞ on $B(q,\mathrm{SO})$ is homologically trivial. We also have that $B(q,\mathrm{SO})$ is an H–space and hence abelian. Thus $B(q,\mathrm{SO}) \simeq \mathrm{hofib}\,\rho^{B(q,\mathrm{SO}(-))}$ for every $q \geq 2$ and it is a nonunital E_∞–ring space by Theorem 2.3. This line of argument can also be used to prove the analogous statement for the commutative $\mathbb{I}$–rig $\boldsymbol{n} \mapsto B(q, O(n))$. □

As remarked in [1, Theorem 6.3], the natural map $\Omega B(q,G) \to \Omega BG$ admits a splitting up to homotopy. It is given by a factorization of the usual homotopy equivalence $G \to \Omega BG$. Indeed we have that $\Sigma G = F_1 B(q,G) = F_1 BG$, where F_1 denotes the first layer in the usual filtration of the geometric realization of these simplicial spaces. Hence, the adjoint of $\Sigma G \to BG$ factors through $\Omega B(q,G)$. Note that this splitting does not in general admit a delooping; see [1, Section 6] for a counterexample. Nevertheless, we have the following theorem. Here $E(q,G)$ denotes the pull-back of the universal G–bundle EG over BG. It is homotopy equivalent to the homotopy fiber of the inclusion $B(q,G) \to BG$.

Theorem 4.2 *For all $q \geq 2$, and $G = U$, SU, SO, O and Sp, there is a homotopy split fibration of infinite loop spaces*

$$E(q,G) \to B(q,G) \to BG.$$

In particular there is a splitting of spaces

$$B(q,G) \simeq BG \times E(q,G).$$

Both are natural in the entry q, meaning that both are compatible with the filtration maps.

In order to prove the theorem, we will need to know the fundamental group of $B(q,G)$ for the groups in question. We have the following general result:

Lemma 4.3 *Let G be a topological group with a CW–structure. Assume $\pi_0(G)$ is abelian and that the natural homomorphism $G \to \pi_0(G)$ splits. Then, for all $q \geq 2$,*

$$\pi_1(B(q,G)) = \pi_0(G).$$

Proof Consider $\Sigma G = F_1 B(q,G) = F_1 BG$. As the 1–skeleton of the realization of a (good) simplicial space is contained in the first filtration [11, Proposition 11.4], any map from S^1 to $B(q,G)$ will factor through ΣG. Hence the map $\Sigma G \to B(q,G)$ is surjective on fundamental groups.

The fundamental group of a suspension ΣX for any space X has fundamental group the free group over the set $\pi_0(X) - \{1\}$; hence we have

$$\pi_1(\Sigma G) = F\big(g \mid g \in \pi_0(G) - \{1\}\big).$$

The inclusion $\Sigma G \to BG$ induces the surjective map of fundamental groups $\pi_1(\Sigma G) \to \pi_0(G)$ which sends a generator g to the *element* $g \in \pi_0(G)$ and, more generally, the word $g_1 \bullet \cdots \bullet g_k$ to the product of the elements $g_1 \cdots g_k$. To see this geometrically, consider $\pi_0(G)$ as a subgroup of G, and note that the 2–simplex (g, h) defines a homotopy from the 2–letter word $g \bullet h$ to the product element gh.

We now note that, as $\pi_0(G)$ is abelian, the 2–simplex (g, h) is contained in $B_2(q, G)$ for $q \geq 2$. Hence all the above relations are already satisfied in $\pi_1(B(q, G))$. As the factorization $\pi_1(\Sigma G) \to \pi_1(B(q, G)) \to \pi_1(BG)$ is surjective, the result follows. $\square$

Proof of Theorem 4.2 As $EG_\infty \simeq *$, for every $q \geq 2$ we have a homotopy fibration sequence $E(q, G_\infty) \to B(q, G_\infty) \to BG_\infty$. As the map on the right is a map of infinite loop spaces, the homotopy fiber $E(q, G_\infty)$ is an infinite loop space. It remains to show that it splits.

Let G_n denote one of the groups $U(n)$, $\mathrm{SU}(n)$, $\mathrm{SO}(n)$, $O(n)$ or $\mathrm{Sp}(n)$, so that $G_\infty = \mathrm{colim}_n\, G_n$ denotes the group U, SU, SO, O or Sp, respectively. For each fixed $q \geq 2$, the assignment $\boldsymbol{n} \mapsto \Omega B(q, G_n)$ defines a commutative $\mathbb{I}$–rig with μ given by block sum and π given by tensor product of matrices. In the same way the assignment $\boldsymbol{n} \mapsto \Omega BG_n$ also defines a commutative $\mathbb{I}$–rig and the inclusion map $\Omega B(q, G_n) \to \Omega BG_n$ defines a morphism of commutative $\mathbb{I}$–rigs.

We claim that the commutative $\mathbb{I}$–rigs G_-, $\Omega B(q, G_-)$ and ΩBG_- satisfy the hypotheses of Theorem 3.1. Indeed, except in the case $G = O$, the group $G_n \simeq \Omega BG_n$ is path-connected for every $n \geq 0$ and, as $\pi_0(\Omega B(q, G_n)) \cong \pi_1(B(q, G_n))$ is trivial by Lemma 4.3, $\Omega B(q, G_n)$ is also path-connected. When $G = O$,

$$\pi_0(\Omega B(q, O(n))) = \pi_1 B(q, O(n)) = \mathbb{Z}/2\mathbb{Z}$$

for each $n \geq 1$ by Lemma 4.3. The multiplication in $\pi_0 \Omega B(q, O(n))$ is compatible with direct sum and stabilization. This checks the second condition in Theorem 3.1.

Except in the cases $G = \mathrm{SO}$ or $G = O$, the action of Σ_n is homologically trivial as conjugation by any element in the path component of the identity is trivial, up to homotopy, and G_n is path-connected. This implies that Σ_∞ acts homologically trivially on G_∞, $\Omega B(q, G_\infty)$ and ΩBG_∞. The same conclusion can be obtained for $G = \mathrm{SO}$ or $G = O$ using a similar argument as in the proof of Theorem 4.1. Hence the first condition from Theorem 3.1 holds.

To verify the third condition, observe that the commutator group of $\pi_1(\Omega B(q, G_n)) \cong \pi_2(B(q, G_n))$ is trivial, as this group is abelian in all cases. Finally, $\Omega B(q, G_\infty)$ is an abelian space since it is a loop space and hence in particular an H–space.

By Theorem 3.1 we thus have maps of E_∞–spaces

$$G_\infty \to \Omega B(q, G_\infty) \to \Omega BG_\infty$$

whose composition is a homotopy equivalence. Taking classifying spaces is compatible with E_∞–space structures and hence the above splitting deloops to give the splitting of the theorem. □

We have concentrated so far on compact groups such as $O(n)$ and $U(n)$, although the methods clearly extend to other linear groups. Using some results by Pettet and Souto [16] and Bergeron [3] we can prove the following theorem:

Theorem 4.4 *Suppose that G is the group of complex or real points in a reductive linear algebraic group (defined over $\mathbb{R}$ in the real case). Let $K \subset G$ be a maximal compact subgroup. Then the inclusion map $i\colon B(q, K) \to B(q, G)$ is a homotopy equivalence for every $q \geq 2$.*

Proof By [3, Theorem I] it follows that the inclusion map $i_n\colon B_n(q, K) \to B_n(q, G)$ is a homotopy equivalence for all $q \geq 2$ and all $n \geq 0$. Thus the inclusion map induces a simplicial map $i_*\colon B_*(q, K) \to B_*(q, G)$ that is a levelwise homotopy equivalence. Since G is assumed to be the group of complex or real points in a reductive linear algebraic group (defined over $\mathbb{R}$ in the real case), we can identify G with a Zariski closed subgroup of $\mathrm{SL}_N(\mathbb{C})$ for some $N \geq 0$. Also, for every $n \geq 0$ we can see the space $B_n(q, G)$ as an algebraic variety since it is defined in terms of iterated commutators of elements in G and such equations can be defined in terms of polynomial functions. Moreover, the subspace $S_n^1(q, G) \subset B_n(q, G)$ consisting of all n–tuples in $B_n(q, G)$ for which at least one of the coordinates is equal to 1_G is an algebraic subvariety of $B_n(q, G)$. By the semialgebraic triangulation theorem (see [8, Section 1]) it follows that $B_n(q, G)$ has the structure of a CW–complex in such a way that $S_n^1(q, G)$ is a subcomplex. In particular, it follows that the pair $(B_n(q, G), S_n^1(q, G))$ is a strong NDR pair. This proves that $B_*(q, G)$ is a proper simplicial space. The same is true for $B_*(q, K)$. Using the gluing lemma — for example see [12, Theorem A.4] — we obtain the result of the theorem. □

Our tools can also be used to obtain a similar filtration for the infinite loop space defining algebraic K–theory for any discrete ring R. Indeed, suppose that R is a discrete ring with unit and let $q \geq 2$. Consider the commutative $\mathbb{I}$–rig $B(q, \mathrm{GL}_-(R))$ defined by $\boldsymbol{n} \mapsto B(q, \mathrm{GL}_n(R))$. As before the morphisms are induced by the natural inclusions and the conjugation action of Σ_n on $B(q, \mathrm{GL}_n(R))$. The multiplication map

$$\mu_{\boldsymbol{n},\boldsymbol{m}}\colon B(q, \mathrm{GL}_n(R)) \times B(q, \mathrm{GL}_m(R)) \to B(q, \mathrm{GL}_{n+m}(R))$$

is also given by the block sum and π by tensor product of matrices. Note that Theorem 3.3 applies to give

$$\operatorname{hocolim}_{\mathbb{I}} B(q, \mathrm{GL}_{-}(R)) \simeq \operatorname{hofib} \rho^{B(q,\mathrm{GL}_{-}(R))}.$$

By Theorem 2.3, this space has the structure of a nonunital E_∞–ring space. This way we obtain a filtration of nonunital E_∞–ring spaces:

$$\operatorname{hocolim}_{\mathbb{I}} B(2, \mathrm{GL}_{-}(R)) \subset \cdots \subset \operatorname{hocolim}_{\mathbb{I}} B(q, \mathrm{GL}_{-}(R)) \subset \cdots \subset \operatorname{hocolim}_{\mathbb{I}} B\mathrm{GL}_{-}(R).$$

As is well known, the conjugation action of Σ_n on $B\mathrm{GL}_n(R)$ is homologically trivial. It follows from Theorems 3.1 and 3.3 that we have an equivalence

$$B\mathrm{GL}_\infty(R)^+ \simeq \operatorname{hofib} \rho^{B\mathrm{GL}_{-}(R)} \simeq \operatorname{hocolim}_{\mathbb{I}} B\mathrm{GL}_{-}(R).$$

Thus the above gives a filtration of nonunital E_∞–ring spaces with final space weakly homotopy equivalent to the algebraic K–theory of R. However, unlike the case of $B\mathrm{GL}_n(R)$, we do not know whether the conjugation action of Σ_n on $B(q, \mathrm{GL}_n(R))$ is homologically trivial, and we expect that the natural map

$$B(q, \mathrm{GL}_\infty(R)) \to \operatorname{hocolim}_{\mathbb{I}} B(q, \mathrm{GL}_{-}(R))$$

is not a homology isomorphism.

In a similar way we can obtain a filtration of $Q(\mathbb{S}^0)$. For this note that the conjugation action of Σ_n on $B\Sigma_n$ is homologically trivial. Therefore, by the Barratt–Priddy–Quillen theorem, the level zero component of $Q(\mathbb{S}^0)$ is equivalent to the homotopy colimit over $\mathbb{I}$ of the classifying spaces of the symmetric groups:

$$Q_0(\mathbb{S}^0) \simeq (B\Sigma_\infty)^+ \simeq \operatorname{hofib} \rho^{B\Sigma_{-}} \simeq \operatorname{hocolim}_{\mathbb{I}} B\Sigma_{-}.$$

Consider the commutative $\mathbb{I}$–rig $B(q, \Sigma_{-})$ defined by $\boldsymbol{n} \mapsto B(q, \Sigma_n)$. The structural maps are given by conjugation of Σ_n and inclusions in an analogous way as above. Then by Theorem 2.2 we have a filtration of nonunital E_∞–ring spaces

$$\operatorname{hocolim}_{\mathbb{I}} B(2, \Sigma_{-}) \subset \cdots \subset \operatorname{hocolim}_{\mathbb{I}} B(q, \Sigma_{-}) \subset \cdots \subset \operatorname{hocolim}_{\mathbb{I}} B\Sigma_{-} \simeq Q_0(\mathbb{S}^0).$$

As in the case of $B(q, \mathrm{GL}_n(R))$, the conjugation action of Σ_n on $B(q, \Sigma_n)$ may fail to be homologically trivial (for example this is the case for the conjugation action of Σ_3 on $B(2, \Sigma_3)$; see [1]). The conditions of Theorem 3.3 are satisfied but the homotopy types of the spaces $\operatorname{hocolim}_{\mathbb{I}} B(q, \Sigma_{-}) \simeq \operatorname{hofib} \rho^{B(q,\Sigma_{-})}$ remain to be determined.

Corollary 4.5 *The spaces*

$$\operatorname{hocolim}_{\mathbb{I}} B(q, \mathrm{GL}_-(R)) \simeq \operatorname{hofib} \rho^{B(q,\mathrm{GL}_-(R))},$$
$$\operatorname{hocolim}_{\mathbb{I}} B(q, \Sigma_-) \simeq \operatorname{hofib} \rho^{B(q,\Sigma_-)}$$

provide filtrations of nonunital E_∞–ring spaces with final target the classical nonunital E_∞–ring spaces $B\mathrm{GL}_\infty(R)^+$ and $Q_0(\mathbb{S}^0)$.

5 Transitional nilpotence, bundles and K–theory

In this section we extend the notions of transitionally commutative bundles and commutative K–theory as defined in [2] to more general q–nilpotent notions for $q \geq 2$, reflecting the filtration induced by the descending central series of the free groups. We will show that these geometrically defined theories are represented by the infinite loop spaces $\mathbb{Z} \times B(q, U)$.

Definition 5.1 For a CW–complex X a principal G–bundle $\pi\colon E \to X$ is said to have *transitional nilpotency class* at most q if there exists an open cover $\{U_i\}_{i\in I}$ of X such that the bundle $\pi\colon E \to X$ is trivial over each U_i and for every $x \in X$ the group generated by the collection $\{\rho_{i,j}(x)\}_{i,j}$ is a group of nilpotency class at most q. Here $\rho_{i,j}\colon U_i \cap U_j \to G$ denotes the transition functions, and i and j run through all indices in I for which $x \in U_i \cap U_j$. The minimum of all such numbers q is said to be transitional nilpotency class of $\pi\colon E \to X$.

The principal G–bundle $p_q\colon E(q, G) \to B(q, G)$ is universal for all principal G–bundles with transitional nilpotency class less than q.

Theorem 5.2 *Assume that G is an algebraic subgroup of $\mathrm{GL}_N(\mathbb{C})$ for some $N \geq 0$, X is a finite CW–complex and that $\pi\colon E \to X$ is a principal G–bundle over X. Then, for any $q \geq 2$, the classifying map $f\colon X \to BG$ of π factors through $B(q, G)$ (up to homotopy) if and only if π has transitional nilpotency class less than q.*

Proof The case $q = 2$ was treated in [2, Theorem 2.2] and in fact this theorem is true for any Lie group in this case. The proof goes through verbatim also for $q > 2$ using the fact that when G is an algebraic subgroup of $\mathrm{GL}_N(\mathbb{C})$, then the simplicial space $B_*(q, G)$ is proper, as was pointed out in the proof of Theorem 4.4. □

As $[\Sigma X, BG] = [X, \Omega BG]$ and the canonical map $\Omega B(q, G) \to \Omega BG$ always admits a splitting up to homotopy, any principal G–bundle on a suspension ΣX has transitional nilpotency class less than q for all q. However, the nilpotency structure is not unique in general, not even up to isomorphism in the sense of the following definition:

Definition 5.3 Let $\pi_0\colon E_0 \to X$ and $\pi_1\colon E_1 \to X$ be two principal G–bundles with transitional nilpotency class less than q. We say that these bundles are q–transitionally isomorphic if there exists a principal G–bundle $p\colon E \to X \times [0,1]$ with transitional nilpotency class less than q such that $\pi_0 = p_{|p^{-1}(X\times\{0\})}$ and $\pi_1 = p_{|p^{-1}(X\times\{1\})}$.

A complex vector bundle $\pi\colon E \to X$ is said to have transitional nilpotency class less than q if the corresponding frame bundle, under a fixed Hermitian metric on E, has transitional nilpotency class less than q. Theorem 4.2 can then be interpreted to say that any vector bundle is stably of transitional nilpotency class less than q for all $q \geq 2$, and there is a functorial choice of such a structure. The set $\mathrm{Vect}_{q\text{–nil}}(X)$ of q–transitionally isomorphism classes of complex vector bundles over X with transitional nilpotency class less than q is a monoid under the direct sum of vector bundles. The q–nilpotent K–theory of X is defined as the associated Grothendieck group.

Definition 5.4 $K_{q\text{–nil}}(X) := \mathrm{Gr}(\mathrm{Vect}_{q\text{–nil}}(X))$.

Tensor products induce a natural multiplication on $K_{q\text{–nil}}(X)$ just as in classical K–theory.

Theorem 5.5 *For any finite CW–complex X there is a natural isomorphism of rings*

$$K_{q\text{–nil}}(X) \cong [X, \mathbb{Z} \times B(q,U)].$$

Hence, it is the zeroth term of a multiplicative generalized cohomology theory.

Proof Let X be a finite CW–complex. By working one path-connected component at a time, we may assume without loss of generality that X is path-connected. By Theorem 5.2,

$$\mathrm{Vect}_{q\text{–nil}}(X) = \Big[X, \bigsqcup_{n\geq 0} B(q,U(n))\Big]$$

as abelian monoids, where the addition is induced by direct sum of matrices on the right hand side. Any injection $\mathbb{N}\times\mathbb{N}\to\mathbb{N}$ induces a linear injection $\mathbb{C}^\infty\times\mathbb{C}^\infty\to\mathbb{C}^\infty$, which in turn induces an H–space product on $\mathbb{Z}\times B(q,U)$. The natural inclusions $B(q,U(n)) \to B(q,U)$ define a map

$$\Big[X, \bigsqcup_{n\geq 0} B(q,U(n))\Big] \to [X, \mathbb{Z}\times B(q,U)].$$

As the symmetric groups act by homotopy equivalences on $B(q,U)$, we see that the above map is compatible with the product structure on both sets, ie it is a map of monoids. By the universal property of the Grothendieck construction, this map factors through a unique map of abelian groups

$$K_{q\text{–nil}}(X) \to [X, \mathbb{Z} \times B(q, U)].$$

As X is compact, any map $X \to B(q, U)$ factors through some $B(q, U(n))$ for some large enough n. Hence the above map is surjective.

To prove that it is injective, suppose that the image of $[A] - [B] \in K_{q\text{–nil}}(X)$ in $[X, \mathbb{Z} \times B(q, U)]$ is zero. Let us write $f_B\colon X \to B(q, U)$ for the image of a map representing B in the colimit $B(q, U) = \operatorname{colim}_{n \in \mathbb{N}} B(q, U(n))$. Since $B(q, U)$ is a grouplike H–space, the induced product on $\operatorname{Map}(X, B(q, U))$ is also a grouplike H–space structure. Let $f_{B'}\colon X \to B(q, U)$ be a homotopy inverse for f_B under this product. Since X is compact, we may factor $f_{B'}$ through a finite stage of the colimit and find a corresponding bundle B' over X with transitional nilpotency class less than q which is classified by the map $f_{B'}$. It follows that $B \oplus B'$ is stably q–transitionally isomorphic to the trivial bundle ϵ_k of rank $k = \dim B + \dim B'$. By our assumption, we see that the image of $[A \oplus B'] - [\epsilon_k]$ in $[X, \mathbb{Z} \times B(q, U)]$ is also zero. This means that $A \oplus B'$ is stably q–transitionally isomorphic to a trivial bundle, say $A \oplus B' \oplus \epsilon_t \cong \epsilon_{k+t}$. We then have the relation

$$[A] - [B] = [A \oplus B' \oplus \epsilon_t] - [\epsilon_{k+t}] = 0$$

in $K_{q\text{–nil}}(X)$, which completes the proof. □

This answers the question raised in [2] for $q = 2$. Moreover, we have a sequence of cohomology theories and maps between them,

$$K_{\mathrm{com}}(X) = K_{2\text{–nil}}(X) \to K_{3\text{–nil}}(X) \to \cdots \to K_{q\text{–nil}}(X) \to \cdots \to K(X).$$

By Theorem 4.2, topological K–theory splits off q–nilpotent K–theory for all $q \geq 2$. These theories are not well understood and would seem to warrant further attention. For example in [2] it was shown that $K_{\mathrm{com}}(\mathbb{S}^i) \cong K(\mathbb{S}^i)$ for $0 \leq i \leq 3$, but that $K_{\mathrm{com}}(\mathbb{S}^4) \neq K(\mathbb{S}^4)$.

We leave it to the reader to formulate q–nilpotent versions of real and hermitian K–theory.

References

[1] **A Adem**, **F R Cohen**, **E Torres Giese**, *Commuting elements, simplicial spaces and filtrations of classifying spaces*, Math. Proc. Cambridge Philos. Soc. 152 (2012) 91–114 MR

[2] **A Adem**, **J M Gómez**, *A classifying space for commutativity in Lie groups*, Algebr. Geom. Topol. 15 (2015) 493–535 MR

[3] **M Bergeron**, *The topology of nilpotent representations in reductive groups and their maximal compact subgroups*, Geom. Topol. 19 (2015) 1383–1407 MR

[4] **A J Berrick**, *The plus-construction and fibrations*, Quart. J. Math. Oxford Ser. 33 (1982) 149–157 MR

[5] **A D Elmendorf**, **M A Mandell**, *Rings, modules, and algebras in infinite loop space theory*, Adv. Math. 205 (2006) 163–228 MR

[6] **Z Fiedorowicz**, **C Ogle**, *Algebraic K–theory and configuration spaces*, Topology 29 (1990) 409–418 MR

[7] **S Gritschacher**, *Modules over operads and iterated loop spaces*, Transfer thesis, University of Oxford (2014)

[8] **H Hironaka**, *Triangulations of algebraic sets*, from "Algebraic geometry" (R Hartshorne, editor), Proc. Sympos. Pure Math. 29, Amer. Math. Soc., Providence, RI (1975) 165–185 MR

[9] **J A Lind**, *Diagram spaces, diagram spectra and spectra of units*, Algebr. Geom. Topol. 13 (2013) 1857–1935 MR

[10] **M A Mandell**, **J P May**, *Equivariant orthogonal spectra and S–modules*, Mem. Amer. Math. Soc. 755, Amer. Math. Soc., Providence, RI (2002) MR

[11] **J P May**, *The geometry of iterated loop spaces*, Lectures Notes in Mathematics 271, Springer, Berlin (1972) MR

[12] **J P May**, *E_∞ spaces, group completions, and permutative categories*, from "New developments in topology" (G Segal, editor), London Math. Soc. Lecture Note Ser. 11, Cambridge Univ. Press (1974) 61–93 MR

[13] **J P May**, *E_∞ ring spaces and E_∞ ring spectra*, Lecture Notes in Mathematics 577, Springer, Berlin (1977) MR

[14] **J P May**, *The spectra associated to permutative categories*, Topology 17 (1978) 225–228 MR

[15] **D McDuff**, **G Segal**, *Homology fibrations and the "group-completion" theorem*, Invent. Math. 31 (1975/76) 279–284 MR

[16] **A Pettet**, **J Souto**, *Commuting tuples in reductive groups and their maximal compact subgroups*, Geom. Topol. 17 (2013) 2513–2593 MR

[17] **O Randal-Williams**, *"Group-completion", local coefficient systems and perfection*, Q. J. Math. 64 (2013) 795–803 MR

[18] **S Sagave**, **C Schlichtkrull**, *Diagram spaces and symmetric spectra*, Adv. Math. 231 (2012) 2116–2193 MR

[19] **C Schlichtkrull**, *Units of ring spectra and their traces in algebraic K–theory*, Geom. Topol. 8 (2004) 645–673 MR

[20] **C Schlichtkrull**, *The homotopy infinite symmetric product represents stable homotopy*, Algebr. Geom. Topol. 7 (2007) 1963–1977 MR

[21] **S Schwede**, *S–modules and symmetric spectra*, Math. Ann. 319 (2001) 517–532 MR

[22] **G Segal**, *Categories and cohomology theories*, Topology 13 (1974) 293–312 MR

Department of Mathematics, University of British Columbia
1984 Mathematics Road, Room 121, Vancouver BC V6T 1Z2, Canada

Departmento de Matemáticas, Universidad Nacional de Colombia
Medellín, AA 3840, Colombia

Department of Mathematics, Reed College
3203 SE Woodstock Blvd., Portland, OR 97202, United States

Mathematical Institute, Oxford University
Oxford, OX2 6GG, United Kingdom

adem@math.ubc.ca, jmgomez0@unal.edu.co, john.alexander.lind@gmail.com, tillmann@maths.ox.ac.uk

Received: 2 September 2015 Revised: 16 September 2016

Algebraic & Geometric Topology 17 (2017) 895–915

Stable functorial decompositions of $F(\mathbb{R}^{n+1}, j)^{+} \wedge_{\Sigma_j} X^{(j)}$

JIE WU
ZIHONG YUAN

We first construct a functorial homotopy retract of $\Omega^{n+1}\Sigma^{n+1}X$ for each natural coalgebra-split sub-Hopf algebra of the tensor algebra. Then, by computing their homology, we find a collection of stable functorial homotopy retracts of $F(\mathbb{R}^{n+1}, j)^{+}\wedge_{\Sigma_j} X^{(j)}$.

55P35; 55P48, 55P65

1 Introduction

In the 1970s, Snaith [12] proved iterated loop suspensions of a space can be split stably into simpler pieces. This is called the Snaith splitting. In detail, let X be a path-connected CW–complex, with $X^{(j)}$ the j–fold self smash product of X. Let $F(\mathbb{R}^{n+1}, j)$ be the j^{th} configuration space of $\mathbb{R}^{n+1}$ and Σ_j be the symmetric group on j letters. Let $D_j(X)$ denote the smash product $F(\mathbb{R}^{n+1}, j)^{+}\wedge_{\Sigma_j} X^{(j)}$. There is a homotopy equivalence

$$\Sigma^{\infty}\Omega^{n+1}\Sigma^{n+1}X \simeq \bigvee_{j=0}^{\infty}\Sigma^{\infty}F(\mathbb{R}^{n+1}, j)^{+}\wedge_{\Sigma_j} X^{(j)} = \bigvee_{j=0}^{\infty}\Sigma^{\infty}D_j(X).$$

Subsequently, it was shown that similar splittings can be applied to a more general space CX; see Cohen, May and Taylor [4; 5] and May and Taylor [8].

A few years later, Bödigheimer [2] showed a unified form of all these splittings. Let K be a finite complex, K_0 a subcomplex and X a connected CW–complex. Let M be a smooth, parallelizable n–manifold with a submanifold M_0 such that $(M, M_0) \simeq (K, K_0)$. For the space $\mathrm{Map}(K, K_0; \Sigma^n X)$ of based maps from K/K_0 to $\Sigma^n X$, there is a stable splitting

$$\Sigma^{\infty}\mathrm{Map}(K, K_0; \Sigma^n X) \simeq \bigvee_{j=1}^{\infty}\Sigma^{\infty}D_k(M, M_0; X),$$

where $D_k(M, M_0; X)$ for $k \geqslant 1$ are simpler pieces constructed from the labeled configuration space $C(M, M_0; X)$.

Snaith splitting is one kind of stable splitting. Recently, the techniques of stable splittings have been applied to toric topology. For instance, Bahri, Bendersky, Cohen and

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.895

Gitler [1] found various stable splittings of polyhedral product functors. Dobrinskaya [6] proved that the loop space of the polyhedral product shares similar decompositions as the Snaith splitting.

Here we study further functorial decompositions of the Snaith splitting. More precisely, we will focus on the functorial homotopy decompositions of $F(\mathbb{R}^{n+1}, j)^+ \wedge_{\Sigma_j} X^{(j)}$. When $n = 0$, we have $F(\mathbb{R}^{n+1}, j)^+ \wedge_{\Sigma_j} X^{(j)} = X^{(j)}$. Selick and the first author [11] showed that if $p = 2$ and $\overline{H}_*(X; \mathbb{Z}/p)$ has a nontrivial Steenrod operation then the irreducible functorial decomposition component of $X^{(j)}$ and the 2–row Young diagram with distinct row numbers are in one-to-one correspondence. In this paper, we will study the case when $n > 0$.

The main idea driving this paper comes from functorial homotopy decompositions of $\Omega\Sigma X$: For each natural coalgebra-split sub-Hopf algebra (see Definition 2.2), there is a functorial homotopy retract of $\Omega\Sigma X$ with the inclusion an Ω–map; see Li, Lei and Wu [7] and Selick and Wu [10]. Among all the natural coalgebra-split sub-Hopf algebras, we mainly focus on a special one. Let $L_m^{\max}$ be the maximal T_m–projective submodule functor of the free Lie algebra functor L_m (see Section 2.1). For a graded (alternatively ungraded) $\mathbb{Z}/p$–module V, the tensor algebra $T(L_m^{\max}(V))$ generated by $L_m^{\max}(V)$ is a natural coalgebra-split sub-Hopf algebra (Proposition 2.3). Following from Section 2.3, there is geometric realization of $L_m^{\max}(V)$, denoted by $L_m^{\max}(X)$, such that $\Omega\Sigma L_m^{\max}(X)$ is a functorial homotopy retract of $\Omega\Sigma X$. Furthermore, the inclusion is an Ω–map.

For a space $\Sigma^n X$, we have that $\Omega\Sigma L_m^{\max}(\Sigma^n X)$ is a functorial homotopy retract of $\Omega\Sigma^{n+1} X$ with the inclusion an Ω–map. Applying the loop functor n times, we can obtain a functorial homotopy retract of $\Omega^{n+1}\Sigma^{n+1} X$ with the functorial the homotopy inclusion an Ω^{n+1}–map. It can be shown that this retract is a $(n+1)$–iterated loop suspension (Lemma 3.1). Now a natural question is: what is the relation between the Snaith splitting of the retract and the Snaith splitting of the original $(n+1)$–iterated loop suspension? To answer this question, we have the following main result:

Theorem 1.1 *Let X be a 1–connected p–local suspension of finite type. For the natural coalgebra-split sub-Hopf algebra $T(L_m^{\max}(V))$, there is an n^{th} desuspension $\Sigma^{-n} L_m^{\max} \Sigma^n X$ of the topological space $L_m^{\max}(\Sigma^n X)$ and a sufficient large integer t such that $\Sigma^t D_j(\Sigma^{-n} L_m^{\max} \Sigma^n X)$ is a functorial homotopy retract of $\Sigma^t D_{jm}(X)$.*

This article is organized as follows. In Section 2, we give a brief introduction about natural coalgebra-split sub-Hopf algebras of the tensor algebra, functorial homotopy retracts of $\Omega\Sigma X$ and the homology of $\Omega^{n+1}\Sigma^{n+1} X$. Section 3 constructs natural homotopy retracts of $\Omega^{n+1}\Sigma^{n+1} X$ from natural coalgebra-split sub-Hopf algebras of the tensor algebra. In Section 4, we compute the homology image of $\Sigma^{-n} L_m^{\max} \Sigma^n X$ in the homology $\Omega^{n+1}\Sigma^{n+1}X$. In Section 5, a collection of the functorial stable homotopy retract of $F(\mathbb{R}^{n+1}, j)^+ \wedge_{\Sigma_j} X^{(j)}$ is constructed. Additionally, the proof of Theorem 1.1 is given in this section. An example is given in Section 6.

2 Preliminaries

Let $\Bbbk = \mathbb{Z}/p$ be the ground ring; p is a prime. All topological spaces are assumed to be p–local CW–complexes. All homology is taken with the coefficients $\mathbb{Z}/p$ unless otherwise stated.

2.1 T_n–projective module

Let V be a graded (ungraded) $\Bbbk$–module. Let $T(V)$ be the tensor algebra generated by V, namely

$$T(V) = \sum_{n=0}^{\infty} V^{\otimes n}.$$

A Hopf algebra structure can be given over $T(V)$ by setting V to be primitive. Let $T_n(V) = V^{\otimes n}$. Then T and T_n can be viewed as functors from the category of graded (ungraded) $\Bbbk$–modules to the category of graded (ungraded) $\Bbbk$–modules.

Let M and N be functors from the category of graded (ungraded) $\Bbbk$–modules to the category of graded (ungraded) $\Bbbk$–modules. M is a *submodule functor* of N if $M(V) \subseteq N(V)$ for each graded (ungraded) $\Bbbk$–module V, and M is a *retract* of N if there are natural transformations $i\colon M \to N$ and $r\colon N \to M$ of $\Bbbk$–modules such that $r \circ s = \mathrm{id}\colon M \to M$. A retract of T_n is related to a $\Bbbk(\Sigma_n)$–projective module (see [7, Proposition 2.10]). Hence, if M is a retract of T_n, we also call it *T_n–projective*.

Let $L(V)$ be the free Lie algebra generated by V. Then L is a submodule functor of T. Let $L_n(V) = L(V) \cap T_n(V)$. From Selick and the first author [10], there exists a submodule functor $L_n^{\max}$ of L_n with the following properties:

Proposition 2.1 [10, Section 6] (1) *$L_n^{\max}$ is T_n–projective.*

(2) *Each T_n–projective submodule functor of L_n is a retract of $L_n^{\max}$.*

Up to isomorphism, $L_n^{\max}$ is the *maximal T_n–projective submodule functor* of L_n.

2.2 Coalgebra-split sub-Hopf algebras

A coalgebra-split sub-Hopf algebra is a retract of $T(V)$ with additional Hopf algebra and coalgebra structures.

Definition 2.2 Let B be a submodule functor of T. We say $B(V)$ is *a natural coalgebra-split sub-Hopf algebra* of $T(V)$ if:

(1) $B(V)$ is a natural sub-Hopf algebra of $T(V)$ with natural inclusion of Hopf algebras $j_V\colon B(V) \to T(V)$.

(2) There is a natural coalgebra transformation $r_V\colon T(V) \to B(V)$ with $r_V \circ j_V = \mathrm{id}_{B(V)}$.

If $B(V)$ is a natural coalgebra-split sub-Hopf algebra defined as above, the natural maps j_V and r_V are called an *associated natural inclusion* and *associated natural retraction* of $B(V)$, respectively.

A natural coalgebra-split sub-Hopf algebra is a tensor algebra. Let $Q(V)$ be the set of indecomposable elements of $B(V)$; this is a $\Bbbk$–submodule of $B(V)$. We have a natural isomorphism of Hopf algebras

$$B(V) \cong T(Q(V)).$$

Define the maps k_V and ψ_V as the canonical inclusion and projection

$$k_V\colon Q(V) \to T(Q(V)) \cong B(V),$$
$$\psi_V\colon B(V) \cong T(Q(V)) \to Q(V).$$

These definitions imply the following commutative diagrams:

$$\begin{tikzcd}
Q(V) \arrow[rr, "i_V"] \arrow[dr, "k_V"'] & & T(V) \\
& B(V) \arrow[ur, "j_V"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
Q(V) & & T(V) \arrow[ll, "\phi_V"'] \arrow[dl, "r_V"] \\
& B(V) \arrow[ul, "\psi_V"] &
\end{tikzcd}$$

Here j_V is a Hopf algebra homomorphism, r_V is a coalgebra homomorphism, $r_V \circ j_V = \mathrm{id}_{B(V)}$, the maps k_V and ψ_V are homomorphisms of $\Bbbk$–modules, and i_V and ϕ_V are defined as the compositions of the other two maps in the triangle.

If $B(V)$ is a sub-Hopf algebra of $T(V)$ only, then properties of $Q(V)$ can imply a coalgebra-split structure of $B(V)$.

Proposition 2.3 [7, Theorem 5.2] *Let $B(V)$ be a natural sub-Hopf algebra of $T(V)$. Then the following statements are equivalent:*

(1) *$B(V)$ is a natural coalgebra-split sub-Hopf algebra of $T(V)$.*

(2) *Each $Q_n(V) = Q(V) \cap T_n(V)$ is naturally equivalent to a T_n–projective subfunctor of L_n.*

(3) *Each Q_n is T_n–projective.*

Since $L_n^{\max}$ is a T_n–projective subfunctor of L_n, Proposition 2.3 implies $T(L_n^{\max}(V))$ is a natural coalgebra-split sub-Hopf algebra of $T(V)$.

2.3 Functorial homotopy retracts of $\Omega\Sigma X$

Let A and B be functors from the (homotopy) category of path-connected p–local CW–complexes to the (homotopy) category of spaces. Let $\mathcal{C}$ be a subcategory of the (homotopy) category of path-connected p–local CW–complexes. A is a *functorial homotopy retract* of B over $\mathcal{C}$ if, for each object X in $\mathcal{C}$, there are natural maps $i_X\colon A(X) \to B(X)$ and $r_X\colon B(X) \to A(X)$ such that $r_X \circ i_X \simeq \mathrm{id}_{A(X)}$. The homotopy need not be natural. The maps i_X and r_X are called an *associated natural inclusion* and *associated natural retraction* of A, respectively.

The functorial homotopy retracts of $\Omega\Sigma X$ are related to natural coalgebra-split sub-Hopf algebras of $T(V)$. Let X be a CW–complex. X is a *p–local suspension of finite type* if X is homotopic equivalent to $\Sigma Y_{(p)}$, the suspension of the p–localization of a finite CW–complex Y. Let $B(V)$ be a natural coalgebra-split sub-Hopf algebra of $T(V)$ and $Q(V)$ be the set of indecomposable elements of $B(V)$. A functorial homotopy retract of $\Omega\Sigma X$ can be constructed from $B(V)$ and $Q(V)$.

Theorem 2.4 [10, Theorem 1.1; 13, Theorem 3.3] *Let X be a 1–connected p–local suspension of finite type. Let $B(V)$ be a natural coalgebra-split sub-Hopf algebra of $T(V)$ with associated natural inclusion $j_V\colon B(V) \to T(V)$, and $Q(V)$ the set of indecomposable elements of $B(V)$. Then there is a functorial space $Q(X)$ with a natural map $i_X\colon Q(X) \to \Omega\Sigma X$ such that:*

(1) *$\Omega\Sigma Q(X)$ is a natural homotopy retract of $\Omega\Sigma X$ with associated natural inclusion $\Omega\tilde{\imath}_X$, where $\tilde{\imath}_X\colon \Sigma Q(X) \to \Sigma X$ is the adjoint of $i_X\colon Q(X) \to \Omega\Sigma X$:*

$$\begin{tikzcd} Q(X) \arrow[rr, "i_X"] \arrow[dr] & & \Omega\Sigma X \\ & \Omega\Sigma Q(X) \arrow[ur, "\Omega\tilde{\imath}_X"'] & \end{tikzcd} \tag{1}$$

Here the map $Q(X) \to \Omega\Sigma Q(X)$ is the canonical suspension map.

(2) *$Q(X)$ has a wedge decomposition. In detail, there are elements $\lambda_m \in \mathbb{Z}(\Sigma_m)$ for $m \geqslant 2$ such that $Q(X) = \bigvee_{m=2}^{\infty} Q_m(X)$, where $Q_m(X) = \mathrm{hocolim}_{\lambda_m} X^{(m)}$. Here Σ_m acts on $X^{(m)}$ by permuting factors.*

(3) *$\bar{H}_*(Q(X)) \cong Q(\bar{H}_*(X))$ and $H_*(\Omega\Sigma Q(X)) \cong B(\bar{H}_*(X))$. Furthermore, the induced diagram from diagram (1) satisfies $(\Omega\tilde{\imath}_X)_* = j_{\bar{H}_*(X)}$:*

$$\begin{tikzcd} Q(\overline{H}_*(X)) \arrow[rr, "i_{X*}"] \arrow[dr] & & T(\overline{H}_*(X)) \\ & B(\overline{H}_*(X)) \arrow[ur, "(\Omega\tilde{\imath}_X)_*"'] & \end{tikzcd}$$

In following discussions, we denote the map $\Omega\tilde{\imath}_X$ by j_X. It follows from the theorem that $\Omega\Sigma Q(X)$ is a functorial homotopy retract of $\Omega\Sigma X$ with an associated natural inclusion $j_X\colon \Omega\Sigma Q(X) \to \Omega\Sigma X$ which is a loop map.

2.4 Homology of $\Omega^{n+1}\Sigma^{n+1}X$

Let X be a connected CW–complex. All homology is taken with $\mathbb{Z}/p$–coefficients. The homology of $\Omega^{n+1}\Sigma^{n+1}X$ can be formulated by H_*X, Dyer–Lashof operations Q^i, Browder operations λ_n (we will also use $[-,-]_n$), a function ξ_n and a function ζ_n. The function ζ_n is defined for $p>2$ only.

To formulate the homology of $\Omega^{n+1}\Sigma^{n+1}X$, a set T_nX will be defined first. For convenience, we list the construction of T_nX for $p>2$ only in the following. The case for $p=2$ is similar.

Let $V=\overline{H}_*X$. An element $x\in V$ is a λ_n–product of *weight* 1 ($\omega(x)=1$); the weight of $[a,b]_n$ is defined by $\omega([a,b]_n)=\omega(a)+\omega(b)$. We say $x\in V$ is a *basic λ_n–product of weight* 1. Assume the basic λ_n–product of weight $j<k$ has been defined and totally ordered; the *basic λ_n–product of weight k* is of the form $[a,b]_n$ such that:

(1) $\omega([a,b]_n)=k$.

(2a) a and b are basic λ_n–products, with $a<b$. If $b=[c,d]_n$ for c and d basic then $a\geqslant c<d$.

(2b) If a is a basic λ_n–product of weight 1 and $n+\mathrm{degree}(a)$ is odd, then $[a,a]_n$ is also a basic λ_n–product of weight 2.

Let $I=(\varepsilon_1,s_1,\dots,\varepsilon_k,s_k)$ be a $2k$–tuple of integers with $s_j\geqslant\varepsilon_j$ and $\varepsilon=0$ or 1. I is *admissible* if $ps_j-\varepsilon_j\geqslant s_{j-1}$ for $2\leqslant j\leqslant k$. Define functions e, d, l and b as follows:

(i) **Excess** $e(I)=2s_1-\varepsilon_1-\sum_{j=2}^k[2s_j(p-1)-\varepsilon_j]$.

(ii) **Degree** $d(I)=\sum_{j=1}^k[2s_j(p-1)-\varepsilon_j]$.

(iii) **Length** $l(I)=k$.

(iv) $b(I)=\varepsilon_1$.

If $I = \varnothing$, then let $e(I) = \infty$ and $d(I) = l(I) = b(I) = 0$.

For $I = (\varepsilon_1, s_1, \dots, \varepsilon_k, s_k)$, let $Q^I y = \beta^{\varepsilon_1} Q^{s_1} \cdots \beta^{\varepsilon_k} Q^{s_k} y$. Define the set $T_n X$ by

$$T_n X = \big\{ Q^I y \mid y \text{ a basic } \lambda_n\text{–product, } I \text{ admissible, } e(I) + b(I) > |y|, \\ \text{if } I = (\varepsilon_1, s_1, \dots, \varepsilon_k, s_k), \text{ then } s_k \leqslant \tfrac{1}{2}(n+q) \big\}.$$

Here we denote $\xi_n x$ by $Q^{(n+q)/2} x$ and $\zeta_n x$ by $\beta Q^{(n+q)/2} x$ for $x \in H_q X$, and $|y|$ is the degree of y.

For a prime p, the homology $H_* \Omega^{n+1} \Sigma^{n+1} X$ is a functor of $H_* X$, denoted by $W_n H_* X$. On the other hand, let $A T_n X$ be the free commutative algebra generated by the set $T_n X$. We have the following theorem:

Theorem 2.5 [3, Theorem 3.1, Lemma 3.8] *For a connected X, there is an isomorphism of algebras*

$$W_n H_* X \cong A T_n X.$$

Remark Here we use $W_n H_* X$ as another notation for $H_* \Omega^{n+1} \Sigma^{n+1} X$. In fact, it can be defined independently as an $A R_n \Lambda_n$–Hopf algebra with conjugation (see [3, Section 2]).

There is a weight filtration defined on $W_n H_* X$. For an element $Q^I y$ in $T_n X$, let its *weight* $\omega(Q^I y)$ be defined by

$$\omega(Q^I y) = p^{l(I)} \omega(y),$$

where $l(I)$ is the length of the tuple I and $\omega(y)$ is the weight of the basic λ_n–product y. Since $H_* \Omega^{n+1} \Sigma^{n+1} X$ is the commutative algebra generated by $T_n X$, we can define the weight of the product $Q^I y \cdot Q^{I'} y'$ as

$$\omega(Q^I y \cdot Q^{I'} y') = \omega(Q^I y) + \omega(Q^{I'} y').$$

This makes $H_* \Omega^{n+1} \Sigma^{n+1} X$ a filtered algebra by defining the filtration as

$$F_k W_n H_* X = \{ x \in H_* \Omega^{n+1} \Sigma^{n+1} X \mid \omega(x) \leqslant k \}.$$

Let $E_k W_n H_* X = F_k W_n H_* X / F_{k-1} W_n H_* X$. There is a geometric realization of $E_k W_n H_* X$.

Proposition 2.6 [3, Section 4] $\overline{H}_*(F(\mathbb{R}^{n+1}, k)^+ \wedge_{\Sigma_k} X^{(k)}) \cong E_k W_n H_* X.$

2.5 Homology suspensions and transgressions

The *homology suspension* is defined as the homomorphism

$$\sigma_* = p \circ \partial^{-1} \colon \overline{H}_*(\Omega B) \xleftarrow[\cong]{\partial} H_{*+1}(PB, \Omega B) \xrightarrow{p_*} H_{*+1}(B),$$

where $p\colon PB \to B$ is the map $p(u) = u(1)$. The transgression is the differential map in the Serre spectral sequences. Fix a fibration $F \to E \to B$ with connected B and F; in the associated Serre spectral sequence, the transgression τ is the differential

$$d_n\colon E^n_{n,0} \to E^n_{0,n-1}.$$

Some general properties of σ_* and τ are listed as follows:

Proposition 2.7 [9, Propositions 6.10 and 6.11] (1) *Let $f\colon X \to \Omega Y$ be a pointed map and $\tilde{f}\colon \Sigma X \to Y$ be its adjoint; then the homology suspension σ_* and the suspension $\Sigma_*\colon H_*X \to H_{*+1}\Sigma X$ form a commutative diagram:*

$$\begin{tikzcd}
\bar{H}_{q-1}(X) \arrow[r, "f_*"] \arrow[d, "\Sigma_*"] & \bar{H}_{q-1}(\Omega Y) \arrow[d, "\sigma_*"] \\
\bar{H}_q(\Sigma X) \arrow[r, "\tilde{f}_*"] & H_q(Y)
\end{tikzcd}$$

(2) *If B is simply connected, then in the Serre spectral sequence of $\Omega B \to PB \to B$ there is a commutative diagram:*

$$\begin{tikzcd}
E^q_{q,0} \arrow[r, "d^q", "\cong"'] \arrow[d, hook] & E^q_{0,q-1} \\
H_q(B) & H_{q-1}(F) \arrow[l, "\sigma_*"'] \arrow[u, two heads]
\end{tikzcd}$$

In particular, the image of σ_* is transgressive.

Consider the relation between τ and the Browder operation $[-,-]_n$; we have:

Proposition 2.8 *If X is connected, then in the Serre spectral sequence of*

$$\Omega^{n+1}\Sigma^{n+1}X \to P\Omega^n\Sigma^{n+1}X \to \Omega^n\Sigma^{n+1}X$$

we have

$$\tau\big([sx_1,\dots,[sx_{k-1},sx_k]_{n-1}]_{n-1}\big) = [x_1,\dots,[x_{k-1},x_k]_n]_n,$$
$$\tau Q^I sx = (-1)^{d(I)} Q^I x,$$

where sx is the image of $x \in H_*X$ under the isomorphism $\Sigma_*\colon H_*X \to H_{*+1}\Sigma X$.

This proposition is implicit in the proof of [3, Theorem 3.2].

3 Functorial homotopy retracts of $\Omega^{n+1}\Sigma^{n+1}X$

Let $B(V)$ be a natural coalgebra-split sub-Hopf algebra of $T(V)$ and $Q(V)$ the set of indecomposable elements of $B(V)$. Let X be a 1–connected p–local suspension of finite type. It follows from Theorem 2.4 that $\Omega\Sigma Q(\Sigma^n X)$ is a functorial homotopy retract of $\Omega\Sigma(\Sigma^n X)$. By applying the loop functor n times, we can obtain that $\Omega^{n+1}\Sigma Q(\Sigma^n X)$ is a homotopy retract of $\Omega^{n+1}\Sigma^{n+1}X$ and the natural inclusion

$$\Omega^n j_{\Sigma^n X}\colon \Omega^{n+1}\Sigma Q(\Sigma^n X)\hookrightarrow \Omega^{n+1}\Sigma^{n+1}X$$

is an Ω^{n+1}–map. If X is a co-H–space, the space $Q(\Sigma^n X)$ can be desuspended n times:

Lemma 3.1 *If X is a co-H–space, then there is a space $\bar{Q}(X)$ such that $Q(\Sigma^n X)$ is naturally homotopic to $\Sigma^n\bar{Q}(X)$.*

Proof Since $Q(X)=\bigvee_{m=2}^{\infty}Q_m(X)$, it is sufficient to prove $Q_m(\Sigma^n X)$ can be desuspended n times. Let $X^{(m)}$ be the m–fold self smash product of X. The definition of $Q_m(\Sigma^n X)$ implies a homotopy commutative diagram:

$$\begin{tikzcd}
(\Sigma^n X)^{(m)} \arrow[r, "\phi"] & (\Sigma^n X)^{(m)} \\
\Sigma^{mn}X^{(m)} \arrow[u, "\text{shuffling isomorphism}"] \arrow[r, "\Sigma^{mn}\bar{\phi}"] & \Sigma^{mn}X^{(m)} \arrow[u, "\text{shuffling isomorphism}"']
\end{tikzcd} \tag{2}$$

Here

$$\begin{aligned}
\phi=\lambda_m&=\sum_{\sigma\in\Sigma_m}k_\sigma\sigma\colon(\Sigma^n X)^{(m)}\to(\Sigma^n X)^{(m)},\\
\bar{\phi}&=\sum_{\sigma\in\Sigma_m}k_\sigma\sigma(-1)^{n^2\,\mathrm{Sign}\,\sigma}\colon X^{(m)}\to X^{(m)},
\end{aligned} \tag{3}$$

and the vertical maps are the natural shuffling homeomorphisms.

Let $\bar{Q}_m(X)=\mathrm{hocolim}_{\bar{\phi}}X^{(m)}$. It is obvious that

$$\Sigma^{mn}\bar{Q}_m(X)\simeq \mathrm{hocolim}_{\Sigma^{mn}\bar{\phi}}\Sigma^{mn}X^{(m)}\cong \mathrm{hocolim}_{\phi}(\Sigma^n X)^{(m)}=Q_m(\Sigma^n X).$$

Thus,

$$Q(\Sigma^n X)=\bigvee_{m=2}^{\infty}Q_m(\Sigma^n X)=\bigvee_{m=2}^{\infty}\Sigma^{mn}\bar{Q}_m(X)=\Sigma^n\bigvee_{m=2}^{\infty}\Sigma^{n(m-1)}\bar{Q}_m(X).$$

It is clear that all homotopy equivalences are natural. □

Remark This lemma shows that $\bigvee_{m=2}^{\infty}\Sigma^{n(m-1)}\bar{Q}_m(X)$ is the n^{th} desuspension of $Q(\Sigma^n X)$. For convenience, in later discussion, $\Sigma^{-n}Q(\Sigma^n X)$ is used to denote the space $\bigvee_{m=2}^{\infty}\Sigma^{n(m-1)}\bar{Q}_m(X)$. Similarly, we use $\Sigma^{-n}Q_m\Sigma^n X$ to denote $\Sigma^{n(m-1)}\bar{Q}_m(X)$.

For the space $\Sigma^{-n}Q(\Sigma^n X)$, there is a natural inclusion

$$\Sigma^{-n}Q\Sigma^n X \to \Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}Q\Sigma^n X) \xrightarrow{\Omega^n j_{\Sigma^n X}} \Omega^{n+1}\Sigma^{n+1}X.$$

Up to homotopy, this map is the adjoint map of

$$Q\Sigma^n X \to \Omega\Sigma(Q\Sigma^n X) \xrightarrow{j_{\Sigma^n X}} \Omega\Sigma(\Sigma^n X).$$

This composition is exactly the functorial map $i_Y\colon Q(Y) \to \Omega\Sigma Y$, where $Y = \Sigma^n X$. In summary, we have the following theorem:

Theorem 3.2 *Let X be a 1–connected p–local suspension of finite type. If $B(V)$ is a natural coalgebra-split sub-Hopf algebra of $T(V)$ and $Q(V)$ is the set of indecomposable elements of $B(V)$, then there exists a functorial homotopy retract $\Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}Q\Sigma^n X)$ with a natural inclusion*

$$i\colon \Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}Q\Sigma^n X) \to \Omega^{n+1}\Sigma^{n+1}X,$$

which is an Ω^{n+1}–map. Furthermore,

$$\overline{H}_*(\Sigma^{-n}Q\Sigma^n X) \cong Q(\overline{H}_*(\Sigma^n X)).$$

4 $\Sigma^{-n}L^{\max}\Sigma^n X$ and its homology image in $\Omega^{n+1}\Sigma^{n+1}X$

Let $L_m^{\max}$ be the maximal T_m–projective submodule functor of L_m. The tensor algebra $T(L_m^{\max}(V))$ is a natural coalgebra-split sub-Hopf algebra with the set of indecomposable elements $L_m^{\max}(V)$. Then we have two spaces $L_m^{\max}(X)$ and $\Sigma^{-n}L_m^{\max}\Sigma^n X$. Furthermore, $\Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}L_m^{\max}\Sigma^n X)$ is a functorial homotopy retract of $\Omega^{n+1}\Sigma^{n+1}X$. The inclusion map is

(4) $$\tilde{\imath}_{n,X}\colon \Sigma^{-n}L_m^{\max}\Sigma^n X \to \Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}L_m^{\max}\Sigma^n X) \xrightarrow{\Omega^n j_{\Sigma^n X}} \Omega^{n+1}\Sigma^{n+1}X,$$

which is the adjoint of the map

$$i_{n,X}\colon L_m^{\max}\Sigma^n X \to \Omega\Sigma(L_m^{\max}\Sigma^n X) \xrightarrow{j_{\Sigma^n X}} \Omega\Sigma(\Sigma^n X).$$

To analyze the homology image of $\Sigma^{-n}L_m^{\max}\Sigma^n X$ in $\Omega^{n+1}\Sigma^{n+1}X$, we need to compute

$$(\tilde{\imath}_{n,X})_*\colon H_*\Sigma^{-n}L_m^{\max}\Sigma^n X \to H_*\Omega^{n+1}\Sigma^{n+1}X.$$

From the properties of the homology suspension σ_* (see Proposition 2.7), we obtain a commutative diagram

$$\begin{tikzcd} H_*\Sigma^{-n}L_m^{\max}\Sigma^n X \arrow[r] \arrow[d, "\Sigma_*^{(n)}"] & H_*\Omega^n\Sigma^n(\Sigma^{-n}L_m^{\max}\Sigma^n X) \arrow[r, "(\Omega^n i_{n,X})_*"] \arrow[d, "\sigma_*^{(n)}"] & H_*\Omega^{n+1}\Sigma^{n+1}X \arrow[d, "\sigma_*^{(n)}"] \\ H_{*+n}L_m^{\max}(\Sigma^n X) \arrow[r, equal] & H_{*+n}L_m^{\max}(\Sigma^n X) \arrow[r, "(i_{n,X})_*"] & H_{*+n}\Omega\Sigma(\Sigma^n X) \end{tikzcd}$$

where $\Sigma_*^{(n)}$ and $\sigma_*^{(n)}$ mean n–fold compositions.

For $x \in H_*X$, denote the image of x under the isomorphism $\Sigma_*\colon H_*X \to H_{*+1}\Sigma X$ by sx. Consequently, $s^n x$ is used to denote $\Sigma_*^{(n)}(x)$. Let $[x_1, x_2, \dots, x_m]_n$ be an arbitrary λ_n–product of weight m formed by elements $x_1, \dots, x_m$. For an element $[s^n x_1, s^n x_2, \dots, s^n x_m]_0$ in $H_{*+n}L_m^{\max}(\Sigma^n X)$, with $x_i \in H_*X$, denote its inverse image under the isomorphism

$$\Sigma_*^{(n)}\colon H_*\Sigma^{-n}L_m^{\max}\Sigma^n X \to H_{*+n}L_m^{\max}(\Sigma^n X)$$

by $s^{-n}[s^n x_1, s^n x_2, \dots, s^n x_m]_0$.

For the map $\tilde{\imath}_{n,X}$, we have the following lemma:

Lemma 4.1 *Under the homomorphism*

$$(\tilde{\imath}_{n,X})_*\colon H_*(\Sigma^{-n}L_m^{\max}\Sigma^n X) \to H_*(\Omega^{n+1}\Sigma^{n+1}X),$$

$s^{-n}[s^n x_1, s^n x_2, \dots, s^n x_m]_0$ *is mapped to* $[x_1, x_2, \dots, x_m]_n$*, with* $x_i \in H_*X$.

Proof We prove this lemma by induction on n. For $n = 1$, there is a commutative diagram:

$$\begin{tikzcd} H_*\Sigma^{-1}L_m^{\max}\Sigma X \arrow[r] \arrow[d, "\Sigma_*"] & H_*\Omega\Sigma(\Sigma^{-1}L_m^{\max}\Sigma X) \arrow[r, "(\Omega i_{1,X})_*"] \arrow[d, "\sigma_*"] & H_*\Omega^2\Sigma^2 X \arrow[d, "\sigma_*"] \\ H_{*+1}L_m^{\max}(\Sigma X) \arrow[r, equal] & H_{*+1}L_m^{\max}(\Sigma X) \arrow[r, "(i_{1,X})_*"] & H_{*+1}\Omega\Sigma^2 X \end{tikzcd}$$

The bottom row is the natural inclusion

$$(i_{1,X})_*\colon L_m^{\max}(sH_*X) \hookrightarrow T(sH_*X).$$

The upper row is exactly $(\tilde{\imath}_{1,X})_*$. Since the first map of the upper row is a natural inclusion, we only need to prove

$$(\Omega i_{1,X})_*(s^{-1}[sx_1, sx_2, \dots, sx_m]_0) = [x_1, x_2, \dots, x_m]_1.$$

To prove this, we consider a natural commutative diagram of Serre path fibrations

$$\begin{tikzcd}
\Omega L_m^{\max}(\Sigma X) \arrow[r] \arrow[d] & PL_m^{\max}(\Sigma X) \arrow[r] \arrow[d] & L_m^{\max}(\Sigma X) \arrow[d] \\
\Omega^2\Sigma^2 X \arrow[r] & P\Omega\Sigma^2 X \arrow[r] & \Omega\Sigma^2 X
\end{tikzcd}$$

which implies a natural morphism of Serre spectral sequences. Therefore, for the transgression τ, there is an equality by naturality,

$$\tau \circ (i_{1,X})_* = (\Omega i_{1,X})_* \circ \tau.$$

In the Serre spectral sequence of the path fibration

$$\Omega^2\Sigma^2 X \to P\Omega\Sigma^2 X \to \Omega\Sigma^2 X,$$

we have the equality (see Proposition 2.8)

$$\tau[sx_1, \dots, sx_m]_0 = [x_1, \dots, x_m]_1.$$

Hence,

$$\begin{aligned}
(\Omega i_{1,X})_*(s^{-1}[sx_1, sx_2, \dots, sx_m]_0) &= (\Omega i_{1,X})_* \circ \tau([sx_1, sx_2, \dots, sx_m]_0)) \\
&= \tau \circ (i_{1,X})_*([sx_1, sx_2, \dots, sx_m]_0) \\
&= \tau([sx_1, sx_2, \dots, sx_m]_0) \\
&= [x_1, \dots, x_m]_1.
\end{aligned}$$

Now assume this lemma is true for $n < k$. For $n = k$, there is a commutative diagram:

$$\begin{tikzcd}
H_*\Sigma^{-k}L_m^{\max}\Sigma^k X \arrow[r] \arrow[d, "\Sigma_*"] & H_*\Omega^k\Sigma^k(\Sigma^{-k}L_m^{\max}\Sigma^k X) \arrow[r, "(\Omega^k i_{k,X})_*"] \arrow[d, "\sigma_*"] & H_*\Omega^{k+1}\Sigma^{k+1}X \arrow[d, "\sigma_*"] \\
H_{*+1}\Sigma^{1-k}L_m^{\max}(\Sigma^k X) \arrow[r] \arrow[d, "\Sigma_*^{(k-1)}"] & H_{*+1}\Omega^{k-1}L_m^{\max}(\Sigma^k X) \arrow[r, "(\Omega^{k-1}i_{k-1,\Sigma X})_*"] \arrow[d, "\sigma_*^{(k-1)}"] & H_{*+1}\Omega^k\Sigma^{k+1}X \arrow[d, "\sigma_*^{(k-1)}"] \\
H_{*+k}L_m^{\max}\Sigma^k X \arrow[r, equal] & H_{*+k}L_m^{\max}\Sigma^k X \arrow[r, "(i_{k,X})_*"] & H_{*+k}\Omega\Sigma^{k+1}X
\end{tikzcd}$$

The composition of the second row is $(\tilde{\imath}_{k-1,\Sigma X})_*$. By induction,

$$(\tilde{\imath}_{k-1,\Sigma X})_*(s^{1-k}[s^k x_1, s^k x_2, \dots, s^k x_m]_0) = [sx_1, sx_2, \dots, sx_m]_{k-1}.$$

The horizontal rows of left commutative squares are natural inclusions. So, the above identity implies

$$(\Omega^{k-1}i_{k-1,\Sigma X})_*(s^{1-k}[s^k x_1, s^k x_2, \dots, s^k x_m]_0) = [sx_1, sx_2, \dots, sx_m]_{k-1}.$$

Note that we need to prove

$$(\Omega^k i_{k,X})_*(s^{-k}[s^k x_1, s^k x_2, \dots, s^k x_m]_0) = [x_1, x_2, \dots, x_m]_k.$$

It follows from the commutative diagram

$$\begin{tikzcd}
\Omega^k L_m^{\max}(\Sigma^k X) \arrow[r] \arrow[d, "\Omega^k i_{k,X}"] & P\Omega^{k-1} L_m^{\max}(\Sigma^k X) \arrow[r] \arrow[d] & \Omega^{k-1} L_m^{\max}(\Sigma^k X) \arrow[d, "\Omega^{k-1} i_{k-1,\Sigma X}"] \\
\Omega^{k+1}\Sigma^{k+1} X \arrow[r] & P\Omega^k \Sigma^{k+1} X \arrow[r] & \Omega^k \Sigma^{k+1} X
\end{tikzcd}$$

and the induced Serre spectral sequences that

$$(\Omega^k i_{k,X})_* \circ \tau = \tau \circ (\Omega^{k-1} i_{k-1,\Sigma X})_*.$$

Thus,

$$\begin{aligned}
(\Omega^k i_{k,X})_*&(s^{-k}[s^k x_1, s^k x_2, \dots, s^k x_m]_0) \\
&= (\Omega^k i_{k,X})_* \circ \tau(s^{1-k}[s^k x_1, s^k x_2, \dots, s^k x_m]_0) \\
&= \tau \circ (\Omega^{k-1} i_{k-1,\Sigma X})_*(s^{1-k}[s^k x_1, s^k x_2, \dots, s^k x_m]_0) \\
&= \tau([s x_1, s x_2, \dots, s x_m]_{k-1}) \\
&= [x_1, \dots, x_m]_k.
\end{aligned}$$

This completes the proof. □

5 Further decompositions of the Snaith splitting

Fix an integer $n \geqslant 0$. The space $\Omega^{n+1}\Sigma^{n+1} X$ has the Snaith splitting

$$\Sigma^\infty \Omega^{n+1}\Sigma^{n+1} X \simeq \bigvee_{j=0}^{\infty} \Sigma^\infty F(\mathbb{R}^{n+1}, j)^+ \wedge_{\Sigma_j} X^{(j)} = \bigvee_{j=0}^{\infty} \Sigma^\infty D_j(X).$$

Here $F(\mathbb{R}^{n+1}, j)$ is the j^{th} configuration space of $\mathbb{R}^{n+1}$, and $D_j(X)$ is the smash product $F(\mathbb{R}^{n+1}, j)^+ \wedge_{\Sigma_j} X^{(j)}$. From the above splitting, $D_j(X)$ is a natural stable retract of $\Omega^{n+1}\Sigma^{n+1} X$. The homology of $D_j(X)$ (see Proposition 2.6) is

$$\bar{H}_*(D_j(X)) \cong F_j W_n H_* X / F_{j-1} W_n H_* X = E_j W_n H_* X.$$

In other words, $\bar{H}_* D_j(X)$ consists of the homology classes in $H_*\Omega^{n+1}\Sigma^{n+1} X$ with weight j.

It follows from Theorem 3.2 that $\Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n} L_m^{\max}\Sigma^n X)$ is a functorial homotopy retract of $\Omega^{n+1}\Sigma^{n+1} X$. Hence we can apply the Snaith splitting to both spaces and compare $D_j(\Sigma^{-n} L_m^{\max}\Sigma^n X)$ with $D_q(X)$ for nonnegative integers j, m and q.

Proof of Theorem 1.1 In Lemma 3.1, we have proved that the n^{th} desuspension $\Sigma^{-n}L_m^{\max}\Sigma^n X$ of $L_m^{\max}(\Sigma^n X)$ exists. The left part of the main theorem will be proved in two steps. First, the stable case will be proved. We claim that there are stable maps

$$\Sigma^\infty D_j(\Sigma^{-n}L_m^{\max}\Sigma^n X) \underset{\psi}{\overset{\phi}{\rightleftarrows}} \Sigma^\infty D_{jm}X$$

such that

$$\psi_* \circ \phi_* = \mathrm{id},$$

that is:

$$\begin{tikzcd} H_*(\Sigma^\infty D_j\Sigma^{-n}L_m^{\max}\Sigma^n X) \arrow[r, "\phi_*"] \arrow[dr, equal, "\mathrm{id}"'] & H_*\Sigma^\infty D_{jm}X \arrow[d, "\psi_*"] \\ & H_*\Sigma^\infty D_j\Sigma^{-n}L_m^{\max}\Sigma^n X \end{tikzcd}$$

Recall that $\Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}L_m^{\max}\Sigma^n X)$ is a natural homotopy retract of $\Omega^{n+1}\Sigma^{n+1}X$, ie there exist maps

$$\Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}L_m^{\max}\Sigma^n X) \underset{h}{\overset{g}{\rightleftarrows}} \Omega^n\Sigma^n X$$

such that

$$h \circ g \simeq \mathrm{id}\,.$$

Furthermore, g is an Ω^{n+1}–map. In fact, g can be chosen to be $\Omega^n j_{\Sigma^n X}$ (see (4)). Applying the Snaith splitting, we have a diagram as follows:

$$\begin{tikzcd} \Sigma^\infty\Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}L_m^{\max}\Sigma^n X) \arrow[r, shift left, "\Sigma^\infty g"] \arrow[d, "\wr"] & \Sigma^\infty\Omega^{n+1}\Sigma^{n+1}X \arrow[l, shift left, "\Sigma^\infty h"] \arrow[d, "\wr"] \\ \bigvee_{j\geqslant 1}\Sigma^\infty D_j(\Sigma^{-n}L_m^{\max}\Sigma^n X) \arrow[d, shift left, "p'_j"] & \bigvee_{q\geqslant 1}\Sigma^\infty D_q X \arrow[d, shift left, "p_q"] \\ \Sigma^\infty D_j\Sigma^{-n}L_m^{\max}\Sigma^n X \arrow[u, shift left, "s'_j"] & \Sigma^\infty D_q X \arrow[u, shift left, "s_q"] \end{tikzcd} \tag{5}$$

where p'_q and p_q are the canonical projections to the q^{th} components, and s'_q and s_q are the canonical inclusions from the q^{th} component to the whole spaces.

Next, consider their induced maps on homology. Recall that

$$H_*\Sigma^\infty\Omega^{n+1}\Sigma^{n+1}X \cong H_*\Omega^{n+1}\Sigma^{n+1}X \cong \bigoplus_{q=1}^{\infty} H_*D_q X \cong \bigoplus_{q=1}^{\infty} H_*\Sigma^\infty D_q X$$

and

$$\overline{H}_* D_q X = E_q W_n H_* X.$$

Hence $(p_q)_*$ is isomorphic to the canonical projection from the direct sum to the q^{th} summand, and $(s_q)_*$ is isomorphic to the canonical inclusion from the q^{th} summand to the whole direct sum. That is,

$$\bigoplus_{q=1}^{\infty} E_q W_n H_* X \underset{(s_q)_*}{\overset{(p_q)_*}{\rightleftarrows}} E_q W_n H_* X.$$

Thus, we obtain a diagram of homology:

$$\begin{tikzcd}
W_n H_*(\Sigma^{-n} L_m^{\max} \Sigma^n X) \arrow[r, shift left, "(\Sigma^\infty g)_*"] \arrow[d, "\cong"] & W_n H_* X \arrow[l, shift left, "(\Sigma^\infty h)_*"] \arrow[d, "\cong"] \\
\bigoplus_{j\geqslant 1} E_j W_n H_*(\Sigma^{-n} L_m^{\max} \Sigma^n X) \arrow[d, shift left, "(p'_j)_*"] & \bigoplus_{q\geqslant 1} E_q W_n H_* X \arrow[d, shift left, "(p_q)_*"] \\
E_j W_n H_* \Sigma^{-n} L_m^{\max} \Sigma^n X \arrow[u, shift left, "(s'_j)_*"] & E_q W_n H_* X \arrow[u, shift left, "(s_q)_*"]
\end{tikzcd}$$

Now the claim below is obvious:

$$(s_h)_* \circ (p_h)_*|_{\overline{H}_* D_q X} = \begin{cases} 0 & \text{if } h \neq q, \\ \mathrm{id}_{\overline{H}_* D_q X} & \text{if } h = q. \end{cases}$$

Let us consider the composition

$$E_j W_n H_* \Sigma^{-n} L_m^{\max} \Sigma^n X \xrightarrow{(s'_j)_*} W_n H_*(\Sigma^{-n} L_m^{\max} \Sigma^n X) \xrightarrow{(\Sigma^\infty g)_*} W_n H_* X.$$

An element of $E_j W_n H_* \Sigma^{-n} L_m^{\max} \Sigma^n X$ can be written as

$$Q^{I_1} y_1(z_1, \dots, z_m) \cdot Q^{I_2} y_2(z_1, \dots, z_m) \cdots Q^{I_k} y_k(z_1, \dots, z_m),$$

where $y_i(z_1, \dots, z_m)$ $(1 \leqslant i \leqslant k)$ are basic λ_n–products formed by $z_1, \dots, z_m$ for $z_i \in H_* \Sigma^{-n} L_m^{\max} \Sigma^n X$, and the product $Q^{I_1} y_1 \cdots Q^{I_k} y_k$ is a homology class of $H_* \Omega^{n+1} \Sigma^{n+1} X$ of weight j. That g is an Ω^{n+1}–map implies that $g_* Q^I = Q^I g_*$ and $g_*[x, y]_n = [g_* x, g_* y]_n$. Thus,

$$g_*(Q^{I_i} y_i(z_1, \dots, z_m)) = Q^{I_i} y_i(g_* z_1, \dots, g_* z_m).$$

By Lemma 4.1, for an element $s^{-n}[s^n x_1, \dots, s^n x_m]_0$ in $H_* \Sigma^{-n} L_m^{\max} \Sigma^n X$, with $x_i \in \overline{H}_* X$, we have

$$g_*(s^{-n}[s^n x_1, \dots, s^n x_m]_0) = [x_1, \dots, x_m]_n.$$

It follows that g_*z_i is of weight m for each element $z_i \in H_*\Sigma^{-n}L_m^{\max}\Sigma^n X$. Thus, the weight of $Q^{I_i}y_i(g_*z_1,\dots,g_*z_m)$ is equal to the weight of $Q^{I_i}y_i$ multiplied by m. Finally,

$$((\Sigma^\infty g)_*\circ(s'_j)_*)(E_jW_nH_*\Sigma^{-n}L_m^{\max}\Sigma^n X)\subseteq E_{jm}W_nH_*X.$$

Now let $\phi_{j,q}=p_q\circ\Sigma^\infty g\circ s'_j$ and $\psi_{j,q}=p'_j\circ\Sigma^\infty h\circ s_q$. We can obtain that

$$(\psi_{j,q})_*\circ(\phi_{j,q})_*=(p'_j)_*\circ(\Sigma^\infty h)_*\circ((s_q)_*\circ(p_q)_*)\circ(\Sigma^\infty g)_*\circ(s'_j)_*.$$

Since

$$\mathrm{Im}((\Sigma^\infty g)_*\circ(s'_j)_*)\subseteq E_{jm}W_nH_*X,$$

we have:

(1) If $q\neq jm$, then

$$(s_q)_*\circ(p_q)_*|_{E_{jm}H_*X}=0.$$

Thus $(\psi_{j,q})_*\circ(\phi_{j,q})_*=0$.

(2) If $q=jm$, then

$$(s_q)_*\circ(p_q)_*|_{E_{jm}H_*X}=\mathrm{id}.$$

Thus $(\psi_{j,q})_*\circ(\phi_{j,q})_*=(p'_j)_*\circ(\Sigma^\infty h)_*\circ(\Sigma^\infty g)_*\circ(s'_j)_*=\mathrm{id}$.

Let $\phi=\phi_{j,jm}$ and $\psi=\psi_{j,jm}$. The discussion above implies that

$$\psi_*\circ\phi_*=\mathrm{id}.$$

This completes the proof of step one.

In step two, it will be proved that the stable maps ϕ and ψ can be induced from unstable maps. Recall diagram (5).

There are an integer t_1 and a map

$$\overline{p}_q\colon \Sigma^{t_1}\Omega^{n+1}\Sigma^{n+1}X\to\Sigma^{t_1}D_qX$$

such that

$$\Sigma^\infty\overline{p}_q\colon \Sigma^\infty\Omega^{n+1}\Sigma^{n+1}X\to\Sigma^\infty D_qX$$

is homotopic to the map p_q [4, Theorem 7.1]. Similarly, we have a map

$$\overline{s}_q\colon \Sigma^{t_2}D_qX\to\Sigma^{t_2}\Omega^{n+1}\Sigma^{n+1}X$$

for some integer t_2. This map induces the stable map s_q. Similarly, we can obtain maps $\overline{p}'_j$ and $\overline{s}'_j$ inducing maps p'_j and s'_j for integers t_3 and t_4, respectively:

$$\overline{p}'_j\colon \Sigma^{t_3}\Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}L_m^{\max}\Sigma^nX)\to\Sigma^{t_3}D_j(\Sigma^{-n}L_m^{\max}\Sigma^nX),$$
$$\overline{s}'_j\colon \Sigma^{t_4}D_j(\Sigma^{-n}L_m^{\max}\Sigma^nX)\to\Sigma^{t_4}\Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}L_m^{\max}\Sigma^nX).$$

Let $t=\max\{t_1,t_2,t_3,t_4\}$. There are four maps $\overline{p}_j$, $\overline{s}_j$, $\overline{p}_j'$ and $\overline{s}_j'$ up to Σ^t. For simplicity, we still denote them by $\overline{p}_j$, $\overline{s}_j$, $\overline{p}_j'$ and $\overline{s}_j'$. Then there is a diagram:

$$\begin{tikzcd}
\Sigma^t\Omega^{n+1}\Sigma^{n+1}(\Sigma^{-n}L_m^{\max}\Sigma^nX) \arrow[r, shift left, "\Sigma^t g"] \arrow[d, shift left, "\overline{p}_j'"] & \Sigma^t\Omega^{n+1}\Sigma^{n+1}X \arrow[l, shift left, "\Sigma^t h"] \arrow[d, shift left, "\overline{p}_q"] \\
\Sigma^t D_j\Sigma^{-n}L_m^{\max}\Sigma^nX \arrow[u, shift left, "\overline{s}_j'"] & \Sigma^t D_qX \arrow[u, shift left, "\overline{s}_q"]
\end{tikzcd}$$

Define two maps $\overline{\phi}$ and $\overline{\psi}$ as follows:

$$\overline{\phi}=\overline{p}_{jm}\circ\Sigma^t g\circ\overline{s}_j'\colon \Sigma^t D_j(\Sigma^{-n}L_m^{\max}\Sigma^nX)\to\Sigma^t D_{jm}X,$$
$$\overline{\psi}=\overline{p}_j'\circ\Sigma^t h\circ\overline{s}_{jm}\colon \Sigma^t D_{jm}X\to\Sigma^t D_j(\Sigma^{-n}L_m^{\max}\Sigma^nX).$$

The map $\overline{\psi}\circ\overline{\phi}$ induces an identity on the homology:

$$(\overline{\psi})_*\circ(\overline{\phi})_*=(\Sigma^\infty\overline{\psi})_*\circ(\Sigma^\infty\overline{\phi})_*=\psi_*\circ\phi_*=\mathrm{id}.$$

By the Whitehead theorem, we have $\overline{\psi}\circ\overline{\phi}$ is a homotopy equivalence. It follows that

$$(\overline{\psi}\circ\overline{\phi})^{-1}\circ\overline{\psi}\circ\overline{\phi}\simeq\mathrm{id}.$$

The maps

$$\overline{\phi}\colon \Sigma^t D_j(\Sigma^{-n}L_m^{\max}\Sigma^nX)\to\Sigma^t D_{jm}X,$$
$$(\overline{\psi}\circ\overline{\phi})^{-1}\circ\overline{\psi}\colon \Sigma^t D_{jm}X\to\Sigma^t D_j(\Sigma^{-n}L_m^{\max}\Sigma^nX),$$

imply that $\Sigma^t D_j(\Sigma^{-n}L_m^{\max}\Sigma^nX)$ is a homotopy retract of $\Sigma^t D_{jm}X$. Note that we assume all spaces are CW–complexes, thus all constructions are natural up to homotopy. This completes the proof of step two. □

From the proof, we can obtain a corollary for the stable case.

Corollary 5.1 *Let X be a 1–connected p–local suspension of finite type. For the natural coalgebra-split sub-Hopf algebra $T(L_m^{\max}(V))$, the spectrum $\Sigma^\infty D_j(\Sigma^{-n}L_m^{\max}\Sigma^nX)$ is a functorial stable homotopy retract of $\Sigma^\infty D_{jm}(X)$. In other words, there are maps*

$$\Sigma^\infty D_j(\Sigma^{-n}L_m^{\max}\Sigma^nX)\underset{\psi}{\overset{\phi}{\rightleftarrows}}\Sigma^\infty D_{jm}X\qquad\textit{such that }\psi\circ\phi\simeq\mathrm{id}.$$

6 Example

Let X be a p–local 2–cell complex. Denote the Steenrod algebra by A. Let $V=\overline{H}_*(X;\mathbb{Z}/p)$. Assume that there are two generators u and v in V such that $P^1_*v=u$, where P^1_* is the dual operation of Steenrod operation P^1. Furthermore, assume that the degrees of u and v are both odd; denote them by $|u|$ and $|v|$, respectively.

Recall $\Sigma^{-1}L^{\max}_p\Sigma X$ is a stable functorial homotopy retract of D_pX. Thus, we have a stable functorial homotopy decomposition

$$D_pX \overset{s}{\simeq} (\Sigma^{-1}L^{\max}_p\Sigma X) \vee M_pX.$$

In the following, the homology of this decomposition and the A–module structure of each piece for $p=5$ will be computed.

6.1 Additive basis

In $H_*\Omega^2\Sigma^2X$, denote the 1–bracket (of Browder operation) $[x_1,\dots,[x_{m-1},x_m]_1,]_1$ by $[x_1,\dots,x_m]_1$. The basic 1–bracket (ie basic λ_1–product) with weight no greater than 5 are

$$\begin{aligned} u<v<[u,v]_1<[u,u,v]_1<[v,u,v]_1<[u,u,u,v]_1<[v,u,u,v]_1<[v,v,u,v]_1\\ <[u,u,u,u,v]_1<[v,u,u,u,v]_1<[v,v,u,u,v]_1<[v,v,v,u,v]_1\\ <[[u,v]_1,[u,u,v]_1]_1<[[u,v]_1,[v,u,v]_1]_1. \end{aligned}$$

Since $|u|$ and $|v|$ are odd, $[u,u]_1$ and $[v,v]_1$ are trivial. All the basic 1–products above are of odd degrees.

Recalling Proposition 2.6, we have the following additive basis of $\overline{H}_*D_pX$:

$$\begin{aligned} &u\cdot[u,u,u,v]_1,\ u\cdot[v,u,u,v]_1,\ u\cdot[v,v,u,v]_1,\ v\cdot[u,u,u,v]_1,\ v\cdot[v,u,u,v]_1,\\ &v\cdot[v,v,u,v]_1,\ [u,v]_1\cdot[u,u,v]_1,\ [u,v]_1\cdot[v,u,v]_1,\ u\cdot v\cdot[u,u,v]_1,\ u\cdot v\cdot[v,u,v]_1,\\ &[u,u,u,u,v]_1,\ [v,u,u,u,v]_1,\ [v,v,u,u,v]_1,\ [v,v,v,u,v]_1,\\ &[[u,v]_1,[u,u,v]_1]_1,\ [[u,v]_1,[v,u,v]_1]_1,\ \zeta_1u,\ \zeta_1v,\ \xi_1u,\ \xi_1v. \end{aligned}\tag{6}$$

In $H_*\Omega^2\Sigma^2X$, the first two rows of this basis are decomposable. The others are indecomposable.

6.2 Module structures over the Steenrod algebra

Let $P^r_*\colon H_*X\to H_{*-2r(p-1)}X$ be the dual operation of the Steenrod operation P^r. We have a right A–module structure on $\overline{H}_*D_5X$. For convenience, we still write the Steenrod operation P^r_* on the left.

There is a new additive basis of $\overline{H}_*D_5X$ which is invariant under Steenrod operations (see [13, Proposition 5.2]):

$$\begin{aligned}(7)\quad & u\cdot[u,u,u,v]_1,\ \ u\cdot[v,u,u,v]_1,\ \ u\cdot[v,v,u,v]_1,\ \ 2v\cdot[u,u,u,v]_1-u\cdot[v,u,u,v]_1,\\ & 2v\cdot[v,u,u,v]_1+u\cdot[v,v,u,v]_1,\ \ v\cdot[v,v,u,v]_1,\ \ [u,v]_1\cdot[u,u,v]_1,\\ & [u,v]_1\cdot[v,u,v]_1,\ u\cdot v\cdot[u,u,v]_1,\ \ u\cdot v\cdot[v,u,v]_1,\ \ -[u,u,u,u,v]_1,\\ & -[v,u,u,u,v]_1+[[u,v]_1,[u,u,v]_1]_1,\ \ 2[v,v,u,u,v]_1-[[u,v]_1,[v,u,v]_1]_1,\\ & -[v,v,v,u,v]_1,\ \ [[u,v]_1,[u,u,v]_1]_1,\ \ [[u,v]_1,[v,u,v]_1]_1,\\ & \zeta_1 u,\ \ \zeta_1 v,\ \ \xi_1 u,\ \ \xi_1 v.\end{aligned}$$

For $x \in \overline{H}_* D_p X$, let $A\langle x\rangle$ be the right A–module generated by x. Define A–modules M_i for $1 \leqslant i \leqslant 5$ as follows:

(1) $M_1 = A\langle [[u,v]_1,[v,u,v]_1]_1\rangle$, with

$$[[u,v]_1,[v,u,v]_1]_1 \xrightarrow{P^1_*} [[u,v]_1,[u,u,v]_1]_1.$$

(2) $M_2 = A\langle u\cdot v\cdot[v,u,v]_1\rangle$, with

$$u\cdot v\cdot[v,u,v]_1 \xrightarrow{P^1_*} u\cdot v\cdot[u,u,v]_1.$$

(3) $M_3 = A\langle [u,v]_1\cdot[v,u,v]_1\rangle$, with

$$[u,v]_1\cdot[v,u,v]_1 \xrightarrow{P^1_*} [u,v]_1\cdot[u,u,v]_1.$$

(4) $M_4 = A\langle \xi_1 v\rangle$. The diagram shows the additive basis of M_4:

$$\begin{tikzcd}
\xi_1 v \arrow[d, "P^1_*"] \arrow[r, "\beta"] \arrow[ddddd, bend left=60, "P^5_*"] & \zeta_1 v \arrow[ddddd, "P^5_*"] \\
-[v,v,v,u,v]_1 \arrow[d, "P^1_*"] & \\
2[v,v,u,u,v]_1-[[u,v]_1,[v,u,v]_1]_1 \arrow[d, "P^1_*"] & \\
-[v,u,u,u,v]_1+[[u,v]_1,[u,u,v]_1]_1 \arrow[d, "P^1_*"] & \\
-[u,u,u,u,v]_1 & \\
\xi_1 u \arrow[r, "\beta"] & \zeta_1 u
\end{tikzcd}$$

(5) $M_5 = A\langle v\cdot[v,v,u,v]_1, u\cdot[v,v,u,v]_1\rangle$, with:

$$\begin{tikzcd}
v\cdot[v,v,u,v]_1 \arrow[d, "P^1_*"] & \\
2v\cdot[v,u,u,v]_1 + u\cdot[v,v,u,v]_1 \arrow[d, "P^1_*"] & u\cdot[v,v,u,v]_1 \arrow[d, "P^1_*"] \\
2v\cdot[u,u,u,v]_1 - u\cdot[v,u,u,v]_1 \arrow[d, "P^1_*"] & u\cdot[v,u,u,v]_1 \arrow[d, "P^1_*"] \\
u\cdot[u,u,u,v]_1 \arrow[r, equal] & u\cdot[u,u,u,v]_1
\end{tikzcd}$$

It is obvious that there is an isomorphism of right A–modules

$$\overline{H}_* D_5 X \cong M_1 \oplus M_2 \oplus M_3 \oplus M_4 \oplus M_5.$$

6.3 $\Sigma^{-1} L_p^{\max} \Sigma X$ and $M_p X$

$L_5^{\max}(V)$ has a basis $[[u,v],[u,u,v]]$, $[[u,v],[v,u,v]]$ [10, Proposition 11.6]. It follows from Lemma 4.1 that this basis is mapped by the map

$$i_*\colon \overline{H}_*\Sigma^{-1} L_p^{\max}\Sigma X \to \overline{H}_*\Omega^2\Sigma^2 X \quad \text{to} \quad [[u,v]_1,[u,u,v]_1]_1,\ \ [[u,v]_1,[v,u,v]_1]_1.$$

Thus we can obtain the homology of $\Sigma^{-1} L_5^{\max}\Sigma X$ and $M_5 X$. The following equations are isomorphisms of right A–modules:

$$\overline{H}_*(\Sigma^{-1} L_5^{\max}\Sigma X) \cong M_1, \quad \overline{H}_* M_5 X \cong M_2 \oplus M_3 \oplus M_4 \oplus M_5.$$

Remark As a right A–module, $\overline{H}_* M_p X$ is splittable, so it is natural to ask whether $M_p X$ is splittable as a topological space, particularly whether the functorial homotopy decomposition exists or not.

Acknowledgements The authors would like to thank the referee most warmly for valuable comments and suggestions that have improved the manuscript. Wu is partially supported by the Singapore Ministry of Education research grant (AcRF Tier 1 WBS No. R-146-000-222-112) and a grant (No. 11329101) of NSFC of China.

References

[1] **A Bahri**, **M Bendersky**, **F R Cohen**, **S Gitler**, *The polyhedral product functor: a method of decomposition for moment-angle complexes, arrangements and related spaces*, Adv. Math. 225 (2010) 1634–1668 MR

[2] **C-F Bödigheimer**, *Stable splittings of mapping spaces*, from "Algebraic topology" (H R Miller, D C Ravenel, editors), Lecture Notes in Math. 1286, Springer, Berlin (1987) 174–187 MR

[3] **F R Cohen**, *The homology of* $\mathcal{C}_{n+1}$*–spaces,* $n \geqslant 0$, from "The homology of iterated loop spaces" (F R Cohen, T J Lada, J P May, editors), Lecture Notes in Mathematics 533, Springer, Berlin (1976) 207–351 MR

[4] **F R Cohen**, **J P May**, **L R Taylor**, *Splitting of certain spaces* CX, Math. Proc. Cambridge Philos. Soc. 84 (1978) 465–496 MR

[5] **F R Cohen**, **J P May**, **L R Taylor**, *Splitting of some more spaces*, Math. Proc. Cambridge Philos. Soc. 86 (1979) 227–236 MR

[6] **N Dobrinskaya**, *Loops on polyhedral products and diagonal arrangements*, preprint (2009) arXiv

[7] **J Y Li**, **F C Lei**, **J Wu**, *Module structure on Lie powers and natural coalgebra-split sub-Hopf algebras of tensor algebras*, Proc. Edinb. Math. Soc. 54 (2011) 467–504 MR

[8] **J P May**, **L R Taylor**, *Generalized splitting theorems*, Math. Proc. Cambridge Philos. Soc. 93 (1983) 73–86 MR

[9] **J McCleary**, *A user's guide to spectral sequences*, 2nd edition, Cambridge Studies in Advanced Mathematics 58, Cambridge University Press (2001) MR

[10] **P Selick**, **J Wu**, *On natural coalgebra decompositions of tensor algebras and loop suspensions*, Mem. Amer. Math. Soc. 701, Amer. Math. Soc., Providence, RI (2000) MR

[11] **P Selick**, **J Wu**, *On functorial decompositions of self-smash products*, Manuscripta Math. 111 (2003) 435–457 MR

[12] **V P Snaith**, *A stable decomposition of* $\Omega^n S^n X$, J. London Math. Soc. 7 (1974) 577–583 MR

[13] **Z H Yuan**, *Functorial homotopy decompositions of iterated loop suspensions*, PhD thesis, National University of Singapore (2014) `http://scholarbank.nus.edu.sg/bitstream/handle/10635/118255/Yuan_Zihong_PhD_Thesis.pdf`

Department of Mathematics, National University of Singapore
Block S17 (SOC1), 06-02, 10 Lower Kent Ridge Road, Singapore 119076, Singapore

Institute for Infocomm Research
1 Fusionopolis Way, #21-01 Connexis South Tower, Singapore 138632, Singapore

`matwuj@nus.edu.sg`, `yuanzhchina@gmail.com`

`http://www.math.nus.edu.sg/~matwujie`

Received: 9 September 2015 Revised: 18 April 2016

Geometry & Topology Publications, an imprint of **mathematical sciences publishers** msp

Algebraic & Geometric Topology 17 (2017) 917–956

Spin, statistics, orientations, unitarity

THEO JOHNSON-FREYD

A topological quantum field theory is *hermitian* if it is both oriented and complex-valued, and orientation-reversal agrees with complex conjugation. A field theory *satisfies spin-statistics* if it is both spin and super, and 360°–rotation of the spin structure agrees with the operation of flipping the signs of all fermions. We set up a framework in which these two notions are precisely analogous. In this framework, field theories are defined over $\mathrm{VECT}_{\mathbb{R}}$, but rather than being defined in terms of a single tangential structure, they are defined in terms of a bundle of tangential structures over $\mathrm{Spec}(\mathbb{R})$. Bundles of tangential structures may be étale-locally equivalent without being equivalent, and hermitian field theories are nothing but the field theories controlled by the unique nontrivial bundle of tangential structures that is étale-locally equivalent to Orientations. This bundle owes its existence to the fact that $\pi_1^{\text{ét}}(\mathrm{Spec}(\mathbb{R})) = \pi_1 \mathrm{BO}(\infty)$. We interpret Deligne's "existence of super fiber functors" theorem as implying that $\pi_2^{\text{ét}}(\mathrm{Spec}(\mathbb{R})) = \pi_2\,\mathrm{BO}(\infty)$ in a categorification of algebraic geometry in which symmetric monoidal categories replace commutative rings. One finds that there are eight bundles of tangential structures étale-locally equivalent to Spins, one of which is distinguished; upon unpacking the meaning of a field theory with that distinguished tangential structure, one arrives at a field theory that is both hermitian and satisfies spin-statistics. Finally, we formulate in our framework a notion of reflection-positivity and prove that if an étale-locally-oriented field theory is reflection-positive then it is necessarily hermitian, and if an étale-locally-spin field theory is reflection-positive then it necessarily both satisfies spin-statistics and is hermitian. The latter result is a topological version of the famous spin-statistics theorem.

14A22, 57R56, 81T50

0 Introduction

The main result of this article is a topological version of the spin-statistics theorem. The usual spin-statistics theorem (see Streater and Wightman [25]) asserts that in a unitary quantum field theory on Minkowskian spacetime, the fields of the theory live in a supervector space, the even (or bosonic) fields are integer spin representations of the Lorentz group, and the odd (or fermionic) fields are half-integer spin representations. In

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.917

other words, the spin of a particle agrees with its parity. Here unitarity is actually two conditions: a hermiticity condition (asserting that the determinant-(-1) component of the Lorentz group acts complex-antilinearly) and a reflection-positivity condition related to the requirement that the Hamiltonian of the quantum field theory have positive spectrum.

To formulate a version in the functorial setting of topological quantum field theory, we need

- to have orientations and spin structures on our source bordism category,
- to have complex supervector spaces in our target category, but to be able to talk about complex-antilinear maps as well as antisuper maps (ie maps that treat even and odd parts differently),
- to be able to link these structures on source and target categories.

We will solve all three problems by introducing generalizations of oriented and spin $\mathbb{R}$–linear field theories (we generally drop the words "topological" and "quantum") that we call étale-locally-oriented and étale-locally-spin. Étale-locally-oriented and étale-locally-spin field theories admit a natural notion of "reflection-positivity" (defined in terms of a certain "integration" map taking in an étale-locally-oriented or -spin field theory and producing an unoriented $\mathbb{R}$–linear field theory). With this technology in place, our main result is the following version of the spin-statistics theorem:

Theorem 0.1 *Every once-extended étale-locally-spin reflection-positive topological quantum field theory is hermitian (hence unitary) and satisfies spin-statistics.*

By definition, a field theory is *unextended* if it is defined in codimensions 0 and 1, and *once-extended* if it is defined in codimensions 0, 1, and 2. Corollary 4.8, which we prove only in outline, extends Theorem 0.1 to more-than-once-extended field theories. Freed and Hopkins prove a similar spin-statistics theorem in [12, Theorem 11.3], but there are notable differences between the approach used there and the one used in this paper.

As a warm-up to Theorem 0.1, in Section 1 we develop in detail the notions of étale-local orientation and reflection-positivity in the context of unextended field theories. The following analog of Theorem 0.1 follows almost immediately from the definitions:

Theorem 0.2 *Every unextended étale-locally-oriented reflection-positive field theory is hermitian.*

The parallel between Theorem 0.1 and Theorem 0.2 is an indication of the second main theme of this paper, which is to argue that hermiticity and spin-statistics phenomena arise from the same source. Note also that we reverse part of the logic from the standard spin-statistics theorem: as usually presented, hermiticity is a required assumption in order to imply spin-statistics; in our version, hermiticity and spin-statistics are both forced by reflection-positivity.

In order to define étale-locally-oriented manifolds, we consider local structures on manifolds that range over not (as in the case of orientations) sets, but schemes over $\mathbb{R}$. There are precisely two local structures that are étale-locally-over-$\mathrm{Spec}(\mathbb{R})$ isomorphic to orientations. The two versions of étale-local-orientations are usual-orientations and hermitian structures; the latter are characterized by the property that the scheme of hermitian structures on a point is $\mathrm{Spec}(\mathbb{C})$ and that the restriction map

$$\{\text{hermitian structures on } [0,1]\} \to \{\text{hermitian structures on } \{0,1\}\}$$

is the "antidiagonal" map $\mathrm{Spec}(\mathbb{C}) \to \mathrm{Spec}(\mathbb{C}) \times_{\mathrm{Spec}(\mathbb{R})} \mathrm{Spec}(\mathbb{C})$ sending λ to $(\lambda, \bar{\lambda})$. Hermitian structures owe their existence to the fact that the absolute Galois group of $\mathbb{R}$ happens to be the same as the group $\pi_0 \mathrm{O}(\infty)$ of connected components of the orthogonal group.

Each étale-local-orientation leads to a version of étale-locally-oriented field theory: in addition to the usual (unextended) oriented bordism category $\mathrm{BORD}^{\mathrm{Or}}_{d-1,d}$, there is a hermitian bordism category $\mathrm{BORD}^{\mathrm{Her}}_{d-1,d}$ which is not a category but rather a stack of categories over $\mathrm{Spec}(\mathbb{R})$; the two types of field theories are symmetric monoidal functors of stacks of categories $\mathrm{BORD}^{\mathrm{Or}}_{d-1,d} \to \mathrm{VECT}_{\mathbb{R}}$ and $\mathrm{BORD}^{\mathrm{Her}}_{d-1,d} \to \mathrm{VECT}_{\mathbb{R}}$, where $\mathrm{VECT}_{\mathbb{R}}$ is enhanced to the stack of categories QCOH. As such, our notion of étale-locally-oriented field theory involves infusing both the source and target categories with $\mathbb{R}$–algebraic geometry. The two versions unpack to $\mathbb{R}$–linear oriented field theories and to hermitian field theories in the usual sense.

Our definition of étale-locally-spin structures requires a categorification of (some basic notions from) real algebraic geometry. We begin this program in Section 2. Our main contribution here is to categorify the notion of field and to interpret Deligne's existence of fiber functors [9] as asserting that the categorified algebraic closure of $\mathbb{R}$ is not $\mathbb{C}$ but rather the category $\mathrm{SUPERVECT}_{\mathbb{C}}$ of complex supervector spaces. (As we will use a slight modification of the main result of [9], we include a complete proof.)

Remark 0.3 As is already apparent, we will be working both with fields in the sense of commutative algebra and field theories in the sense of physics, and English includes an unfortunate terminological conflict. We don't have a good solution to this problem, but will stick to the following convention: "field" used as a noun means "field in the sense of algebra"; "field theory" means "(classical or quantum) functorial topological field theory in the sense of physics".

We also prove that the extension $\mathrm{VECT}_{\mathbb{R}} \hookrightarrow \mathrm{SUPERVECT}_{\mathbb{C}}$ is Galois, and use this fact to categorify the notion of étale-local. There are precisely eight types of étale-locally-spin structures, of which one is distinguished by the following coincidence: the categorified absolute Galois group of $\mathbb{R}$ is canonically equivalent to the Picard groupoid $\pi_{\leq 1}\mathrm{O}(\infty)$. This distinguished version incorporates both hermiticity and spin-statistics phenomena. In summary, we find that the second row of the following table is a categorification of the first:

algebraic closure	tangential structure	Galois group	physical phenomenon
$\mathbb{R} \hookrightarrow \mathbb{C}$	$\mathrm{SO}(d) \hookrightarrow \mathrm{O}(d)$	$\mathrm{Gal}(\mathbb{C}/\mathbb{R}) = \pi_0\mathrm{O}(\infty)$	hermiticity
$\mathrm{VECT}_{\mathbb{R}} \hookrightarrow \mathrm{SUPERVECT}_{\mathbb{C}}$	$\mathrm{Spin}(d) \to \mathrm{O}(d)$	$\mathrm{Gal}(\mathrm{SUPERVECT}_{\mathbb{C}}/\mathbb{R}) = \pi_{\leq 1}\mathrm{O}(\infty)$	spin-statistics

Our categorification result suggests the following conjecture:

Conjecture 0.4 *There is an infinitely categorified version of commutative algebra, and in it the infinitely categorified absolute Galois group of* $\mathbb{R}$ *is* $\mathrm{O}(\infty)$.

Remark 0.5 The papers Ganter and Kapranov [13] and Kapranov [15] suggest that rather than $\mathrm{O}(\infty)$, it is the sphere spectrum that controls supermathematics. Very low homotopy groups cannot distinguish between various important spectra. The connection with topological quantum field theory focused on in this paper provides a reason to prefer $\mathrm{O}(\infty)$.

We prove Theorem 0.1 in Section 3, which also contains examples of various types of étale-locally-spin field theories. We end the paper in Section 4 by outlining how to extend our étale-locally-structured cobordism categories to the fully-extended ∞–categorical world of Lurie [20].

1 Oriented, hermitian, and unitary field theories

This section serves as an extended warm-up to the remainder of the paper. We will develop in a 1-categorical setting the notions of "étale-locally-oriented" and "reflection-positive" and prove Theorem 0.2, which asserts that étale-locally-oriented reflection-positive topological quantum field theories are necessarily hermitian.

The functorial framework for quantum field theory, as formulated by Atiyah and Segal in [1; 23], is well-known. Fix a dimension d and construct a symmetric monoidal category $\mathrm{BORD}_{d-1,d}$ whose objects are $(d-1)$–dimensional closed smooth manifolds, morphisms are d–dimensional smooth cobordisms up to isomorphism, and the symmetric monoidal structure is disjoint union. An (unextended) *unoriented* or *unstructured* $\mathbb{R}$*–linear* d*–dimensional functorial topological quantum field theory* is a symmetric monoidal functor $\mathrm{BORD}_{d-1,d} \to \mathrm{VECT}_{\mathbb{R}}$. We will henceforth drop the words "functorial topological quantum".

In general, one does not care simply about unstructured field theories. Let MAN_d denote the site of d–dimensional (possibly open) manifolds and local diffeomorphisms, with covers the surjections. If $\mathcal{X}$ is a category with limits, an $\mathcal{X}$*–valued local structure* is a sheaf $\mathcal{G}\colon \mathrm{MAN}_d \to \mathcal{X}$. A local structure is *topological* if it takes isotopic (among local diffeomorphisms) maps of manifolds to equal morphisms in $\mathcal{X}$.

The reason for considering local structures valued in general categories is because, in examples, the collection of $\mathcal{G}$–structures on a manifold M is not just a set but carries more algebraic or analytic structure. For example, Stolz and Teichner [24] require local structures valued in supermanifolds. We will focus on the case when $\mathcal{G}$ is valued in the category $\mathrm{SCH}_{\mathbb{R}}$ of schemes over $\mathbb{R}$. (In fact, all of our examples will take values in the subcategory $\mathrm{AFSCH}_{\mathbb{R}}$ of affine schemes.)

The following is an easy exercise:

Lemma 1.1 *Suppose $d \geq 1$. There are precisely two isotopy classes of local diffeomorphisms $\mathbb{R}^d \to \mathbb{R}^d$ (the identity and orientation-reversal), and so if $\mathcal{G}$ is an $\mathcal{X}$–valued topological local structure, then $\mathcal{G}(\mathbb{R}^d)$ has an action by $\mathbb{Z}/2$. The assignment $\mathcal{G} \mapsto \mathcal{G}(\mathbb{R}^d)$ gives an equivalence of categories between the category of $\mathcal{X}$–valued topological local structures and the category $\mathcal{X}^{\mathbb{Z}/2}$ of objects in $\mathcal{X}$ equipped with a $\mathbb{Z}/2$–action.* □

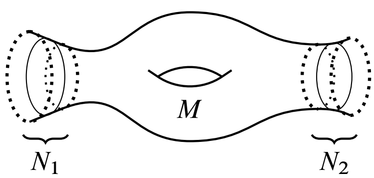
Example 1.2 The topological local structure $\mathcal{G}_X$ corresponding to a $\mathbb{Z}/2$–set $X \in \mathrm{SETS}^{\mathbb{Z}/2}$ can be constructed as follows. For any manifold M, let $\mathrm{Or}_M \to M$ denote the orientation double cover; then $\mathcal{G}_X(M) = \mathrm{maps}_{\mathbb{Z}/2}(\mathrm{Or}_M, X)$, where $\mathrm{maps}_{\mathbb{Z}/2}$ denotes continuous $\mathbb{Z}/2$–equivariant functions. If $\mathcal{X}$ has limits, then for $X \in \mathcal{X}^{\mathbb{Z}/2}$ the formula $\mathrm{maps}_{\mathbb{Z}/2}(\mathrm{Or}_M, X)$ continues to make sense, and again defines the topological local structure corresponding to X.

The most important example is when $X = \mathbb{Z}/2$ is the *trivial* $\mathbb{Z}/2$*–torsor* given by the translation action of $\mathbb{Z}/2$ on itself. Then $\mathcal{G}_{\mathbb{Z}/2} = \mathrm{Or}$ is the sheaf $\mathrm{Or}(M) = \{\text{orientations of } M\}$.

Given a SETS–valued topological local structure $\mathcal{G}$, there is a $\mathcal{G}$*–structured bordism category* $\mathrm{BORD}^{\mathcal{G}}_{d-1,d}$, an object of which consists of a closed $(d-1)$–manifold N

together with an element of $\mathcal{G}(N \times \mathbb{R})$, and whose morphisms are d–dimensional cobordisms similarly equipped with $\mathcal{G}$–structure. If $\mathcal{G}$ is a $\mathrm{S{\scriptsize ETS}}$–valued topological local structure, a *$\mathcal{G}$–structured $\mathbb{R}$–linear d–dimensional field theory* is a symmetric monoidal functor $\mathrm{B{\scriptsize ORD}}^{\mathcal{G}}_{d-1,d} \to \mathrm{V{\scriptsize ECT}}_{\mathbb{R}}$. It will be useful to unpack the construction of $\mathrm{B{\scriptsize ORD}}^{\mathcal{G}}_{d-1,d}$ in order to have a more explicit description of $\mathcal{G}$–structured field theories. The following logic is used in [20, Section 3.2] to reduce the "$\mathcal{G}$–structured cobordism hypothesis" to the unstructured case; see also [22, Section 3.5].

Let $\mathrm{S{\scriptsize PANS}}(\mathrm{S{\scriptsize ETS}})$ denote the symmetric monoidal category whose objects are sets and whose morphisms are isomorphism classes of *correspondences*, ie diagrams of shape $X \leftarrow A \to Y$; composition is by fibered product and the symmetric monoidal structure is by cartesian product. A *$\mathcal{G}$–structured classical field theory* is a symmetric monoidal functor $\mathrm{B{\scriptsize ORD}}^{\mathcal{G}}_{d-1,d} \to \mathrm{S{\scriptsize PANS}}(\mathrm{S{\scriptsize ETS}})$. Every $\mathrm{S{\scriptsize ETS}}$–valued topological local structure $\mathcal{G}$ defines an unstructured classical field theory $\widetilde{\mathcal{G}}\colon \mathrm{B{\scriptsize ORD}}_{d-1,d} \to \mathrm{S{\scriptsize PANS}}(\mathrm{S{\scriptsize ETS}})$:



$$\overset{\widetilde{\mathcal{G}}}{\longmapsto}\quad \begin{tikzcd}[column sep=tiny] & \{\mathcal{G}\text{–structures on } M\} \arrow[dl, "\text{restrict}"'] \arrow[dr, "\text{restrict}"] & \\ \{\mathcal{G}\text{–structures on } N_1\} & & \{\mathcal{G}\text{–structures on } N_2\} \end{tikzcd}$$

Functoriality for $\widetilde{\mathcal{G}}\colon \mathrm{B{\scriptsize ORD}}_{d-1,d} \to \mathrm{S{\scriptsize PANS}}(\mathrm{S{\scriptsize ETS}})$ follows from the sheaf axiom for $\mathcal{G}$.

Unpacking the definitions results in the following:

Lemma 1.3 *Let $\mathrm{S{\scriptsize PANS}}(\mathrm{S{\scriptsize ETS}}; \mathrm{V{\scriptsize ECT}}_{\mathbb{R}})$ denote the symmetric monoidal category whose objects are pairs (X, V) where $X \in \mathrm{S{\scriptsize ETS}}$ and V is a vector bundle over X, and for which a morphism from (X, V) to (Y, W) is an isomorphism class of diagrams $X \xleftarrow{f} A \xrightarrow{g} Y$ together with a vector bundle map $f^*V \to g^*W$. Then a $\mathcal{G}$–structured field theory is the same data as a choice of lift:*

$$\begin{tikzcd} & \mathrm{S{\scriptsize PANS}}(\mathrm{S{\scriptsize ETS}}; \mathrm{V{\scriptsize ECT}}_{\mathbb{R}}) \arrow[d, "\text{forget the } \mathrm{V{\scriptsize ECT}}_{\mathbb{R}}\text{–data}"] \\ \mathrm{B{\scriptsize ORD}}_{d-1,d} \arrow[r, "\widetilde{\mathcal{G}}"] \arrow[ur, dashed] & \mathrm{S{\scriptsize PANS}}(\mathrm{S{\scriptsize ETS}}) \end{tikzcd}$$ □

Suppose that $\mathcal{G}$ is a topological local structure valued not in $\mathrm{S{\scriptsize ETS}}$ but in $\mathrm{S{\scriptsize CH}}_{\mathbb{R}}$. Our strategy will be to take Lemma 1.3 as the model for the definition of $\mathcal{G}$–structured field theory. To do this, note that $\mathrm{V{\scriptsize ECT}}_{\mathbb{R}}$ is naturally an object of $\mathbb{R}$–algebraic geometry. Indeed, there is a stack of categories on $\mathrm{S{\scriptsize CH}}_{\mathbb{R}}$, namely $\mathrm{Q{\scriptsize COH}}\colon \mathrm{Spec}(A) \mapsto \mathrm{M{\scriptsize OD}}_A$, whose category of global sections is nothing but $\mathrm{QCOH}(\mathrm{Spec}(\mathbb{R})) = \mathrm{VECT}_{\mathbb{R}}$. We can therefore define:

Definition 1.4 Let $\mathcal{G}$ be a topological local structure valued in schemes over $\mathbb{R}$, thought of as a classical field theory

$$\widetilde{\mathcal{G}}\colon \mathrm{BORD}_{d-1,d} \to \mathrm{SPANS}(\mathrm{SCH}_{\mathbb{R}}).$$

Let $\mathrm{SPANS}(\mathrm{SCH}_{\mathbb{R}};\mathrm{QCOH})$ denote the symmetric monoidal category whose objects are pairs (X,V) where X is a scheme over $\mathbb{R}$ and $V \in \mathrm{QCOH}(X)$, in which a morphism from (X,V) to (Y,W) is (an isomorphism class of) a correspondence of schemes $X \xleftarrow{f} A \xrightarrow{g} Y$ together with a map of quasicoherent sheaves $f^*V \to g^*W$, in which composition is by fibered product, and in which the symmetric monoidal structure is $\times_{\mathrm{Spec}(\mathbb{R})}$. A *$\mathcal{G}$–structured field theory* is a choice of lift:

$$\begin{tikzcd} & \mathrm{SPANS}(\mathrm{SCH}_{\mathbb{R}};\mathrm{QCOH}) \arrow[d, "\text{forget the QCOH–data}"] \\ \mathrm{BORD}_{d-1,d} \arrow[ur, dashed] \arrow[r, "\widetilde{\mathcal{G}}"] & \mathrm{SPANS}(\mathrm{SCH}_{\mathbb{R}}) \end{tikzcd}$$

Any topological local structure $\mathcal{G}$ valued in SETS defines a topological local structure, which we will also call $\mathcal{G}$, valued in $\mathrm{SCH}_{\mathbb{R}}$, via the symmetric monoidal inclusion $\mathrm{SETS} \hookrightarrow \mathrm{SCH}_{\mathbb{R}}$, $S \mapsto S \times \mathrm{Spec}(\mathbb{R})$. In this case, the notion of $\mathcal{G}$–structured field theory from Definition 1.4 agrees with the usual notion in terms of symmetric monoidal functors $\mathrm{BORD}^{\mathcal{G}}_{d-1,d} \to \mathrm{VECT}_{\mathbb{R}}$, since $\mathrm{QCOH}(S \times \mathrm{Spec}(\mathbb{R})) = \{\text{real vector bundles on } S\}$.

We will focus on four examples of topological local structures $\mathcal{G}$ valued in $\mathrm{SCH}_{\mathbb{R}}$, two of which come from topological local structures valued in SETS. We will unpack a bit about the values of $\mathcal{G}$–structured field theories in all four cases to make everything explicit.

Example 1.5 An *unstructured* or *unoriented* field theory is a "$\mathrm{Spec}(\mathbb{R})$–structured" one, where $\mathrm{Spec}(\mathbb{R})(M) = \mathrm{Spec}(\mathbb{R})$ for all manifolds M. Let Z be an unstructured field theory. If M a closed d–dimensional manifold, then $Z(M) \in \mathcal{O}(\mathrm{Spec}(\mathbb{R})) = \mathbb{R}$. If N is a closed $(d-1)$–dimensional manifold, then $Z(N) \in \mathrm{QCOH}(\mathrm{Spec}(\mathbb{R})) = \mathrm{VECT}_{\mathbb{R}}$. Consider the *macaroni* cobordisms $N \times \supset\colon N \sqcup N \to \varnothing$ and $N \times \subset\colon \varnothing \to N \sqcup N$. The first defines a symmetric pairing $Z(N \times \supset)\colon Z(N) \otimes Z(N) \to \mathbb{R}$ and the second a symmetric copairing $\mathbb{R} \to Z(N) \otimes Z(N)$. The *zig-zag equations* $S = /$ and $Z = \backslash$ require this pairing and copairing to be inverse to each other, and are equivalent to making $V = Z(N)$ into a symmetrically self-dual vector space over $\mathbb{R}$, ie we have $\varphi\colon V \xrightarrow{\sim} V^*$ with $\varphi^* \circ \varphi = \mathrm{id}_V$.

Example 1.6 An *oriented* field theory is one with topological local structure $\mathrm{Or} = \mathcal{G}_{\mathbb{Z}/2}$ from Example 1.2, thought of as being valued in $\mathrm{SCH}_{\mathbb{R}}$ via $S \mapsto S \times \mathrm{Spec}(\mathbb{R})$. Orientations are distinguished among all topological local structures by Lemma 1.1: they correspond to the trivial $\mathbb{Z}/2$–torsor. We will review the basic structure enjoyed by an oriented field theory Z.

Let M be a connected closed d–dimensional manifold. Then $Z(M)$ is a function on $\mathrm{Or}(M) \times \mathrm{Spec}(\mathbb{R})$. If M is nonorientable, then $\mathrm{Or}(M) = \varnothing$ and $Z(M)$ is no data. If M is orientable, then $\mathrm{Or}(M) \times \mathrm{Spec}(\mathbb{R}) \cong \mathrm{Spec}(\mathbb{R}) \sqcup \mathrm{Spec}(\mathbb{R})$, the two points corresponding to the two orientations of M, and $Z(M)$ is an element of $\mathcal{O}(\mathrm{Spec}(\mathbb{R}) \sqcup \mathrm{Spec}(\mathbb{R})) = \mathbb{R} \times \mathbb{R}$, ie a pair of numbers (indexed by the two orientations of M).

Suppose now that N is a closed connected $(d-1)$–dimensional manifold. Again if N is nonorientable, $\mathrm{Or}(N)$ is empty and Z assigns no data. If N is orientable, $Z(N)$ is a sheaf on $\mathrm{Or}(N) \times \mathrm{Spec}(\mathbb{R}) \cong \mathrm{Spec}(\mathbb{R}) \sqcup \mathrm{Spec}(\mathbb{R})$, ie a pair (V, V') of real vector spaces, one for each orientation of N. These vector spaces are not independent. Rather, the macaroni cobordisms $N \times \supset\colon N \sqcup N \to \varnothing$ and $N \times \subset\colon \varnothing \to N \sqcup N$ each admit two orientations, which induce orientations of their boundaries such that the two copies of N have opposite orientations. Some definition-unpacking shows that the data of $Z(N \times \supset)$ is nothing but a linear map $V \otimes_{\mathbb{R}} V' \to \mathbb{R}$, and the data of $Z(N \times \subset)$ is a linear map $\mathbb{R} \to V \otimes_{\mathbb{R}} V'$. The zig-zag equations assert that $Z(N \times \supset)$ and $Z(N \times \subset)$ make V and V' into dual vector spaces.

Example 1.7 Lemma 1.1 distinguishes a second topological local structure valued in $\mathrm{SCH}_{\mathbb{R}}$. Specifically, there is a canonical nontrivial $\mathbb{Z}/2$–torsor over $\mathrm{Spec}(\mathbb{R})$, namely $\mathrm{Spec}(\mathbb{C})$ with the complex conjugation action. We will suggestively write $\mathrm{Her}\colon \mathrm{MAN}_d \to \mathrm{SCH}_{\mathbb{R}}$ for this topological local structure, and call $\mathrm{Her}(M)$ the scheme of *hermitian structures* on M. One easily sees that for any manifold M,

$$\mathrm{Her}(M) = \mathrm{Or}(M) \times_{\mathbb{Z}/2} \mathrm{Spec}(\mathbb{C}),$$

where $\mathbb{Z}/2$ acts on $\mathrm{Or}(M)$ by orientation-reversal and on $\mathrm{Spec}(\mathbb{C})$ by complex conjugation, and $\times_{\mathbb{Z}/2}$ denotes the coequalizer of these actions. A hermitian field theory is *étale-locally-oriented* in the sense that Her and Or are both valued in schemes étale over $\mathrm{Spec}(\mathbb{R})$ and are étale-locally isomorphic as topological local structures over $\mathrm{Spec}(\mathbb{R})$, since they pull back to isomorphic topological local structures along $\mathrm{Spec}(\mathbb{C}) \to \mathrm{Spec}(\mathbb{R})$. Since there are precisely two $\mathbb{Z}/2$–torsors over $\mathrm{Spec}(\mathbb{R})$, there are precisely two topological local structures étale-locally isomorphic to Or, ie precisely two kinds of étale-locally-oriented field theory.

We now justify the name "hermitian". Suppose that Z is a Her–structured field theory and M is a closed d–dimensional manifold. If M is not orientable, then $\mathrm{Her}(M) = \varnothing$ is the empty scheme and $Z(M)$ is no data. If M is orientable and nonempty, then $\mathrm{Her}(M)$ is noncanonically isomorphic to the disjoint union of $2^{|\pi_0 M|-1}$ copies of $\mathrm{Spec}(\mathbb{C})$. In particular, if M is connected and orientable, then either orientation of M determines an isomorphism $\mathrm{Her}(M) \cong \mathrm{Spec}(\mathbb{C})$. Thus, either choice of orientation identifies $Z(M) \in \mathcal{O}(\mathrm{Her}(M))$ with a complex number. The two choices of orientation determine isomorphisms that differ by complex conjugation. So one can think of Z as assigning to every oriented manifold a complex number, subject to the condition that orientation-reversal agrees with complex conjugation. Finally, if $M = \varnothing$, then $\mathrm{Her}(M) = \mathrm{Spec}(\mathbb{R})$ and $Z(M) = 1$.

Suppose now that N is a closed connected $(d-1)$–dimensional manifold. Again, if N is nonorientable, then $\mathrm{Her}(N) = \varnothing$ and $Z(N)$ is no data. If N is orientable, $Z(N)$ is a vector bundle on $\mathrm{Her}(N) \cong \mathrm{Spec}(\mathbb{C})$, ie a complex vector space. The values of the macaroni $Z(N \times \supset)$ and $Z(N \times \subset)$ now are bundles of linear maps over $\mathrm{Her}(N \times \supset) \cong \mathrm{Her}(N \times \subset) \cong \mathrm{Spec}(\mathbb{C})$. The domain and codomain of $Z(N \times \supset)$ are given by pulling back $Z(N \sqcup N)$ and $Z(\varnothing)$ along the restrictions

$$\mathrm{Her}(N \times \supset) \to \mathrm{Her}(N \sqcup N) = \mathrm{Her}(N) \times_{\mathrm{Spec}\,\mathbb{R}} \mathrm{Her}(N) \quad \text{and} \quad \mathrm{Her}(N \times \supset) \to \mathrm{Her}(\varnothing),$$

and similarly for $Z(N \times \subset)$. Unpacking gives

$$Z(N \times \supset) \in \hom_{\mathbb{R}}\big(Z(N \times \{\mathrm{pt}\}) \otimes_{\mathbb{R}} Z(N \times \{\mathrm{pt}\}), \mathbb{R}\big) \otimes_{\mathbb{R}} \mathbb{C},$$
$$Z(N \times \subset) \in \hom_{\mathbb{R}}\big(\mathbb{R}, Z(N \times \{\mathrm{pt}\}) \otimes_{\mathbb{R}} Z(N \times \{\mathrm{pt}\})\big) \otimes_{\mathbb{R}} \mathbb{C}.$$

The restriction map

$$\mathrm{Spec}(\mathbb{C}) = \mathrm{Her}(N \times \supset) \to \mathrm{Her}(N) \times_{\mathrm{Spec}\,\mathbb{R}} \mathrm{Her}(N) = \mathrm{Spec}(\mathbb{C}) \times_{\mathrm{Spec}(\mathbb{R})} \mathrm{Spec}(\mathbb{C})$$

is the antidiagonal map $\lambda \mapsto (\lambda, \bar{\lambda})$, and so $Z(N \times \supset)$ is a sesquilinear pairing on $Z(N)$. It follows from the zig-zag equations that $Z(N \times \supset)$ and $Z(N \times \subset)$ identify the $\mathbb{C}$–linear dual vector space $Z(N)^*$ to $Z(N) \in \mathrm{V{\scriptsize ECT}}_{\mathbb{C}}$ with the complex conjugate space $\overline{Z(N)}$. Finally, the symmetry of $N \times \supset$ translates into the requirement that the sesquilinear pairing on $Z(N)$ is symmetric, or equivalently the isomorphism $\varphi \colon Z(N)^* \xrightarrow{\sim} \overline{Z(N)}$ satisfies $\bar{\varphi}^* \circ \varphi = \mathrm{id}$. It is in this sense that hermitian field theories are "hermitian".

Example 1.8 In addition to $\mathrm{Her} \colon \mathrm{M{\scriptsize AN}}_d \to \mathrm{S{\scriptsize CH}}_{\mathbb{R}}$, there is another topological local structure whose value on $\mathbb{R}^d$ is $\mathrm{Spec}(\mathbb{C})$, namely the one corresponding via Lemma 1.1 to $\mathrm{Spec}(\mathbb{C})$ with the trivial $\mathbb{Z}/2$–action. We will simply call this topological local structure $\mathrm{Spec}(\mathbb{C})$. It satisfies $\mathrm{Spec}(\mathbb{C})(M) = \mathrm{Spec}(\mathbb{C})^{\pi_0 M}$ for every manifold M. When one unpacks the notion of $\mathrm{Spec}(\mathbb{C})$–structured field theory, one finds that they are nothing but *complex-linear* unstructured field theories. For example, the values of $\mathrm{Spec}(\mathbb{C})$–structured field theories on closed connected $(d-1)$– and d–dimensional manifolds are objects of $\mathrm{QCOH}(\mathrm{Spec}(\mathbb{C})) = \mathrm{VECT}_{\mathbb{C}}$ and elements of $\mathcal{O}(\mathrm{Spec}(\mathbb{C})) = \mathbb{C}$, respectively.

Example 1.7 provided one of two reasons why hermitian field theories are distinguished: they correspond to the unique nontrivial torsor over $\mathrm{Spec}(\mathbb{R})$ for the group $\mathbb{Z}/2 = \pi_0 \hom_{\mathrm{MAN}_d}(\mathbb{R}^d, \mathbb{R}^d)$. Theorem 0.2 provides the second reason, by asserting that of the two types of étale-locally-oriented field theories, only hermiticity is compatible with reflection-positivity. We now define reflection-positivity and prove Theorem 0.2.

Definition 1.9 A d–dimensional unstructured (ie $\mathrm{Spec}(\mathbb{R})$–structured) field theory $Z\colon \mathrm{BORD}_{d-1,d} \to \mathrm{VECT}_{\mathbb{R}}$ is *reflection-positive* if the nondegenerate symmetric pairing $Z(N \times \supset)\colon Z(N) \otimes Z(N) \to \mathbb{R}$ is positive-definite for every closed $(d-1)$–dimensional manifold N.

Most of the physics literature, including Atiyah's original definition of functorial topological field theory from [1], includes hermiticity directly in the definition of quantum field theory. As such, reflection-positivity is usually posed as the requirement that the hermitian form on the complex vector space $Z(N)$ should be positive-definite. For nontopological quantum field theories defined on Minkowski $\mathbb{R}^{d-1,1}$, reflection-positivity is a stronger condition assuring the existence of an analytic continuation to imaginary time $\mathbb{R}^{d-1} \times i\mathbb{R}_{\geq 0}$, and reflection refers to reflection in the time axis. Positive-definiteness of the Hilbert space is what remains when interpreting this stronger condition for topological field theories.

From the point of view of this paper, the nonhermitian version of reflection-positivity in Definition 1.9 is the most primitive. The hermitian version arises as follows. Suppose first that Z is not hermitian but oriented. One can produce an unstructured field theory $\int_{\mathrm{Or}} Z$ from Z by integrating out the choice of orientation:

$$\int_{\mathrm{Or}} Z\colon M \mapsto \int_{\sigma \in \mathrm{Or}(M)} Z(M, \sigma).$$

Here the integral is a finite sum of numbers when M is d–dimensional and a finite direct sum when $\dim M < d$. In particular, for N a connected $(d-1)$–dimensional manifold, $(\int_{\mathrm{Or}} Z)(N) = Z(N) \oplus Z(N)^*$ with the obvious symmetric pairing.

Let Z^* denote the orientation-reversal of the field theory Z. There is a canonical equivalence $\int_{\mathrm{Or}} Z \cong \int_{\mathrm{Or}} Z^*$. It follows that $\int_{\mathrm{Or}}$ makes sense not just for oriented field theories but for any étale-locally-oriented field theory. Indeed, suppose Z is not oriented but hermitian. Using the isomorphism $\mathrm{Her} \times_{\mathrm{Spec}(\mathbb{R})} \mathrm{Spec}(\mathbb{C}) \cong \mathrm{Or} \times \mathrm{Spec}(\mathbb{C})$, one sees that the base-changed field theory $Z_{\mathbb{C}} = Z \otimes_{\mathbb{R}} \mathbb{C}$ is naturally oriented and $\mathbb{C}$–linear, and so $\int_{\mathrm{Or}} Z_{\mathbb{C}}$ makes sense as a $\mathbb{C}$–linear unstructured field theory. But the hermiticity of Z defines a Galois action on $\int_{\mathrm{Or}} Z_{\mathbb{C}}$, describing how it descends to an $\mathbb{R}$–linear unstructured field theory $\int_{\mathrm{Or}} Z$. One finds that, for Z a hermitian field theory and N a connected $(d-1)$–dimensional manifold, $(\int_{\mathrm{Or}} Z)(N)$ is nothing but the underlying real vector space of the hermitian vector space $Z(N)$; the symmetric pairing is twice the real part of the hermitian pairing on $Z(N)$.

The usual notion of reflection-positivity is then captured by the following:

Definition 1.10 An étale-locally-oriented field theory Z is *reflection-positive* if the unoriented field theory $\int_{\mathrm{Or}} Z$ is reflection-positive. A field theory is *unitary* if it is reflection-positive and hermitian.

With this notion, the proof of Theorem 0.2 is immediate:

Proof of Theorem 0.2 If V is a nonzero real vector space, $V \oplus V^*$ is never positive-definite. □

Remark 1.11 One can also integrate a $\mathrm{Spec}(\mathbb{C})$–structured field theory to a $\mathrm{Spec}(\mathbb{R})$–structured one. One finds that $\int_{\mathrm{Spec}(\mathbb{C})} c = 2\,\mathrm{Re}(c)$ for $c \in \mathcal{O}(\mathrm{Spec}(\mathbb{C}))$, and that the integral of a complex vector space $V \in \mathrm{V{\small ECT}}_{\mathbb{C}}$ is the underlying real vector space of V. If Z is a $\mathrm{Spec}(\mathbb{C})$–structured field theory, then $Z(N \times \supset)$ is a $\mathbb{C}$–linear symmetric pairing on the complex vector space $Z(N)$, and $\int_{\mathrm{Spec}(\mathbb{C})} Z(N \times \supset)$ is twice its real part, thought of as a symmetric pairing on the real vector space $\int_{\mathrm{Spec}(\mathbb{C})} Z(N)$. The real part of a complex-linear symmetric pairing is never positive-definite.

2 A categorified Galois extension

Section 1 illustrated the important role that algebraic geometry and Galois theory play in explaining the origin of hermitian phenomena in quantum field theory. The goal of this section and the next is to tell a similar story concerning super phenomena of fermions and spinors. Explicitly, $\mathbb{C}$ appeared because it is the algebraic closure of $\mathbb{R}$. This section will explain that $\mathrm{S{\small UPER}V{\small ECT}}_{\mathbb{C}}$ is the categorified algebraic closure of $\mathrm{V{\small ECT}}_{\mathbb{R}}$. This is essentially Deligne's "existence of super fiber functors" theorem from [9]. We state this result as Theorem 2.7 and provide details of its proof, as our phrasing is somewhat different from that of [9].

A convenient setting for categorified $\mathbb{R}$–linear algebra is provided by the bicategory $\mathrm{P{\small RES}}_{\mathbb{R}}$ of $\mathbb{R}$–linear locally presentable categories, $\mathbb{R}$–linear cocontinuous functors, and natural transformations: direct sums play the role of addition and quotients play the role of subtraction. Two of the many ways that $\mathrm{P{\small RES}}_{\mathbb{R}}$ is convenient are that it admits all limits and colimits [5] and that it has a natural symmetric monoidal structure $\boxtimes = \boxtimes_{\mathbb{R}}$ satisfying a hom-tensor adjunction [16]. The unit object for $\boxtimes$ is $\mathrm{V{\small ECT}}_{\mathbb{R}}$. Basic examples of $\mathbb{R}$–linear locally presentable categories include the categories $\mathrm{M{\small OD}}_A$ of A–modules for any $\mathbb{R}$–algebra A; the tensor product enjoys $\mathrm{M{\small OD}}_A \boxtimes \mathrm{M{\small OD}}_B \simeq \mathrm{M{\small OD}}_{A\otimes B}$.

Definition 2.1 A *categorified commutative* $\mathbb{R}$*–algebra* is a symmetric monoidal object in $\mathrm{P{\small RES}}_{\mathbb{R}}$.

We embed noncategorified commutative $\mathbb{R}$–algebras among categorified commutative $\mathbb{R}$–algebras with the following lemma, whose proof is a straightforward exercise (see [8, Proposition 2.3.9]):

Lemma 2.2 *The assignment taking a commutative* $\mathbb{R}$*–algebra* R *to the categorified commutative* $\mathbb{R}$*–algebra* $(\mathrm{M{\small OD}}_R, \otimes_R)$ *and an* $\mathbb{R}$*–algebra homomorphism* $f\colon R \to S$ *to extension of scalars* $\otimes_R S\colon \mathrm{M{\small OD}}_R \to \mathrm{M{\small OD}}_S$ *defines a fully faithful embedding of the category of commutative* $\mathbb{R}$*–algebras into the bicategory of categorified commutative* $\mathbb{R}$*–algebras.* □

We turn now to categorifying the notion of algebraic closure. Algebraic closures of fields are determined by a weak universal property ranging over only finite-dimensional algebras. Summarizing the story over $\mathbb{R}$, we have:

Lemma 2.3 (0) $\mathbb{C}$ *is a nonzero finite-dimensional commutative* $\mathbb{R}$*–algebra.*

(1) *Every map* $\mathbb{C} \to A$ *of nonzero finite-dimensional commutative* $\mathbb{R}$*–algebras is an injection.*

(2) *If* A *is a nonzero finite-dimensional commutative* $\mathbb{R}$*–algebra, then there exists a map* $A \to \mathbb{C}$ *of commutative* $\mathbb{R}$*–algebras.*

(3) *Items* (0)–(2) *determine* $\mathbb{C}$ *uniquely up to nonunique isomorphism.* □

Of course, (0)–(1) are equivalent to the statement that $\mathbb{C}$ is a *field*, and (2) is equivalent to the statement that $\mathbb{C}$ is *algebraically closed*. We categorify these notions in turn.

Definition 2.4 A *strongly generating set* in an $\mathbb{R}$–linear locally presentable category $\mathcal{C}$ is a set of objects in $\mathcal{C}$ that generate $\mathcal{C}$ under colimits. The category $\mathcal{C}$ is *finite-dimensional* if it admits a finite strongly generating set $\{C_1, \dots, C_n\}$ such that all hom-spaces between generators $\hom(C_i, C_j)$ are finite-dimensional and moreover every generator C_i is *compact projective* in $\mathcal{C}$, in the sense that $\hom(C_i, -)\colon \mathcal{C} \to \mathrm{VECT}_{\mathbb{R}}$ is cocontinuous.

A categorified commutative $\mathbb{R}$–algebra $(\mathcal{C}, \otimes_{\mathcal{C}}, \dots)$ is *finite-dimensional* as a categorified commutative $\mathbb{R}$–algebra if the underlying $\mathbb{R}$–linear category of $\mathcal{C}$ is finite-dimensional and moreover every projective object $P \in \mathcal{C}$ is dualizable.

Compact projectivity, sometimes called tininess, is a strong but reasonable finiteness condition to impose on an object. There are many definitions of projectivity that agree for abelian categories but diverge for locally presentable but not necessarily abelian categories; ours is one of the stronger possible choices. If $\mathcal{C}$ is a finite-dimensional $\mathbb{R}$–linear locally presentable category, then $\mathcal{C}$ is automatically equivalent to the category MOD_A of modules for a finite-dimensional associative algebra A (eg one can take $A = \mathrm{End}(\bigoplus_i C_i)$).

Finite-dimensionality as a categorified algebra is stronger than just finite-dimensionality of the underlying category. The condition that compact projectivity implies dualizability expresses a compatibility between internal and external notions of finite-dimensionality in a symmetric monoidal category, which otherwise might badly diverge [18]. Indeed, $P \in \mathrm{MOD}_A$ is compact projective exactly when the functor $\otimes_{\mathbb{R}} P\colon \mathrm{VECT}_{\mathbb{R}} \to \mathrm{MOD}_A$ has a right adjoint of the form $\otimes_A P^\vee$ for some left A–module $P^\vee$, whereas, for $(\mathcal{C}, \otimes_{\mathcal{C}}, \dots)$ a symmetric monoidal category, $P \in \mathcal{C}$ is dualizable when the functor $\otimes P\colon \mathcal{C} \to \mathcal{C}$ has a right adjoint of the form $\otimes P^*$ for some $P^* \in \mathcal{C}$.

To check that $(\mathcal{C}, \otimes_{\mathcal{C}}, \dots)$ is finite-dimensional as a categorified commutative $\mathbb{R}$–algebra, it suffices to check that the underlying $\mathbb{R}$–linear category $\mathcal{C}$ is finite-dimensional and that each generator C_i is dualizable.

Definition 2.4 explains how to categorify item (0) from Lemma 2.3. With it in hand, we may categorify the notion of algebraically closed field by following items (1)–(2):

Definition 2.5 A *categorified field* is a nonzero finite-dimensional categorified commutative $\mathbb{R}$–algebra $(\mathcal{C}, \otimes_{\mathcal{C}}, \dots)$ such that every 1–morphism $(\mathcal{C}, \otimes_{\mathcal{C}}, \dots) \to (\mathcal{D}, \otimes_{\mathcal{D}}, \dots)$ of nonzero categorified commutative $\mathbb{R}$–algebras is faithful and injective on isomorphism classes of objects.

A finite-dimensional categorified field $(\mathcal{C}, \otimes, \dots)$ is *algebraically closed* if for every nonzero finite-dimensional categorified commutative $\mathbb{R}$–algebra $(\mathcal{B}, \otimes, \dots)$, there exists a 1–morphism $F\colon (\mathcal{B}, \otimes_{\mathcal{B}}, \dots) \to (\mathcal{C}, \otimes_{\mathcal{C}}, \dots)$ of categorified commutative $\mathbb{R}$–algebras.

Lemma 2.6 *A finite-dimensional commutative $\mathbb{R}$–algebra R is a field if and only if $(\mathrm{MOD}_R, \otimes_R, \dots)$ is a categorified field.*

Proof It is clear that if $(\mathrm{MOD}_R, \otimes_R, \dots)$ is a categorified field, then R is a field, simply by using the faithfulness assumption and 1–morphisms to categorified algebras of the form $(\mathrm{MOD}_S, \otimes_S, \dots)$.

Conversely, suppose R is a field and $F\colon \mathrm{MOD}_R \to \mathcal{C}$ is any $\mathbb{R}$–linear functor. Suppose F is not faithful. Then there is a nonzero morphism $f\colon X \to Y$ in MOD_R with $F(f) = 0$. Using the fact that in MOD_R all exact sequences split, one can show that $F(\mathrm{im}(f)) = 0$, from which it follows that $F(R) = 0$. If F is symmetric monoidal, $F(R) \cong \mathbb{1}_{\mathcal{C}}$ is the monoidal unit in $\mathcal{C}$, and so $\mathcal{C}$ is the zero category. This verifies the faithfulness condition in Definition 2.5.

Suppose that $(\mathcal{C}, \otimes_{\mathcal{C}}, \dots)$ is a finite-dimensional categorified commutative algebra over $\mathbb{R}$, and let $\mathbb{1}_{\mathcal{C}}$ denote its monoidal unit. Any $\lambda \in \mathrm{End}_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}})$ defines a natural endomorphism of the identity functor on $\mathcal{C}$ via

$$\lambda|_X = \lambda \otimes \mathrm{id}_X\colon X = \mathbb{1}_{\mathcal{C}} \otimes_{\mathcal{C}} X \to \mathbb{1}_{\mathcal{C}} \otimes_{\mathcal{C}} X = X,$$

and clearly $\lambda|_{\mathbb{1}} = \lambda$. Since $\mathcal{C}$ is finite-dimensional, it is equivalent to MOD_A for a finite-dimensional associative algebra A; then the algebra of natural endomorphisms of the identity functor is nothing but the center $Z(A) \subseteq A$. It follows that $\mathrm{End}_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}}) \subseteq Z(A)$ is finite-dimensional. Suppose that $\mathbb{1}_{\mathcal{C}} \in \mathcal{C}$ corresponded to an infinite-dimensional A–module M_A. Then $\mathrm{End}_A(M_A) = \mathrm{End}_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}})$ would be infinite-dimensional, as it is the subalgebra of $\mathrm{End}_{\mathbb{R}}(M)$ cut out by finitely many equations (imposing compatibility with multiplication by a basis in the finite-dimensional algebra A). It follows that $\mathbb{1}_{\mathcal{C}}$ corresponds to a finite-dimensional A–module, and so $\mathbb{1}_{\mathcal{C}}$ is a *compact* object in $\mathcal{C}$ in the sense that $\hom_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}}, -)\colon \mathcal{C} \to \mathrm{VECT}_{\mathbb{R}}$ preserves infinite direct sums.

If R is a field, every object in MOD_R is isomorphic to $R^{\oplus\alpha}$ for some cardinal α. Let $F\colon (\mathrm{MOD}_R, \otimes_R, \dots) \to \mathcal{C}$ be a cocontinuous symmetric monoidal functor. On objects it takes $R \in \mathrm{MOD}_R$ to $\mathbb{1}_{\mathcal{C}}$, and so takes $R^{\oplus\alpha}$ to $\mathbb{1}_{\mathcal{C}}^{\oplus\alpha}$. Since $\mathbb{1}_{\mathcal{C}}$ is compact, $\hom_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}}, \mathbb{1}_{\mathcal{C}}^{\oplus\alpha}) = \mathrm{End}_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}})^{\oplus\alpha}$ is $(\dim(\mathrm{End}_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}})) \times \alpha)$–dimensional over $\mathbb{R}$. Since $\dim(\mathrm{End}_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}})) < \infty$, the cardinal α is determined by the cardinal $\dim(\mathrm{End}_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}})) \times \alpha$. This verifies the injectivity-on-objects condition in Definition 2.5. □

We now describe the categorified algebraic closure of $\mathbb{R}$. Recall that the symmetric monoidal category $\mathrm{SUPERVECT}_{\mathbb{C}}$ of *supervector spaces* over $\mathbb{C}$ is by definition equivalent as a monoidal category, but not as a symmetric monoidal category, to the category $\mathrm{REP}_{\mathbb{C}}(\mathbb{Z}/2)$ of complex representations of the group $\mathbb{Z}/2$. Let $\mathbb{J}$ denote the sign representation, also called the *odd line*. In $\mathrm{REP}_{\mathbb{C}}(\mathbb{Z}/2)$, the symmetry $\mathbb{J} \otimes \mathbb{J} \to \mathbb{J} \otimes \mathbb{J}$ is multiplication by $+1$; in $\mathrm{SUPERVECT}_{\mathbb{C}}$ the symmetry is -1. The rest of the symmetry is determined from this law by the axioms of a symmetric monoidal category.

The following is, with just a few changes of context, the main result of [9]; because of these few changes, we review the proof.

Theorem 2.7 $\textsc{SuperVect}_{\mathbb{C}}$ *is the unique (up to nonunique equivalence) finite-dimensional algebraically closed categorified field over* $\mathbb{R}$*.*

Proof To show that $\textsc{SuperVect}_{\mathbb{C}}$ is a categorified field, one proceeds as in the proof of Lemma 2.6. We need the following additional observation. Let $F\colon \textsc{SuperVect}_{\mathbb{C}} \to \mathcal{C}$ be a morphism of finite-dimensional categorified commutative $\mathbb{R}$–algebras, and let $\mathbb{J}_{\mathcal{C}} = F(\mathbb{J})$ denote the image of the odd line. Then $\mathbb{J}_{\mathcal{C}}$ has self-braiding -1 whereas $\mathbb{1}_{\mathcal{C}}$ has self-braiding $+1$, from which it follows that $\mathbb{1}_{\mathbb{C}}$ and $\mathbb{J}_{\mathbb{C}}$ are not isomorphic. On the other hand, tensoring with $\mathbb{J}_{\mathcal{C}}$ induces an autoequivalence of $\mathcal{C}$, and so $\mathbb{J}_{\mathcal{C}}$, like $\mathbb{1}_{\mathcal{C}}$, is compact and nonzero. From these facts, it follows that F is faithful and that one can recover the isomorphism type of an object $V = \mathbb{1}^{\oplus\alpha} \oplus \mathbb{J}^{\oplus\beta} \in \textsc{SuperVect}_{\mathbb{C}}$ from the vector space $\hom_{\mathcal{C}}(\mathbb{1}_{\mathcal{C}} \oplus \mathbb{J}_{\mathcal{C}}, F(V))$.

We next verify that, assuming $\textsc{SuperVect}_{\mathbb{C}}$ is algebraically closed, it is the unique such category. Suppose that $\mathcal{C}$ is another algebraically closed finite-dimensional categorified field over $\mathbb{R}$. Then there are symmetric monoidal functors $\mathcal{C} \to \textsc{SuperVect}_{\mathbb{C}}$ and $\textsc{SuperVect}_{\mathbb{C}} \to \mathcal{C}$, both faithful and injective on objects. Their composition $\textsc{SuperVect}_{\mathbb{C}} \to \mathcal{C} \to \textsc{SuperVect}_{\mathbb{C}}$ is full and essentially surjective as it necessarily takes $\mathbb{1}$ to $\mathbb{1}$ and $\mathbb{J}$ to $\mathbb{J}$. Thus the functor $\mathcal{C} \to \textsc{SuperVect}_{\mathbb{C}}$ is essentially surjective and full (fullness uses that $\mathcal{C} \to \textsc{SuperVect}_{\mathbb{C}}$ is injective on objects).

Finally, we prove that $\textsc{SuperVect}_{\mathbb{C}}$ is algebraically closed. Let $\mathcal{C}$ be a nonzero finite-dimensional categorified commutative $\mathbb{R}$–algebra. We must construct a 1–morphism $\mathcal{C} \to \textsc{SuperVect}_{\mathbb{C}}$. By including $\mathcal{C} \to \mathcal{C} \boxtimes_{\mathbb{R}} \textsc{SuperVect}_{\mathbb{C}}$ if necessary, we may assume without loss of generality that $\mathcal{C}$ receives a 1–morphism $\textsc{SuperVect}_{\mathbb{C}} \to \mathcal{C}$. As above, we will denote the images under this 1–morphism of $\mathbb{1}, \mathbb{J} \in \textsc{SuperVect}_{\mathbb{C}}$ by $\mathbb{1}_{\mathcal{C}}, \mathbb{J}_{\mathcal{C}}$, respectively.

We will need the following notion. Let λ be a partition of $n \in \mathbb{N}$ and V_λ the corresponding irrep of the symmetric group $\mathbb{S}_n$. Recall that, for any $\mathbb{C}$–linear symmetric monoidal category $(\mathcal{C}, \otimes, \dots)$ containing direct sums and splittings of idempotents, the *Schur functor* $S_\lambda\colon \mathcal{C} \to \mathcal{C}$ is the (nonlinear) functor $X \mapsto (X^{\otimes n} \otimes V_\lambda)_{\mathbb{S}_n}$, where $\mathbb{S}_n$ acts on $X^{\otimes n}$ via the symmetry on $\mathcal{C}$, and $(-)_{\mathbb{S}_n}$ denotes the functor of coinvariants. S_λ is natural for symmetric monoidal $\mathbb{C}$–linear functors.

Choose a strong projective generator $P \in \mathcal{C}$. (In the notation of Definition 2.4, one can for example take $P = \bigoplus_i C_i$.) Then the underlying category of $\mathcal{C}$ is equivalent to the category of $\mathrm{End}_{\mathcal{C}}(P)$–modules, and the subcategory of compact objects of $\mathcal{C}$

is the abelian category of finite-dimensional $\mathrm{End}_{\mathcal{C}}(P)$–modules. In particular, every compact object has finite length. As shown in the proof of Lemma 2.6, $\mathbb{1}_{\mathcal{C}}$ is compact, from which it follows that all dualizable objects are compact. Since P is dualizable by assumption, $P^{\otimes n}$ is also dualizable and hence compact.

We claim that there exists some λ such that $S_\lambda(P) = 0$. Indeed, suppose that there were not. Then, as in [9, Paragraph 1.20], the isomorphism $P^{\otimes n} \cong \bigoplus_{|\lambda|=n} V_\lambda \otimes S_\lambda(P)$ would imply that

$$\mathrm{length}(P^{\otimes n}) \geq \sum_{|\lambda|=n} \dim V_\lambda \geq \Big(\sum (\dim V_\lambda)^2\Big)^{1/2} = (n!)^{1/2},$$

which grows more quickly than any geometric series. Suppose that $X, Y, M \in \mathcal{C}$ are compact objects and E is an extension of X by Y. Then, as in [9, Lemma 4.8], right exactness of the tensor functor implies

$$\mathrm{length}(E \otimes M) \leq \mathrm{length}(E \otimes X) + \mathrm{length}(E \otimes Y).$$

From this, the lengths of the tensor products of simple objects, and the fact that finite-dimensional algebras admit only finitely many simple modules, one can bound the growth of $\mathrm{length}(P^{\otimes n})$ by some geometric series.

Given a commutative algebra object $A \in \mathcal{C}$, let $\mathbb{1}_A$ and $\mathbb{J}_A$ denote the images of $\mathbb{1}_{\mathcal{C}}$ and $\mathbb{J}_{\mathcal{C}}$ under the extension-of-scalars functor $\otimes A\colon \mathcal{C} \to \{A\text{–modules in } \mathcal{C}\}$. Note that $\otimes_A$ makes $\{A\text{–modules in } \mathcal{C}\}$ into a categorified commutative $\mathbb{R}$–algebra. Following [9, Proposition 2.9], we will find a nonzero commutative algebra $A \in \mathcal{C}$ such that $P \otimes A \cong \mathbb{1}_A^{\oplus r} \oplus \mathbb{J}_A^{\oplus s}$ for some $r, s \in \mathbb{N}$. Supposing we have done so, let R be the commutative superalgebra whose even part is $\mathrm{End}(\mathbb{1}_A) \cong \mathrm{End}(\mathbb{J}_A)$ and whose odd part is $\hom(\mathbb{J}_A, \mathbb{1}_A) \cong \hom(\mathbb{1}_A, \mathbb{J}_A)$, ie the "endomorphism superalgebra" of $\mathbb{1}_A$. Since P is a compact projective generator of $\mathcal{C}$ and $P \otimes A \cong \mathbb{1}_A^{\oplus r} \oplus \mathbb{J}_A^{\oplus s}$, the symmetric monoidal category $\{A\text{–modules in } \mathcal{C}\}$ is strongly generated as a category by $\mathbb{1}_A$ and $\mathbb{J}_A$, and so is equivalent to the category $\mathrm{S{\scriptsize UPER}M{\scriptsize OD}}_R$ of R–modules in $\mathrm{S{\scriptsize UPER}V{\scriptsize ECT}}_{\mathbb{C}}$; this equivalence is then manifestly symmetric monoidal.

Suppose by induction that we have found a nonzero commutative algebra object $A \in \mathcal{C}$ such that $P \otimes A \cong \mathbb{1}_A^{\oplus r'} \oplus \mathbb{J}_A^{\oplus s'} \oplus P'$ for some $P' \in \{A\text{–modules in } \mathcal{C}\}$. Then P' is a summand of a dualizable object and hence dualizable. If $\mathrm{Sym}^n P' = \bigwedge^n P' = 0$ for all sufficiently large n, then $P' = 0$ by [9, Corollary 1.7 and Lemma 1.17]. If on the other hand $\mathrm{Sym}^n P' \neq 0$ for all n (resp. $\bigwedge^n P' \neq 0$ for all n), then [9, Lemma 2.8], which does not assume the category to be rigid, constructs a nonzero A–algebra A' such that $P' \otimes A' \cong \mathbb{1}_{A'} \oplus P''$ (resp. $P' \otimes A' \cong \mathbb{J}_{A'} \oplus P''$). We iterate, continually splitting off $\mathbb{1}_A$s and $\mathbb{J}_A$s. The iteration must terminate as otherwise $S_\lambda(P) \neq 0$ for all λ [9, Corollary 1.9].

Thus we have found a nonzero commutative superalgebra R and a morphism $\mathcal{C} \to \textsc{SuperMod}_R$ of categorified commutative $\mathbb{R}$–algebras. We can choose a field $\mathbb{L}$ that receives a map from R and extend scalars further so as to build a linear cocontinuous symmetric monoidal functor $\mathcal{C} \to \textsc{SuperVect}_{\mathbb{L}}$. Moreover, since $\mathrm{End}_{\mathcal{C}}(P)$ is finite-dimensional over $\mathbb{C}$, the functor $\mathcal{C} \to \textsc{SuperVect}_{\mathbb{L}}$ factors through $\textsc{SuperVect}_{\mathbb{K}}$ for some intermediate field $\mathbb{C} \subseteq \mathbb{K} \subseteq \mathbb{L}$ which is finite-dimensional over $\mathbb{C}$. But since $\mathbb{C}$ is algebraically closed, the only such field is $\mathbb{K} = \mathbb{C}$. □

Remark 2.8 The fact that $\textsc{SuperVect}_{\mathbb{C}}$ is algebraically closed explains its central role in categorified representation theory [17; 13].

Remark 2.9 The categorified algebraic closure of $\overline{\mathbb{F}}_p$ is not yet known. When $p > 2$, Ostrik [21] conjectures that the answer is a characteristic-p version of quantum SU(2) at level $p-2$ called $\textsc{Ver}_p$. Etingof has conjectured that the categorified algebraic closure of $\overline{\mathbb{F}}_2$ is a nonsemisimple characteristic-2 version of $\textsc{SuperVect}$, described by the triangular Hopf algebra $\overline{\mathbb{F}}_2[x]/(x^2)$ with $\Delta(x) = 1 \otimes x + x \otimes 1$ and R–matrix $R = 1 \otimes 1 + x \otimes x$.

We now use the algebraic closure $\textsc{Vect}_{\mathbb{R}} \to \textsc{SuperVect}_{\mathbb{C}}$ to categorify the notion of torsor over $\mathrm{Spec}(\mathbb{R})$. We first show that $\textsc{Vect}_{\mathbb{R}} \to \textsc{SuperVect}_{\mathbb{C}}$ is "Galois". Let $(\mathcal{C}, \otimes_{\mathcal{C}}, \dots)$ be a categorified commutative $\mathbb{R}$–algebra. A *$\mathcal{C}$–module* is an $\mathbb{R}$–linear locally presentable category $\mathcal{V} \in \textsc{Pres}_{\mathbb{R}}$ together with an action of $\mathcal{C}$ on $\mathcal{V}$ which is cocontinuous in each variable. A morphism of finite-dimensional $\mathcal{C}$–modules is a cocontinuous strong module functor. Since $\mathcal{C}$ is commutative, the bicategory $\mathcal{M}\textsc{od}_{\mathcal{C}}$ of finite-dimensional $\mathcal{C}$–modules carries a symmetric monoidal structure $\boxtimes_{\mathcal{C}}$. See for example Definitions 2.1, 2.6 and 3.2 of [10].

Let $(\mathcal{C}, \otimes_{\mathcal{C}}, \dots) \to (\mathcal{D}, \otimes_{\mathcal{D}}, \dots)$ be a 1–morphism of categorified commutative $\mathbb{R}$–algebras. Such a map makes $(\mathcal{D}, \otimes_{\mathcal{D}}, \dots)$ into a commutative algebra object in $\mathcal{M}\textsc{od}_{\mathcal{C}}$. Let $\mathrm{Aut} = \mathrm{Aut}_{\mathcal{C}}(\mathcal{D})$ denote the group of $\mathcal{C}$–linear symmetric monoidal automorphisms of $\mathcal{D}$. We will denote by $\mathcal{M}\textsc{od}_{\mathcal{D} \rtimes \mathrm{Aut}}$ the bicategory of $\mathcal{D}$–modules equipped with a $\mathcal{C}$–linear Aut–action such that the $\mathcal{D}$–action is Aut–equivariant. It is a symmetric monoidal bicategory with symmetric monoidal structure given by the tensor product of underlying $\mathcal{D}$–modules. The scalar extension functor $\boxtimes_{\mathcal{C}}\mathcal{D}\colon \mathcal{M}\textsc{od}_{\mathcal{C}} \to \mathcal{M}\textsc{od}_{\mathcal{D}}$ factors canonically through $\mathcal{M}\textsc{od}_{\mathcal{D} \rtimes \mathrm{Aut}}$:

$$\mathcal{M}\textsc{od}_{\mathcal{C}} \xrightarrow{\boxtimes_{\mathcal{C}}\mathcal{D}} \mathcal{M}\textsc{od}_{\mathcal{D} \rtimes \mathrm{Aut}} \xrightarrow{\text{forget}} \mathcal{M}\textsc{od}_{\mathcal{D}}.$$

The functor $\boxtimes_{\mathcal{C}}\mathcal{D}\colon \mathcal{M}\textsc{od}_{\mathcal{C}} \to \mathcal{M}\textsc{od}_{\mathcal{D} \rtimes \mathrm{Aut}}$ has a right adjoint $(-)^{\mathrm{Aut}}\colon \mathcal{M}\textsc{od}_{\mathcal{D} \rtimes \mathrm{Aut}} \to \mathcal{M}\textsc{od}_{\mathcal{C}}$ given by taking the Aut–fixed points of a module $\mathcal{V} \in \mathcal{M}\textsc{od}_{\mathcal{D} \rtimes \mathrm{Aut}}$.

Definition 2.10 An extension of categorified fields $(\mathcal{C}, \otimes_{\mathcal{C}}, \dots) \to (\mathcal{D}, \otimes_{\mathcal{D}}, \dots)$ is *Galois* if

$$\boxtimes_{\mathcal{C}}\mathcal{D}\colon \mathcal{M}\mathrm{OD}_{\mathcal{C}} \leftrightarrows \mathcal{M}\mathrm{OD}_{\mathcal{D}\rtimes \mathrm{Aut}} :(-)^{\mathrm{Aut}}$$

is an equivalence of bicategories.

We will prove:

Theorem 2.11 *The extension* $\mathrm{VECT}_{\mathbb{R}} \to \mathrm{SUPERVECT}_{\mathbb{C}}$ *is Galois.*

Remark 2.12 For comparison, the extensions

$$\mathrm{VECT}_{\overline{\mathbb{F}}_p} \to \mathrm{VER}_p \quad \text{and} \quad \mathrm{VECT}_{\overline{\mathbb{F}}_2} \to \mathrm{REP}(\mathbb{F}[x]/(x^2))$$

from Remark 2.9 are not Galois (except for when $p = 3$). Indeed, the latter is "purely inseparable", and the maximal "separable" subextension of $\mathrm{VECT}_{\overline{\mathbb{F}}_p} \to \mathrm{VER}_p$ is $\mathrm{SUPERVECT}_{\overline{\mathbb{F}}_p}$.

We henceforth write $\mathrm{GAL}(\mathbb{R}) = \mathrm{Aut}_{\mathbb{R}}(\mathrm{SUPERVECT}_{\mathbb{C}})$, and call it the *categorified absolute Galois group of* $\mathbb{R}$. We first calculate it:

Lemma 2.13 *The categorified absolute Galois group of* $\mathbb{R}$ *is* $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$.

Proof Since categorified commutative $\mathbb{R}$–algebras form a bicategory, $\mathrm{GAL}(\mathbb{R})$ is a group object in homotopy 1–types. A symmetric monoidal autoequivalence of $\mathrm{SUPERVECT}_{\mathbb{C}}$ consists of a functor $F\colon \mathrm{SUPERVECT}_{\mathbb{C}} \to \mathrm{SUPERVECT}_{\mathbb{C}}$ and some compatible isomorphisms. We can canonically trivialize the isomorphisms $F(\mathbb{1}) \cong \mathbb{1}$ and $F(\mathbb{1} \otimes X) \cong \mathbb{1} \otimes F(X)$, and so the only remaining datum is an isomorphism $\phi\colon F(\mathbb{J} \otimes \mathbb{J}) \xrightarrow{\sim} F(\mathbb{J}) \otimes F(\mathbb{J})$, of which there are $\mathbb{C}^\times$–many. The functor F admits symmetric monoidal natural automorphisms that are trivial on $\mathbb{1}$ but act on $\mathbb{J}$ by $\alpha \in \mathbb{C}^\times$. Under such an automorphism, the map ϕ transforms to $\phi\alpha^2$. Thus we find that

$$\mathrm{Aut}(F) \cong \ker\bigl(\mathbb{C}^\times \xrightarrow{\alpha\mapsto\alpha^2} \mathbb{C}^\times\bigr) \cong \mathbb{Z}/2.$$

F induces an automorphism of $\mathbb{C} = \mathrm{End}(\mathbb{1})$. If this is the identity, then F is monoidally equivalent to the identity; otherwise, F is monoidally equivalent to extension of scalars along the complex conjugation map $\mathbb{C} \to \mathbb{C}$. Thus $\pi_0(\mathrm{GAL}(\mathbb{R})) \cong \mathbb{Z}/2$, and the computation above shows that each connected component is a $\mathrm{B}(\mathbb{Z}/2)$. These fit together via the Galois action of $\mathbb{Z}/2$ on

$$\ker\bigl(\mathbb{C}^\times \xrightarrow{\alpha\mapsto\alpha^2} \mathbb{C}^\times\bigr),$$

and so $\mathrm{GAL}(\mathbb{R})$ is a split extension $\mathbb{Z}/2 \ltimes \mathrm{B}(\mathbb{Z}/2)$. Direct calculation verifies that it is the trivial extension; one can also show via standard techniques that there are no nontrivial split extensions of $\mathbb{Z}/2$ by $\mathrm{B}(\mathbb{Z}/2)$. $\square$

The nontrivial element in $\pi_1 \mathrm{B}(\mathbb{Z}/2)$ acts on $\mathrm{SUPERVECT}_{\mathbb{C}}$ as the natural transformation of the identity commonly called $(-1)^f$, where f stands for "fermion number".

Proof of Theorem 2.11 The bicategory $\mathcal{M}\mathrm{OD}_{\mathrm{VECT}_{\mathbb{R}}}$ is nothing but $\mathrm{PRES}_{\mathbb{R}}$ itself. Given $\mathcal{V} \in \mathrm{PRES}_{\mathbb{R}}$, its image under $\boxtimes_{\mathbb{R}} \mathrm{SUPERVECT}_{\mathbb{C}}$ in $\mathcal{M}\mathrm{OD}_{\mathrm{SUPERVECT}_{\mathbb{C}} \rtimes \mathrm{Aut}}$ can be described as follows. The objects of $\mathcal{V} \boxtimes_{\mathbb{R}} \mathrm{SUPERVECT}_{\mathbb{C}}$ are formal direct sums $V_0 \oplus \mathbb{J} V_1$ where V_0 and V_1 are objects of $\mathcal{V}$. The morphisms are

$$\hom(V_0 \oplus \mathbb{J} V_1, W_0 \oplus \mathbb{J} W_1) = \hom_{\mathcal{V}}(V_0, W_0) \otimes \mathbb{C} \oplus \hom_{\mathcal{V}}(V_1, W_1) \otimes \mathbb{C}.$$

$\mathrm{SUPERVECT}_{\mathbb{C}}$ acts on $\mathcal{V} \boxtimes_{\mathbb{R}} \mathrm{SUPERVECT}_{\mathbb{C}}$ in the obvious way. The action of $\mathrm{Aut}_{\mathbb{R}}(\mathrm{SUPERVECT}_{\mathbb{C}}) = \mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$ is via complex conjugation and $(-1)^f$, just as it is on $\mathrm{SUPERVECT}_{\mathbb{C}}$. The fixed-points of this action are therefore the purely even objects — those of the form $V_0 \oplus \mathbb{J}0$ — equipped with a $\mathbb{C}$–antilinear involutive automorphism of V_0. The fact that $\mathbb{R} \to \mathbb{C}$ is Galois then implies that the composition $(-)^{\mathrm{Aut}} \circ (\boxtimes_{\mathbb{R}} \mathrm{SUPERVECT}_{\mathbb{C}})$ is equivalent to the identity.

It remains to verify that $(\boxtimes_{\mathbb{R}} \mathrm{SUPERVECT}_{\mathbb{C}}) \circ (-)^{\mathrm{Aut}}$ is equivalent to the identity. Let $\mathcal{V}$ be a $\mathrm{SUPERVECT}_{\mathbb{C}}$–module. Then $\mathcal{V}$ comes equipped with an endofunctor $\mathbb{J}\otimes\colon \mathcal{V} \to \mathcal{V}$, given by the action of the odd line $\mathbb{J} \in \mathrm{SUPERVECT}_{\mathbb{C}}$, satisfying $(\mathbb{J}\otimes)^2 \cong \mathrm{id}$, and for each $X \in \mathrm{VECT}_{\mathbb{C}}$ an endofunctor $X\otimes\colon \mathcal{V} \to \mathcal{V}$. The data of an Aut–action on $\mathcal{V}$ compatible with these actions consists of: an endofunctor $V \mapsto \overline{V}$, squaring to the identity, such that $\overline{X \otimes V} \cong \overline{X} \otimes \overline{V}$ for $X \in \mathrm{VECT}_{\mathbb{C}}$; and a natural automorphism θ of the identity functor, squaring to the identity, such that $\theta_{\mathbb{J}\otimes V} = -\mathrm{id}_{\mathbb{J}} \otimes \theta_V$. Let's say that $V \in \mathcal{V}$ is *purely even* if $\theta_V = +1$ and *purely odd* if $\theta_V = -1$. Then every $V \in \mathcal{V}$ canonically decomposes into a direct sum $V = V_0 \oplus V_1$ of purely even and purely odd submodules. The Aut–fixed points are the purely even submodules $V = V_0$ equipped with isomorphisms $V \cong \overline{V}$. Note that $\mathbb{J}\otimes$ interchanges purely even and purely odd objects, and so

$$\begin{aligned}\mathcal{V} &\simeq \{\text{purely even objects in } \mathcal{V}\} \boxplus \{\text{purely odd objects in } \mathcal{V}\} \\ &\simeq \{\text{purely even objects in } \mathcal{V}\} \boxtimes_{\mathbb{C}} \mathrm{SUPERVECT}_{\mathbb{C}}.\end{aligned}$$

Finally, since $\mathbb{R} \to \mathbb{C}$ is Galois, restricting from $\{\text{purely even objects in } \mathcal{V}\}$ to those with $V \cong \overline{V}$ gives an $\mathbb{R}$–linear category whose tensor product with $\mathbb{C}$ is exactly $\{\text{purely even objects in } \mathcal{V}\}$. $\square$

Remark 2.14 Theorem 2.11 implies that the full list of categorified field extensions of $\mathbb{R}$ consists of the familiar categories $\mathrm{VECT}_{\mathbb{R}}$, $\mathrm{VECT}_{\mathbb{C}}$, $\mathrm{SUPERVECT}_{\mathbb{R}}$ and $\mathrm{SUPERVECT}_{\mathbb{C}}$, and a less-familiar category that deserves to be called $\mathrm{SUPERVECT}_{\mathbb{H}}$. The first four are the fixed-points for the obvious subgroups $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$, $\mathrm{B}(\mathbb{Z}/2)$, $\mathbb{Z}/2$ and $\{1\}$ of the categorified Galois group $\mathrm{GAL}(\mathbb{R})$. The last is the fixed-points for the nonobvious inclusion $\mathbb{Z}/2 \hookrightarrow \mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$ which is the identity on the first component and the nontrivial map $\mathbb{Z}/2 \to \mathrm{B}(\mathbb{Z}/2)$ on the second component (corresponding to the nontrivial class in $\mathrm{H}^2(\mathrm{B}(\mathbb{Z}/2); \mathbb{Z}/2)$). As a category, $\mathrm{SUPERVECT}_{\mathbb{H}} \simeq \mathrm{VECT}_{\mathbb{R}} \boxplus \mathrm{MOD}_{\mathbb{H}}$, hence the name. The monoidal structure involves the Morita equivalence $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} \simeq \mathbb{R}$.

We can now categorify the usual classification of torsors in terms of Galois actions.

Definition 2.15 Let G be a finite Picard groupoid. A *categorified G–torsor over $\mathbb{R}$* is a nonzero G–equivariant categorified commutative $\mathbb{R}$–algebra $\mathcal{T}$ such that the functor

$$\mathcal{T} \boxtimes_{\mathbb{R}} \mathcal{T} \to \mathrm{maps}(G, \mathcal{T}), \quad t_1 \boxtimes t_2 \mapsto \big(g \mapsto (g \triangleright t_1) \otimes t_2\big)$$

is an equivalence, where $\mathrm{maps}(G, \mathcal{T})$ denotes the categorified commutative algebra of $\mathcal{T}$–valued functors on the underlying groupoid of G, $\triangleright$ denotes the action of G on $\mathcal{T}$, and $\otimes$ denotes the multiplication in $\mathcal{T}$.

Proposition 2.16 *Let* $\mathrm{GAL}(\mathbb{R}) = \mathrm{Aut}_{\mathbb{R}}(\mathrm{SUPERVECT}_{\mathbb{C}})$ *denote the categorified absolute Galois group of* $\mathbb{R}$. *For each finite categorified group* G, *there is a natural-in-G equivalence*

$$\{\text{categorified } G\text{–torsors over } \mathbb{R}\} \simeq \mathrm{maps}(\mathrm{B}\,\mathrm{GAL}(\mathbb{R}), \mathrm{B}G).$$

The proof is just as in the uncategorified situation:

Proof Let $\mathcal{T}$ be a categorified G–torsor over $\mathbb{R}$. Then $\mathcal{T}' = \mathcal{T} \boxtimes_{\mathbb{R}} \mathrm{SUPERVECT}_{\mathbb{C}}$ is a G–torsor over $\mathrm{SUPERVECT}_{\mathbb{C}}$. Since $\mathrm{SUPERVECT}_{\mathbb{C}}$ is algebraically closed, we can choose a symmetric monoidal functor $F\colon \mathcal{T}' \to \mathrm{SUPERVECT}_{\mathbb{C}}$. Let $\boxtimes'$ denote $\boxtimes_{\mathrm{SUPERVECT}_{\mathbb{C}}}$. The equivalence $\mathcal{T}' \boxtimes' \mathcal{T}' \mapsto \mathrm{maps}(G, \mathcal{T}')$ making $\mathcal{T}'$ into a torsor over $\mathrm{SUPERVECT}_{\mathbb{C}}$ fits into a commutative square

$$\begin{tikzcd}
\mathcal{T}' \boxtimes' \mathcal{T}' \arrow[r, "\sim"] \arrow[d, "\mathrm{id} \boxtimes' F"] & \mathrm{maps}(G, \mathcal{T}') \arrow[d, "F \circ"] \\
\mathcal{T}' \arrow[r, "t \mapsto (g \mapsto F(g \triangleright t))"] & \mathrm{maps}(G, \mathrm{SUPERVECT}_{\mathbb{C}})
\end{tikzcd}$$

in which the downward arrows are both equivalent to $\boxtimes_{\mathcal{T}'}\mathrm{S{\small UPER}V{\small ECT}}_{\mathbb{C}}$. It follows that $\mathcal{T}'$ is a trivial G–torsor over $\mathrm{S{\small UPER}V{\small ECT}}_{\mathbb{C}}$.

Theorem 2.11 then provides an equivalence of homotopy 2–types

$$\{\text{categorified } G\text{–torsors over } \mathbb{R}\} \\ \simeq \big\{\mathrm{G{\small AL}}(\mathbb{R})\text{–actions on the trivial } G\text{–torsor over } \mathrm{S{\small UPER}V{\small ECT}}_{\mathbb{C}} \\ \text{compatible with the action on } \mathrm{S{\small UPER}V{\small ECT}}_{\mathbb{C}}\big\}.$$

But $\mathrm{Aut}_{\mathbb{R}}(\mathrm{maps}(G, \mathrm{S{\small UPER}V{\small ECT}}_{\mathbb{C}})) \simeq G \times \mathrm{G{\small AL}}(\mathbb{R})$, and the equivariance requirement is equivalent to the requirement that the morphism $\mathrm{G{\small AL}}(\mathbb{R}) \to G \times \mathrm{G{\small AL}}(\mathbb{R})$ is the identity on the second component. Therefore we are left with maps $\mathrm{G{\small AL}}(\mathbb{R}) \to G$ up to equivalences given by inner automorphism. □

3 Spin and spin-statistics field theories

With the categorified Galois extension $\mathrm{V{\small ECT}}_{\mathbb{R}} \to \mathrm{S{\small UPER}V{\small ECT}}_{\mathbb{C}}$ from Section 2 in hand, we are equipped to categorify the story from Section 1. The uncategorified story related orientations with hermiticity; the categorified story will relate spin and statistics.

Recall that a *spin structure* on a d–dimensional manifold M is a $\mathrm{Spin}(d)$–principal bundle $P \to M$ together with an isomorphism $P \times_{\mathrm{Spin}(d)} \mathbb{R}^d \cong \mathrm{T}M$. The collection $\mathrm{Spins}(M)$ of spin structures on M is not naturally a set, but rather a groupoid. We therefore extend without further comment the notion of *topological local structure* valued in a bicategory $\mathcal{X}$ to be a sheaf $\mathrm{M{\small AN}}_d \to \mathcal{X}$ that takes homotopies between maps in $\mathrm{M{\small AN}}_d$ to isomorphisms between maps in $\mathcal{X}$ and homotopies between homotopies to equalities between isomorphisms. Generalizing Lemma 1.1, we have:

Lemma 3.1 *Let $\mathcal{X}$ be a bicategory with limits. Topological local structures on $\mathrm{M{\small AN}}_d$ valued in $\mathcal{X}$ are equivalent to objects of $\mathcal{X}$ equipped with an action by the Picard groupoid $\pi_{\leq 1}\hom_{\mathrm{M{\small AN}}_d}(\mathbb{R}^d, \mathbb{R}^d) = \pi_{\leq 1}\mathrm{O}(d)$. When $d \geq 3$, this Picard groupoid is canonically equivalent to $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$.* □

Remark 3.2 The existence of an identification $\pi_{\leq 1}\mathrm{O}(d) \cong \mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$, $d \geq 3$, is the same as the standard assertion that the k-invariant connecting $\pi_1\mathrm{BO}(\infty)$ and $\pi_2\mathrm{BO}(\infty)$ vanishes. However, the group $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$ admits a nontrivial group automorphism, given by the identity on each factor and the nontrivial group map $\mathbb{Z}/2 \to \mathrm{B}(\mathbb{Z}/2)$, corresponding to the nontrivial element of $\mathrm{H}^2(\mathrm{B}(\mathbb{Z}/2); \mathbb{Z}/2) = \mathbb{Z}/2$, mixing the factors. Thus there are two inequivalent identifications $\pi_{\leq 1}\mathrm{O}(\infty) \cong \mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$. To pick one is the same as to pick a splitting of the projection

$\pi_{\leq 1}\mathrm{O}(d) \to \pi_0\mathrm{O}(d) = \mathbb{Z}/2$. There is a canonical choice: the "stable" splitting $\mathbb{Z}/2 \to \mathrm{O}(d)$ sending the nontrivial element of $\mathbb{Z}/2$ to the matrix

$$\begin{pmatrix} -1 & & & \\ & +1 & & \\ & & +1 & \\ & & & \ddots \end{pmatrix},$$

called "T" in the physics literature. Corresponding to the two splittings $\mathbb{Z}/2 \to \pi_{\leq 1}\mathrm{O}(d)$ are two projections $\pi_{\leq 1}\mathrm{O}(d) \to \mathrm{B}(\mathbb{Z}/2)$, the kernels of which are the two *pin* groups $\mathrm{Pin}^{\pm}(d)$.

Example 3.3 We recall two standard facts about spin structures. First, given any spin structure on a d–dimensional manifold M, let $P \to M$ denote the corresponding $\mathrm{Spin}(d)$–bundle. Then $P \times_{\mathrm{Spin}(d)} \mathrm{Pin}^+(d)$ is a $\mathrm{Pin}^+(d)$–bundle over M with a distinguished sheet. The other sheet of $P \times_{\mathrm{Spin}(d)} \mathrm{Pin}^+(d)$ also defines a $\mathrm{Spin}(d)$–bundle over M, corresponding to the *orientation-reversal* of the original spin structure. Second, any spin structure on M admits a square-1 automorphism which acts on the bundle $P \to M$ by multiplication by the nontrivial central element of $\mathrm{Spin}(d)$ coming from 360°–rotation in $\mathrm{SO}(d)$. The mapping cylinder of this automorphism is the product spin manifold $M \times ℓ$, where ℓ denotes the nontrivial-rel-boundary spin structure on the interval $[0,1]$. We will also use the name "ℓ" to denote the automorphism of the spin structure. Together, orientation-reversal and ℓ define an action of $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$ on $\mathrm{Spins}(M)$.

When $M = \mathbb{R}^d$, the orientation reversal and 360°–rotation action of $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$ witness $\mathrm{Spins}(\mathbb{R}^d)$ as the trivial $(\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2))$–torsor. When $d \geq 3$, orientation-reversal and 360°–rotation make up the full group $\pi_{\leq 1}\mathrm{O}(d) = \mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$, and so Spins is the topological local structure corresponding to the trivial $\pi_{\leq 1}\mathrm{O}(d)$–torsor via Lemma 3.1. When $d < 3$, the canonical inclusion $X \mapsto \begin{pmatrix} X & \\ & 1 \end{pmatrix}$ of $\mathrm{O}(d)$ into $\mathrm{O}(3)$ provides an action of $\pi_{\leq 1}\mathrm{O}(d)$ on $\pi_{\leq 1}\mathrm{O}(3) \cong \mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$, which in turn corresponds to the topological local structure Spins.

We now move to an algebrogeometric setting in which there are interesting topological local structures that are étale-locally equivalent to Spins in the way that Her was étale-locally equivalent to Or. Ordinary algebraic geometry does not suffice, since $\mathrm{Spec}(\mathbb{C})$ is étale-contractible in the ordinary sense. Instead, since groupoids are a categorification of sets, we work with a categorification of schemes:

Definition 3.4 The bicategory $\mathrm{CatAffSch}_{\mathbb{R}}$ of *categorified affine schemes over* $\mathbb{R}$ is opposite to the bicategory of categorified commutative $\mathbb{R}$–algebras in the sense of

Definition 2.1. We will write $\mathrm{Spec}(\mathcal{C})$ for the categorified affine scheme corresponding to a categorified commutative algebra $\mathcal{C}$.

Lemma 2.2 provides a fully faithful inclusion of the category $\mathrm{AFFSCH}_{\mathbb{R}}$ of uncategorified affine schemes into $\mathrm{CATAFFSCH}_{\mathbb{R}}$; in particular, we identify $\mathrm{Spec}(\mathbb{R})$ with $\mathrm{Spec}(\mathrm{VECT}_{\mathbb{R}})$. The details of notions like "nonaffine categorified scheme" and "categorified étale topology" have yet to be worked out, and are the subject of joint work in progress by A Chirvasitu, E Elmanto and the author. Theorem 2.11 suggests that $\mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}}) \to \mathrm{Spec}(\mathbb{R})$ is a "categorified étale cover" and Theorem 2.7 suggests that $\mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})$ is "categorified étale contractible". In particular, we will say that categorified affine schemes X and Y are *étale-locally equivalent* if their pullbacks $X \times_{\mathrm{Spec}(\mathbb{R})} \mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})$ and $Y \times_{\mathrm{Spec}(\mathbb{R})} \mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})$ are equivalent as categorified affine schemes over $\mathrm{SUPERVECT}_{\mathbb{C}}$. This in particular implies that for any Picard groupoid G, the geometric notion of categorified G–torsors over $\mathrm{Spec}(\mathbb{R})$, defined as G–objects over $\mathrm{Spec}(\mathbb{R})$ étale-locally equivalent to G acting on itself, agrees with the algebraic notion from Definition 2.15, which by Lemma 2.13 and Proposition 2.16 are classified by maps $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2) \to G$.

Now note the following coincidence: there is a canonical equivalence $\pi_{\leq 1}\mathrm{O}(d) \cong \mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2) \cong \mathrm{GAL}(\mathbb{R})$, and hence a canonical categorified $\pi_{\leq 1}\mathrm{O}(d)$–torsor, when $d \geq 3$. This torsor is nothing but the categorified affine scheme $\mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})$ equipped with its $\mathrm{GAL}(\mathbb{R})$–action.

Definition 3.5 The sheaf of *hermitian spin-statistics structures* is the sheaf

$$\mathrm{HerSpinStats}\colon \mathrm{MAN}_d \to \mathrm{CATAFFSCH}_{\mathbb{R}}$$

such that $\mathrm{HerSpinStats}(\mathbb{R}^d) = \mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})$, on which $\pi_{\leq 1}\mathrm{O}(d) \cong \mathrm{GAL}(\mathbb{R})$ acts via the Galois action.

Lemma 3.6 *For any manifold M,*

$$\mathrm{HerSpinStats}(M) = \frac{\mathrm{Spins}(M) \times \mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})}{\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)},$$

where $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$ acts on $\mathrm{Spins}(M)$ by orientation-reversal and by ➰ from Example 3.3, and it acts on $\mathrm{SUPERVECT}_{\mathbb{C}}$ by complex conjugation and by $(-1)^f$ from Lemma 2.13. $\square$

Lemma 3.6 begins to justify the phrase "hermitian spin-statistics structure" from Definition 3.5: that orientation-reversal acts by complex conjugation is the essence of hermiticity, and that ꝭ acts by $(-1)^f$ is a version of spin-statistics as it is used in physics.

To further justify the name, we should study hermitian spin-statistics field theories directly. The definition of hermitian spin-statistics field theory will be a direct analog of hermitian field theory from Section 1.

Let $\mathrm{BORD}_{d-2,d-1,d}$ denote the once-extended d–dimensional bordism bicategory constructed by Schommer-Pries in [22]. Given any topological local structure valued in groupoids $\mathcal{G}\colon \mathrm{MAN}_d \to \mathrm{GPOIDS}$, [22] also explains how to build a symmetric monoidal bicategory $\mathrm{BORD}^{\mathcal{G}}_{d-2,d-1,d}$ of bordisms with $\mathcal{G}$–structure. A *once-extended $\mathcal{G}$–structured field theory* is then a symmetric monoidal functor $Z\colon \mathrm{BORD}^{\mathcal{G}}_{d-2,d-1,d} \to \mathcal{V}$ for some symmetric monoidal bicategory $\mathcal{V}$ of "categorified vector spaces".

We will take $\mathcal{V} = \mathrm{ALG}_{\mathbb{R}}$ to be the symmetric monoidal "Morita" bicategory of associative algebras, bimodules and intertwiners. Just as $\mathrm{VECT}_{\mathbb{R}}$ had a natural extension to the stack QCOH of categories over $\mathrm{SCH}_{\mathbb{R}}$, so too $\mathrm{ALG}_{\mathbb{R}}$ has a natural extension allowing for "bundles" or "sheaves" of algebras over any categorified affine scheme: given a categorified commutative $\mathbb{R}$–algebra $\mathcal{C}$, set $\mathrm{ALG}(\mathrm{Spec}(\mathcal{C})) = \mathrm{ALG}(\mathcal{C})$ to be the symmetric monoidal bicategory of algebra objects in $\mathcal{C}$, bimodule objects in $\mathcal{C}$, and intertwiners in $\mathcal{C}$. Although we have not defined, and will not use, any topology on $\mathrm{CATAFFSCH}_{\mathbb{R}}$, and so cannot say precisely what it means to be a stack of bicategories, it is not hard to find a bicategory object internal to $\mathrm{CATAFFSCH}_{\mathbb{R}}$ that represents $\mathrm{ALG}(-)$, and so $\mathrm{ALG}(-)$ is certainly a stack of bicategories in any subcanonical topology.

Remark 3.7 The Eilenberg–Watts theorem [11; 26] identifies $\mathrm{ALG}_{\mathbb{R}}$ with the full subbicategory of $\mathrm{PRES}_{\mathbb{R}}$ whose objects admit a compact projective generator. The correct target for once-extended *nontopological* quantum field theory is more likely the larger $\mathrm{PRES}_{\mathbb{R}}$. But it is reasonable to expect that every *topological* field theory factors through $\mathrm{ALG}_{\mathbb{R}}$, since it is expected that only categories equivalent to MOD_A, $A \in \mathrm{ALG}_{\mathbb{R}}$, are sufficiently dualizable (see [6]). Indeed, one should expect more: topological field theories should factor through the subbicategory of $\mathrm{ALG}_{\mathbb{R}}$ whose objects are finite-dimensional algebras and whose morphisms are finite-dimensional bimodules. This subbicategory is equivalent to the bicategory $\mathcal{M}\mathrm{OD}_{\mathrm{PRES}_{\mathbb{R}}}$ of finite-dimensional $\mathrm{PRES}_{\mathbb{R}}$–modules from Section 2. More generally, for $\mathcal{C}$ a finite-dimensional categorified commutative ring, the bicategory $\mathcal{M}\mathrm{OD}_{\mathcal{C}}$ of finite-dimensional $\mathcal{C}$–modules is a subbicategory of $\mathrm{ALG}(\mathcal{C})$, which is a subbicategory of the bicategory of all $\mathcal{C}$–modules.

Definition 3.8 Let $\mathrm{SPANS}_2(\mathrm{CATAFFSCH}_{\mathbb{R}})$ denote the symmetric monoidal bicategory whose objects are categorified affine schemes, 1–morphisms are spans $X \leftarrow A \to Y$, and 2–morphisms are spans-between-spans:

$$\begin{tikzcd} & A \arrow[dl] \arrow[dr] & \\ X & M \arrow[u] \arrow[d] & Y \\ & B \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Composition is by fibered product, and the symmetric monoidal structure is the cartesian product in $\mathrm{CATAFFSCH}_{\mathbb{R}}$. Let $\mathcal{G}$ be a topological local structure valued in $\mathrm{CATAFFSCH}_{\mathbb{R}}$; it defines a symmetric monoidal functor

$$\widetilde{\mathcal{G}}\colon \mathrm{BORD}_{d-2,d-1,d} \to \mathrm{SPANS}_2(\mathrm{CATAFFSCH}_{\mathbb{R}}).$$

Let $\mathrm{SPANS}_2(\mathrm{CATAFFSCH}_{\mathbb{R}};\mathrm{ALG})$ be the symmetric monoidal bicategory whose objects are a categorified affine scheme X together with an algebra $V \in \mathrm{ALG}(X)$, whose 1–morphisms are spans $X \xleftarrow{f} A \xrightarrow{g} Y$ together with a bimodule between f^*V and g^*W in $\mathrm{ALG}(A)$, and whose 2–morphisms are spans of spans together with an intertwiner between pulled-back bimodules. A *$\mathcal{G}$–structured field theory* is a lift:

$$\begin{tikzcd} & \mathrm{SPANS}_2(\mathrm{CATAFFSCH}_{\mathbb{R}};\mathrm{ALG}) \arrow[d, "\text{forget the ALG–data}"] \\ \mathrm{BORD}_{d-2,d-1,d} \arrow[ur, dashed] \arrow[r, "\widetilde{\mathcal{G}}"] & \mathrm{SPANS}_2(\mathrm{CATAFFSCH}_{\mathbb{R}}) \end{tikzcd}$$

Example 3.9 We now continue to justify the name "hermitian spin-statistics" from Lemma 3.6. Let Z be a d–dimensional HerSpinStats–structured field theory. We will unpack its values on various manifolds.

Suppose first that M is a closed d–dimensional manifold. Considered as an element of $\mathrm{BORD}_{d-2,d-1,d}$, M is an endo-2–morphism of the identity 1–morphism of the unit object. Then $Z(M)$ is an endo-2–morphism of the identity 1–morphism of the unit object in $\mathrm{ALG}(\mathrm{HerSpinStats}(M))$, ie a function $Z(M) \in \mathcal{O}(\mathrm{HerSpinStats}(M))$. Any choice of spin structure for M determines a map $\mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}}) \to \mathrm{HerSpinStats}(M)$, and these maps together cover $\mathrm{HerSpinStats}(M)$ as the spin structure varies over M. Thus the data of $Z(M)$ is the data of an element of $\mathcal{O}(\mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})) = \mathbb{C}$ for each spin structure on M. By the construction of HerSpinStats from Lemma 3.6, two spin structures on M with reversed orientation lead to complex-conjugate values of $Z(M)$. This is a manifestation of the hermiticity of Z.

To see spin-statistics phenomena, consider next the case of a closed $(d-1)$–dimensional manifold N. Then $Z(N)$ is an endo-1–morphism of the unit object in the category

$\mathrm{A}\textsc{lg}(\mathrm{HerSpinStats}(N))$, ie a vector bundle on $\mathrm{HerSpinStats}(N)$. Again, any spin structure on N allows this vector bundle to be pulled back to a vector bundle on $\mathrm{Spec}(\mathrm{S}\textsc{uper}\mathrm{V}\textsc{ect}_{\mathbb{C}})$, and so $Z(N)$ assigns a complex supervector space to each spin structure on N. In addition to the hermiticity requirement that orientation-reversed spin structures map to complex-conjugate supervector spaces, there is another relation between these supervector spaces and the spin structures. Indeed, fix a spin structure σ on N, and let $Z(N,\sigma)$ denote the corresponding complex supervector space. Consider the spin cobordism $(N,\sigma)\times \circlearrowleft$. This spin structure picks out a particular map $\mathrm{Spec}(\mathrm{S}\textsc{uper}\mathrm{V}\textsc{ect}_{\mathbb{C}}) \to \mathrm{HerSpinStats}(N\times[0,1])$, along which $Z(N\times[0,1])$ pulls back to a map $Z((N,\sigma)\times \circlearrowleft)\colon Z(N,\sigma)\to Z(N,\sigma)$. But $(N,\sigma)\times \circlearrowleft$ is simply the mapping cylinder of the 360°–rotation of σ, and Lemma 3.6 identifies 360°–rotation with $(-1)^f$. All together, we find that $Z((N,\sigma)\times \circlearrowleft)$ is required to evaluate to $(-1)^f\colon Z(N,\sigma)\to Z(N,\sigma)$.

Similar discussion applies also in codimension-2, and hermitian spin-statistics field theories unpack to spin field theories $\mathrm{B}\textsc{ord}^{\mathrm{Spins}}_{d-2,d-1,d} \to \mathrm{A}\textsc{lg}(\mathrm{S}\textsc{uper}\mathrm{V}\textsc{ect}_{\mathbb{C}})$ such that the actions of $\mathbb{Z}/2\times \mathrm{B}(\mathbb{Z}/2)$ on the source and target categories are intertwined. The phrase "spin-statistics" refers to the identification $\circlearrowleft = (-1)^f$. In a spin field theory the (-1)–eigenstates of $\circlearrowleft$ are called *spinors* and in a super field theory the (-1)–eigenstates of $(-1)^f$ are called *fermions*, so "spin-statistics" can be equivalently described as the assertion that the classes of spinors and fermions agree.

By construction, HerSpinStats is an *étale-locally-spin* topological local structure in the sense that

$$\mathrm{HerSpinStats}\times_{\mathrm{Spec}(\mathbb{R})}\mathrm{Spec}(\mathrm{S}\textsc{uper}\mathrm{V}\textsc{ect}_{\mathbb{C}}) \quad\text{and}\quad \mathrm{Spins}\times\mathrm{Spec}(\mathrm{S}\textsc{uper}\mathrm{V}\textsc{ect}_{\mathbb{C}})$$

are equivalent. Since $\mathrm{G}\textsc{al}(\mathbb{R}) = \mathbb{Z}/2\times\mathrm{B}(\mathbb{Z}/2)$ and Spins corresponds to the trivial $\mathbb{Z}/2\times\mathrm{B}(\mathbb{Z}/2)$–torsor, Proposition 2.16 asserts that the set of inequivalent topological local structures étale-locally-equivalent to Spins is equivalent to

$$\pi_0 \operatorname{maps}\bigl(\mathrm{B}(\mathbb{Z}/2\times\mathrm{B}(\mathbb{Z}/2)), \mathrm{B}(\mathbb{Z}/2\times\mathrm{B}(\mathbb{Z}/2))\bigr),$$

which can be easily computed as

$$\mathrm{H}^1(\mathrm{B}(\mathbb{Z}/2);\mathbb{Z}/2)\times\mathrm{H}^2(\mathrm{B}(\mathbb{Z}/2);\mathbb{Z}/2)\times\mathrm{H}^1(\mathrm{B}^2(\mathbb{Z}/2);\mathbb{Z}/2)\times\mathrm{H}^2(\mathrm{B}^2(\mathbb{Z}/2);\mathbb{Z}/2) \cong (\mathbb{Z}/2)^3,$$

and so there are exactly eight different choices. Whether the corresponding field theories are oriented or hermitian is controlled by the component $\mathbb{Z}/2\to\mathbb{Z}/2$ relating complex conjugation with orientation-reversal. Whether the field theories are spin or spin-statistics is controlled by the component $\mathrm{B}(\mathbb{Z}/2)\to\mathrm{B}(\mathbb{Z}/2)$ relating $(-1)^f$ with $\circlearrowleft$.

But once these choices are made, there is still the choice of map $\mathbb{Z}/2 \to \mathrm{B}(\mathbb{Z}/2)$ — the possible choices are parametrized by $\mathrm{H}^2(\mathbb{Z}/2;\mathbb{Z}/2) \cong \mathbb{Z}/2$ — which adjusts how orientation-reversal behaves on fermions.

There are also topological local structures $\mathcal{G}$ satisfying $\mathcal{G}(\mathbb{R}^d) = \mathrm{Spec}(\textsc{SuperVect}_{\mathbb{C}})$ but in which part or all of $\pi_{\leq 1}\mathrm{O}(d)$ acts trivially, analogous to the $\mathbb{C}$–linear unstructured field theories from Example 1.8. We now illustrate a few of the possible choices to emphasize that spin and statistics are not intrinsically linked, even in the presence of hermiticity. We will then prove Theorem 0.1 showing that spin and statistics are linked when an extra reflection-positivity hypothesis is imposed. In order to construct examples of field theories with various topological local structures, we focus on the case when $d = 2$, since then we can use Schommer-Pries's classification of 2–dimensional field theories from [22].

Example 3.10 A *hermitian spin field theory* is an $\mathbb{R}$–linear field theory with local structure $\mathrm{Spins} \times_{\mathbb{Z}/2} \mathrm{Spec}(\mathbb{C})$. Unpacking the definition, a hermitian spin field theory is a nonsuper $\mathbb{C}$–linear spin field theory such that orientation-reversal agrees with complex conjugation. In terms of simultaneously-spin-and-super field theories, $\circlearrowleft$ acts nontrivially but $(-1)^f$ acts trivially.

Two-dimensional $\mathbb{C}$–linear spin field theories in $\textsc{Alg}$ are classified by finite-dimensional complex semisimple algebras A equipped with a trivialization $\varphi\colon A^* \otimes_A A^* \xrightarrow{\sim} A$ of A–A bimodules, where A^* denotes the linear dual bimodule to A, such that the two maps $\varphi \otimes \mathrm{id}\colon A^* \otimes_A A^* \otimes_A A^* \xrightarrow{\sim} A \otimes_A A^* = A^*$ and $\mathrm{id} \otimes \varphi\colon A^* \otimes_A A^* \otimes_A A^* \xrightarrow{\sim} A^* \otimes_A A = A^*$ agree. The hermiticity requirement unpacks to having a ($\mathbb{C}$–antilinear) *stellar* structure, ie a Morita equivalence $A^{\mathrm{op}} \cong \overline{A}$, where $\overline{A}$ is the complex-conjugate algebra, satisfying certain requirements [22, Section 3.8.6]. Stellar structures are the Morita-equivariant version of $*$–structures, and any $*$–structure defines a stellar structure. Hermiticity requires that φ be real.

For example, we can take $A = \mathbb{C}$ with its standard $*$–algebra structure, and choose the trivialization $\varphi\colon \mathbb{C} = \mathbb{C}^* \otimes_{\mathbb{C}} \mathbb{C}^* \xrightarrow{\sim} \mathbb{C}$ to be multiplication by -1. Either trivialization $\pm\sqrt{-1}\colon \mathbb{C}^* \to \mathbb{C}$ presents the $\mathbb{C}$–linear field theory defined by A as the underlying spin field theory of an oriented field theory over $\mathbb{C}$. But as a hermitian spin theory, the field theory defined by A is fundamentally spin, since neither $\pm\sqrt{-1}$ is real.

Example 3.11 A *hermitian super field theory* is an $\mathbb{R}$–linear field theory with local structure $\mathrm{Or} \times_{\mathbb{Z}/2} \mathrm{Spec}(\textsc{SuperVect}_{\mathbb{C}}) \cong \mathrm{Her} \times \mathrm{Spec}(\textsc{SuperVect}_{\mathbb{R}})$, ie an oriented field theory valued in $\textsc{SuperVect}_{\mathbb{C}}$ such that orientation-reversal agrees with complex conjugation. In terms of simultaneously-spin-and-super field theories, $(-1)^f$ acts nontrivially but $\circlearrowleft$ acts trivially.

Two-dimensional hermitian super field theories are classified by symmetric Frobenius stellar superalgebras. In particular, every symmetric Frobenius $*$–superalgebra determines a hermitian super field theory. Consider the complex superalgebra $\mathbb{C}\mathrm{liff}(2) = \mathbb{C}\langle x, y\rangle/(x^2 = y^2 = 1,\ [x, y] = 0)$, where x and y are odd. It admits a $*$–structure in which $x^* = x\sqrt{-1}$ and $y^* = y\sqrt{-1}$. Then xy is imaginary and even, and $\mathbb{C}\mathrm{liff}(2)$ admits a symmetric Frobenius $*$–superalgebra structure in which $\mathrm{tr}(xy) = \sqrt{-1}$ and $\mathrm{tr}(1) = \mathrm{tr}(x) = \mathrm{tr}(y) = 0$.

As a complex Frobenius superalgebra, $\mathbb{C}\mathrm{liff}(2)$ is Morita-equivalent to $\mathbb{C}$, and so the $\mathbb{C}$–linear oriented super field theory defined by $\mathbb{C}\mathrm{liff}(2)$ is the superification of a purely bosonic theory. But the Morita equivalence $\mathbb{C}\mathrm{liff}(2) \simeq \mathbb{C}$ is not compatible with the stellar structure, and so the corresponding hermitian super field theory defined by $\mathbb{C}\mathrm{liff}(2)$ is fundamentally super.

Example 3.12 Two-dimensional spin-statistics field theories are classified by finite-dimensional semisimple "twisted-symmetric" Frobenius superalgebras. Specifically, let A be a finite-dimensional semisimple superalgebra arising as $Z(\{\mathrm{pt}\})$ for some 2–dimensional field theory. Then 360° rotation acts by the dual bimodule $Z(\circlearrowright) = {}_A A^*_A$. Let ${}_A(-1)^f_A$ denote the bimodule A with actions $a \triangleright m \triangleleft b = am(-1)^{|b|}b$; it is the bimodule corresponding to the algebra automorphism $(-1)^f \colon A \to A$. The spin-statistics data "$\circlearrowright = (-1)^f$" then corresponds to a bimodule isomorphism $\phi\colon {}_A A^*_A \xrightarrow{\sim} {}_A(-1)^f_A$.

Consider the trace $\mathrm{tr}(a) = \langle \phi^{-1}(1_A), a\rangle$, where $\langle\ ,\ \rangle\colon A^* \otimes A \to \mathbb{C}$ denotes the canonical pairing. This trace is not symmetric. In a symmetric Frobenius superalgebra, the trace should satisfy $\mathrm{tr}(ab) = (-1)^{|a|\cdot|b|}\,\mathrm{tr}(ba)$. Instead, the trace pairing above satisfies $\mathrm{tr}(ab) = (-1)^{|a|\cdot(|b|+1)}\,\mathrm{tr}(ba) = \mathrm{tr}(ba)$, where the second equality follows from the fact that tr, being an even map, vanishes on odd elements. Thus not the superalgebra A but rather the underlying nonsuper algebra $\mathrm{Forget}(A)$ is symmetric Frobenius.

Real spin-statistics field theories are classified by twisted-symmetric Frobenius superalgebras in $\mathrm{S\small{UPER}V\small{ECT}}_{\mathbb{R}}$. Hermitian spin-statistics field theories are classified by twisted-symmetric Frobenius stellar superalgebras in $\mathrm{S\small{UPER}V\small{ECT}}_{\mathbb{C}}$, where the isomorphism ϕ is real.

The Clifford algebras $\mathbb{C}\mathrm{liff}(n) = \mathbb{C}\langle x_1, \dots, x_n\rangle/([x_j, x_k] = 2\delta_{jk})$ admit twisted-symmetric Frobenius $*$–superalgebra structures. As in Example 3.11, we can give $\mathbb{C}\mathrm{liff}(n)$ a $*$–structure by declaring $x_j^* = x_j\sqrt{-1}$. When n is odd, there is an isomorphism of superalgebras $\mathbb{C}\mathrm{liff}(n) \cong \mathbb{C}\mathrm{liff}(1) \otimes \mathrm{Mat}_{\mathbb{C}}(2^{(n-1)/2})$, where $\mathrm{Mat}_{\mathbb{C}}(m)$ is the purely-even algebra of $m \times m$ complex matrices, and so we can define the trace $\mathrm{tr}\colon \mathbb{C}\mathrm{liff}(n) \to \mathbb{C}$ to be the matrix trace on the even part (and to vanish on the odd part). This tr is twisted-symmetric and real and so defines a 2–dimensional spin-statistics hermitian field theory.

When n is even, the isomorphism $\mathrm{Forget}(\mathbb{C}\mathrm{liff}(n)) \cong \mathrm{Mat}_{\mathbb{C}}(2^{n/2})$ defines a twisted-symmetric Frobenius structure on $\mathbb{C}\mathrm{liff}(n)$. When n is even, $\mathbb{C}\mathrm{liff}(n)$ also admits a nontwisted symmetric Frobenius structure; Example 3.11 describes the case $n = 2$.

Example 3.13 A *twisted-hermitian spin field theory* is like a hermitian spin field theory except that rather than the canonical action of $\mathbb{Z}/2$ on Spin_s, we twist the action by the nontrivial map $\mathbb{Z}/2 \to \mathrm{B}(\mathbb{Z}/2)$. This unpacks to the requirement that the trivialization φ in Example 3.10 be purely imaginary.

A *twisted-hermitian super field theory* is like a hermitian super field theory except that rather than the canonical action of $\mathbb{Z}/2$ on $\mathrm{SUPERVECT}_{\mathbb{C}}$, we twist the action by the nontrivial map $\mathbb{Z}/2 \to \mathrm{B}(\mathbb{Z}/2)$. These are classified not by symmetric Frobenius stellar superalgebras, but by symmetric Frobenius *twisted-stellar* superalgebras. These are defined analogously to stellar superalgebras but with one modification. For any superalgebra A, consider the superalgebra A' defined by $x \cdot' y = (-1)^{|x|\cdot|y|} xy$. A stellar structure on A includes a Morita equivalence between the opposite superalgebra A^{op} and the complex conjugate superalgebra $\overline{A}$. A twisted-stellar structure instead makes A^{op} equivalent to $\overline{A}'$. A special case is that of twisted-$*$–superalgebras. In a $*$–superalgebra, $x \mapsto x^*$ must be an algebra antiautomorphism, which in $\mathrm{SUPERVECT}_{\mathbb{C}}$ means that $(xy)^* = (-1)^{|x|\cdot|y|} y^* x^*$. In a twisted-$*$–superalgebra, we have instead $(xy)^* = y^* x^*$ for elements of arbitrary parity. Examples of twisted-$*$–superalgebras include $\mathbb{C}\mathrm{liff}(n)$ for arbitrary n with $x_i^* = x_i$.

The nontrivial automorphism of $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$ mentioned in Remark 3.2 defines a second $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$–torsor over $\mathrm{Spec}(\mathbb{R})$ with total space $\mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})$. The corresponding topological local structure controls *twisted-hermitian spin-statistics field theories*. The twisted-$*$–superalgebras $\mathbb{C}\mathrm{liff}(n)$ with their twisted-symmetric Frobenius structures from Example 3.12 provide examples of twisted hermitian spin-statistics field theories.

Twisted-real spin-statistics field theories are classified by twisted-symmetric Frobenius algebra objects in the category $\mathrm{SUPERVECT}_{\mathbb{H}}$ from Remark 2.14.

We now extend the notion of reflection-positivity from Definitions 1.9 and 1.10 to the étale-locally-spin case. Following the physics literature, and in disagreement with Freed and Hopkins [12], we declare that reflection-positivity of an extended field theory can be detected in codimension one:

Definition 3.14 An extended unstructured field theory $Z\colon \mathrm{BORD}_{d-2,d-1,d} \to \mathrm{ALG}_{\mathbb{R}}$ is *reflection-positive* if its restriction $Z|_{\mathrm{BORD}_{d-1,d}}\colon \mathrm{BORD}_{d-1,d} \to \mathrm{VECT}_{\mathbb{R}}$ to an unextended field theory is reflection-positive in the sense of Definition 1.9, ie if the symmetric pairing $Z(N \times \supset)\colon Z(N)^{\otimes 2} \to \mathbb{R}$ is positive-definite for every closed $(d-1)$–dimensional manifold N.

In Definition 1.10 we defined reflection-positivity for étale-locally-oriented field theories in terms of integration over the space of étale-local orientations. We now extend that logic to étale-locally-spin field theories. Consider first the case when Z is spin. For any manifold M, $\mathrm{Spins}(M)$ is a finite groupoid, and so Baez and Dolan [4] define an integration map

$$\int_{\mathrm{Spins}(M)}\colon \mathcal{O}(\mathrm{Spins}(M)) \to \mathbb{R}, \quad f \mapsto \sum_{x \in \pi_0 \mathrm{Spins}(M)} f(x)/|\pi_1(\mathrm{Spins}(M), x)|.$$

When V is a bundle over $\mathrm{Spins}(M)$, $\int_{\mathrm{Spins}(M)} V$ is the space of coinvariants of V.

Example 3.15 Given a two-dimensional nonhermitian spin field theory Z corresponding to the algebra $Z(\{\mathrm{pt}\}) = A$ and trivialization $\varphi\colon A^* \otimes_A A^* \xrightarrow{\sim} A$, one can compute the unstructured field theory $\int_{\mathrm{Spins}} Z$ in two steps. First, one can integrate over the fibers of the projection $\mathrm{Spins} \to \mathrm{Or}$. The corresponding oriented field theory $\int_{\mathrm{Spins}\,/\,\mathrm{Or}} Z$ is controlled by the symmetric Frobenius algebra $B = A \oplus A^*$ with multiplication $(a\oplus\alpha)\cdot(b\oplus\beta) = (ab+\varphi(\alpha\otimes\beta))\oplus(a\beta+\alpha b)$ and Frobenius structure $\mathrm{tr}(a\oplus\alpha) = \alpha(1)$. Second, one can integrate over the choice of orientation, producing the unstructured field theory controlled by $B \oplus B^{\mathrm{op}}$ with the obvious algebraic $*$–structure.

The construction $\int_{\mathrm{Spins}}$ makes sense for any étale-locally-spin field theory: if Z has local structure $\mathcal{G}$ where $\mathcal{G}(\{\mathrm{pt}\})$ is a categorified $\mathbb{Z}/2 \times \mathrm{B}(\mathbb{Z}/2)$–torsor, then the base change $Z_{\mathcal{G}}$ of Z along $\mathcal{G}(\{\mathrm{pt}\}) \to \mathrm{Spec}(\mathbb{R})$ is a Galois-equivariant spin field theory over $\mathcal{G}(\{\mathrm{pt}\})$; thus $\int_{\mathrm{Spins}} Z_{\mathcal{G}}$ is a Galois-equivariant unstructured field theory and so descends to $\mathrm{Spec}(\mathbb{R})$.

Example 3.16 Suppose that Z is a two-dimensional spin-statistics field theory, either hermitian or oriented. In order to treat both oriented and hermitian field theories, we first study the $\mathbb{C}$–linear spin-statistics field theory $Z_{\mathbb{C}} = Z \otimes_{\mathbb{R}} \mathbb{C}$.

As in Example 3.12, $Z_{\mathbb{C}}$ is determined by a finite-dimensional semisimple $\mathbb{C}$–linear superalgebra A together with a bimodule isomorphism $\phi\colon {}_AA^*_A \xrightarrow{\sim} {}_A(-1)^f_A$. Let $\tilde{Z}$ denote the $\textsc{SuperVect}_{\mathbb{C}}$–valued spin field theory determined by A together with the isomorphism

$$\varphi = \phi \otimes \phi\colon A^* \otimes_A A^* \xrightarrow{\sim} (-1)^f \otimes_A (-1)^f \cong A.$$

Integrate $\tilde{Z}$ to a $\textsc{SuperVect}_{\mathbb{C}}$–valued oriented field theory $\int_{\mathrm{Spins}\,/\,\mathrm{Or}} \tilde{Z}$ controlled by the superalgebra algebra $B = A \oplus A^*$. Then $\tilde{Z}$ canonically descends to a nonsuper $\mathbb{C}$–linear oriented field theory $\int_{\mathrm{Spins}\,/\,\mathrm{Or}} Z_{\mathbb{C}}$. Indeed, the isomorphism $\phi\colon A^* \xrightarrow{\sim} (-1)^f_A$

identifies B with the semidirect product for the parity-reversal action $A \rtimes \mathbb{Z}/2 = A \oplus A\epsilon$ where $\epsilon = \epsilon^{-1}$ is even and $\epsilon a = (-1)^{|a|} a \epsilon$. In particular, the bimodule ${}_B(-1)^f_B$ is canonically trivialized by $b \mapsto b\epsilon$. Let $\operatorname{Forget}(B)$ denote the underlying nonsuper algebra of B. The trivialization ${}_B(-1)^f_B \cong {}_B B_B$ determines a Morita equivalence, namely $B/(1-\epsilon) \oplus \Pi B/(1+\epsilon)$, between B and $\operatorname{Forget}(B)$. The nonsuper functor $\int_{\mathrm{Spins}\,/\,\mathrm{Or}} Z_{\mathbb{C}}$ assigns $\operatorname{Forget}(B)$ to the point.

Finally, because the entire construction is equivariant under complex conjugation, if Z was real, then $\int_{\mathrm{Spins}\,/\,\mathrm{Or}} Z_{\mathbb{C}}$ naturally descends to a real oriented field theory, and if Z was hermitian, then $\int_{\mathrm{Spins}\,/\,\mathrm{Or}} Z_{\mathbb{C}}$ is naturally hermitian. Let us describe the hermitian case, as it is the more interesting one. In terms of algebras, if Z was hermitian, then A is stellar. By declaring that ϵ is real, B also becomes stellar, and hence so too is the Morita-equivalent purely even algebra $\operatorname{Forget}(B)$. This stellar structure defines $\int_{\mathrm{Spins}\,/\,\mathrm{Or}} Z$ as a hermitian field theory. In most examples, the stellar structure on A comes from a $*$–structure. In this case, B is also $*$. After tracing through the equivalences, one finds that the induced $*$–structure on $\operatorname{Forget}(B)$ is $b \mapsto b^* \epsilon^{|b|}$.

Remark 3.17 One can also understand Example 3.16 in terms of categories of modules. The Morita class of the superalgebra $A = \tilde{Z}(\{\mathrm{pt}\})$ is determined by the supercategory $\mathcal{A} = \mathrm{S}\textsc{uper}\mathrm{M}\textsc{od}_A$. $\tilde{Z}(\circlearrowright)$ defines an action of $\mathbb{Z}/2$ on $\mathcal{A}$, and $\mathcal{B} = \mathrm{S}\textsc{uper}\mathrm{M}\textsc{od}_B$ is the supercategory of fixed points for this action. Being supercategories, $\mathcal{A}$ and $\mathcal{B}$ carry endo-superfunctors $(-1)^f_{\mathcal{A}}$ and $(-1)^f_{\mathcal{B}}$ which are the identity on objects and even morphisms but act by $(-1)^f$ on odd morphisms. The spin-statistics data $\tilde{Z}(\circlearrowright) \cong (-1)^f_{\mathcal{A}}$ provides a trivialization of $(-1)^f_{\mathcal{B}}$. This is precisely the data needed to factor $\mathcal{B} \simeq \mathcal{B}_{\mathrm{ev}} \boxtimes \mathrm{S}\textsc{uper}\mathrm{V}\textsc{ect}_{\mathbb{C}}$, where $\mathcal{B}_{\mathrm{ev}}$ is the plain category consisting of the even objects of $\mathcal{B}$, ie objects for which the trivialization $(-1)^f_{\mathcal{B}} \cong \mathrm{id}$ acts as the identity.

The Morita equivalence between B and $\operatorname{Forget}(B)$ in Example 3.16 identifies $\mathcal{B}_{\mathrm{ev}}$ with $\mathrm{M}\textsc{od}_{\operatorname{Forget}(B)}$. A straightforward calculation shows that $\mathcal{B}_{\mathrm{ev}}$ is also the underlying nonsuper category $\mathcal{A}_0$ of $\mathcal{A}$, ie the one with the same objects and even morphisms but with odd morphisms forgotten. Since the restriction to $\mathcal{A}_0$ of $(-1)^f_{\mathcal{A}}$ is trivial, and since we started with an isomorphism of superfunctors $(-1)^f_{\mathcal{A}} \cong \circlearrowright$, on the category $\mathcal{A}_0$ we have $\circlearrowright \cong \mathrm{id}$. This is another way to see that $\mathcal{A}_0 = \mathcal{B}_{\mathrm{ev}}$ defines an oriented field theory.

Definition 3.18 An étale-locally-spin field theory Z is *reflection-positive* if the unstructured field theory $\int_{\mathrm{Spins}} Z = \int_{\mathrm{Or}} \int_{\mathrm{Spins}\,/\,\mathrm{Or}} Z$ is reflection-positive.

We can now prove Theorem 0.1, which asserts that all extended étale-locally-spin reflection-positive field theories are hermitian and satisfy spin-statistics.

Proof of Theorem 0.1 An étale-locally-spin field theory either satisfies spin-statistics or is spin-but-not-super. It suffices to show that if Z is a nonzero spin-but-not-super field theory then it is not reflection-positive; hermiticity will follow from Theorem 0.2. (A field theory is *zero* if it sends to the zero object all nonempty cobordisms. The zero field theory is vacuously reflection-positive and makes sense for all topological local structures.)

Suppose that Z is a nonzero spin-but-not-super field theory and consider the $\mathbb{C}$–linear spin-but-not-super field theory $Z_{\mathbb{C}} = Z \otimes_{\mathbb{R}} \mathbb{C}$. Let P be a connected oriented $(d-2)$–dimensional manifold and let $\mathrm{Spins}/\mathrm{Or}(P)$ denote the groupoid of spin structures on P compatible with the chosen orientation. Since Z is nonzero, we can find P such that $Z(P,\sigma) \neq 0$ for at least one $\sigma \in \mathrm{Spins}/\mathrm{Or}(P)$. Then in particular $\mathrm{Spins}/\mathrm{Or}(P) \neq \varnothing$ and so $\mathrm{Spins}/\mathrm{Or}(P)$ is a torsor for $\mathrm{B}(\mathbb{Z}/2) \times \mathrm{H}^1(P;\mathbb{Z}/2)$.

Each choice of $\sigma \in \mathrm{Spins}/\mathrm{Or}(P)$ determines a dimensional reduction of $Z_{\mathbb{C}}$ to the two-dimensional $\mathbb{C}$–linear spin-but-not-super field theory $Z_{\mathbb{C}}(- \times (P,\sigma))$. By the classification of two-dimensional field theories [22], $A = Z_{\mathbb{C}}(\{\mathrm{pt}\} \times (P,\Sigma))$ is a finite-dimensional semisimple algebra over $\mathbb{C}$, and so up to Morita equivalence we can assume $A = \mathbb{C}^{\oplus n}$ for some n. A bimodule isomorphism $A^* \otimes_A A^* \xrightarrow{\sim} A$ cannot permute the direct summands, and so the field theory $Z_{\mathbb{C}}(- \times (P,\sigma))$ is equivalent to a direct sum $\bigoplus_{i=1}^n Y_\sigma^{(i)}$ of complex-linear spin field theories each of which satisfies $A^{(i)} = Y_\sigma^{(i)}(\{\mathrm{pt}\}) = \mathbb{C}$.

The two-dimensional spin-but-not-super field theory $Y_\sigma^{(i)}$ then satisfies

$$\int_{\mathrm{Spins}/\mathrm{Or}} Y_\sigma^{(i)}(\{\mathrm{pt}\}) = A^{(i)} \oplus (A^{(i)})^* = \mathbb{C}[x]/(x^2=1)$$

with $\mathrm{tr}(a+bx) = b$, and the complex Hilbert space is $\int_{\mathrm{Spins}/\mathrm{Or}} Y_\sigma^{(i)}(S^1) = \mathbb{C}^2$ with purely off-diagonal inner product. Thus

$$\int_{\mathrm{Spins}/\mathrm{Or}} Z(P \times S^1) = \int_{\sigma \in \mathrm{Spins}/\mathrm{Or}(P)} \bigoplus\nolimits_i \mathbb{C}^2$$

is a nonzero direct sum of Hilbert spaces with purely off-diagonal inner product. Such an inner product cannot be positive-definite. □

Example 3.19 The hermitian spin-statistics field theory Z_n defined by $\mathbb{C}\mathrm{liff}(n)$ from Example 3.12 is reflection-positive. Indeed, when n is odd, Example 3.16 implies that the hermitian field theory $\int_{\mathrm{Spins}/\mathrm{Or}} Z_n$ is controlled by

$$\mathrm{Forget}(\mathbb{C}\mathrm{liff}(n) \rtimes \mathbb{Z}/2) \cong \mathrm{Mat}_{\mathbb{C}}(2^{(n+1)/2}).$$

As discussed before Definition 1.10, the Hilbert space $\big(\int_{\mathrm{Spins}} Z_n\big)(S^1)$ is then the underlying real vector space of $\big(\int_{\mathrm{Spins}\,/\,\mathrm{Or}} Z_n\big)(S^1) = \mathbb{C}$ equipped with the real part of its hermitian pairing, which comes in turn from the $*$–structure on $\mathrm{Mat}_{\mathbb{C}}(2^{(n+1)/2})$. But $\langle v, v\rangle = |v|^2\langle 1, 1\rangle = |v|^2\,\mathrm{tr}(1) = |v|^2 2^{(n+1)/2} > 0$, so Z_n is reflection-positive. When n is even, $\big(\int_{\mathrm{Spins}} Z_n\big)(S^1) \cong \mathbb{C}^{\oplus 2}$ with its positive-definite hermitian form, where the first copy of $\mathbb{C}$ comes from "a boson on S^1 with its trivial spin structure" and the second from "a fermion on S^1 with its nontrivial spin structure".

When n is even, $\mathbb{C}\mathrm{liff}(n)$ also admits a symmetric Frobenius structure, and so defines a hermitian nonspin super field theory Z_n'. We can mimic Remark 1.11 and integrate over $\mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})$. The corresponding Hilbert space $\big(\int_{\mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})} Z_n'\big)(S^1)$ is again a copy of $\mathbb{C}^{\oplus 2}$, but this time with the indefinite hermitian inner product.

4 Extension to higher categories

This last section explains how to extend the ideas in this paper to the higher-categorical setting championed by Lurie [20]. We will assume familiarity with (∞, n)–categories and give only an outline of the necessary constructions. Following the by-now standard notation in the ∞–categorical literature, we let SPACES denote the ∞–category of topological spaces. For the remainder of this paper, let MAN_d denote the $(\infty, 1)$–category coming from the topological category of d–dimensional smooth manifolds and local diffeomorphisms. Given an $(\infty, 1)$–category $\mathcal{X}$ with limits, a *topological local structure on d–dimensional manifolds valued in $\mathcal{X}$* is a sheaf $\mathcal{G}\colon \mathrm{MAN}_d \to \mathcal{X}$. We will not specify precisely the meaning of "sheaf"; one version is spelled out by Ayala [2]. (Although that paper begins with "geometric" local structures, its main theorem asserts that the cobordism category it constructs from a geometric local structure $\mathcal{F}$ depends only on the corresponding topological local structure $\tau\mathcal{F}$.) We will care most about the case when $\mathcal{X}$ is an ∞–topos, for example the ∞–topos of sheaves of spaces on a site like $\mathrm{SCH}_{\mathbb{R}}$ or $\mathrm{CATAFFSCH}_{\mathbb{R}}$.

Generalizing Lemma 1.1 and Lemma 3.1, the following standard fact follows from the existence of good open covers together with the homotopy equivalence $\mathrm{O}(d) \simeq \hom_{\mathrm{MAN}_d}(\mathbb{R}^d, \mathbb{R}^d)$; see Ayala and Francis [3]:

Lemma 4.1 *The $(\infty, 1)$–category of $\mathcal{X}$–valued topological local structures on d–dimensional manifolds is equivalent to the $(\infty, 1)$–category $\mathcal{X}^{\mathrm{O}(d)}$ of $\mathcal{X}$–objects equipped with an action by the topological group $\mathrm{O}(d)$, the equivalence being given by sending a sheaf $\mathcal{G}\colon \mathrm{MAN}_d \to \mathcal{X}$ to $\mathcal{G}(\mathbb{R}^d) \in \mathcal{X}$.* □

The sheaf corresponding to an object $X \in \mathcal{X}^{\mathrm{O}(d)}$ can be constructed as follows. Given any d–manifold $M \in \mathrm{MAN}_d$, define $X(M) = \mathrm{maps}_{\mathrm{O}(d)}(\mathrm{Fr}_M, X) \in \mathcal{X}$, where $\mathrm{Fr}_M \to M$ denotes the frame bundle, $\mathrm{maps}_{\mathrm{O}(d)}$ denotes $\mathrm{O}(d)$–equivariant maps, and $\mathcal{X}$ is tensored over SPACES since it is an ∞–topos. Then $X(\mathbb{R}^d) \simeq X$ in $\mathcal{X}^{\mathrm{O}(d)}$. A special case is when $X \in \mathcal{X}$ is equipped with the trivial $\mathrm{O}(d)$ action. Then $X(M) \simeq \mathrm{maps}(M, X)$ is the *classical topological sigma-model with target X*.

Example 4.2 Let $G \to \mathrm{O}(d)$ be a map of topological groups. A *G–tangential structure* on a d–dimensional manifold M is a G–principal bundle $P \to M$ with an equivalence $P \times_G \mathrm{O}(d) \simeq \mathrm{Fr}_M$ of $\mathrm{O}(d)$–bundles. The sheaf $\mathrm{MAN}_d \to \mathrm{SPACES}$ of G–tangential structures is classified by the quotient $\mathrm{O}(d)/G$ with its natural $\mathrm{O}(d)$–action.

We now explain how to build an (∞, d)–category $\mathrm{BORD}_d^{\mathcal{G}} = \mathrm{BORD}_{0,\dots,d}^{\mathcal{G}}$ of "$\mathcal{G}$–structured bordisms" for each topological local structure $\mathcal{G}$. For a suitable target $\mathcal{V}$, a *fully-extended $\mathcal{G}$–structured quantum field theory* will be a symmetric monoidal functor $\mathrm{BORD}_d^{\mathcal{G}} \to \mathcal{V}$; the precise statement is in Definition 4.4. We let BORD_d denote the "unstructured" bordism category whose construction is thoroughly outlined by Lurie [20], and for which all details have been provided by Calaque and Scheimbauer [7]. We will not review the construction of BORD_d itself.

Let $\mathcal{Y}$ be an $(\infty, 1)$–category with finite limits; for example, $\mathcal{Y} = \mathcal{X}$ an ∞–topos. For each d, Haugseng [14] builds from $\mathcal{Y}$ a symmetric monoidal (∞, d)–category $\mathrm{SPANS}_d(\mathcal{Y})$. (The case when $\mathcal{Y} = \mathrm{SPACES}$ is outlined, under the name FAM_d, in [20].) The objects of $\mathrm{SPANS}_d(\mathcal{Y})$ are those of $\mathcal{Y}$, but a 1–morphism from X to Y in $\mathrm{SPANS}_d(\mathcal{Y})$ is a *span* $X \leftarrow A \to Y$ in $\mathcal{Y}$, and higher morphisms are spans-between-spans. Following [14], we call symmetric monoidal functors $\mathrm{BORD}_d \to \mathrm{SPANS}_d(\mathcal{Y})$ *classical* (unstructured, fully extended) field theories valued in $\mathcal{Y}$.

Every $\mathcal{Y}$–valued topological local structure $\mathcal{G}$ determines a classical field theory $\widetilde{\mathcal{G}}$ (and the celebrated cobordism hypothesis of [20] implies that all classical field theories arise from topological local structures). Indeed, given a k–dimensional manifold M for $k \le d$, set $\widetilde{\mathcal{G}}(M) = \mathcal{G}(M \times \mathbb{R}^{d-k})$; if M has boundary, first glue on a "collar" $M \rightsquigarrow M \cup_{\partial M} (\partial M \times \mathbb{R}_{\geq 0})$. Then if M is a cobordism from N_1 to N_2, the restriction maps $\mathcal{G}(M) \to \mathcal{G}(N_1)$ and $\mathcal{G}(M) \to \mathcal{G}(N_2)$ make $\mathcal{G}(M)$ into a span, and functoriality for the assignment $\widetilde{\mathcal{G}}\colon \mathrm{BORD}_d \to \mathrm{SPANS}_d(\mathcal{Y})$ follows from the sheaf axiom for $\mathcal{G}$.

Remark 4.3 In the model of BORD_d from [7], k–morphisms are not k–dimensional manifolds, but rather d–dimensional manifolds properly submersed over $\mathbb{R}^{d-k}$. When using that model, one can directly define $\widetilde{\mathcal{G}}\colon \mathrm{BORD}_d \to \mathrm{SPANS}_d(\mathcal{Y})$ simply as $\widetilde{\mathcal{G}}(M) = \mathcal{G}(M)$.

Let $\{\mathrm{pt}\} \in \mathcal{Y}$ denote the terminal object and $\mathcal{Y}_{\{\mathrm{pt}\}/}$ the "undercategory" of *pointed objects* $\{\mathrm{pt}\} \to X$ in $\mathcal{Y}$. The logic of [20] is to construct $\mathrm{BORD}_d^{\mathcal{G}}$ for $\mathcal{G}$ a SPACES–valued topological local structure and then observe that there is a pullback square of symmetric monoidal (∞, d)–categories, where the vertical arrows are the obvious forgetful functors:

$$\begin{tikzcd} \mathrm{BORD}_d^{\mathcal{G}} \arrow[r] \arrow[d, "\text{Forget } \mathcal{G}"'] \arrow[dr, phantom, "\ulcorner", very near start] & \mathrm{SPANS}_d(\mathrm{SPACES}_{\{\mathrm{pt}\}/}) \arrow[d, "\text{forget the pointing}"] \\ \mathrm{BORD}_d \arrow[r, "\mathcal{G}"] & \mathrm{SPANS}_d(\mathrm{SPACES}) \end{tikzcd}$$

Indeed, a $\mathcal{G}$–structured manifold M is nothing but a manifold M together with a pointing of the space $\mathcal{G}(M)$. We will reverse the logic and interpret the above pullback square as the definition of $\mathrm{BORD}_d^{\mathcal{G}}$. Some care must be taken when replacing SPACES by an ∞–topos $\mathcal{X}$, as in general very few objects $X \in \mathcal{X}$ admit "global" points $\{\mathrm{pt}\} \to X$. The correct approach is to work with symmetric monoidal (∞, d)–categories "internal to $\mathcal{X}$"; for the definition, see Haugseng [14] and Li-Bland [19].

Definition 4.4 Let $\mathcal{X}$ be an ∞–topos. By [19, Theorem 4.3], the symmetric monoidal (∞, d)–category $\mathrm{SPANS}_d(\mathcal{X})$ constructed in [14] underlies an internal symmetric monoidal (∞, d)–category in $\mathcal{X}$, which in an abuse of notation we will also call $\mathrm{SPANS}_d(\mathcal{X})$; the same argument implies also that $\mathrm{SPANS}_d(\mathcal{X}_{\{\mathrm{pt}\}/})$ is naturally an internal symmetric monoidal (∞, d)–category in $\mathcal{X}$. Via the unique topos map $\mathrm{SPACES} \to \mathcal{X}$, also view BORD_d as an internal symmetric monoidal (∞, d)–category in $\mathcal{X}$.

Let $\mathcal{G}\colon \mathrm{MAN}_d \to \mathcal{X}$ be an $\mathcal{X}$–valued topological local structure and $\widetilde{\mathcal{G}}\colon \mathrm{BORD}_d \to \mathrm{SPANS}_d(\mathcal{X})$ the corresponding classical field theory. It extends canonically to a functor of internal symmetric monoidal (∞, d)–categories. The (∞, d)–category $\mathrm{BORD}_d^{\mathcal{G}}$ of *$\mathcal{G}$–structured bordisms* is by definition the following pullback of internal symmetric monoidal (∞, d)–categories:

$$\begin{tikzcd} \mathrm{BORD}_d^{\mathcal{G}} \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & \mathrm{SPANS}_d(\mathcal{X}_{\{\mathrm{pt}\}/}) \arrow[d, "\text{forget the pointing}"] \\ \mathrm{BORD}_d \arrow[r, "\mathcal{G}"] & \mathrm{SPANS}_d(\mathcal{X}) \end{tikzcd}$$

Let $\mathcal{V}$ be a symmetric monoidal (∞, d)–category internal to $\mathcal{X}$. A *$\mathcal{G}$–structured field theory valued in $\mathcal{V}$* is a functor $\mathrm{BORD}_d^{\mathcal{G}} \to \mathcal{V}$ of internal symmetric monoidal (∞, d)–categories.

The need to work with internal categories in Definition 4.4 is in some sense unavoidable: "functors internal to $\mathcal{X}$" is the appropriate language with which to impose that a functor be "smooth" for families parametrized by objects of $\mathcal{X}$. But one can also describe $\mathcal{G}$–structured field theories "externally" in terms of the lifting problems in Definition 1.4 and Definition 3.8. Given an ∞–topos $\mathcal{X}$ and a symmetric monoidal (∞, d)–category $\mathcal{V}$ internal to $\mathcal{X}$, the papers [14; 19] construct a symmetric monoidal (∞, d)–category $\mathrm{SPANS}_d(\mathcal{X}; \mathcal{V})$ whose k–morphisms are "bundles of k–morphisms in $\mathcal{V}$ over k–fold spans in $\mathcal{X}$". Such a notion makes sense exactly because $\mathcal{V}$ is internal to $\mathcal{X}$: by definition, a *bundle of k–morphisms in $\mathcal{V}$ over $X \in \mathcal{X}$* is a map from X to the $\mathcal{X}$–object of k–morphisms in $\mathcal{V}$. After unpacking adjunctions, one finds:

Proposition 4.5 *Let $\mathcal{X}$ be an ∞–topos, $\mathcal{G}\colon \mathrm{MAN}_d \to \mathcal{X}$ a topological local structure, and $\mathcal{V}$ a symmetric monoidal (∞, d)–category internal to $\mathcal{X}$. Then the data of a $\mathcal{G}$–structured field theory $\mathrm{BORD}_d^{\mathcal{G}} \to \mathcal{V}$ is the same as the data of a lift:*

$$\begin{tikzcd} & \mathrm{SPANS}_d(\mathcal{X}; \mathcal{V}) \arrow[d, "\text{forget the } \mathcal{V}\text{–data}"] \\ \mathrm{BORD}_d \arrow[r, "\widetilde{\mathcal{G}}"] \arrow[ur, dashed] & \mathrm{SPANS}_d(\mathcal{X}) \end{tikzcd}$$

□

Corollary 4.6 *Let $\mathcal{X}$ be an ∞–topos and $\mathcal{V}$ an internal-to-$\mathcal{X}$ symmetric monoidal (∞, d)–category with duals in the sense of* [14]. *Let $\mathcal{G}\colon \mathrm{MAN}_d \to \mathcal{X}$ be a topological local structure, and $\mathcal{G}(\{\mathrm{pt}\}) = \mathcal{G}(\mathbb{R}^d)$ the corresponding object in $\mathcal{X}^{\mathrm{O}(d)}$. Assuming the cobordism hypothesis, $\mathcal{G}$–structured field theories valued in $\mathcal{V}$ are classified by $\mathrm{O}(d)$–equivariant bundles of $\mathcal{V}$–objects over $\mathcal{G}(\{\mathrm{pt}\})$.* □

We conclude by extending the examples from this paper. Note that under Lemma 4.1, the sheaves Or and Spins of orientations and spin structures correspond, respectively, to the actions of $\mathrm{O}(d)$ on the 0– and 1–truncations $\pi_{\leq 0}\mathrm{O}(\infty)$ and $\pi_{\leq 1}\mathrm{O}(\infty)$, or equivalently to the trivial torsors for these groups. Any ∞–topos $\mathcal{X}$ admits a notion of torsor for topological groups: $X \in \mathcal{X}^G$ is a *G–torsor* if the map $G \times X \to X \times X$, $(g, x) \mapsto (gx, x)$ is an equivalence. An $\mathcal{X}$–valued topological local structure $\mathcal{G}\colon \mathrm{MAN}_d \to \mathcal{X}$ is *locally* Or (resp. *locally* Spins) if $\mathcal{G}(\mathbb{R}^d)$ is a torsor for $\pi_{\leq 0}\mathrm{O}(\infty)$ (resp. $\pi_{\leq 1}\mathrm{O}(\infty)$). Suppose $\mathcal{X}$ is the ∞–topos of sheaves (valued in SPACES) on some site (with some subcanonical topology) containing the category $\mathrm{AFFSCH}_{\mathbb{R}}$ of affine schemes over $\mathbb{R}$. Then there is a canonical $\mathcal{X}$–valued topological local structure Her: $\mathrm{MAN}_d \to \mathcal{X}$ whose value on $\mathbb{R}^d$ is $\mathrm{Spec}(\mathbb{C})$. If $\mathcal{X}$ is the ∞–topos of sheaves on some site containing $\mathrm{CATAFFSCH}_{\mathbb{R}}$, then similarly there is a canonical topological local structure HerSpinStats: $\mathbb{R}^d \mapsto \mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})$.

For $\mathcal{V}$ a suitable target symmetric monoidal (∞, d)–category internal to $\mathcal{X}$, we can then define *hermitian* and *hermitian spin-statistics* field theories as being Her– and HerSpinStats–structured field theories in the sense of Definition 4.4. Note that the details of the ∞–topos $\mathcal{X}$ are largely irrelevant: given Proposition 4.5, what matters for hermitian and spin-statistics field theories are the symmetric monoidal (∞, d)–categories of X–points of $\mathcal{V}$ for X ranging over the possible values $\mathrm{Spec}(\mathbb{R})$, $\mathrm{Spec}(\mathbb{C})$, $\mathrm{Spec}(\mathrm{SUPERVECT}_{\mathbb{C}})$, $\ldots$ of Her and HerSpinStats.

One standard criterion for deciding whether a proposed target $\mathcal{V}$ is suitable is that "near the top" $\mathcal{V}$ should look like VECT. More precisely, any symmetric monoidal (∞, d)–category $\mathcal{V}$ determines a symmetric monoidal $(\infty, 1)$–category $\Omega^{d-1}\mathcal{V}$ of endomorphisms of the identity $(d-2)$–morphism on the identity $(d-3)$–morphism on $\ldots$ on the unit object in $\mathcal{V}$. The passage $\mathcal{V} \mapsto \Omega^{d-1}\mathcal{V}$ makes sense also for internal categories. The "looks like VECT near the top" criterion then says that for R a commutative $\mathbb{R}$–algebra, the $\mathrm{Spec}(R)$–points of $\Omega^{d-1}\mathcal{V}$ should be MOD_R, and that for $\mathcal{C}$ a categorified commutative $\mathbb{R}$–algebra, the $\mathrm{Spec}(\mathcal{C})$–points of $\Omega^{d-1}\mathcal{V}$ should be $\mathcal{C}$ itself. This assures, for example, that if $Z\colon \mathrm{BORD}_d^{\mathrm{Her}} \to \mathcal{V}$ is a fully-extended hermitian field theory, then its restriction

$$Z|_{\mathrm{BORD}^{\mathrm{Her}}_{d-1,d}}\colon \mathrm{BORD}^{\mathrm{Her}}_{d-1,d} = \Omega^{d-1}\mathrm{BORD}_d \to \Omega^{d-1}\mathcal{V} = \mathrm{VECT}$$

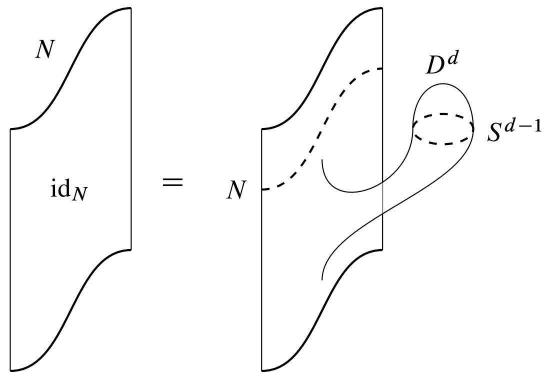
unpacks to a hermitian unextended field theory in the sense of Definition 1.4.

An extended field theory $Z\colon \mathrm{BORD}_d \to \mathcal{V}$ is *reflection-positive* if the unextended field theory $Z|_{\mathrm{BORD}_{d-1,d}}$ is reflection-positive in the sense of Definition 1.9. (This is different from the notion in [12] of reflection-positivity for extended field theories, which requires extra positivity data to be specified in high codimension.) Definition 1.10 and Definition 3.18 then apply to extended hermitian and spin-statistics field theories.

One could worry that restricting a field theory just to its top part is too much loss of information. The following observation is due to Chris Schommer-Pries:

Lemma 4.7 *Let $\mathcal{V}$ be some symmetric monoidal (∞, d)–category with a zero object, and $Z\colon \mathrm{BORD}_d^{\mathcal{G}} \to \mathcal{V}$ be a $\mathcal{G}$–structured extended field theory for some topological local structure $\mathcal{G}$. Suppose that the unextended field theory $Z|_{\mathrm{BORD}^{\mathcal{G}}_{d-1,d}}$ is* zero *in the sense that it vanishes on all nonempty inputs.* (Symmetric monoidality forces $Z(\varnothing)$ to be the unit object of $\Omega^{d-1}\mathcal{V}$.) *Then Z is zero.*

Proof A k–morphism F is zero if and only if its identity $(k+1)$–morphism id_F is zero. It therefore suffices to show that for N an arbitrary $\mathcal{G}$–structured $(d-1)$–dimensional cobordism, $Z(N \times [0,1])\colon Z(N) \to Z(N)$ is the zero d–morphism. But the $\mathcal{G}$–structured cobordism $N \times [0,1]$ can be factored through $N \sqcup S^{d-1}$ where the sphere S^{d-1} is given the $\mathcal{G}$–structure that extends to the disk D^d:



By assumption, $Z(S^{d-1}) = 0$, since S^{d-1} is closed, and so

$$Z(N \sqcup S^{d-1}) \cong Z(N) \otimes Z(S^{d-1}) = 0.$$

Only a zero morphism can factor through a zero object, and so $Z(N \times [0,1]) = 0$. □

Only the zero field theory is compatible with multiple topological local structures. Lemma 4.7 assures that if a $\mathcal{G}$–structured fully extended field theory Z is not zero, then neither is its restriction $Z|_{\mathrm{BORD}^{\mathcal{G}}_{d-2,d-1,d}}$ to a once-extended theory, and so $Z|_{\mathrm{BORD}^{\mathcal{G}}_{d-2,d-1,d}}$ detects the local structure $\mathcal{G}$. Along with Theorem 0.1, we conclude:

Corollary 4.8 *Reflection-positive étale-locally-spin fully-extended field theories are necessarily unitary and satisfy spin-statistics.* □

Acknowledgements I would like to thank K Costello and D Gaiotto for asking the questions that led to this note, D Freed for emphasizing the importance of positivity in the spin-statistics theorem, A Chirvasitu and E Elmanto for many discussions about Galois theory and locally presentable categories, and C Schommer-Pries for providing the proof of Lemma 4.7. I would like particularly to thank the referee for their considerable engagement with the organization and presentation of the ideas in this paper; their comments and suggestions led to many improvements. Some of this work occurred while I was a visitor to the Perimeter Institute for Theoretical Physics in Waterloo, Ontario, where I was treated to generous hospitality and a great work environment. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development & Innovation. This research is also supported by the NSF grant DMS-1304054.

References

[1] **M Atiyah**, *Topological quantum field theories*, Inst. Hautes Études Sci. Publ. Math. 68 (1988) 175–186 MR

[2] **D Ayala**, *Geometric cobordism categories*, PhD thesis, Stanford University (2009) MR Available at `http://search.proquest.com/docview/304999097`

[3] **D Ayala**, **J Francis**, *Factorization homology of topological manifolds*, J. Topol. 8 (2015) 1045–1084 MR

[4] **J C Baez**, **J Dolan**, *From finite sets to Feynman diagrams*, from "Mathematics unlimited—2001 and beyond" (B Engquist, W Schmid, editors), Springer, Berlin (2001) 29–50 MR

[5] **G J Bird**, *Limits in* 2*–categories of locally presentable categories*, PhD thesis, University of Sydney (1984) Available at `http://maths.mq.edu.au/~street/BirdPhD.pdf`

[6] **M Brandenburg**, **A Chirvasitu**, **T Johnson-Freyd**, *Reflexivity and dualizability in categorified linear algebra*, Theory Appl. Categ. 30 (2015) 808–835 MR

[7] **D Calaque**, **C Scheimbauer**, *A note on the* (∞, n)*–category of cobordisms*, preprint (2015) arXiv

[8] **A Chirvasitu**, **T Johnson-Freyd**, *The fundamental pro-groupoid of an affine* 2*–scheme*, Appl. Categ. Structures 21 (2013) 469–522 MR

[9] **P Deligne**, *Catégories tensorielles*, Mosc. Math. J. 2 (2002) 227–248 MR

[10] **C L Douglas**, **C Schommer-Pries**, **N Snyder**, *The balanced tensor product of module categories*, preprint (2014) arXiv

[11] **S Eilenberg**, *Abstract description of some basic functors*, J. Indian Math. Soc. 24 (1960) 231–234 MR

[12] **D S Freed**, **M J Hopkins**, *Reflection positivity and invertible topological phases*, preprint (2016) arXiv

[13] **N Ganter**, **M Kapranov**, *Symmetric and exterior powers of categories*, Transform. Groups 19 (2014) 57–103 MR

[14] **R Haugseng**, *Iterated spans and "classical" topological field theories*, preprint (2014) arXiv

[15] **M Kapranov**, *Supergeometry in mathematics and physics*, preprint (2015) arXiv

[16] **G M Kelly**, *Basic concepts of enriched category theory*, London Math. Soc. Lecture Note Ser. 64, Cambridge University Press (1982) MR

[17] **A Kleshchev**, *Linear and projective representations of symmetric groups*, Cambridge Tracts in Mathematics 163, Cambridge University Press (2005) MR

[18] **L G Lewis, Jr**, *When projective does not imply flat, and other homological anomalies*, Theory Appl. Categ. 5 (1999) 202–250 MR

[19] **D Li-Bland**, *The stack of higher internal categories and stacks of iterated spans*, preprint (2015) arXiv

[20] **J Lurie**, *On the classification of topological field theories*, from "Current developments in mathematics" (D Jerison, B Mazur, T Mrowka, W Schmid, R Stanley, S-T Yau, editors), Int. Press, Somerville, MA (2009) 129–280 MR

[21] **V Ostrik**, *On symmetric fusion categories in positive characteristic*, preprint (2015) arXiv

[22] **C J Schommer-Pries**, *The classification of two-dimensional extended topological field theories*, PhD thesis, University of California, Berkeley (2009) MR Available at `https://arxiv.org/abs/1112.1000`

[23] **G Segal**, *The definition of conformal field theory*, from "Topology, geometry and quantum field theory" (U Tillmann, editor), London Math. Soc. Lecture Note Ser. 308, Cambridge University Press (2004) 421–577 MR

[24] **S Stolz**, **P Teichner**, *Supersymmetric field theories and generalized cohomology*, from "Mathematical foundations of quantum field theory and perturbative string theory" (H Sati, U Schreiber, editors), Proc. Sympos. Pure Math. 83, Amer. Math. Soc., Providence, RI (2011) 279–340 MR

[25] **R F Streater**, **A S Wightman**, *PCT, spin and statistics, and all that*, W A Benjamin, New York (1964) MR

[26] **C E Watts**, *Intrinsic characterizations of some additive functors*, Proc. Amer. Math. Soc. 11 (1960) 5–8 MR

Department of Mathematics, Northwestern University
2033 Sheridan Road, Evanston, IL 60208, United States

`theojf@math.northwestern.edu`

`http://math.northwestern.edu/~theojf`

Received: 28 October 2015 Revised: 6 June 2016

Algebraic & Geometric Topology 17 (2017) 957–982

Hopf ring structure on the mod p cohomology of symmetric groups

LORENZO GUERRA

We describe a Hopf ring structure on $\bigoplus_{n\geq 0} H^*(\Sigma_n;\mathbb{Z}_p)$, discovered by Strickland and Turner, where Σ_n is the symmetric group of n objects and p is an odd prime. We also describe an additive basis on which the cup product is explicitly determined, compute the restriction to modular invariants and determine the action of the Steenrod algebra on our Hopf ring generators. For $p=2$ this was achieved in work of Giusti, Salvatore and Sinha, of which this work is an extension.

20J06

1 Introduction

Let Σ_n be the symmetric group of n objects. Strickland and Turner [8] proved that, for a multiplicative cohomology theory E, the group $A = E\big(\coprod_{n\geq 0} B(\Sigma_n)\big)$ has the structure of a Hopf ring (ie, it admits a coproduct Δ, two products $\odot$ and $\cdot$ and an antipode η, which make it a ring object in the category of coalgebras). Equivalently, the following conditions hold:

- $(A,\Delta,\cdot)$ is a bialgebra.
- $(A,\Delta,\odot,\eta)$ is a Hopf algebra.
- If $\Delta(x)=\sum_i x_i'\otimes x_i''$, then

$$x\cdot(y\odot z)=\sum_i\big[(-1)^{\dim(x_i'')\dim(y)}(x_i'\cdot y)\odot(x_i''\cdot z)\big].$$

Explicitly, the structural maps are defined as follows. The obvious monomorphisms $i_{n,m}\colon \Sigma_n\times\Sigma_m\to\Sigma_{n+m}$ determine the maps $B(\Sigma_n)\times B(\Sigma_m)\to B(\Sigma_{m+n})$, homotopy equivalent to finite coverings. Passing to cohomology and taking their direct sum yields the coproduct Δ. Additionally, $i_{n,m}$ also determines a transfer homomorphism $\mathrm{tr}_{n,m}\colon H^*(\Sigma_n;\mathbb{Z}_p)\otimes H^*(\Sigma_m;\mathbb{Z}_p)\to H^*(\Sigma_{n+m};\mathbb{Z}_p)$. The product $\odot$ is given by $\bigoplus_{n,m\geq 0}\mathrm{tr}_{n,m}$. The product $\cdot$ is the usual cup product. Finally, η is induced by the additive inverse of the sphere spectrum by applying the extended power functor and then cohomology (see [8, pages 140–142]).

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.957

Giusti, Salvatore and Sinha [3] have studied this structure for the ordinary cohomology with coefficients in $\mathbb{Z}_2$ and constructed the following:

- An explicit presentation, in terms of generators and relations, of this Hopf ring.
- An additive basis for the mod 2 cohomology of the symmetric groups in which the products $\cdot$ and $\odot$ and the coproduct Δ defined above can be computed by an explicit rule.

In this presentation, the relations involve only the $\odot$ product. For this reason, all the relations for the cup product in the cohomology of symmetric groups follow, in the mod 2 case, from Hopf ring distributivity. In addition, the authors calculated the restriction to the Dickson invariants and the action of the Steenrod algebra on these groups.

The purpose of this paper is to study the algebraic structure of the cohomology rings $H^*(\Sigma_n;\mathbb{Z}_p)$, where p is an odd prime, as well as the derivation of the mod p analogs of Giusti, Salvatore and Sinha's results. In particular, following their work, we will write a presentation of the Hopf ring $H^*\bigl(\coprod_{n\geq 0} B(\Sigma_n);\mathbb{Z}_p\bigr)$.

The generalizations to the mod p case required overcoming some complications in calculations, especially at odd degrees and when dealing with the more complicated coefficients arising in the description of the Steenrod algebra action. The main differences with the mod 2 case are the following:

- To obtain their Hopf ring presentation, Giusti, Salvatore and Sinha needed to relate the linear duals of $\cdot$, $\odot$ and Δ to the Dyer–Lashof operations. Then they used Nakaoka's description of $H_*(\Sigma_n;\mathbb{Z}_2)$ and dualized to obtain results in cohomology. In the mod p case the need to treat the Bockstein homomorphism separately yields a more complicated structure for the dual of the Dyer–Lashof algebra, which is not a polynomial algebra as in the mod 2 case. This forces us, in the presentation of the Hopf ring $\bigoplus_n H^*(\Sigma_n;\mathbb{Z}_p)$, to use more generators and some nontrivial relations involving the cup product.
- Consider in the cohomology groups $H^*(\Sigma_{2^n};\mathbb{Z}_2)$ the linear duals of the Dyer–Lashof operations with respect to the Nakaoka monomial basis in homology. It is known that the restriction homomorphism onto the ring of Dickson invariants $D_n=\mathbb{Z}_2[x_1,\dots,x_n]^{\mathrm{GL}_n(\mathbb{Z}_2)}$ maps the subalgebra generated by those dual elements surjectively onto D_n. In [3], the computation of the restriction of the Hopf ring generators to D_n relies on this fact. For mod p coefficients this is no longer true; hence, we needed to use a different technique to achieve this goal.

Apart from this introduction, this paper is organized into five sections. In Section 2 we describe a presentation, with generators and relations, of the mod p cohomology of the symmetric groups as a Hopf ring, obtaining the mod p analog of the main theorem in [3]. In Section 3 we obtain an additive basis with a rule for computing the products. In Section 4 we carry out the calculation of the restriction of our Hopf ring generators to the Dickson–Mùi invariant algebras. This will be crucial to the computation of the Steenrod algebra action, which is explained in Section 5. In Section 6 we use our Hopf ring presentation to describe the cup product structure for $H^*(\Sigma_{p^2};\mathbb{Z}_p)$.

Acknowledgments The author would like to thank Professor Frederick Cohen and Professor Dev Sinha for valuable conversations, Professor Mario Salvetti for his guidance, and the referee for helpful comments and suggestions.

2 Hopf ring structure

In this section, we describe $A=\bigoplus_{n\geq 0} H^*(\Sigma_n;\mathbb{Z}_p)$ as a Hopf ring.

Theorem 2.1 [8, Theorem 3.2] *A, with the coproduct Δ, the two products $\odot$ and $\cdot$ and the antipode η described in the introduction, is a Hopf ring.*

We need to describe the homology $H=\bigoplus_{n\geq 0} H_*(\Sigma_n;\mathbb{Z}_p)$, dual to A. In order to establish the notation, we recall the Dyer–Lashof operations, acting on the homology of the symmetric groups. A complete treatment of these operations can be found in Cohen, Lada and May [2], to which we refer for details and proofs. Given a group G, its classifying space is denoted by $B(G)$, its total space (ie a contractible topological space with a free G–action) by $E(G)$. Suppose that X is a space, and we are given a map $\theta\colon E(\Sigma_p)\times_{\Sigma_p} X^p\to X$, where Σ_p acts on X^p by permuting the p factors. Let π_p be a cyclic group of order p, considered a subgroup of Σ_p in the obvious way. Let W_* be the standard resolution of $\mathbb{Z}_p$ with $\mathbb{Z}_p[\pi_p]$–free modules. We can consider the composition map

$$\Theta\colon H_*(W_*\otimes_{\pi_p} C_*(X)^{\otimes p})\to H_*(E(\Sigma_p)\times_{\Sigma_p} X^p;\mathbb{Z}_p)\xrightarrow{\theta_*} H_*(X;\mathbb{Z}_p).$$

For every $i\geq 0$ and $c\in H_d(X;\mathbb{Z}_p)$, we define

$$Q_i(c)=\Theta(e_i\otimes_{\pi_o} c^{\otimes p})\in H_{i+pd}(X;\mathbb{Z}_p),$$

where e_i is the standard generator of W_i.

When θ arises from an action of an E_∞–operad $\mathcal{C}$ on X, Q_i is different from 0 on $H_q(X;\mathbb{Z}_p)$ only if i is congruent to $q(p-1)$ or to $q(p-1)-1$ modulo $2(p-1)$ and $Q_{k(p-1)-1}(x)=\beta Q_{k(p-1)}(x)$, where β is the homology Bockstein homomorphism. Hence, by making a change of indices and defining

$$Q^i=(-1)^{i+\frac{q(q-1)(p-1)}{4}}\left(\tfrac{1}{2}(p-1)!\right)^q Q_{(2i-q)(p-1)}\colon H_q(X;\mathbb{Z}_p)\to H_{q+2i(p-1)}(X;\mathbb{Z}_p),$$

we see that the Q^i and βQ^i generate all the nontrivial operations.

In the category of $\mathcal{C}$–spaces, these operations also satisfy the following properties (see Cohen, Lada and May [2, Theorem 1.1, page 5]):

- Let $*$ denote the product in the homology of a $\mathcal{C}$–space X. The Q_i are $\mathbb{Z}_p$–linear, natural with respect to maps of $\mathcal{C}$–spaces, $Q_0(x)=x^{*p}$ and $Q_i(1_{H_*(X;\mathbb{Z}_p)})=0$ for $i>0$. Hence the operations Q_i can be regarded as homological derived p^{th} powers.
- The following Cartan formula holds for $x\in H_q(X;\mathbb{Z}_p)$ and $y\in H_{q'}(X;\mathbb{Z}_p)$:
$$Q^r(x*y)=\sum_{i+j=r}Q^i(x)*Q^j(x).$$
- The following Adem relations hold:
$$Q^r\circ Q^s=\sum_i(-1)^{r+i}\binom{(p-1)(i-s)-1}{pi-r}Q^{r+s-i}\circ Q^i \qquad \text{if } r>ps,$$
$$\begin{aligned}Q^r\circ\beta Q^s=&\sum_i(-1)^{r+i}\binom{(p-1)(i-s)}{pi-r}\beta Q^{r+s-i}\circ Q^i\\&-\sum_i(-1)^{r+i}\binom{(p-1)(i-s)-1}{pi-r-1}Q^{r+s-i}\circ\beta Q^i \qquad \text{if } r\geq ps.\end{aligned}$$

By using the Adem relations, we can write an arbitrary composition of k operations $Q_{i_1}\circ\cdots\circ Q_{i_k}$ as a linear combination of sequences $Q_{j_1}\circ\cdots\circ Q_{j_k}$ with nondecreasing j_l. Furthermore, when applied to an even-dimensional class, we can also require that $j_l=\sum_{l<m\leq k}j_m(p-1)$ or $\sum_{l<m\leq k}j_m(p-1)-1 \bmod 2(p-1)$. We call a sequence of nonnegative integers $J=(j_1,\dots,j_k)$ *admissible* if it satisfies the previous two conditions. We call it *strongly admissible* if, in addition, $j_1\neq 0$. To simplify the notation, we write Q_J for $Q_{j_1}\circ\cdots\circ Q_{j_k}$. If we translate to the upper-indices notation, a composition $\beta^{\varepsilon_1}Q^{i_1}\circ\cdots\circ\beta^{\varepsilon_k}Q^{i_k}$ is admissible if and only if $pi_l-\varepsilon_l\geq i_{l-1}$ for all l, and is strongly admissible if and only if, in addition, $i_1-\sum_{l=2}^k[2(p-1)i_l-\varepsilon_l]>0$.

The Dyer–Lashof operations completely describe the structure of $\bigoplus_{n\geq 0}H_*(\Sigma_n;\mathbb{Z}_p)$.

Theorem 2.2 [2, Theorem 4.1, page 40] *Let $\iota\in H_0(\Sigma_1;\mathbb{Z}_p)$ be the homology class of any point in $B(\Sigma_1)$. Let $H=\bigoplus_{n\geq 0}H_*(\Sigma_n;\mathbb{Z}_p)$. Then H, under the product $*$ induced by the inclusions $\Sigma_n\times\Sigma_m\to\Sigma_{n+m}$, is the free graded commutative algebra generated (in appropriate dimensions) by $Q^I(\iota)$ for strongly admissible sequences I. Moreover, the action of the operations Q_i is determined by the properties listed above. In other words, it is isomorphic to the free allowable $\mathcal{R}$–algebra on ι, as defined in* [2, Section I.2].

As a consequence, the basis for this algebra as a $\mathbb{Z}_p$–vector space is given by products of such $Q^I(\iota)$. We call these basis elements *Nakaoka monomials*.

We now define some cohomology classes, which we will prove to be Hopf-ring generators for A.

Definition 2.3 Let the symbol $\vee$ denote the linear dual with respect to the Nakaoka monomial basis of H. Now we define some classes:

$$\alpha_{j,k} = \big[Q^{p^{k-1}-p^{k-1-j}} \circ \cdots \circ Q^{p^j-1} \circ \beta Q^{p^{j-1}} \circ \cdots \circ Q^p \circ Q^1(\iota)\big]^\vee,$$
$$\beta_{j,k,m} = \big[\big(\beta Q^{p^{k-1}-p^{k-1-j}} \circ \cdots \circ Q^{p^{j+1}-p} \circ Q^{p^j-1} \circ \beta Q^{p^{j-1}} \circ \cdots \circ Q^1(\iota)\big)^{*m}\big]^\vee,$$
$$\gamma_{k,m} = \big[\big(Q^{p^{k-1}} \circ \cdots \circ Q^p \circ Q^1(\iota)\big)^{*m}\big]^\vee.$$

Note that $\alpha_{j,k}$ is an odd-dimensional homogeneous element of A, while $\beta_{j,k,m}$ and $\gamma_{k,m}$ are even-dimensional. Note also that we can easily convert the sequences of operations that appear in the definition above into the lower-index notation. For example, $\gamma_{k,m}$ is the linear dual to

$$(-1)^k Q_{2(p-1)}^{\circ k}(\iota)^{*m}.$$

Similarly, the linear duals of $\alpha_{j,k}$ and $\beta_{j,k,m}$ can be written as nonzero multiples of the elements

$$Q_{p-1}^{\circ k-j} \circ Q_{2p-3} \circ Q_{2(p-1)}^{\circ j-1}(\iota) \quad \text{and} \quad \big[Q_{p-2} \circ Q_{p-1}^{\circ j-i-1} \circ Q_{2p-3} \circ Q_{2(p-1)}^{\circ i-1}(\iota)\big]^{*m}.$$

The structure of A with only the transfer product has a nice description that can be obtained with essentially the same proof adopted by Giusti, Salvatore and Sinha in [3], using the fact that the Bockstein homomorphism is a derivation with respect to the cross product.

Theorem 2.4 [3, Theorem 4.13] *For every sequence I of nonnegative integers, $\Delta_\odot(Q_I(\iota)) = Q_I(\iota) \otimes 1 + 1 \otimes Q_I(\iota)$. In other words, $(H, \Delta_\odot, *)$, the Hopf dual of $(A, \Delta, \odot)$, is freely generated under $*$ by elements that are primitive under $\Delta_\odot$. Hence $(A, \odot)$ is the tensor product*

$$\bigotimes_{\substack{\dim(Q^I)\text{ even} \\ k\in\mathbb{N}}} \frac{\mathbb{Z}_p[(Q^I(\iota)^{p^k})^\vee]}{([(Q^I(\iota)^{p^k})^\vee]^p)} \otimes \bigwedge(\{Q^I(\iota)^\vee\}_{\dim(Q^I)\text{ odd}})$$

of a divided power polynomial algebra and an exterior algebra, where the Q^I indexing the tensor products above are the strongly admissible sequences of Dyer–Lashof operations $\beta^{\varepsilon_1} Q^{i_1} \circ \cdots \circ \beta^{\varepsilon_k} Q^{i_k}$. Moreover, the following relations hold:

(1) $\beta_{i,j,m} \odot \beta_{i,j,n} = \binom{n+m}{m} \beta_{i,j,n+m}$,

(2) $\gamma_{k,m} \odot \gamma_{k,n} = \binom{n+m}{m} \gamma_{k,n+m}$.

Thus, as far as the transfer product is concerned, we have relations totally analogous to those described by Sinha, Giusti and Salvatore in the mod 2 case.

However, if the cohomology is taken modulo an odd prime, there are also nontrivial relations for the cup product of the generators, due to the more complicated structure of the dual of the Dyer–Lashof algebra. We state them in the following lemma.

Lemma 2.5 *With the previous notation, the following equalities hold:*

(3) $\alpha_{i,k}\alpha_{j,k} = \gamma_{k,1}\beta_{i,j,p^{k-j}}$ *if* $i < j$.

(4) $\beta_{i,j,p^{k-j}}\alpha_{l,k} = (-1)^{\rho}\beta_{\rho(i),\rho(j),p^{k-\rho(j)}}\alpha_{\rho(l),k}$ *if* i,j,l *are pairwise distinct, where* ρ *is a permutation of the indexes* i,j,l *such that* $\rho(i) < \rho(j)$*, while* $\beta_{i,j,p^{k-j}}\alpha_{l,k} = 0$ *if* i,j,l *are not pairwise distinct.*

(5) $\beta_{i,j,m}\beta_{i',j',m'} = [(-1)^{\rho}]^{m}\beta_{\rho(i),\rho(j),mp^{j-\rho(j)}}\beta_{\rho(i'),\rho(j'),m'p^{j'-\rho(j')}}$ *if we suppose that* $mp^{j} = m'p^{j'}$ *and that* i,j,i',j' *are pairwise distinct, where* ρ *is a permutation of the indexes* i,j,i',j' *such that* $\rho(i) < \rho(j)$ *and* $\rho(i') < \rho(j')$*, while* $\beta_{i,j,m}\beta_{i',j',m'} = 0$ *otherwise.*

Proof This is an almost direct consequence of Cohen, Lada and May [2, Theorem 3.7, page 29]. Explicitly, let $\mathcal{R}$ be the Dyer–Lashof algebra as defined in [2]. Let $\mathcal{R}[k]$ be its k^{th} component, so that $\mathcal{R} = \bigoplus_{k\geq 0}\mathcal{R}[k]$. The evaluation of Dyer–Lashof operations on ι gives a morphism of coalgebras $\varphi_k\colon \mathcal{R}[k] \to H_*(\Sigma_{p^k};\mathbb{Z}_p)$, which dualizes to a map of algebras $\varphi_k^*\colon H^*(\Sigma_{p^k};\mathbb{Z}_p) \to \mathcal{R}[k]^*$.

Because of the theorem from [2] cited above, by definition these relations hold in the linear duals of $R[k]$. We are left to check them on the full set of Nakaoka monomials. When m is a power of p this follows immediately from the bialgebra structure of $(H,*,\Delta_\cdot)$, where $\Delta_\cdot$ is the coproduct dual to the cup product. □

Remark The relations described above can be recalled by the properties of the Bockstein homomorphism β in the duals, namely $\beta^2 = 0$ and the fact that β commutes with the product.

Example We provide a very simple example to show how the previous relations work. In $H^*(\Sigma_{p^2};\mathbb{Z}_p)$, relation (3) reduces to

$$\alpha_{2,1}\alpha_{2,2} = \gamma_{2,1}\beta_{1,2,1}.$$

Instead, since we do not have three distinct indices in $\{1,2\}$, the relations in form (4) can be written as $\beta_{1,2,1}\alpha_{1,2} = 0$ and $\beta_{1,2,1}\alpha_{2,2} = 0$. Similarly, (5) only assures that $\beta_{1,2,1}^2 = 0$.

For $H^*(\Sigma_{p^3};\mathbb{Z}_p)$ the relations which can be obtained by Lemma 2.5 are:

$$\alpha_{1,3}\alpha_{2,3}=\gamma_{3,1}\beta_{1,2,p},\alpha_{1,3}\alpha_{3,3}=\gamma_{3,1}\beta_{1,3,1} \quad \text{and} \quad \alpha_{2,3}\alpha_{3,3}=\gamma_{3,1}\beta_{2,3,1},$$
$$\beta_{1,2,p}\alpha_{1,3}=\beta_{1,2,p}\alpha_{2,3}=\beta_{1,3,1}\alpha_{1,3}=\beta_{1,3,1}\alpha_{3,3}=\beta_{2,3,1}\alpha_{2,3}=\beta_{2,3,1}\alpha_{3,3}=0,$$
$$\beta_{1,2,p}\alpha_{3,3}=-\beta_{1,3,1}\alpha_{2,3}=\beta_{2,3,1}\alpha_{1,3},$$
$$\beta_{1,2,p}^2=\beta_{1,2,p}\beta_{1,3,1}=\beta_{1,2,p}\beta_{2,3,1}=\beta_{1,3}^2=\beta_{1,3,1}\beta_{2,3,1}=\beta_{2,3}^2=0.$$

We now turn to the coproduct in A. Using the fact that this is dual to the product of H the following lemma follows from the definitions.

Lemma 2.6 *The following equalities hold:*

- $\Delta(\alpha_{j,k})=\alpha_{j,k}\otimes 1+1\otimes\alpha_{j,k}$
- $\Delta(\beta_{i,j,m})=\sum_{l=0}^{m}(\beta_{i,j,l}\otimes\beta_{i,j,m-l})$
- $\Delta(\gamma_{k,m})=\sum_{l=0}^{m}(\gamma_{k,l}\otimes\gamma_{k,m-l})$

At this point, we have all the ingredients to describe a presentation of A as a Hopf ring analogous to that of Giusti, Salvatore and Sinha [3, Theorem 1.2].

Theorem 2.7 *As a graded commutative Hopf ring, A is generated by the elements $\alpha_{j,k}$, $\beta_{i,j,m}$ and $\gamma_{k,m}$ as defined above (of suitable dimensions) under the relations* (1)–(5) *as explained in Theorem 2.4 and in Lemma 2.5, together with:*

(6) *The product $\cdot$ between two generators belonging to different components is 0.*

Moreover, the value of Δ on generators is determined by the preceding lemma and the antipode is the multiplication by $(-1)^n$ on the component corresponding to Σ_n.

Proof Let $B=(B;\odot_B,\cdot_B,\Delta_B)$ be the graded commutative Hopf ring generated by elements $\alpha_{j,k}$, $\beta_{i,j,m}$ and $\gamma_{k,m}$ (of suitable degree) with the specified relations. There is an obvious morphism $\psi\colon B\to A$.

One can see that, using (3)–(5), B is generated under $\odot$ only by elements that can be written in one of the two following forms:

$$\prod_j\gamma_{k_j,m_j}\cdot\prod_{a=1}^{r}\beta_{i_{2a-1},i_{2a},lp^{-i_{2a}}},$$
$$\prod_j\gamma_{k_j,m_j}\cdot\prod_{a=1}^{r}\beta_{i_{2a-1},i_{2a},p^{c-i_{2a}}}\beta_{i_{2a-1},i_{2a},p^{c-i_{2a}}}\alpha_{i_{2r+1},c}.$$

Here in the first case $1 \le i_1 < \cdots < i_{2r}$, $p^{i_{2r}} \le l$ and $p^{k_j} m_j = l$, while in the second case $1 \le i_1 < \cdots < i_{2r+1} \le c$ and $m_j p^{k_j} = p^c$. We will always suppose that the k_j are arranged in nonincreasing order. Borrowing the notation from [3], we will call these elements *gathered blocks* or simply *blocks*. By relations (3)–(6), these are all the elements that can be obtained from the generators by applying $\cdot_B$. We will call *Hopf monomials* the objects in the form $b_1 \odot_B \cdots \odot_B b_s$, where every b_j is a gathered block.

Then, using relations (1), (2) and (6) and Hopf distributivity, one can prove that for every gathered block b (of even dimension), we have $b^{\odot p} = 0$. Let us define an algebra

$$C = C_{\mathrm{even}} \otimes C_{\mathrm{odd}},$$

with

$$C_{\mathrm{even}} = \bigotimes_{\substack{d,k \ge 0 \\ b \in H^{2d}(\Sigma_{p^k};\mathbb{Z}_p) \text{ block}}} \frac{\mathbb{Z}_p[b]}{b^p},$$

$$C_{\mathrm{odd}} = \bigwedge\big(\{b : b \in H^{2d+1}(\Sigma_{p^k};\mathbb{Z}_p) \text{ block}, d,k \ge 0\}\big),$$

where $\bigwedge(X)$ indicates the exterior algebra generated by the elements of X (in appropriate degrees). By virtue of the above property, there is a morphism $\chi\colon C \to (B,\odot)$. Moreover, notice that, by Hopf distributivity and our coproduct formula for the generators, we have

$$\prod_j \gamma_{k_j,m_j} \cdot \prod_{a=1}^r \beta_{i_{2a-1},i_{2a},lp^{-i_{2a}}} \odot \prod_j \gamma_{k_j,m'_j} \cdot \prod_{a=1}^r \beta_{i_{2a-1},i_{2a},l'p^{-i_{2a}}} = \binom{l+l'}{l} \prod_j \gamma_{k_j,m_j+m'_j} \cdot \prod_{a=1}^r \beta_{i_{2a-1},i_{2a},(l+l')p^{-i_{2a}}}.$$

Hence, every gathered block can be written uniquely as a nonzero multiple of gathered blocks which lie in components indexed by a power of p. This proves that χ is surjective. Theorem 2.2 and Theorem 2.4 imply that the composition $\psi \circ \chi\colon C \to (A,\odot)$ is an isomorphism, proving the theorem. □

3 Presentation of product structures through an additive basis

In this section we will observe that the previous theorem allows us to obtain an additive basis of A as a $\mathbb{Z}_p$–vector space, similar to that in [3]. In order to describe this basis, we need a preliminary definition.

Definition 3.1 Let

$$b = \gamma_{k_1,m_1} \cdots \gamma_{k_s,m_s} \beta_{i_1,i_2,m'_1} \cdots \beta_{i_{2a-1},i_{2a},m'_a}$$

be an even-dimensional gathered block and $r = 2a$. We define the *profile* of b as the pair $(\underline{k}, \underline{e})$, where $\underline{k} = (k_1, \dots, k_s)$, and we suppose that, as usual, k_j is arranged in nonincreasing order, while $\underline{e} = (i_1, \dots, i_r)$.

For example, the profile of $\gamma_{2,2}^3 \gamma_{1,2p} \beta_{1,2,2}$ is determined by $\underline{k} = (2, 2, 2, 1)$, $\underline{e} = (1, 2)$.

The following result is an easy consequence of the proof of Theorem 2.7

Corollary 3.2 *Consider the set $\mathcal{M}$ of all Hopf monomials $\bigodot_{i=1}^r b_i$ with the property that the gathered blocks b_i of even dimension have pairwise distinct profiles, and the odd-dimensional blocks are pairwise distinct. This is a bigraded basis for A as a $\mathbb{Z}_p$–vector space.*

It must be noted that the pairing between this basis in cohomology and the Nakaoka monomials in homology is not completely understood. Indeed, the necessity to apply the Adem relations to describe the coproduct dual to $\cdot$ in terms of this basis complicates this pairing. For example, if $p = 3$ then $\gamma_{1,3}^4 = (Q^8 \circ Q^4(\iota))^\vee - (Q^9 \circ Q^3(\iota))^\vee$, because the formula for the coproduct $\Delta.$ of $Q^9 \circ Q^3(\iota)$ yields a summand $Q^3 \circ Q^0(\iota) \otimes Q^6 \circ Q^3(\iota)$, which can be written as $-Q^2 \circ Q^1(\iota) \otimes Q^6 \circ Q^3(\iota)$.

It is helpful to give a graphical description of this basis, similar to that obtained in [3]. First, we describe the generators as rectangles:

- $\gamma_{k,n}$ is a hollow rectangle of width np^k and height $2(1 - p^{-k})$.
- $\beta_{j,k,n}$ is a solid rectangle of width np^k and height $2(1 - p^{-j} - p^{-k})$.
- $\alpha_{j,k}$ is a solid rectangle of width p^k and height $2(1 - p^{-j}) - p^{-k}$.

In this way, the area of the rectangle is the homological dimension of the corresponding generator and its width accounts for the component in which the generator lies. Hollow rectangles represent generators whose linear duals in the Nakaoka basis lie in the subalgebra of H generated by sequences of Dyer–Lashof operations $Q^{i_1} \circ \cdots \circ Q^{i_k}(\iota)$ without the Bockstein. In terms of lower-indexed operations, these are written as multiples of $Q_{j_1} \circ \cdots \circ Q_{j_k}(\iota)$ where every j_l is even. These generators behave very similarly to the ones obtained in the mod 2 case. The other generators correspond to solid rectangles. We describe a gathered monomial, which is a product of $\gamma_{k,n}$, $\beta_{j,k,m}$ and possibly $\alpha_{j,k}$ all lying in the same component, as the column obtained by placing the corresponding rectangles on top of each other. A basis element, which is a transfer product of some gathered monomials $b_1, \dots, b_r$, is described by the diagram obtained by arranging the columns corresponding to $b_1, \dots, b_r$ next to each other horizontally. In order to conform to the notation used in [3], we will call these objects *skyline diagrams*. Some examples of skyline diagrams are depicted in Figure 1.

With the aid of this graphical description, we can elucidate the relations (3)–(5) of Lemma 2.5. First, observe that the rectangles of a column associated with a gathered block must satisfy some necessary condition. For example, there must be at most one odd-dimensional solid rectangle. This leads to the following definition.

Definition 3.3 A column made of rectangles with the same width stacked one onto the other is called *admissible* if it is associated with a gathered block.

As we will see at the end of this section, the cup product of two columns is essentially described as a new column obtained by stacking the original ones on top of each other. Hence relation (3) says that, if a column of width l contains two odd-dimensional solid rectangles, we can replace them with a hollow rectangle of height $2(1-l^{-1})$ and another solid rectangle to match the column's height. For the graphical representation of relation (3) see the first example in Figure 1.

Relations (4) and (5) determine how cup products of generators of the form $\beta_{i,j,m}$ and $\alpha_{j,k}$ behave when some indices are permuted. Their graphical interpretation is that if two columns are made only with solid rectangles of which at most one is odd-dimensional, they must be equal up to sign. Given such a column, there are two cases:

- If no admissible all-solid column of the same width and height exists, then it is 0.
- Otherwise it is equal, up to sign, to the (necessarily unique) admissible all-solid column with the same dimensions.

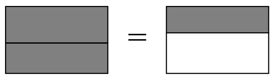
This gives a simple algorithm to write a nonadmissible column as a multiple of an admissible one, which is the graphical counterpart of what we observed in the proof of Theorem 2.7.

With this basis, one can describe the products. For example, in $H^*(\Sigma_{p^2}, \mathbb{Z}_p)$, let x be one of the elements $\gamma_{2,1}$, $\alpha_{1,2}$, $\alpha_{2,2}$ or $\beta_{1,2,1}$. We have $x(\gamma_{1,k} \odot 1_{p(p-k)}) = 0$ for $1 \le k \le p-1$. Indeed, $\Delta(x) = x \otimes 1 + 1 \otimes x$, hence, by Hopf ring distributivity,

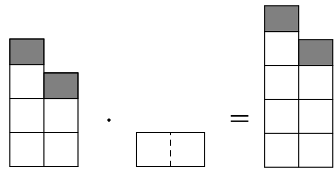
$$x(\gamma_{1,k} \odot 1_{p(p-k)}) = x\gamma_{1,k} \odot 1_{p(p-k)} + \gamma_{1,k} \odot x 1_{p(p-k)} = 0 \odot 1_{p(p-k)} + \gamma_{1,k} \odot 0 = 0.$$

Similarly one can prove that $x(\gamma_{1,k-1} \odot \alpha_{1,1} \odot 1_{p(p-k)}) = 0$ for all $1 \le k \le p$.

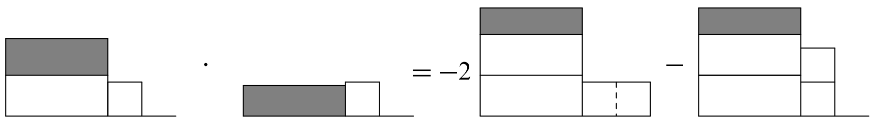
The general case can be derived in the exact same way as described by Giusti, Salvatore and Sinha [3, Section 6] and is indeed a straightforward consequence of the Hopf ring presentation. For this reason, we omit the proofs.



$\alpha_{1,2}\alpha_{2,2} = \gamma_{2,1}\beta_{1,2,1}$



$(\gamma_{1,1}^{i}\alpha_{1,1} \odot \gamma_{1,1}^{j}\alpha_{1,1}) \cdot \gamma_{1,2} = \gamma_{1,1}^{i+1}\alpha_{1,1} \odot \gamma_{1,1}^{j+1}\alpha_{1,1}$



$(\gamma_{2,1}\alpha_{2,2} \odot \gamma_{1,1} \odot 1_p) \cdot (\alpha_{1,2} \odot \gamma_{1,1} \odot 1_p) = -2\gamma_{2,1}^{2}\beta_{1,2,1} \odot \gamma_{1,2} - \gamma_{2,1}^{2}\beta_{1,2,1} \odot \gamma_{1,1}^{2} \odot 1_p$

Figure 1: Examples of calculations using the graphical representation. The size of the rectangles is correct only for $p=3$, but the same calculations with classes understood to be in different degrees are actually true for every p.

We begin with the transfer product, which can be described very easily. Given two Hopf monomials $x = b_1 \odot \cdots \odot b_r$ and $y = b'_1 \odot \cdots \odot b'_s$ in $\mathcal{M}$, the transfer product $x \odot y$ is again a Hopf monomial, but it may have gathered blocks with the same profile. However, two even-dimensional gathered blocks with the same profile can be merged together using the formula

$$(\gamma_{k_1,m_1} \dots \gamma_{k_r,m_r}\beta_{i_1,i_2,n_1} \dots \beta_{i_{2a-1},i_{2a},n_a}) \odot (\gamma_{k_1,m'_1} \dots \beta_{i_{2a-1},i_{2a},n'_a}) \\ = \binom{m_1+m'_1}{m_1}\gamma_{k_1,m_1+m'_1} \dots \beta_{i_{2a-1},i_{2a},n_a+n'_a}.$$

In this way, we can write $x \odot y$ as a multiple of an element of $\mathcal{M}$. Graphically, the transfer product corresponds to placing two skyline diagrams next to each other, merging two columns if they have constituent blocks of the same height and multiplying by $\binom{n+m}{n}$, where n and m are the widths of the two columns.

In order to provide a formula for the coproduct, we need the following:

Definition 3.4 Let $b = \gamma_{l_1,m_1} \cdots \gamma_{l_r,m_r}\beta_{i_1,i_2,n_1} \cdots \beta_{i_{2s-1},i_{2s},n_s}$ be an even-dimensional gathered block. Let $c(b) = p^{l_1}m_1 = \cdots = p^{i_s}n_s$ be the integer corresponding to the component of A in which b lies. We say that a k–tuple $(b_1, \dots, b_k)$ of gathered blocks is a *partition* of b if every b_i has the same profile as b and $\sum_{i=1}^{k} c(b_i) = c(b)$. Some $c(b_i)$ are allowed to be 0, in which case b_i is understood to be 1_0. If b is an odd-dimensional block, a partition is defined in the same way, but we only allow b_i to be equal to 1_0 or to b itself. A partition with $k=2$ is called a *splitting*.

The coproduct of elements of $\mathcal{M}$ can be calculated with the formula

$$\Delta(b_1 \odot \cdots \odot b_s) = \sum (b'_1 \odot \cdots \odot b'_s) \otimes (b''_1 \odot \cdots \odot b''_s).$$

Here the sum is taken over all the possible splittings $\{b'_i, b''_i\}$ of the constituent blocks b_i. In terms of our graphical representation, the coproduct can be described by dividing each rectangle corresponding to $\gamma_{k,n}$ or $\beta_{j,k,n}$ into n equal parts using vertical dashed lines. The coproduct of a skyline diagram is obtained by cutting each column along the dashed lines that cross it from top to bottom and partitioning them into two to create two other skyline diagrams. This must be done in every possible way and all the outcomes must be summed.

The formula for the cup product of two elements of $\mathcal{M}$ is

$$(b_1 \odot \cdots \odot b_r) \cdot (b'_1 \odot \cdots \odot b'_s) = \sum_{(\mathcal{P},\mathcal{P}')} (-1)^{\varepsilon_{\mathcal{P},\mathcal{P}'}} \bigodot_{j=1}^{s} \bigodot_{i=1}^{r} (b_{i,j} b'_{j,i});$$

the sum is over all pairs of sets $\mathcal{P} = \{(b_{i,1}, \dots, b_{i,s})\}_{i=1}^{r}$ and $\mathcal{P}' = \{(b'_{i,1}, \dots, b'_{i,r})\}_{i=1}^{s}$ such that $(b_{i,1}, \dots, b_{i,s})$ is a partition of b_i and $(b'_{i,1}, \dots, b'_{i,r})$ is a partition of b'_i. The number $\varepsilon_{\mathcal{P},\mathcal{P}'}$ is given by

$$\varepsilon_{\mathcal{P},\mathcal{P}'} = \sum_{\substack{1 \le i < j \le s \\ 1 \le k \le r}} \dim(b'_{i,k}) \dim(b_{k,j}) + \sum_{\substack{1 \le h < k \le r \\ 1 \le i \le s}} \dim(b'_{i,h}) \dim(b_{k,i}).$$

The coefficient $(-1)^{\varepsilon_{\mathcal{P},\mathcal{P}'}}$ is due to the skew-commutativity of the product. Since the cup product of two gathered blocks, when it is not zero, is equal up to sign to a gathered block, each summand in the previous formula is zero or can be written, up to sign, as a transfer product of gathered blocks. Thus, omitting all the zero summands and eventually merging together the transfer product of gathered blocks with the same profile as before, we can write the desired cup product as a linear combination of elements of $\mathcal{M}$. Note that one can restrict the sum to the $\mathcal{P}$ and $\mathcal{P}'$ such that $b_{i,j}$ and $b'_{j,i}$ lie in the same component, as the other terms are equal to 0.

Graphically, if we are given two skyline diagrams, in order to compute their cup product, we apply the following algorithm:

(1) Divide the rectangles with vertical dashed lines as explained before.

(2) Divide each diagram into columns using both the boundaries of the rectangles and the vertical dashed lines.

(3) Match each column of the first diagram with a column of the second one in all possible ways up to automorphisms, stack the matched columns one on top of

the other and place these newly constructed columns side by side to make new diagrams.

(4) These diagrams may contain a pair of columns with the same profiles. In this case we must use the transfer product formula to merge them. There may also be nonadmissible columns, that we must write as a multiple of admissible ones via the previously described algorithm.

For clarity, we compute two examples, represented graphically in Figure 1:

- Let $x = \gamma_{1,1}^i \alpha_{1,1} \odot \gamma_{1,1}^j \alpha_{1,1}$ and $y = \gamma_{1,2}$. Since x is made of two columns of width p, the only splitting of y which can yield a nontrivial summand in the formula for the cup product is $(\gamma_{1,1}, \gamma_{1,1})$. Hence

$$x \cdot y = \gamma_{1,1}^i \alpha_{1,1} \gamma_{1,1} \odot \gamma_{1,1}^j \alpha_{1,1} \gamma_{1,1} = \gamma_{1,1}^{i+1} \alpha_{1,1} \odot \gamma_{1,1}^{j+1} \alpha_{1,1}.$$

Working graphically, the rectangle corresponding to y should be divided with a dashed line into two equal parts ($\gamma_{1,1}$). Up to automorphisms, there is only one way to match the columns of x with them. Stacking matched columns is equivalent to adding one hollow rectangle of height $2(1-p^{-1})$ to each column of x.

- Let $x = \gamma_{2,1}\alpha_{2,2} \odot \gamma_{1,1} \odot 1_p$ and $y = \alpha_{1,2} \odot \gamma_{1,1} \odot 1_p$. The only two partitions of x that can yield a nontrivial summand in the cup product are $(\gamma_{2,1}\alpha_{2,2}, \gamma_{1,1}, 1_p)$ and $(\gamma_{2,1}\alpha_{2,2}, 1_p, \gamma_{1,1})$. Thus, by our formula,

$$\begin{aligned} x \cdot y &= \gamma_{2,1}\alpha_{2,2}\alpha_{1,2} \odot \gamma_{1,1}^2 \odot 1_p + \gamma_{2,1}\alpha_{2,2}\alpha_{1,2} \odot \gamma_{1,1} \odot \gamma_{1,1} \\ &= -\gamma_{2,1}^2 \beta_{1,2,1} - 2\gamma_{2,1}^2 \beta_{1,2,1} \odot \gamma_{1,2}. \end{aligned}$$

Graphically, there are two possible matches of the columns of x and y because we only need to ensure that the two largest columns match together. When we stack the two large columns one on top of the other we obtain a nonadmissible column that can be transformed as described in the figure. By stacking the remaining columns in the two possible ways, we obtain the two skyline diagrams on the left. In one diagram, two rectangles with the same height have been merged together, and a coefficient of 2 appears.

4 Restriction to modular invariants

Consider the regular representation of $V_n = \mathbb{Z}_p^n$ (the action of V_n on itself given by the usual $\mathbb{Z}_p$–vector space addition). This gives a map $V_n \to \Sigma_{p^n}$, as the set V_n has cardinality p^n. This section is devoted to the computation of the restriction map $\rho_n \colon H^*(\Sigma_{p^n}; \mathbb{Z}_p) \to H^*(V_n; \mathbb{Z}_p)$, induced by this immersion. This is related to the action of the Steenrod algebra on our Hopf ring generators, as we will see in the next section.

First, recall that $H^*(\mathbb{Z}_p;\mathbb{Z}_p)$ is isomorphic as a $\mathbb{Z}_p$–algebra to $\mathbb{Z}_p[y]\otimes\Lambda(x)$, where x and y are generators of the first and the second cohomology groups, respectively. We will also suppose that $\beta(x)=y$, where β is the cohomology Bockstein. Hence, by the Künneth formula,

$$H^*(V_n;\mathbb{Z}_p)=H^*(\mathbb{Z}_p;\mathbb{Z}_p)^{\otimes n}=\mathbb{Z}_p[y_1,\dots,y_n]\otimes\Lambda(x_1,\dots,x_n).$$

Recall that, by a result in Adem and Milgram [1, Corollary 1.8, page 182] the image of ρ_n is contained in the invariant subalgebra $\big[\mathbb{Z}_p[y_1,\dots,y_n]\otimes\Lambda(x_1,\dots,x_n)\big]^{\mathrm{GL}_n(\mathbb{Z}_p)}$, which was determined by Mùi in [7]. In particular, the product gives a $\mathbb{Z}_p$–vector space isomorphism of the previous algebra with $\mathbb{Z}_p[d_{0,n},\dots,d_{n-1,1}]\otimes M$, where M is the $\mathbb{Z}_p$–vector space with basis $\{R_{n,\underline{s}}:0\le s_1<\dots<s_l<n\}$ indexed by subsets of $\{0,\dots,n-1\}$.

The objects $d_{k,n-k}$ and $R_{n,s_1,\dots,s_l}$ are defined by Mùi in terms of some determinants. More precisely, we can define

$$L_{n,k}=\det\big[y_i^{p^{j-\delta_{j\le k}}}\big]_{1\le i,j\le n},$$

and (letting $\widehat{\cdot}$ denote omission)

$$M_{n,s_1,\dots,s_l}=\frac{1}{l!}\det\begin{bmatrix} x_1 & \dots & x_1 & y_1 & \dots & \widehat{y_1^{p^{s_1}}} & \dots & \widehat{y_1^{p^{s_l}}} & \dots & y_1^{p^{n-1}}\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots\\ x_n & \dots & x_n & y_n & \dots & \widehat{y_n^{p^{s_1}}} & \dots & \widehat{y_n^{p^{s_l}}} & \dots & y_n^{p^{n-1}}\end{bmatrix}.$$

Additionally, we have the equalities

$$d_{k,n-k}=\frac{L_{n,k}}{L_{n,n}}\quad\text{and}\quad R_{n,s_1,\dots,s_l}=M_{n,s_1,\dots,s_l}L_{n,k}^{p-2}.$$

The dimensions of $d_{k,n-k}$ and $R_{n,s_1,\dots,s_l}$ are $2(p^n-p^k)$ and $l+2(p^n-1-\sum_{j=1}^{l}p^{s_j})$, respectively.

Thus, as an algebra, $\big[\mathbb{Z}_p[y_1,\dots,y_n]\otimes\Lambda(x_1,\dots,x_n)\big]^{\mathrm{GL}_n(\mathbb{Z}_p)}$ is generated by these objects $d_{k,n-k}$, which are the classical Dickson invariants, and $R_{n,s_1,\dots,s_l}$ and the product structure are determined by $d_{0,n}$ being a nonzero divisor and the relations

$$R_{n,s_1,\dots,s_l}^2=0\quad\text{and}\quad R_{n,s_1}\dots R_{n,s_l}=(-1)^{\frac{l(l-1)}{2}}R_{n,s_1,\dots,s_l}d_{0,n}^{l-1}.$$

Much is known about these classes. For example, the Steenrod algebra action, which we will need soon, has been determined by Hung and Minh:

Theorem 4.1 [4, page 42] *Let $0 \leq r < p^n$. Let $r = \sum_{i=0}^{n-1} a_i p^i$ be the p–adic expansion of r. We agree that $a_{-1} = 0$ by convention. Then:*

- *$\mathcal{P}^r(d_{s,n-s})$ is 0 unless $a_i \geq a_{i-1}$ for all $0 \leq i < n$, $i \neq s$ and $a_s + 1 \geq a_{s-1}$. In this case it is given by the formula*

$$\lambda_{r,n,s} \prod_{i=0}^{n-1} d_{i,n-i}^{a_i - a_{i-1} + \delta_{i,s}}, \quad \text{where } \delta_{i,s} = \begin{cases} 1 & \text{if } i = s, \\ 0 & \text{otherwise,} \end{cases}$$

and the following formula for $\lambda_{r,n,s}$ holds:

$$\lambda_{r,n,s} = \frac{(p-1)!}{(p-1-a_{n-1})! \prod_{1 \leq i \leq n-1, i \neq s} (a_i - a_{i-1})! (a_s + 1 - a_{s-1})!} (a_s + 1).$$

- *$\mathcal{P}^r(R_{n,s})$ is 0 unless $a_i \in \{0,1\}$, $a_i \geq a_{i-1}$ for all $i \neq s$ and $a_s = 0$. This condition is equivalent to $r = (p-1)^{-1}(p^n + p^s - p^{t_1} - p^{t_2})$ for some $t_1 \leq s < t_2 \leq n$. In this case,*

$$\mathcal{P}^r(R_{n,s}) = R_{n,t_1} d_{t_2,n-t_2} - R_{n,t_2} d_{t_1,n-t_1}.$$

Here, we use the convention that $R_{n,n} = 0$ and $d_{n,0} = 1$.

- *$\mathcal{P}^r(R_{n,s_1,s_2})$ is 0 unless $a_i \in \{0,1\}$, $a_i \geq a_{i-1}$ for $i \neq s_1, s_2$ and $a_{s_1} = a_{s_2} = 0$. This condition is equivalent to $r = (p-1)^{-1}(p^n + p^{s_1} + p^{s_2} - p^{t_1} - p^{t_2} - p^{t_3})$ for some $t_1 \leq s_1 < t_2 \leq s_2 < t_3 \leq n$. In this case, the following formula holds:*

$$\mathcal{P}^r(R_{n,s_1,s_2}) = R_{n,t_1,t_2} d_{t_3,n-t_3} - R_{n,t_1,t_2} d_{t_2,n-t_2} + R_{n,t_2,t_3} d_{t_1,n-t_1}.$$

Again, we agree that $R_{n,s,n} = 0$ and $d_{0,n} = 1$.

Although we will not need this fact, it can be observed that, for $\mathcal{P}^r(d_{s,n-s})$, the coefficients $\lambda_{r,n,s}$ assume a nicer form if we express them as functions of the exponents $e_i = a_i + a_{i-1} + \delta_{i,s}$ that appear in the expression on the right. Explicitly,

$$p\lambda_{r,n,s} = \frac{p!}{\left(p - \sum_{i=0}^{n-1} e_i\right)! \prod_{i=0}^{n-1} e_i!} \sum_{i=0}^{s} e_i.$$

The first factor on the right is the number of choices of disjoint subsets of cardinalities $e_1, \dots, e_{n-1}$ in $\{1, \dots p\}$. After introducing the appropriate notions in Section 5, it will be obvious that $\sum_{i=0}^{n-1} e_i!$ counts the number of factors with an "effective scale" of at least $n - s$.

We now need a preliminary lemma.

Lemma 4.2 *Let $k \in \mathbb{N}$. We define J_k as the k–tuple $(2(p-1),\dots,2(p-1))$. Let $J=(j_1,\dots,j_k)$ be a sequence of nonnegative integers (not necessarily admissible). If $Q_J=\sum_{J' \text{ admissible}} \lambda_{J,J'} Q_{J'}$ is the expansion of Q_J as a linear combination of admissible sequences of operations, then $\lambda_{J,J_k}=0$ unless $J=J_k$.*

Proof We recall that $Q_{J_k}=\pm Q^{p^{k-1}} \circ \cdots \circ Q^p \circ Q^1$ and use upper indices, since Adem relations assume a much better form this way. Given a nonadmissible sequence in $\mathcal{R}$, its expansion in the admissible basis is obtained by iterative applications of the Adem relations. Hence, in order to prove the lemma, it is enough to check that for every $\beta^{\varepsilon} Q^r \beta^{\varepsilon'} Q^s$ with $r > ps-\varepsilon'$, when we apply the suitable Adem relation written as in Section 2, the expression we obtain does not contain a summand in the form $\lambda Q^{p^{l+1}} Q^{p^l}$ for some $\lambda \in \mathbb{Z}_p \setminus \{0\}$. This is obvious if $\varepsilon \neq 0$ or $\varepsilon' \neq 0$. If $\varepsilon=\varepsilon'=0$, then $Q^r \circ Q^s=\sum_i c_i Q^{r+s-i} Q^i$ for some nonzero coefficients c_i only if $pi \geq r$. If there exists $\bar{\imath}$ such that $c_{\bar{\imath}} \neq 0$, $r+s-\bar{\imath}=p^{l+1}$ and $\bar{\imath}=p^l$, then $r+s=p^{l+1}+p^l$ and $r > ps$ implies $r > p^{l+1}$. This is contradictory because $p\bar{\imath}=p^{l+1} < r$. □

We will also need to know how the transfer product behaves with respect to the restriction maps.

Lemma 4.3 *If $x_1 \in H^*(\Sigma_r;\mathbb{Z}_p)$ and $x_2 \in H^*(\Sigma_{p^n-r};\mathbb{Z}_p)$ are Hopf monomials that are different from 1, then $\rho_n^*(x_1 \odot x_2)=0$.*

Proof Recall that the inclusion of V_n in Σ_{p^n} factors through the iterated wreath product $\mathbb{Z}_p \wr (\mathbb{Z}_p \wr \cdots \wr (\mathbb{Z}_p \wr \mathbb{Z}_p)\cdots)$ (see Adem and Milgram [1, page 185]). By construction, the image in $H_*(\Sigma_{p^n};\mathbb{Z}_p)$ of the homology of this subgroup is given by Dyer–Lashof operations of length n. Hence, $H_*(V_n)$ maps onto the linear span of these classes, which are primitive with respect to $\Delta_\odot$. As a consequence, they must pair trivially with $x_1 \odot x_2$. □

We are now ready to describe the action of ρ_n on the generators, which is the analog of [3, Corollary 7.6] but is proved using a different technique.

Proposition 4.4 *The following formulas hold:*

$$\begin{aligned}
\rho_{j+k}(\alpha_{j,j+k}) &= (-1)^j R_{j+k,k}, \\
\rho_{j+k}(\beta_{i,j,p^k}) &= (-1)^{k+i} R_{j+k,k,k+j-i}, \\
\rho_{j+k}(\gamma_{j,p^k}) &= (-1)^j d_{k,j}.
\end{aligned}$$

Proof To prove the proposition, we will take advantage of the way Steenrod operations are constructed to inductively compute $\rho_j(\gamma_{j,1})$. Then we will use the naturality of the Steenrod action to work out the remaining cases. The core of this idea was originally used by Mann [5] to compute $\mathrm{im}(\rho_j)$. To a certain extent, we follow his reasoning, but we are also able to reconcile this approach with the Hopf ring structure and to describe in simpler terms the classes in the cohomology of Σ_{p^j} which restrict to $d_{l,j-l}$, $R_{j,l}$ and $R_{j,l,m}$.

First we will prove that $\rho_j(\gamma_{j,1}) = (-1)^j d_{0,j}$, or equivalently, by shifting to the lower-index notation, $\rho_j(Q_{J_j}(\iota)^\vee) = d_{0,j}$, where J_j is the j–tuple defined in Lemma 4.2.

Let us identify $H_*(V_j;\mathbb{Z}_p)$ with $H_*(\mathbb{Z}_p;\mathbb{Z}_p) \otimes H_*(V_{j-1};\mathbb{Z}_p)$. The homomorphism $(\rho_n)_*\colon H_*(V_n;\mathbb{Z}_p) \to H_*(\Sigma_{p^n};\mathbb{Z}_p)$ satisfies, for every $x \in H_s(V_{n-1};\mathbb{Z}_p)$ and for every $r \geq 0$, the formula

$$\begin{aligned}(\rho_n)_*(e_r \otimes x) = \nu(s) \textstyle\sum_k (-1)^k Q_{r+2k-s} \circ \mathcal{P}^k_*(x)& \\ - \delta(r)\nu(s-1) \textstyle\sum_k (-1)^k Q_{r+p+(2pk-s)(p-1)} \circ \mathcal{P}^k_* \beta(x).&\end{aligned}$$

Here $\mathcal{P}^k_*$ is the linear dual to the k^{th} Steenrod power $\mathcal{P}^k$,

$$\nu(2j+\varepsilon) = (-1)^j \left(\tfrac{1}{2}(p-1)\right)!^{\varepsilon} \quad \text{and} \quad \delta(2j+\varepsilon) = \varepsilon \quad \text{if } \varepsilon \in \{0,1\}.$$

This is stated in May [6, Proposition 9.1, page 205], where it is used as a preliminary step for the proof of Nishida relations, and is essentially the dualization of the original construction of $\mathcal{P}^k$ made by Steenrod.

Note that, by Lemma 4.2, all the summands in the previous formula pair trivially with $Q_{J_j}(\iota)^\vee$, except possibly those in the form $Q_{r+2k-s} \circ \mathcal{P}^k_*(x)$ with $r+2k-s = 2(p-1)$ and $s - 2k(p-1) = 2(p^{j-1}-1)$. This means that $r = (p^{j-1}-l)(p-1)$ and $s = 2(p^{j-1}-1) + 2k(p-1)$. Hence, dually, we have

$$\rho_j(Q_{J_j}(\iota)^\vee) = \sum_{k=0}^{p^{j-1}-1} (-1)^k \mathcal{P}^k \rho_{j-1}(Q^\vee_{J_{j-1}}) y_j^{(p-1)(p^{j-1}-1)}.$$

This implies by induction on j that the right member is equal to $d_{0,j}$. Explicitly, for $j=1$ the statement is trivial. For $j>1$, by the induction hypothesis, $\rho_j(Q_{J_j}(\iota)^\vee)$ is a $\mathrm{GL}_j(\mathbb{Z}_p)$–invariant polynomial in $H^*(V_{j-1};\mathbb{Z}_p)[y_j]$ whose leading coefficient is $d^p_{0,j-1}$. This must be $d_{0,j}$.

The calculations of $\rho_n(x)$ for γ_{n-k,p^k} with $k>0$, $\alpha_{j,n}$ and $\beta_{i,j,p^{n-j}}$ follow directly from the naturality of the Steenrod powers with respect to the restrictions ρ_n and from the formulas in Cohen, Lada and May [2, Theorem 3.9], which determine the Steenrod action on the dual of $\mathcal{R}[n]$. These formulas are true in $H^*(\Sigma_{p^n};\mathbb{Z}_p)$ only up to summands containing nontrivial transfer products, but they still determine $\rho_n \circ \mathcal{P}^r$ on Hopf ring generators because of Lemma 4.3. Comparison with the Steenrod powers of Mùi invariants as determined by Hung and Minh [4, Theorems B and C] yields the result. □

As a corollary, we obtain a known fact about the image of ρ_n.

Corollary 4.5 [5, Theorem A] *The image of ρ_n in $H^*(V_n;\mathbb{Z}_p)^{\mathrm{GL}_n(\mathbb{Z}_p)}$ is the subalgebra generated by $d_{j,n-j}$, $R_{n,j}$ and $R_{n,i,j}$. This can be described as*

$$\bigoplus_{l=0}^{n} \bigoplus_{0\leq s_1<\cdots<s_l<n} \mathbb{Z}_p[d_{0,n},\dots,d_{n-1,1}]d_{0,n}^{\lceil l/2\rceil,0}R_{n,\underline{s}}.$$

Hence, in general, ρ_k is not surjective.

5 Steenrod algebra action

This section is devoted to the computation of the action of the Steenrod powers on the Hopf ring A. We will achieve this by combining the calculations of Proposition 4.4 with the ideas used by Giusti, Salvatore and Sinha [3, Section 8] for the mod 2 cohomology.

First note that, as in the mod 2 case, the products $\odot$ and $\cdot$, the coproduct Δ and the antipode are induced from stable maps; hence, there are Cartan formulas for all these. This means that A is a Hopf ring over the mod p Steenrod algebra $\mathcal{A}(p)$, so it is sufficient to determine the action of $\beta^\varepsilon\mathcal{P}^l$ ($l\geq 0$ and $\varepsilon\in\{0,1\}$) on the Hopf ring generators $\alpha_{j,j+k}$, β_{i,j,p^k} and γ_{j,p^k}.

In order to describe the Steenrod algebra action on A in terms of our additive basis, we introduce some notation.

Definition 5.1
- The height (ht) of a gathered block b is the number of generators that must be cup-multiplied to obtain b. The height of a Hopf monomial is the largest of the heights of its constituent blocks.
- We define the effective scale (effsc) of a gathered block, which we assume in the form $b=\gamma_{l_1,n_1}\cdots\gamma_{l_r,n_r}\beta_{i_1,i_2,m_1}\cdots\beta_{i_{2s-1},i_{2s},m_s}\alpha_{j,k}^\varepsilon$ ($\varepsilon=0,1$) as the largest of the integers $l_1,\dots,l_r,i_{2s}$ if $\varepsilon=0$, or as k if $\varepsilon=1$. In other words, for $b\in H^*(\Sigma_{p^n};\mathbb{Z}_p)$, effsc$(x)$ is the minimum $k\geq 0$ such that the restriction of x to $\Sigma_{p^k}^{p^{n-k}}$ is not zero. The effective scale of a Hopf monomial is the minimum of the effective scales of its constituent blocks.
- A Hopf monomial is *full-width* if none of its constituent blocks is 1_{Σ_n}.
- We say that a gathered block is of *type A* if all the Hopf ring generators that must be cup-multiplied to obtain it are in the form $\gamma_{l,n}$, except one that is in the form $\alpha_{j,k}$. For example, $\gamma_{1,p^2}^3\alpha_{1,3}$ is of type A. More generally, a Hopf monomial is of type A if all its constituent blocks are of type A.

- We say that a gathered block is of *type B* if all the Hopf ring generators that appear in it are in the form $\gamma_{l,n}$, except one in the form $\beta_{i,j,m}$. For example, $\gamma_{3,1}^5\gamma_{2,p}^2\beta_{1,2,3}$ is of type B. More generally, a Hopf monomial is of type B if all its constituent blocks are of type B.
- We say that a Hopf monomial is of *type C* if it is obtained by applying $\cdot$ and $\odot$ only to elements in the form $\gamma_{l,n}$.

These definitions can be understood graphically. Given a skyline diagram:

- Its height is the maximal number of rectangles stacked one on top of the other that appear in the diagram.
- Its effective scale is the width of the thinnest column among those delineated by the original boundaries and the vertical dashed lines of full height.
- It is full-width if there are no columns of height 0.
- It is of type A if its columns contain exactly one solid rectangle and it is odd-dimensional. It is of type B if its columns contain exactly one solid rectangle and it is even-dimensional, while it is of type C if it is made only of hollow rectangles.

The definitions of height, effective scale and full-width monomial are borrowed from [3] and make sense also for the mod 2 cohomology.

We will also need the following result from Adem and Milgram's book:

Lemma 5.2 [1, Corollary 1.4, page 180] *Let ρ_n and τ_n be the natural restrictions from the cohomology of Σ_{p^n} to $H^*(V_n;\mathbb{Z}_p)$ and $H^*(\Sigma_{p^{n-1}}^p;\mathbb{Z}_p)\cong H^*(\Sigma_{p^{n-1}};\mathbb{Z}_p)^{\otimes p}$, respectively. The following homomorphism, whose components are ρ_n and τ_n, is injective:*

$$H^*(\Sigma_{p^n};\mathbb{Z}_p)\to H^*(V_n;\mathbb{Z}_p)\oplus H^*(\Sigma_{p^{n-1}}^p;\mathbb{Z}_p).$$

This lemma is derived in [1] by proving that elementary abelian subgroups detect the cohomology of Σ_{p^n}, and that all these groups are conjugate to subgroups of $\Sigma_{p^{n-1}}^p$ or to V_n. However, the same result can also be obtained as a consequence of our description. Indeed, the restriction of ρ_n to the linear span of Hopf monomials of height n in $H^*(\Sigma_{p^n};\mathbb{Z}_p)$ is injective by Proposition 4.4. These monomials map trivially to $H^*(\Sigma_{p^{n-1}}^p;\mathbb{Z}_p)$. Recall that a basis for $H^*(\Sigma_{p^{n-1}};\mathbb{Z}_p)^{\otimes p}$ is given by $x_1\otimes\cdots\otimes x_p$, where x_i are Hopf monomials and that τ_n can be identified with the iterated coproduct. Let $x_1\otimes\cdots\otimes x_p$ be such a basis element. By our coproduct formulas, there exists exactly one Hopf monomial $x\in H^*(\Sigma_{p^n};\mathbb{Z}_p)$ of height less than n such that $x_1 \otimes \cdots \otimes x_p$ appears with a nonzero coefficient in the expansion of $\tau_n(x)$. Explicitly, x is the transfer product of the gathered blocks $b \in H^*(\Sigma_m; \mathbb{Z}_p)$ such that, for every $1 \le i \le p$, there is $b_i \in H^*(\Sigma_{m_i}; \mathbb{Z}_p)$ that is a constituent block of x_i with the same profile of b and $\sum_i m_i = m$ (some m_i are allowed to be 0). This implies the lemma.

With these tools, we can obtain formulas for the Steenrod action on A. The idea is to use Theorem 4.1 and Lemma 5.2 and check, case by case, that the two expressions we wish to be equal assume the same value if restricted to $H^*(V_n; \mathbb{Z}_p)$ and $H^*(\Sigma_{p^{n-1}}^p; \mathbb{Z}_p)$.

Lemma 5.3 *$\mathcal{P}^r(\gamma_{n-k,p^k})$ can be expressed as a linear combination of full-width Hopf monomials of Type C with a height of at most p and an effective scale of at least $n-k$.*

Following the notation used by Giusti, Salvatore and Sinha [3], we will call these monomials the *outgrowth monomials* of γ_{n-k,p^k}. We denote the set of such monomials by $\mathrm{Outgrowth}(\gamma_{n-k,p^k})$.

Proof of Lemma 5.3 The proof will follow that of [3, Theorem 8.3]. We proceed by induction on k. First, assume $k=0$. By Theorem 4.1 and Lemma 5.2, $\mathcal{P}^r(\gamma_{n,1})$ must restrict to

$$\begin{cases} 0 & \text{on } H^*(\Sigma_{p^{n-1}}^p; \mathbb{Z}_p), \\ (-1)^n \lambda_{r,n,0} \prod_{i=0}^{n-1} d_{i,n-i}^{a_i - a_{i-1} + \delta_{i,0}} & \text{on } H^*(V_n; \mathbb{Z}_p). \end{cases}$$

Hence, it must be a multiple of $\prod_{i=0}^{n-1} \gamma_{n-i,p^i}^{a_i - a_{i-1} + \delta_{i,0}}$. This is the only full-width Hopf monomial of Type C, of degree $2(p^n-1) + 2r(p-1)$ with a height of at most p and an effective scale of at least n.

For $k > 0$, since $\tau_n(\gamma_{n-k,p^k}) = \gamma_{n-k,p^{k-1}}^{\otimes p}$, using the external Cartan formula, we have

$$\tau_n(\mathcal{P}^r(\gamma_{n-k,p^k})) = \sum_{r_1+\cdots+r_p=r} \mathcal{P}^{r_1}(\gamma_{n-k,p^{k-1}}) \otimes \cdots \otimes \mathcal{P}^{r_p}(\gamma_{n-k,p^{k-1}}).$$

By induction, this is a linear combination of elements $x_1 \otimes \cdots \otimes x_p$, where each x_i is an outgrowth monomial of $\gamma_{n-k,p^{k-1}}$. Recall that, for each $x_1 \otimes \cdots \otimes x_p$ in this form, there exists a unique Hopf monomial $x \in H^*(\Sigma_{p^n}; \mathbb{Z}_p)$ with $\mathrm{effsc}(x) < n$ whose restriction to $H^*(\Sigma_{p^{n-1}}^p; \mathbb{Z}_p)$ has a nonzero multiple of $x_1 \otimes \cdots \otimes x_p$ as a summand. We have described x explicitly above. Moreover, we have $\mathrm{effsc}(x) \le n-1$ and $x \in \mathrm{Outgrowth}(\gamma_{n-k,p^k})$ since height and the fact of being full-width are preserved by the coproduct, and the minimum of the effective scales of x_i must be equal to $\mathrm{effsc}(x)$. A Hopf monomial $x \notin \mathrm{Outgrowth}(\gamma_{n-k,p^k})$ with an effective scale less than n cannot appear in the expression of $\mathcal{P}^r(\gamma_{n-k,p^{k-1}})$, because this would yield summands in

$\tau_n(\mathcal{P}^r(\gamma_{n-k,p^k}))$ that are not tensor products of elements in $\mathrm{Outgrowth}(\gamma_{n-k,p^{k-1}})$. If a Hopf monomial with an effective scale equal to n appear, this must, once again, be an outgrowth monomial of γ_{n-k,p^k}. Otherwise, by applying the restriction to $H^*(V_n;\mathbb{Z}_p)$, we would contradict Theorem 4.1. □

Thus,

$$\mathcal{P}^r(\gamma_{n-k,p^k}) = \sum_{\substack{x\in \mathrm{Outgrowth}(\gamma_{n-k,p^k}) \\ \deg(x)=2(p^n-p^k)+2r(p-1)}} c_{n,k,x}x.$$

We are left to determine the coefficients $c_{n,k,x}$. Note that, by restricting to $H^*(V_n;\mathbb{Z}_p)$, using Proposition 4.4 and comparing with the formula in Theorem 4.1, we can directly determine $c_{n,k,x}$ when

$$x = \prod_{i=0}^{n-1} \gamma_{n-i,p^i}^{a_i-a_{i-1}+\delta_{i,k}}$$

is the unique term made by a single gathered block. Explicitly,

$$c_{n,k,\prod_{i=0}^{n-1}\gamma_{n-i,p^i}^{a_i-a_{i-1}+\delta_{i,k}}} = (-1)^{n-k+\sum_{i=0}^{n-1}(a_i-a_{i-1}+\delta_{i,s})(n-i)}\lambda_{r,n,k},$$

where $r=\sum_i a_i p^i$.

In general, let $x=b_1\odot\cdots\odot b_l\in H^*(\Sigma_{p^n};\mathbb{Z}_p)$ be the transfer product of l gathered blocks with pairwise distinct profiles. We assume that $b_i\in H^*(\Sigma_{p^{n_i}m_i};\mathbb{Z}_p)$ with $\mathrm{effsc}(b_i)=n_i$. As a notational convention, given a block b, we denote the (necessarily unique) block which has the same profile and lies in $H^*(\Sigma_{p^{\mathrm{effsc}(b)}};\mathbb{Z}_p)$ by b'. The restriction of x to the cohomology of $\prod_{i=1}^l \Sigma_{p^{n_i}}^{m_i}$ is the symmetrization of the class $b_1'^{\otimes m_1}\otimes\cdots\otimes b_l'^{\otimes m_l}$. By observing that

$$\gamma_{n-k,p^k}|_{\prod_i \Sigma_{p^{n_i}}^{m_i}} = \bigotimes_i \gamma_{n-k,p^{k-n+n_i}}^{m_i},$$

we obtain, by application of the naturality of the Steenrod operations and of the external Cartan formula for $\mathcal{P}^r$ as above, that $c_{n,k,x}=\prod_{i=1}^l c_{n-n_i,n_i-n+k,b_i'}$. This reduces the computation of $c_{n,k,x}$ to the previous particular case.

We summarize our calculations in the following proposition.

Proposition 5.4 *Let $0\le k<n$. Let $b=\prod_{i=0}^{n-1}\gamma_{n-i,p^i}^{e_i}\in \mathrm{Outgrowth}(\gamma_{n-k,p^k})$ be the gathered block with an effective scale of n. We define*

$$\begin{aligned} c_{n,k,b} &= (-1)^{n-k+\sum_i e_i(n-i)}\lambda_{(p-1)^{-1}[\sum_i 2(p^n-p^i)-2(p^n-p^k)],n,k} \\ &= (-1)^{n-k+\sum_i e_i(n-i)}\frac{(p-1)!}{(p-\mathrm{ht}(b))!\prod_{i=1}^{n-1} e_i!}\sum_{i=1}^k e_i. \end{aligned}$$

Let $x \in \mathrm{Outgrowth}(\gamma_{n-k,p^k})$ be a general outgrowth monomial. Then $x = b_1 \odot \cdots \odot b_s$, with $b_i \in H^*(\Sigma_{l_i};\mathbb{Z}_p)$ that are gathered blocks with pairwise distinct profiles. We define

$$c_{n,k,x} = \prod_{i=1}^{l} c^{l_i}_{\mathrm{effsc}(b_i),k-n+\mathrm{effsc}(b_i),b_i'}.$$

Then

$$\mathcal{P}^r(\gamma_{n-k,p^k}) = \sum_{\substack{x \in \mathrm{Outgrowth}(\gamma_{n-k,p^k}) \\ \deg(x)=2(p^n-p^k+r(p-1))}} c_{n,k,x}x.$$

Remark Note that, with our proof, we do not need to check inductively that the coefficients agree when we restrict to $H^*(\Sigma_{p^{n-1}};\mathbb{Z}_p)$ because this is automatically satisfied. However, this can be proved "manually" by observing that $\lambda_{r,n,s} = \lambda_{rp^k,n+k,s+k}$. Because of this, for a block $b = \prod_{i=0}^{n-1} \gamma_{n-i,i}^{a_i-a_{i-1}+\delta_{i,k}} \in H^*(\Sigma_{p^n};\mathbb{Z}_p)$, not necessarily with $\mathrm{effsc}(b) = n$, we have

$$c_{n,k,b} = (-1)^{n-k+\sum_i (a_i-a_{i-1}+\delta_{i,k})(n-i)} \lambda_{\sum_i a_i p^i,n,k},$$

agreeing with Theorem 4.1. More generally, given a gathered block

$$b = \prod\nolimits_i \gamma_{n-i,mp^i}^{a_i-a_{i-1}+\delta_{i,k}}$$

in $H^*(\Sigma_{p^n m})$ with $\mathrm{effsc}(b) = n$ and given two partitions $(p^{k_1},\dots,p^{k_1})$, $(p^{k_1'},\dots p^{k_{l'}'})$ of m with powers of p, the following equality holds in $\mathbb{Z}_p$:

$$\prod_{j=1}^{l} \lambda_{\sum_i a_i p^{i+k_j},n+k_j,k_j} = \prod_{j=1}^{l'} \lambda_{\sum_i a_i p^{i+k_j'},n+k_j',k_j'}.$$

This implies that the desired coefficients agree in the restriction to $H^*(\Sigma^p_{p^{n-1}};\mathbb{Z}_p)$.

The computation of $\mathcal{P}^r(\alpha_{j,k})$ and $\mathcal{P}^r(\beta_{i,j,p^k})$ can be done in the same way. Before stating the final results we define the analogous notion of outgrowth monomials for $\alpha_{i,j}$ and β_{i,j,p^k} as the full-width monomials of height one or two with an effective scale of at least j, of types A and B respectively. We will denote the set of such monomials by $\mathrm{Outgrowth}(\alpha_{i,j})$ and $\mathrm{Outgrowth}(\beta_{i,j,p^k})$, respectively.

Proposition 5.5 *Let $1 \le j \le n$. For $x = \gamma_{n-u,p^u}\alpha_{n-t,n} \in \mathrm{Outgrowth}(\alpha_{j,n})$, we define $c'_{n,j,x} = (-1)^{j+t+u}(\delta_{t \le n-j} - \delta_{u \le n-j})$. Here, we allow $u = n$ with the convention that $\gamma_{0,p^n} = 1$. Then*

$$\mathcal{P}^r(\alpha_{j,n}) = \sum_{\substack{x \in \mathrm{Outgrowth}(\alpha_{j,n}) \\ \deg(x)=2((p-1)r+p^n-p^{n-j})-1}} c'_{n,j,x}x.$$

Let $1 \le i < j \le n$ and let $k = n - j$. Given a gathered block $b = \gamma_{n-v,v}\beta_{n-u,n-t,p^t}$ in Outgrowth(β_{i,j,p^k})*, define*

$$c''_{n,i,j,b} = (-1)^{i+k+t+u+v}(\delta_{v>k+j-i} - \delta_{u>k+j-i})\delta_{t\le k+j-i}(\delta_{t\le k} - \delta_{v\le k})\delta_{u>k}.$$

For a general outgrowth monomial $x = b_1 \odot \cdots \odot b_l$ with $b_s \in H^*(\Sigma_{m_s};\mathbb{Z}_p)$ and effsc$(b_s) = n_s$ *for all $1 \le s \le l$, we define $c''_{n,i,j,x} = \prod_{s=1}^{l}(c''_{n_s,i,j,b_s})^{m_s}$. Then*

$$\mathcal{P}^r(\beta_{i,j,p^k}) = \sum_{\substack{x\in \mathrm{Outgrowth}(\beta_{i,j,p^k})\\ \deg(x)=2((p-1)r+p^n-p^{n-j}-p^{n-i})}} c''_{n,i,j,x}x.$$

In this result, the coefficients $c'_{n,j,x}$ and $c''_{n,i,j,x}$ are always equal to -1, 0 or 1.

We close this section with a proposition that describes the action of the Bockstein homomorphism β on Hopf ring generators. This clearly determines β on the whole Hopf ring and follows easily from [6, Theorem 3.9, page 33].

Proposition 5.6 *The following formulas hold:*

- $\beta(\alpha_{j,k}) = \gamma_{k,1}$ *if $j = k$ and is equal to 0 otherwise.*
- $\beta(\beta_{i,j,p^k}) = -\alpha_{i,j}$ *if $k = 0$ and is equal to 0 otherwise.*
- $\beta(\gamma_{j,p^k}) = 0$.

6 An example

As an example we now extract the cup product structure from our Hopf ring presentation in the case of $H^*(\Sigma_{p^2};\mathbb{Z}_p)$. An equivalent description has been given by Mùi [7] by analyzing the restriction to elementary p–subgroups.

First note that, by our results, an additive basis for $H^*(\Sigma_{p^2};\mathbb{Z}_p)$ is given by

$$\mathcal{B} = \left\{\gamma_{2,1}^a\gamma_{1,p}^b\alpha_{1,2}^{\varepsilon_1}\alpha_{2,2}^{\varepsilon_2}\beta_{1,2,1}^{\varepsilon_3} : a,b,\varepsilon_i \ge 0, \sum_{i=1}^{3}\varepsilon_i \le 1\right\} \cup \left\{\bigodot_{i=1}^{p}\gamma_{1,1}^{t_i}\alpha_{1,1}^{\varepsilon_i} : t_i \ge 0, \varepsilon_i \in \{0,1\} \text{ not all factors equal}\right\}.$$

The elements of the last set are to be considered up to permutations of the p factors. More simply, we can order the basis for $H^*(\Sigma_p;\mathbb{Z}_p)$ with the rule

$$\gamma_{1,1}^a\alpha_{1,1}^{\varepsilon} > \gamma_{1,1}^b\alpha_{1,1}^{\delta} \iff (a > b) \vee (a = b \wedge \varepsilon > \delta),$$

where we agree that, in the last set, $\gamma_{1,1}^{t_1}\alpha_{1,1}^{\varepsilon_1} \geq \cdots \geq \gamma_{1,1}^{t_p}\alpha_{1,1}^{\varepsilon_p}$ in the given order. It will be useful to order the set of the basis elements in this form with the product order.

We now write the generators and relations in $H^*(\Sigma_{p^2};\mathbb{Z}_p)$ as a ring. We define:

- $x_1 = \gamma_{2,1}$
- $x_2 = \alpha_{1,2}$
- $x_3 = \alpha_{2,2}$
- $x_4 = \beta_{1,2,1}$
- $y_i = \gamma_{1,i} \odot 1_{p^2-pi}$ for $1 \leq i \leq p-1$
- $y_p = \gamma_{1,p}$
- $z_i = \gamma_{1,i-1} \odot \alpha_{1,1} \odot 1_{p^2-pi}$ for $1 \leq i \leq p$

There are equalities $x_2x_3 = x_1x_4$, $x_2x_4 = 0$, $x_3x_4 = 0$ and $x_4^2 = 0$ coming directly from relations (3)–(5) in our Hopf ring presentation. Moreover, we have seen as an example in Section 2 that, for every $1 \leq i \leq 4$, we have $x_i y_j = 0$ for $1 \leq j \leq p-1$ and $x_i z_j = 0$ for $1 \leq j \leq p$. These will be our cup product generators and relations.

Proposition 6.1 *Consider the unital ring*

$$S = \frac{U(x_1, x_2, x_3, x_4, y_1, y_2, y_3, z_1, z_2, z_3)}{I},$$

where $U(X)$ is the free associative skew-commutative algebra generated by the elements of X (in appropriate dimensions) and let $I \subseteq S$ be the bilateral ideal generated by the relations above. There is an isomorphism $\varphi\colon S \to H^*(\Sigma_{p^2};\mathbb{Z}_p)$.

Proof There is a φ defined in the obvious way because we have checked that these relations hold in this cohomology ring. We now prove that φ is bijective. First, let $w = \bigodot_{j=1}^{p} \gamma_{1,1}^{t_j}\alpha_{1,1}^{\varepsilon_j}$, with the factors $\gamma_{1,1}^{t_j}\alpha_{1,1}^{\varepsilon_j}$ ordered from largest to smallest with respect to the product ordering. Consider the set $\mathcal{P}$ of elements $\{(b_{k,1},\dots,b_{k,p})\}_{k=1}^{2}$ where $(b_{1,1},\dots b_{1,p})$ is a partition of $\gamma_{1,i}$ and $(b_{2,1},\dots,b_{2,p})$ is a partition of $1_{p(p-i)}$. Moreover, let $\mathcal{P}'$ be the set of elements $\{(b'_{j,1},b'_{j,2})\}_{j=1}^{p}$ where $(b'_{j,1},b'_{j,2})$ is a splitting of $\gamma_{1,1}^{t_j}$. Using our rule for the cup product explained at the end of Section 2, we have

$$y_i \cdot w = \sum_{\mathcal{P},\mathcal{P}'} \bigodot_{j=1}^{p} b_{1,j}b'_{j,1} \odot \bigodot_{j=1}^{p} b_{2,j}b'_{j,2}.$$

Observe that a partition $(b_{1,1},\dots,b_{1,p})$ of $\gamma_{1,k}$ corresponds to a partition $k_1,\dots,k_p$ of the natural number k with nonnegative integers. Explicitly, the correspondence is given by $b_{1,j} = \gamma_{1,k_j}$. Similarly, a partition $(b_{2,1},\dots,b_{2,p})$ of $1_{p(p-i)}$ corresponds to a partition $h_1,\dots,h_p$ of $p(p-i)$ by the rule $b_{2,j} = 1_{h_j}$. The only splittings of $\gamma_{1,1}^a \alpha_{1,1}^\varepsilon$ are $(\gamma_{1,1}^a \alpha_{1,1}^\varepsilon, 1_0)$ and $(1_0, \gamma_{1,1}^a \alpha_{1,1}^\varepsilon)$.

Hence, we can write explicitly $y_i \cdot (\bigodot_{j=1}^p \gamma_{1,1}^{t_j})$ as a linear combination of elements of our basis $\mathcal{B}$, and we obtain

$$y_i \cdot w = \lambda_i \bigodot_{k=1}^{p} \gamma_{1,1}^{t_k+\delta_{k\le i}} \alpha_{1,1}^{\varepsilon_k} + \cdots$$

for some $\lambda_i \in \mathbb{Z}_p \setminus \{0\}$, where $\dots$ stands for terms that are smaller than the previous one in the considered ordering. With the same reasoning we can prove that

$$z_i \cdot w = \eta_i \bigodot_{k=1}^{p} \gamma_{1,1}^{t_k+\delta_{k<i}} \alpha_{1,1}^{\varepsilon_k+\delta_{\min\{h\ge i:\varepsilon_h=0\}}(k)} + \cdots,$$

where $\eta_i \in \mathbb{Z}_p \setminus \{0\}$ and $\dots$ has the same meaning as before.

An additive basis for S is given by

$$\mathcal{B}' = \Big\{x_1^a y_p^b x_2^{\varepsilon_1} x_3^{\varepsilon_2} x_4^{\varepsilon_3} : a,b,\varepsilon_i \ge 0,\ \sum_{i=1}^{3} \varepsilon_i \le 1\Big\} \cup \Big\{\prod_{i=1}^{p} y_i^{t_i} \prod_{i=1}^{p} z_i^{\varepsilon_i} : t_i \ge 0, \varepsilon_i \in \{0,1\}\Big\}.$$

By induction, using the previous formulas, the expansion in the basis $\mathcal{B}$ of the cohomology class $\varphi\big(\prod_{i=1}^p y_i^{t_i} \prod_{i=1}^p z_i^{\varepsilon_i}\big)$ (with $t_i \ge 0$ and $\varepsilon_i \in \{0,1\}$) is in the form

$$\varphi\Big(\prod_{i=1}^{p} y_i^{t_i} \prod_{i=1}^{p} z_i^{\varepsilon_i}\Big) = \lambda_{\underline{t},\underline{\varepsilon}} \bigodot_{i=1}^{p} \gamma_{1,1}^{\sum_{k=i}^p t_k + \sum_{k=i+1}^p \varepsilon_k} \alpha_{1,1}^{\varepsilon_i} + \cdots,$$

where, again, $\lambda_{\underline{t},\underline{\varepsilon}} \ne 0$ in $\mathbb{Z}_p$ and $\dots$ stands for smaller terms. This implies that the matrix associated with the $\mathbb{Z}_p$–linear function

$$\varphi\colon \operatorname{Span}\Big\{\prod_{i=1}^{p} y_i^{t_i} \prod_{i=1}^{p} z_i^{\varepsilon_i}\Big\} \to \operatorname{Span}\Big\{\bigodot_{i=1}^{p} \gamma_{1,1}^{t_i} \alpha_{1,1}^{\varepsilon_i}\Big\}$$

with respect to the two bases considered above (if we properly order their elements) is triangular, with all nonzero entries on the diagonal. Hence, $\varphi\colon A \to H^*(\Sigma_{p^2};\mathbb{Z}_p)$ must be an isomorphism. □

References

[1] **A Adem**, **R J Milgram**, *Cohomology of finite groups*, 2nd edition, Grundl. Math. Wissen. 309, Springer, Berlin (2004) MR

[2] **F R Cohen**, **T J Lada**, **J P May**, *The homology of iterated loop spaces*, Lecture Notes in Mathematics 533, Springer, Berlin (1976) MR

[3] **C Giusti**, **P Salvatore**, **D Sinha**, *The mod-2 cohomology rings of symmetric groups*, J. Topol. 5 (2012) 169–198 MR

[4] **N H V Hung**, **P A Minh**, *The action of the mod p Steenrod operations on the modular invariants of linear groups*, Vietnam J. Math. 23 (1995) 39–56 MR

[5] **B M Mann**, *The cohomology of the symmetric groups*, Trans. Amer. Math. Soc. 242 (1978) 157–184 MR

[6] **J P May**, *A general algebraic approach to Steenrod operations*, from "The Steenrod algebra and its applications" (F P Peterson, editor), Lecture Notes in Mathematics 168, Springer, Berlin (1970) 153–231 MR

[7] **H Mui**, *Modular invariant theory and cohomology algebras of symmetric groups*, J. Fac. Sci. Univ. Tokyo Sect. IA Math. 22 (1975) 319–369 MR

[8] **N P Strickland**, **P R Turner**, *Rational Morava E–theory and DS^0*, Topology 36 (1997) 137–151 MR

Scuola Normale Superiore
Piazza dei Cavalieri 7, I-56126 Pisa, Italy

lorenzo.guerra@sns.it

Received: 9 November 2015 Revised: 16 September 2016

Algebraic & Geometric Topology 17 (2017) 983–1020

Rational SO(2)–equivariant spectra

DAVID BARNES
J P C GREENLEES
MAGDALENA KĘDZIOREK
BROOKE SHIPLEY

We prove that the category of rational SO(2)–equivariant spectra has a simple algebraic model. Furthermore, all of our model categories and Quillen equivalences are monoidal, so we can use this classification to understand ring spectra and module spectra via the algebraic model.

55N91, 55P42, 55P60

1 Introduction

Rational equivariant cohomology theories This paper is a contribution to the study of equivariant cohomology theories, and gives a rather complete analysis for one class of theories. To start with, G–equivariant cohomology theories are represented by G–spectra, so that the category of G–equivariant cohomology theories and stable natural transformations between them is equivalent to the homotopy category of G–spectra, and it is natural to study the homotopy theory of G–spectra. One cannot expect a complete analysis of either cohomology theories or spectra integrally, but if we rationalize, things are greatly simplified, whilst valuable geometric and group-theoretic structures remain. Henceforth we restrict attention to rational cohomology theories and rational spectra without further comment. The general conjecture states that there is a nice algebraic model for rational G–spectra and, more precisely, a graded abelian category $\mathcal{A}(G)$ and a Quillen equivalence

$$G\text{–spectra} \simeq d\mathcal{A}(G),$$

where $d\mathcal{A}(G)$ consists of differential objects of $\mathcal{A}(G)$. The category $\mathcal{A}(G)$ is of injective dimension equal to the rank of G and of a form that is easy to use in calculations. Of course one would like the Quillen equivalence to reflect as much structure as possible. The conjecture is known for quite a number of groups in some form, and we refer to Greenlees and Shipley [16] for a summary of what is known. In the present paper we are concerned with the specific case of the circle group, and with giving a zigzag of Quillen equivalences which are symmetric monoidal.

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.983

The circle group We will entirely focus on the circle group, because it plays a critical role in understanding the case of all other infinite compact Lie groups. As an added benefit, it is significantly simpler than any other group, so we can focus on the critical features without being distracted by extraneous complication. We refer to the group as $\mathrm{SO}(2)$, because we have in mind as first applications its role as a subgroup of $O(2)$ (in Barnes [5]) and $\mathrm{SO}(3)$ (in Kędziorek [20]).

Our main result is as follows:

Main Theorem *The model category of rational* $\mathrm{SO}(2)$*–equivariant spectra is Quillen equivalent to the algebraic model* $d\mathcal{A}(\mathrm{SO}(2))_{\mathrm{dual}}$*. Furthermore, these Quillen equivalences are all symmetric monoidal, hence the homotopy category of rational* $\mathrm{SO}(2)$*–equivariant spectra and the homotopy category of the algebraic model* $D(\mathcal{A}(\mathrm{SO}(2)))$ *are equivalent as symmetric monoidal categories.*

The algebraic model is described in Section 2 below.

Rings and commutative rings Our main theorem establishes a zigzag of symmetric monoidal Quillen equivalences between the symmetric monoidal model category of rational $\mathrm{SO}(2)$–spectra and the symmetric monoidal model category $d\mathcal{A}(\mathrm{SO}(2))_{\mathrm{dual}}$. In particular we may use Schwede and Shipley [24, Theorem 3.12] to see that the model category of ring spectra is Quillen equivalent to the category of monoids in $\mathcal{A}(\mathrm{SO}(2))$. This means that a ring object R_a in $d\mathcal{A}(\mathrm{SO}(2))$ corresponds to a ring object R_{top} in $\mathrm{SO}(2)$–spectra in a homotopy-invariant fashion. Furthermore, the category of R_a–modules is Quillen equivalent to the category of R_{top}–modules.

However, it is essential to emphasize that if R_a is commutative, it does not follow that R_{top} will be a commutative $\mathrm{SO}(2)$–ring spectrum. The reason is that the correspondence between R_a and R_{top} is not simply applying the symmetric monoidal functors. Instead it involves derived functors and hence fibrant and cofibrant approximations. These approximations are only in the category of rings rather than in the category of commutative rings. This is inevitable, since for example the ring spectrum $R_{\mathrm{top}} = \widetilde{E}\mathcal{F}$ corresponds to a small and explicit commutative ring R_a, but it is well known — see McClure [22] — that $\widetilde{E}\mathcal{F}$ is not a commutative ring in orthogonal $\mathrm{SO}(2)$–spectra.

Greenlees [9] showed that if C is a generalized elliptic curve over a $\mathbb{Q}$–algebra, there is an associated $\mathrm{SO}(2)$–spectrum EC representing elliptic cohomology. Indeed the proof proceeds by writing down an object EC_a in $\mathcal{A}(\mathrm{SO}(2))$, and taking $EC = EC_{\mathrm{top}}$ to be the corresponding $\mathrm{SO}(2)$–spectrum. It is transparent from the construction that EC_a is a commutative ring in $\mathcal{A}(\mathrm{SO}(2))$, and it is a consequence of the present work that EC is a ring spectrum. As commented above, this does not prove that EC is a commutative ring spectrum, though it is easily verified to be homotopy commutative and compatible with the homotopy ring structure used by Ando and Greenlees [1].

Contribution of this paper To place the contribution of this paper in the study of rational SO(2)–spectra, we need to give a little history. A description of the homotopy category of rational SO(2)–spectra was given by Greenlees [8]. This took the form of an equivalence of homotopy categories

$$\mathrm{Ho}(\mathrm{SO}(2)\text{–spectra}) \simeq D(\mathcal{A}(\mathrm{SO}(2)))$$

for the abelian category $\mathcal{A}(\mathrm{SO}(2))$ (described in Section 2 below) without giving a Quillen equivalence of model categories inducing it. Since $\mathcal{A}(\mathrm{SO}(2))$ is rather simple and of injective dimension 1, this gives a practical means for calculating the space $[X, Y]_*^{\mathrm{SO}(2)}$ of maps for arbitrary (rational) SO(2)–spectra X and Y up to extension. Since $\mathcal{A}(\mathrm{SO}(2))$ is (in a sense that will appear later) evenly graded, the extensions split, and so [8] gives a complete description of the category Ho(SO(2)–spectra). Unfortunately, [8] claimed that the above equivalence of homotopy categories is an equivalence of triangulated categories, but there is a gap in this argument. Patchkoria, who noticed this gap, gave in [23] (and more recent work) an illuminating systematic analysis of lifting equivalences of homotopy categories to ones that preserve triangulations and other structures. The purported argument for $\mathcal{A}(\mathrm{SO}(2))$ in [8] fits into Patchkoria's formalism, but does not satisfy the conditions necessary to apply Patchkoria's results. Shipley [26] showed that if the claimed triangulated equivalence of homotopy categories in [8] existed, it would lift to a Quillen equivalence of model categories. Work then began to give an algebraic model for the homotopy category of G–spectra for a torus G (eventually leading to Greenlees and Shipley [16]); it was soon apparent that the only way to approach this is to first prove a Quillen equivalence between G–spectra and $d\mathcal{A}(G)$ and then deduce the equivalence of homotopy categories as a consequence. This general project has taken some time, and has a complicated history of its own [13; 14; 15; 16], but the special case of the circle is much simpler than the general case, and easily explained. The underlying strategy applied in [16] is the same as that adopted here for the circle group, but there are some significant differences of implementation adopted from Barnes [2; 5] and Kędziorek [19; 20].

Meanwhile, work began on the group $O(2)$ (culminating in the model of Barnes [5]) and the group SO(3) (culminating in the model of Kędziorek [19]). Those models depended on the Quillen equivalence for SO(2); they originally built upon Greenlees and Shipley [16], but the technical context adopted here has advantages for them. The proof for the general torus is considerably more complicated than that for the circle, principally because SO(2) has only two connected subgroups (namely the trivial group and the whole group) rather than infinitely many for higher-dimensional tori.

Accordingly, it is much easier to see the essential structure of the argument in the case of the circle. It is therefore desirable to give a separate account for $\mathrm{SO}(2)$ to show the simplicity of the argument, and to provide the input to the work on $O(2)$ and $\mathrm{SO}(3)$.

Perhaps a more important reason for publishing a separate account for $\mathrm{SO}(2)$ is that at present we can prove more for the circle group than for a general torus. The category of G–spectra is a monoidal model category, and $\mathcal{A}(G)$ is a monoidal abelian category. One would like to have a monoidal equivalence between G–spectra and $d\mathcal{A}(G)$. Of course this requires more care and some more delicate analysis than a simple Quillen equivalence. As the first step, one needs a monoidal model structure on $d\mathcal{A}(G)$. The abelian category $\mathcal{A}(G)$ does not have enough projectives and the injective model structure on $d\mathcal{A}(G)$ used in earlier work is certainly not monoidal. On the other hand, for $G = \mathrm{SO}(2)$, constructing a monoidal model structure on $d\mathcal{A}(\mathrm{SO}(2))$ is the primary task of Barnes [4]. This result relies on some explicit constructions in $\mathcal{A}(\mathrm{SO}(2))$ from [8] that are not made explicit in Greenlees [10; 11] for higher tori. It is expected that a similar construction will work for other groups, but additional work will be necessary. Once a monoidal model structure is defined on $d\mathcal{A}(G)$, one would need to ensure that all Quillen pairs making up the equivalence are monoidal. At present, this is only accessible for $G = \mathrm{SO}(2)$.

The Hasse–Tate isotropy square The overarching strategy for building an algebraic model is to break the category of $\mathrm{SO}(2)$–spectra into parts, give an algebraic model of each part and then assemble an algebraic model for all spectra from the algebraic models of the parts.

To analyse an individual $\mathrm{SO}(2)$–spectrum it is natural to use isotropy separation, to assemble the spectrum from information at the family $\mathcal{F}$ of finite subgroups and the information at $\mathrm{SO}(2)$ itself. This can be implemented using the Tate square

$$\begin{tikzcd}
X \arrow[r] \arrow[d] & X \wedge \widetilde{E}\mathcal{F} \arrow[d] \\
F(E\mathcal{F}_+, X) \arrow[r] & F(E\mathcal{F}_+, X) \wedge \widetilde{E}\mathcal{F}
\end{tikzcd}$$

which expresses X as the homotopy pullback of its $\mathcal{F}$–completion, $F(E\mathcal{F}_+, X)$, and its localization away from $\mathcal{F}$, namely $X \wedge \widetilde{E}\mathcal{F}$, over the Tate object, $F(E\mathcal{F}_+, X) \wedge \widetilde{E}\mathcal{F}$. Thus X is the homotopy pullback of a punctured square diagram (ie a diagram of shape $\mathscr{P} = (\bullet \to \bullet \leftarrow \bullet)$). The basic idea is to do this at the level of model categories. We would like to assemble the category of all $\mathrm{SO}(2)$–spectra from the category of $\mathcal{F}$–complete objects and objects localized away from $\mathcal{F}$. The way we do it here is to take suitable model categories of $\mathcal{F}$–complete spectra, of spectra away from $\mathcal{F}$, and Tate spectra, and then construct a model structure on the category $S^\bullet$–mod of $\mathscr{P}$–diagrams of such objects: a cellularization (K_{top}–cell–$S^\bullet$–mod) of this model category of $\mathscr{P}$–diagrams is then shown to be Quillen equivalent to the original category of SO(2)–spectra essentially using the fact that the Tate square is a homotopy pullback. The machinery of Greenlees and Shipley [15] was built for this purpose.

The alternative, adopted in Greenlees and Shipley [16], is to say that the category of SO(2)–spectra is equivalent to the category of $\mathbb{S}$–modules in SO(2)–spectra, where $\mathbb{S}$ is the sphere spectrum. We then consider the special case of the Tate square in which $X = \mathbb{S}$ and say that $\mathbb{S}$ is the pullback of a diagram of rings, so that the module category of $\mathbb{S}$ is Quillen equivalent to a cellularization of the model category of modules over the $\mathscr{P}$–diagram of rings.

After this, the general strategy in either case is to show the $\mathscr{P}$–diagram of model categories is equivalent to a simpler one that can be made algebraically explicit. In the present paper, several of the monoidal functors are taken from Barnes [2; 5] and Kędziorek [19; 20] and, since we work in a context where $\widetilde{E}\mathcal{F}$ is not a commutative ring, we adopt their methods for the formality argument in Section 4.1.

Summary of the zigzag of Quillen equivalences To illustrate the zigzag of Quillen equivalences in the Main Theorem we present a diagram of key steps:

$$\begin{tikzcd}
 & L_{\mathbb{S}_{\mathbb{Q}}}\mathbb{T}\,\mathrm{Sp} \arrow[d, shift right, "S^\bullet\wedge-"'] & \text{(in } \mathbb{T}\,\mathrm{Sp}) \\
\text{(Proposition 3.2.5)} & K_{\mathrm{top}}\text{–cell–}S^\bullet\text{–mod} \arrow[u, shift right, "\mathrm{pb}"'] \arrow[d, shift left, "(-)^{\mathbb{T}}"] & \text{(in } \mathbb{T}\,\mathrm{Sp}) \\
\text{(Corollary 3.3.6)} & K^{\mathbb{T}}_{\mathrm{top}}\text{–cell–}S^\bullet_{\mathrm{top}}\text{–mod} \arrow[u, shift left, "\zeta_\#"] \arrow[d, leftrightarrow, "\text{zigzag}"', "\text{of Quillen equivalences}"] & \text{(in Sp)} \\
\text{(Corollary 3.4.6)} & K_t\text{–cell–}S^\bullet_t\text{–mod} \arrow[d, leftrightarrow, "\text{zigzag}"', "\text{of Quillen equivalences}"] & \text{(in Ch}(\mathbb{Q})) \\
\text{(Section 4.1)} & K_a\text{–cell–}S^\bullet_a\text{–mod} \arrow[d, shift left, "\Gamma"] & \text{(in Ch}(\mathbb{Q})) \\
\text{(Proposition 4.2.4)} & dA(\mathbb{T})_{\mathrm{dual}} \arrow[u, shift left, "\mu"] & \text{(in Ch}(\mathbb{Q}))
\end{tikzcd}$$

At the top we have our preferred model for rational $\mathbb{T}$–spectra (namely the left Bousfield localization $L_{\mathbb{S}_\mathbb{Q}}\mathbb{T}\,\mathrm{Sp}$ of the category of orthogonal $\mathbb{T}$–spectra at the rational sphere spectrum (Definition 3.2.1)) and at the bottom we have our algebraic model $d\mathcal{A}(\mathbb{T})_{\mathrm{dual}}$.

The first step, moving into categories of $\mathscr{P}$–diagrams, was suggested in the previous subsection, and the other steps will be described in the body of the paper. The reader may wish to refer to this diagram now, but the notation will be introduced as we proceed. In the diagram, left Quillen functors are placed on the left and $\mathbb{T} := \mathrm{SO}(2)$. References to specific results are given on the left, and on the right there is an indication of the ambient category. In the following the subscript "top" indicates that the corresponding object has a topological origin, whereas the subscript "t" indicates that the object is algebraic, but has been produced by applying the results of Shipley [27]. The subscript "a" indicates that the object is algebraic in nature and has an explicit description. The symbols $S^\bullet_{(-)}$ refer to particular $\mathscr{P}$–diagrams of model categories, and the various categories $S^\bullet_{(-)}$–mod are generalizations of the notion of a module over a diagram of rings; see Section 3.1. These model categories are cellularized (ie right Bousfield localized) at the sets of objects $K_{(-)}$, which at every level of the diagram are the derived images of the usual stable generators $\mathbb{T}/H_+$ of $\mathbb{T}$–spectra, where H varies through closed subgroups of $\mathbb{T}$.

Notation From now on we will write $\mathbb{T}$ for the group $\mathrm{SO}(2)$. We also stick to the convention of drawing the left adjoint above the right one in any adjoint pair. We use $\mathrm{Ch}(\mathbb{Q})$ for the category of chain complexes of rational vector spaces, Sp for the category of orthogonal spectra, $G\,\mathrm{Sp}$ for the category of orthogonal G–spectra and Sp^Σ for the category of symmetric spectra.

2 The algebraic model $d\mathcal{A}(\mathbb{T})$

In this section we recall the algebraic category $\mathcal{A}(\mathbb{T})$ as developed by the second author [8]. This category is naturally enriched in graded abelian groups. We use the notation $d\mathcal{A}(\mathbb{T})$ for the category of objects in $\mathcal{A}(\mathbb{T})$ with a differential and call it the algebraic model for rational $\mathbb{T}$–spectra. A nonmonoidal model structure for the category $d\mathcal{A}(\mathbb{T})$ is given in [8]. Work of the first author [4] builds upon this and constructs a monoidal model structure on $d\mathcal{A}(\mathbb{T})$.

We call $\mathcal{A}(\mathbb{T})$ the *abelian model* for rational $\mathbb{T}$–spectra and $d\mathcal{A}(\mathbb{T})$ the *algebraic model* for rational $\mathbb{T}$–spectra. The model structures we construct on $d\mathcal{A}(\mathbb{T})$ are such that $\mathrm{Ho}(d\mathcal{A}(\mathbb{T}))$ is equivalent to the derived category of the abelian model, $D(\mathcal{A}(\mathbb{T}))$, which is equivalent to the homotopy category of rational $\mathbb{T}$–spectra by [8].

2.1 The abelian model $\mathcal{A}(\mathbb{T})$

The abelian model for rational $\mathbb{T}$–spectra is established in [8]. We introduce this category, explain how to turn it into a differential graded category and then define the injective model structure.

Definition 2.1.1 Let $\mathcal{F}$ be the set of finite subgroups of $\mathbb{T}$. Let $\mathcal{O}_\mathcal{F}$ be the graded ring $\prod_{n\geqslant 1}\mathbb{Q}[c_n]$ with c_n of degree -2. Let e_n be the idempotent arising from projection onto factor n. In general, let ϕ be a subset of $\mathcal{F}$ and define e_ϕ to be the idempotent coming from projection onto the factors in ϕ. We let c be the unique element of $\mathcal{O}_\mathcal{F}$ such that $c_n = e_n c$ for all $n \geq 1$.

We use the notation $\mathcal{E}^{-1}\mathcal{O}_\mathcal{F} = \mathrm{colim}_{n\geqslant 1}\,\mathcal{O}_\mathcal{F}[c_1^{-1},\dots,c_n^{-1}]$. It is easy to see that $\mathcal{E}^{-1}\mathcal{O}_\mathcal{F}$ is a ring. The notation arises since this ring can also be described in terms of inverting a certain set of Euler classes. As a vector space, $(\mathcal{E}^{-1}\mathcal{O}_\mathcal{F})_{2k}$ is $\prod_{n\geqslant 1}\mathbb{Q}$ for $k \leqslant 0$ and is $\bigoplus_{n\geqslant 1}\mathbb{Q}$ for $n > 0$.

For any $\mathcal{O}_\mathcal{F}$ module N, we define $\mathcal{E}^{-1}N$ to be $\mathcal{E}^{-1}\mathcal{O}_\mathcal{F}\otimes_{\mathcal{O}_\mathcal{F}} N$.

Definition 2.1.2 We define the *abelian model* $\mathcal{A} = \mathcal{A}(\mathbb{T})$ as follows. Its class of objects is the collection of triples (N, U, β) where N is an $\mathcal{O}_\mathcal{F}$–module, U is a graded rational vector space and

$$\beta\colon N \to \mathcal{E}^{-1}\mathcal{O}_\mathcal{F}\otimes U$$

is an $\mathcal{O}_\mathcal{F}$–module map such that $\mathcal{E}^{-1}\beta$ is an isomorphism.[1] We will often refer to β as the *structure map*.

A map (θ,ϕ) in $\mathcal{A}$ is a commutative square

$$\begin{tikzcd} N \arrow[r, "\beta"] \arrow[d, "\theta"'] & \mathcal{E}^{-1}\mathcal{O}_\mathcal{F}\otimes U \arrow[d, "\mathrm{Id}\otimes\phi"] \\ N' \arrow[r, "\beta'"] & \mathcal{E}^{-1}\mathcal{O}_\mathcal{F}\otimes U' \end{tikzcd}$$

where θ is a map of $\mathcal{O}_\mathcal{F}$–modules and ϕ is a map of graded rational vector spaces.

The relation between this category and rational $\mathbb{T}$–equivariant spectra is given by the following pair of theorems from [8].

Theorem 2.1.3 *The homotopy category of rational $\mathbb{T}$–equivariant spectra is equivalent to the derived category of $\mathcal{A}$.*

[1] The tensor product in the target of β is over $\mathbb{Q}$, which we omit from the notation.

For a rational $\mathbb{T}$–equivariant spectrum X, let $\pi_*^{\mathcal{A}}(X)$ be the following object of $\mathcal{A}$, which is its counterpart in $\mathcal{A}$:

$$\pi_*^{\mathcal{A}}(X) = \big(\pi_*^{\mathbb{T}}(X \wedge DE\mathcal{F}_+) \to \pi_*^{\mathbb{T}}(X \wedge DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F}) \cong \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \pi_*(\Phi^{\mathbb{T}} X)\big).$$

For details of the spectra $DE\mathcal{F}_+$ and $\widetilde{E}\mathcal{F}$ see Definition 3.2.2. The spectrum $\Phi^{\mathbb{T}} X$ is the geometric $\mathbb{T}$–fixed points of X.

There is also an Adams short exact sequence which explains how to calculate maps in the homotopy category of rational $\mathbb{T}$–equivariant spectra:

Theorem 2.1.4 *Let X and Y be rational $\mathbb{T}$–equivariant spectra. Then the sequence below is exact:*

$$0 \to \mathrm{Ext}_{\mathcal{A}}(\pi_*^{\mathcal{A}}(\Sigma X), \pi_*^{\mathcal{A}}(Y)) \to [X, Y]_*^{\mathbb{T}} \to \mathrm{Hom}_{\mathcal{A}}(\pi_*^{\mathcal{A}}(X), \pi_*^{\mathcal{A}}(Y)) \to 0.$$

In [8] a model structure is given for the category of objects in $\mathcal{A}$ that have a differential. We define what it means to have a differential and then introduce the model structure. We will leave the proof that $\mathcal{A}$ has all small limits and colimits to the next subsection (see also [8]).

We can consider $\mathcal{O}_{\mathcal{F}}$ as an object of $\mathrm{Ch}(\mathbb{Q})$ with trivial differential and, as such, it is a commutative algebra in $\mathrm{Ch}(\mathbb{Q})$. An $\mathcal{O}_{\mathcal{F}}$–module in $\mathrm{Ch}(\mathbb{Q})$ is an $\mathcal{O}_{\mathcal{F}}$–module in graded vector spaces N along with maps $d_n\colon N_n \to N_{n-1}$. Note that these maps satisfy the relations

$$d_{n-1} \circ d_n = 0, \quad c d_n = d_{n-2} c.$$

Definition 2.1.5 We define the category $d\mathcal{A} = d\mathcal{A}(\mathbb{T})$ as follows. Its class of objects is the collection of triples (N, U, β) where N is a rational chain complex with an action of $\mathcal{O}_{\mathcal{F}}$, U is a rational chain complex and

$$\beta\colon N \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U$$

is a $\mathcal{O}_{\mathcal{F}}$–module map in $\mathrm{Ch}(\mathbb{Q})$ such that $\mathcal{E}^{-1}\beta$ is an isomorphism.

A map (θ, ϕ) in $d\mathcal{A}$ is then a commutative square as for $\mathcal{A}$ such that θ is a map in the category of $\mathcal{O}_{\mathcal{F}}$–modules in $\mathrm{Ch}(\mathbb{Q})$ and ϕ is a map of $\mathrm{Ch}(\mathbb{Q})$.

We call this category the *algebraic model for rational $\mathbb{T}$–spectra*.

Note that the category $d\mathcal{A}$ is not the same as $\mathrm{Ch}(\mathcal{A})$, since $\mathcal{A}$ is a graded category and in $d\mathcal{A}$ we do not introduce an additional grading; instead we take objects of $\mathcal{A}$ with a differential.

The following result is the subject of [8, Appendix B]:

Proposition 2.1.6 *The category $d\mathcal{A}$ has a model structure where the class of weak equivalences is exactly the class of quasi-isomorphisms. The class of cofibrations is the class of monomorphisms.* (This is called the *injective model structure*. We write $d\mathcal{A}_i$ for this model structure.)

As we shall see shortly, the category $\mathcal{A}$ has a monoidal product which induces a monoidal product on $d\mathcal{A}$. But the injective model structure does not make $d\mathcal{A}$ into a monoidal model category. This failure occurs because of c–torsion, just as the injective model structure on $\mathrm{Ch}(\mathbb{Z})$ is not monoidal due to torsion.

This is a serious defect, as we are unable to compare the monoidal product in $d\mathcal{A}$ to the smash product of $\mathbb{T}$–spectra. This defect is further complicated by the lack of projective objects of $\mathcal{A}$. There is however a cofibrantly generated monoidal model structure on $d\mathcal{A}$ which is Quillen equivalent to the injective model structure. It is constructed in [4] and we recall it in the next subsection.

2.2 The monoidal model structure

This subsection has three aims, namely to prove that $\mathcal{A}$ and $d\mathcal{A}$ have all small limits and colimits (see also [8]), define the monoidal product and recall the dualizable model structure on $d\mathcal{A}$ (see [4]), which is monoidal. To do so, we will need to relate $\mathcal{A}$ to a larger category $\widehat{\mathcal{A}}$, which we introduce next.

We let $\widehat{\mathcal{A}}$ be category of triples $(N, U, \alpha\colon N \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U)$ where N is an $\mathcal{O}_{\mathcal{F}}$–module, U is graded $\mathbb{Q}$–module and the map α is a map of $\mathcal{O}_{\mathcal{F}}$–modules. A map of such diagrams is a commutative diagram as below where θ is a map of $\mathcal{O}_{\mathcal{F}}$–modules, and ϕ is a map of graded $\mathbb{Q}$–modules:

$$\begin{tikzcd} N \arrow[r] \arrow[d, "\theta"] & \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U \arrow[d, "\mathrm{Id}\otimes\phi"] \\ N' \arrow[r] & \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U' \end{tikzcd}$$

Thus $\widehat{\mathcal{A}}$ is $\mathcal{A}$ without the restriction that the structure map of an object should be an isomorphism after $\mathcal{E}$ is inverted. There is an adjunction

$$\iota\colon \mathcal{A} \rightleftarrows \widehat{\mathcal{A}} :\Gamma_h,$$

where ι is the inclusion. The functor ι is full and faithful. The explicit construction of the right adjoint Γ_h, which we call the *torsion functor*, is quite intricate and therefore we leave the details to [8, Section 20.2].

Our first use of the torsion functor Γ_h is to define limits in $\mathcal{A}$. It follows from [8, Section 20.2] that the adjunction (ι, Γ_h) passes to categories with differentials, as does the following definition:

Definition 2.2.1 Let I be some small category and let $\{N_i \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U_i\}$ be the objects of some I–shaped diagram in $\mathcal{A}$. The colimit over I is

$$\mathrm{colim}_i\, N_i \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes (\mathrm{colim}_i\, U_i).$$

The limit is formed by applying the functor ι, taking limits in $\widehat{\mathcal{A}}$ and then applying Γ_h. In more detail, we construct the following pullback square:

$$\begin{tikzcd} M \arrow[r] \arrow[d, "f"'] & \lim(N_i) \arrow[d] \\ \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \lim(U_i) \arrow[r] & \lim(\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U_i) \end{tikzcd}$$

The limit of the I–shaped diagram $\{N_i \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U_i\}$ is $\Gamma_h f$.

Now we turn to the monoidal product of $\mathcal{A}$ and $d\mathcal{A}$.

Definition 2.2.2 For $\beta\colon N \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U$ and $\beta'\colon N' \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U'$ in $d\mathcal{A}$, their *tensor product* is

$$\beta \otimes \beta'\colon N \otimes_{\mathcal{O}_{\mathcal{F}}} N' \to (\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U) \otimes_{\mathcal{O}_{\mathcal{F}}} (\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U') \cong \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes (U \otimes_{\mathbb{Q}} U').$$

The unit of this monoidal product is the object $S^0 = (i\colon \mathcal{O}_{\mathcal{F}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q})$.

This monoidal product is related to the smash product of spectra, as we can see from the short exact sequence of [8],

$$0 \to \pi_*^{\mathcal{A}}(X) \otimes \pi_*^{\mathcal{A}}(Y) \to \pi_*^{\mathcal{A}}(X \wedge Y) \to \Sigma\mathrm{Tor}(\pi_*^{\mathcal{A}}(X), \pi_*^{\mathcal{A}}(Y)) \to 0.$$

This monoidal structure is closed, that is, there is an internal function object describing the $d\mathcal{A}$–object of maps between two objects. This functor is more complicated than the tensor product and requires use of the torsion functor Γ_h.

Definition 2.2.3 Consider two elements of $d\mathcal{A}$,

$$A = (\beta\colon N \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U) \quad \text{and} \quad B = (\beta'\colon N' \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U').$$

The *function object* $F(A,B)$ is the map $\Gamma_h\delta$, where δ is defined by the pullback square:

$$\begin{tikzcd}
Q \arrow[rr, "\delta"] \arrow[dd] & & \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathrm{Hom}_{\mathbb{Q}}(U,U') \arrow[d] \\
& & \mathrm{Hom}_{\mathcal{O}_{\mathcal{F}}}(\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U, \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U') \arrow[d] \\
\mathrm{Hom}_{\mathcal{O}_{\mathcal{F}}}(N,N') \arrow[rr] & & \mathrm{Hom}_{\mathcal{O}_{\mathcal{F}}}(N, \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U')
\end{tikzcd}$$

The monoidal product and function object are related by a natural isomorphism by [8, Lemma 22.6.2]. Let A, B and C be objects of $d\mathcal{A}$; then

$$d\mathcal{A}(A \otimes B, C) \cong d\mathcal{A}(A, F(B,C)).$$

Definition 2.2.4 For $K \in \mathrm{Ch}(\mathbb{Q})$ we define $LK \in d\mathcal{A}$ as

$$LK = (i \otimes \mathrm{Id}_K \colon \mathcal{O}_{\mathcal{F}} \otimes K \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes K).$$

Note that $LK = S^0 \otimes K$, where $S^0 = (i \colon \mathcal{O}_{\mathcal{F}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q})$. For A and B in $d\mathcal{A}$, we define $\mathcal{A}(A,B)_*$ to be the graded set of maps of $\mathcal{A}$ (ignoring the differential). We then equip this graded $\mathbb{Q}$–module with the differential induced by the convention $df_n = d_B f_n + (-1)^{n+1} f_n d_A$. This construction gives a functor

$$R \colon d\mathcal{A} \to \mathrm{Ch}(\mathbb{Q}), \quad RA := \mathcal{A}(S^0, A)_*.$$

The functors L and R form an adjoint pair between $\mathrm{Ch}(\mathbb{Q})$ and $d\mathcal{A}$. Furthermore, they give $d\mathcal{A}$ the structure of a closed $\mathrm{Ch}(\mathbb{Q})$–module in the sense of [18, Section 4.1].

This module structure and the closed monoidal product interact to give $d\mathcal{A}$ a tensor product, a cotensor product and an enrichment over $\mathrm{Ch}(\mathbb{Q})$. Let $K \in \mathrm{Ch}(\mathbb{Q})$ and $A = (\beta \colon N \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U)$ in $d\mathcal{A}$. Their *tensor product* $A \otimes K$ is defined to be $A \otimes LK$. Thus $A \otimes K$ is given by

$$\beta \otimes \mathrm{Id}_K \colon N \otimes_{\mathbb{Q}} K \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes (U \otimes_{\mathbb{Q}} K).$$

There is a *cotensor product* A^K defined to be $F(LK, A)$. The *enrichment* is given by $RF(A,B)$ for A and B in $d\mathcal{A}$. This enrichment, tensor and cotensor are related by the natural isomorphisms

$$d\mathcal{A}(A, B^K) \cong d\mathcal{A}(A \otimes K, B) = d\mathcal{A}(A \otimes LK, B) \cong \mathrm{Ch}(\mathbb{Q})(K, RF(A,B)).$$

Now we are ready to recall the monoidal model structure on $d\mathcal{A}$ from [4] and compare it to several other model categories, in particular the injective model structure on $d\mathcal{A}$

introduced in [8]. This monoidal model structure is defined in terms of the (strongly) dualizable objects of $d\mathcal{A}$.

Definition 2.2.5 An object $A \in \mathcal{A}$ is said to be (strongly) *dualizable* if for any $B \in \mathcal{A}$ the canonical map

$$F(A, S^0) \otimes B \to F(A, B)$$

is an isomorphism. The *functional dual* of an object B is the object $DB = F(B, S^0)$.

Let $\mathcal{P}$ be a set of representatives for the isomorphisms classes of dualizable objects in $\mathcal{A}$. Such a set exists by [4, Corollary 5.8]. The following theorem summarizes [4, Section 6]:

Theorem 2.2.6 *There is a cofibrantly generated model structure on $d\mathcal{A}$ with weak equivalences the class generated by the homology isomorphisms. The generating cofibrations have the form*

$$S^{n-1} \otimes P \to D^n \otimes P$$

for $P \in \mathcal{P}$ and $n \in \mathbb{Z}$, where S^n is the chain complex consisting of one copy of $\mathbb{Q}$ in degree n and 0 elsewhere and D^n consists of two copies of $\mathbb{Q}$ in degrees n and $n-1$ with the identity as the only nontrivial differential.

(We call this model structure the *dualizable model structure* and denote it by $d\mathcal{A}_{\mathrm{dual}}$. The dualizable model structure is proper, symmetric monoidal and satisfies the monoid axiom.)

Since all cofibrations in the dualizable model structure are in particular monomorphisms we get the following comparison with the injective model structure of [8], which we write as $d\mathcal{A}_i$ (see Proposition 2.1.6 for the description of the injective model structure).

Lemma 2.2.7 *The identity functor from $d\mathcal{A}_{\mathrm{dual}}$ to $d\mathcal{A}_i$ is the left adjoint of a Quillen equivalence,*

$$\mathrm{Id}\colon d\mathcal{A}_{\mathrm{dual}} \rightleftarrows d\mathcal{A}_i :\mathrm{Id}.$$

The object S^0 is clearly dualizable. Similarly, if V is a finite-dimensional vector space, then $S^0 \otimes V$ is dualizable. As a consequence, we have the following lemma:

Lemma 2.2.8 *There is a strong symmetric monoidal Quillen pair*

$$L\colon \mathrm{Ch}(\mathbb{Q}) \rightleftarrows d\mathcal{A}_{\mathrm{dual}} :R,$$

where $LV = S^0 \otimes V$ and $RA = \mathcal{A}(S^0, A)_*$ (see Definition 2.2.4). *Thus, $d\mathcal{A}_{\mathrm{dual}}$ is a closed $\mathrm{Ch}(\mathbb{Q})$–model category.*

3 Obtaining an algebraic category

The method of this section is the synthesis of three ideas. The first idea is to use the Hasse–Tate square from the introduction to separate the homotopical information of $\mathbb{T}$–equivariant spectra into pieces where we can remove equivariance without losing any information.

For $\mathbb{T}$–equivariant spectra, the relevant decomposition is to separate the homotopical information coming from finite subgroups from the homotopical information coming from the whole group. For this separation we will need a diagram of model categories rather than a diagram of commutative rings. We establish the categorical foundations in the next subsection and then perform the separation in Section 3.2.

The second is that the correct way to remove equivariance is to take fixed points. The primary example is that taking $\mathbb{T}$–fixed points gives a Quillen equivalence from $DE\mathbb{T}_+$–modules in rational $\mathbb{T}$–equivariant spectra to $DB\mathbb{T}_+$–modules in rational spectra. Here $DE\mathbb{T}_+$ is the Spanier–Whitehead dual of $E\mathbb{T}_+$ in $\mathbb{T}$–spectra and $DB\mathbb{T}_+$ is the Spanier–Whitehead dual of $B\mathbb{T}_+$ in the category of spectra. See Section 3.3.

With the separation complete and equivariance removed, we use the results of [27] to move to an algebraic setting in Section 3.4. That is, we obtain a Quillen equivalence between rational $\mathbb{T}$–spectra and some combined cellularization–localization of an algebraic category.

The next step is to simplify that algebraic category into the algebraic model $\mathcal{A}(\mathbb{T})$, by directly calculating the effects of these cellularizations and localizations. This is the essence of the third idea: to leave any examination of localizations or cellularizations until one is working with an algebraic category. This occurs in Section 4.1, where we simplify the category created by the results of [27] and remove a localization. Finally, in Section 4.2 we remove a cellularization to get to the algebraic model.

3.1 Diagrams of model categories

We will use several model categories that are built from diagrams of model categories. This idea has been studied in some detail in [15]. In this section we introduce the relevant structures and leave most of the proofs to the reference. We will only use one shape of diagram, the pullback diagram $\mathscr{P}$:

$$\bullet \to \bullet \leftarrow \bullet.$$

Pullbacks of model categories are also considered in detail in [7].

Definition 3.1.1 A *$\mathscr{P}$–diagram of model categories* $R^\bullet$ is a pair of Quillen pairs

$$L\colon \mathcal{A} \rightleftarrows \mathcal{B} :R, \quad F\colon \mathcal{C} \rightleftarrows \mathcal{B} :G,$$

with L and F the left adjoints. We will usually draw this as

$$\mathcal{A} \underset{R}{\overset{L}{\rightleftarrows}} \mathcal{B} \underset{G}{\overset{F}{\leftrightarrows}} \mathcal{C}.$$

A standard example comes from a $\mathscr{P}$–diagram of rings $R = (R_1 \xrightarrow{f} R_2 \xleftarrow{g} R_3)$. Using the adjoint pairs of extension and restriction of scalars we obtain a $\mathscr{P}$–diagram of model categories $R^\bullet$:

$$R_1\text{–mod} \underset{f^*}{\overset{R_2 \otimes_{R_1} -}{\rightleftarrows}} R_2\text{–mod} \underset{g^*}{\overset{R_2 \otimes_{R_3} -}{\leftrightarrows}} R_3\text{–mod}.$$

Definition 3.1.2 Given a $\mathscr{P}$–diagram of model categories $R^\bullet$ we can define a new category, $R^\bullet$–mod. The objects of this category are pairs of morphisms $\alpha\colon La \to b$ and $\gamma\colon Fc \to b$ in $\mathcal{B}$. We usually abbreviate a pair $(\alpha\colon La \to b,\ \gamma\colon Fc \to b)$ to a quintuple $(a, \alpha, b, \gamma, c)$. We find this notation suggestive but emphasize that objects of $R^\bullet$–mod are not usually modules over a diagram of rings.

A morphism in $R^\bullet$–mod from $(a, \alpha, b, \gamma, c)$ to $(a', \alpha', b', \gamma', c')$ is a triple of maps $x\colon a \to a'$ in $\mathcal{A}$, $y\colon b \to b'$ in $\mathcal{B}$ and $z\colon c \to c'$ in $\mathcal{C}$ such that we have a commuting diagram in $\mathcal{B}$:

$$\begin{tikzcd}
La \arrow[r, "\alpha"] \arrow[d, "Lx"] & b \arrow[d, "y"] & Fc \arrow[l, "\gamma"'] \arrow[d, "Fz"] \\
La' \arrow[r, "\alpha'"] & b' & Fc' \arrow[l, "\gamma'"']
\end{tikzcd}$$

Note that we could also have defined an object as a sequence $(a, \overline{\alpha}, b, \overline{\gamma}, c)$, where $\overline{\alpha}\colon a \to Rb$ is a map in $\mathcal{A}$ and $\overline{\gamma}\colon c \to Gb$ is a map in $\mathcal{C}$.

We say that a map (x, y, z) in $R^\bullet$–mod is an objectwise cofibration if x is a cofibration of $\mathcal{A}$, y is a cofibration of $\mathcal{B}$ and z is a cofibration of $\mathcal{C}$. We define objectwise weak equivalences similarly.

Lemma 3.1.3 [15, Proposition 3.3] *Consider a $\mathscr{P}$–diagram of model categories $R^\bullet$, with each category cellular and proper,*

$$\mathcal{A} \underset{R}{\overset{L}{\rightleftarrows}} \mathcal{B} \underset{G}{\overset{F}{\leftrightarrows}} \mathcal{C}.$$

The category $R^\bullet$–mod admits a cellular proper model structure with cofibrations and weak equivalences defined objectwise. (This is called the *diagram injective* model structure.)

Whilst there is also a diagram projective model structure, in this paper we only use the diagram injective model structure (and cellularizations thereof) on diagrams of model categories.

Now consider maps of $\mathscr{P}$–diagrams of model categories. Let $R^\bullet$ and $S^\bullet$ be two $\mathscr{P}$–diagrams, where $R^\bullet$ is as above and $S^\bullet$ is

$$\mathcal{A}' \underset{R'}{\overset{L'}{\rightleftarrows}} \mathcal{B}' \underset{G'}{\overset{F'}{\leftrightarrows}} \mathcal{C}'.$$

Now we assume that we have Quillen adjunctions such that $P_2 L$ is naturally isomorphic to $L'P_1$ and $P_2 F$ is naturally isomorphic to $F'P_3$, given by

$$P_1\colon \mathcal{A} \rightleftarrows \mathcal{A}' :Q_1,$$
$$P_2\colon \mathcal{B} \rightleftarrows \mathcal{B}' :Q_2,$$
$$P_3\colon \mathcal{C} \rightleftarrows \mathcal{C}' :Q_3.$$

We then obtain a Quillen adjunction (P, Q) between $R^\bullet$–mod and $S^\bullet$–mod. For example, the left adjoint P takes the object $(a, \alpha, b, \gamma, c)$ to $(P_1 a,\, P_2\alpha,\, P_2 b,\, P_2\gamma,\, P_3 c)$. The commutativity assumptions ensure that this is an object of $S^\bullet$–mod. It is easy to see the following:

Lemma 3.1.4 *If the Quillen adjunctions (P_i, Q_i) are Quillen equivalences then the adjunction (P, Q) between $R^\bullet$–mod and $S^\bullet$–mod is a Quillen equivalence.*

Now we turn to monoidal considerations. There is an obvious monoidal product for $R^\bullet$–mod, provided that each of $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ is monoidal and that the left adjoints L and F are strong monoidal,

$$(a, \alpha, b, \gamma, c) \wedge (a', \alpha', b', \gamma', c') := (a \wedge a', \alpha \wedge \alpha', b \wedge b', \gamma \wedge \gamma', c \wedge c').$$

Let $S_{\mathcal{A}}$ be the unit of $\mathcal{A}$, let $S_{\mathcal{B}}$ be the unit of $\mathcal{B}$ and let $S_{\mathcal{C}}$ be the unit of $\mathcal{C}$. Since L and F are monoidal, we have maps $\eta_{\mathcal{A}}\colon LS_{\mathcal{A}} \to S_{\mathcal{B}}$ and $\eta_{\mathcal{C}}\colon FS_{\mathcal{C}} \to S_{\mathcal{B}}$. The unit of the monoidal product on $R^\bullet$–mod is $(S_{\mathcal{A}}, \eta_{\mathcal{A}}, S_{\mathcal{B}}, \eta_{\mathcal{C}}, S_{\mathcal{C}})$.

It is worth noting that this category has an internal function object when $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are closed monoidal categories and thus itself is closed.

Lemma 3.1.5 *Consider a $\mathscr{P}$–diagram of model categories $R^\bullet$ such that each vertex is a cellular monoidal model category. Assume further that the two adjunctions of the diagram are strong monoidal Quillen pairs. Then $R^\bullet$–mod is also a monoidal model category. If each vertex also satisfies the monoid axiom, so does $R^\bullet$–mod.*

Proof Since the cofibrations and weak equivalences are defined objectwise, the pushout product and monoid axioms hold provided they do so in each model category in the diagram $R^\bullet$. □

We can also extend our monoidal considerations to maps of diagrams. Return to the setting of a map (P, Q) of $\mathscr{P}$–diagrams from $R^\bullet$ to $S^\bullet$ as described above. If we assume that each of the adjunctions (P_1, Q_1), (P_2, Q_2) and (P_3, Q_3) is a symmetric monoidal Quillen equivalence, then we see that (P, Q) is a symmetric monoidal Quillen equivalence.

With these formalities out of the way, we are ready to move from the model category of rational $\mathbb{T}$–spectra to modules over a $\mathscr{P}$–diagram of model categories.

3.2 Isotropy separation

In this subsection we separate the homotopical information of rational $\mathbb{T}$–spectra into three parts. The first part takes care of the homotopical information coming from the finite cyclic subgroups. The second part deals with the homotopical information coming from $\mathbb{T}$. The third part is a comparison term which enforces some compatibility conditions on the two other parts.

We achieve this separation by replacing the category of rational $\mathbb{T}$–spectra with a Quillen equivalent category $S^\bullet$–mod, for $S^\bullet$ a $\mathscr{P}$–diagram of model categories (see Definition 3.2.3).

Before we do that, let us first recall some basic definitions and properties for $\mathbb{T}$–spectra.

Definition 3.2.1 Let $\mathbb{T}\,\mathrm{Sp}$ be the category of $\mathbb{T}$–equivariant orthogonal spectra indexed on a complete $\mathbb{T}$–universe $\mathcal{U}$ considered with the stable model structure.

This model category is monoidal, proper and cellular [21]. The weak equivalences are those maps f such that $\pi_*^H(f)$ is an isomorphism for all closed subgroups H of $\mathbb{T}$.

Following [3, Section 5] and using [21, Theorem IV.6.3], we localize this model category at the rational sphere spectrum $\mathbb{S}_\mathbb{Q}$. That is, we leave the underlying category unchanged and alter the model structure. We call the weak equivalences of the localized model structure *rational equivalences*: a map f is a rational equivalence if $\pi_*^H(f)\otimes\mathbb{Q}$ is an isomorphism for all closed subgroups H of $\mathbb{T}$. We call this model structure the *rational model structure* and use the notation $L_{\mathbb{S}_\mathbb{Q}}\mathbb{T}\,\mathrm{Sp}$.

The localized model category is still proper, cellular, monoidal and stable.

Definition 3.2.2 Let $\mathcal{F}$ be the collection of finite cyclic subgroups of $\mathbb{T}$. There is a universal space for this family, called $E\mathcal{F}$, where, by definition, $E\mathcal{F}^H$ is nonequivariantly contractible for each finite cyclic subgroup H and $E\mathcal{F}^{\mathbb{T}} = \varnothing$. We define $\widetilde{E}\mathcal{F}$ via the cofibre sequence of $\mathbb{T}$–spaces

$$E\mathcal{F}_+ \to S^0 \to \widetilde{E}\mathcal{F}.$$

We define $DE\mathcal{F}_+$ to be $F(E\mathcal{F}_+, N^{\#}\mathbb{S})$. Here $N^{\#}$ is the lax monoidal right adjoint described in [21, Theorem IV.3.9] from EKMM $\mathbb{T}$–equivariant $\mathbb{S}$–modules to $\mathbb{T}\,\mathrm{Sp}$.

Recall that $N^{\#}$ is the right adjoint of a Quillen equivalence when $\mathbb{T}\,\mathrm{Sp}$ is considered with the positive stable model structure (see [21, Chapter IV] for more details). The spectrum $DE\mathcal{F}_+$ is a commutative ring spectrum, which is fibrant in the positive stable model structure on $\mathbb{T}\,\mathrm{Sp}$.

We can use the above cofibre sequence to produce the Hasse–Tate homotopy pullback square of $\mathbb{T}$–equivariant spectra [12, Section 17]:

$$\begin{tikzcd}
\mathbb{S} \arrow[r] \arrow[d] & \widetilde{E}\mathcal{F} \arrow[d] \\
DE\mathcal{F}_+ \arrow[r] & DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F}
\end{tikzcd}$$

To see that it is a homotopy pullback square, note that the homotopy fibres of the top and bottom row are weakly equivalent (where the bottom row is the top one smashed with $DE\mathcal{F}_+$).

We have three model categories:

- $L_{\mathbb{S}_{\mathbb{Q}}}(DE\mathcal{F}_+\text{–mod})$, which captures the behaviour of the finite cyclic groups.
- $L_{\mathbb{S}_{\mathbb{Q}} \wedge \widetilde{E}\mathcal{F}}\mathbb{T}\,\mathrm{Sp}$, which captures the behaviour of $\mathbb{T}$.
- $L_{\mathbb{S}_{\mathbb{Q}} \wedge DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F}}(DE\mathcal{F}_+\text{–mod})$, which captures the interaction of the first two.

Now we can give our diagram of model categories that separates the behaviour of the finite cyclic groups from the rest.

Definition 3.2.3 We define $S^\bullet$ to be the $\mathscr{P}$–diagram of model categories

$$L_{\mathbb{S}_{\mathbb{Q}}}(DE\mathcal{F}_+\text{–mod}) \underset{\mathrm{Id}}{\overset{\mathrm{Id}}{\rightleftarrows}} L_{\mathbb{S}_{\mathbb{Q}} \wedge DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F}}(DE\mathcal{F}_+\text{–mod}) \underset{U}{\overset{DE\mathcal{F}_+ \wedge -}{\leftrightarrows}} L_{\mathbb{S}_{\mathbb{Q}} \wedge \widetilde{E}\mathcal{F}}\mathbb{T}\,\mathrm{Sp}.$$

Since all of the model categories in the diagram are cellular, proper, monoidal model categories, we have a cellular proper stable monoidal model category $S^\bullet$–mod that satisfies the monoid axiom.

Given an $X \in \mathbb{T}\,\mathrm{Sp}$, we have an $S^\bullet$–module

$$S^\bullet \wedge X := (DE\mathcal{F}_+ \wedge X, \mathrm{Id}, DE\mathcal{F}_+ \wedge X, \mathrm{Id}, X).$$

The functor $S^\bullet \wedge -$ has a right adjoint. Let $A = (a, \alpha, b, \gamma, c)$ be an $S^\bullet$–module. Then there are maps in $\mathbb{T}\,\mathrm{Sp}$, namely $a \to b$ and $c \to DE\mathcal{F}_+ \wedge c \to b$, where in the composite the first map is the unit of the adjunction $(DE\mathcal{F}_+ \wedge -, U)$ and the second map is γ. Thus we have a diagram in $\mathbb{T}\,\mathrm{Sp}$: $a \to b \leftarrow c$. We write $\mathrm{pb}\, A$ for the pullback of this diagram in $\mathbb{T}\,\mathrm{Sp}$. We assemble this construction into the following result, the proof of which is entirely routine:

Proposition 3.2.4 *There is a strong symmetric monoidal Quillen adjunction*

$$S^\bullet \wedge -\colon L_{S_\mathbb{Q}}(\mathbb{T}\,\mathrm{Sp}) \rightleftarrows S^\bullet\text{–mod} :\mathrm{pb}\,.$$

We want to turn this adjunction into a Quillen equivalence. To do so, we apply the cellularization principle of [13, Proposition 2.7]. The idea is to cellularize (right Bousfield localize; see also Section 5.1) the right-hand model category so that this adjunction induces a Quillen equivalence. In general, A–cell–$\mathcal{M}$ denotes the cellularization of the model category $\mathcal{M}$ with respect to a set of objects A in $\mathcal{M}$, which we call cells.

The generators for the homotopy category of $L_{\mathbb{S}_\mathbb{Q}}(\mathbb{T}\,\mathrm{Sp})$ are all suspensions and desuspensions of objects of the form $\mathbb{T}/H_+$ for H a subgroup of $\mathbb{T}$. For later purposes (see Section 4.2), we want a set of cells with simpler algebraic models. For every natural $n > 1$, let

$$\sigma_n = \mathbb{T}_+ \wedge_{C_n} e_{C_n} \mathbb{S},$$

where e_{C_n} is the idempotent in the Burnside ring for C_n (cyclic group of order n) corresponding to C_n. By [8, Lemma 2.1.5],

$$\mathbb{T}/C_{n+} = \bigvee_{C_m \subseteq C_n} \sigma_m;$$

hence we know that the set

$$K = \{\Sigma^k \mathbb{S} \mid k \in \mathbb{Z}\} \cup \{\Sigma^k \sigma_n \mid n > 1,\ k \in \mathbb{Z}\}$$

is a set of (cofibrant and homotopically small) generators for $L_{\mathbb{S}_\mathbb{Q}}(\mathbb{T}\,\mathrm{Sp})$.

Let K_{top} be the set of images of the objects from K under the functor $S^\bullet \wedge -$ (up to isomorphism). The elements of this set K_{top} will be called *basic cells*.

To apply the cellularization principle of [13] we need to know that these cells are homotopically small (this is also known as small or compact; see Definition 5.1.4).

First note that if X is homotopically small in $\mathbb{T}\,\mathrm{Sp}$ then it is so in $L_{\mathbb{S}_\mathbb{Q}}\mathbb{T}\,\mathrm{Sp}$ (since rationalization is a smashing localization).

Now consider the three elements of $S^\bullet$–mod

$$(*, *, DE\mathcal{F}_+ \wedge X, *, *), \quad (*, *, DE\mathcal{F}_+ \wedge X, \mathrm{Id}, X), \quad (DE\mathcal{F}_+ \wedge X, \mathrm{Id}, DE\mathcal{F}_+ \wedge X, *, *).$$

It is routine to check that these are cofibrant and homotopically small whenever X is cofibrant and homotopically small in $\mathbb{T}\,\mathrm{Sp}$. Finally, let X be cofibrant in $\mathbb{T}\,\mathrm{Sp}$. There is a homotopy pushout diagram, where the final term is $S^\bullet \wedge X$:

$$\begin{tikzcd}
(*, *, DE\mathcal{F}_+ \wedge X, *, *) \arrow[r] \arrow[d] & (*, *, DE\mathcal{F}_+ \wedge X, \mathrm{Id}, X) \arrow[d] \\
(DE\mathcal{F}_+ \wedge X, \mathrm{Id}, DE\mathcal{F}_+ \wedge X, *, *) \arrow[r] & (DE\mathcal{F}_+ \wedge X, \mathrm{Id}, DE\mathcal{F}_+ \wedge X, \mathrm{Id}, X)
\end{tikzcd}$$

Homotopically small objects are preserved by homotopy pushouts (consider the associated cofibre sequence). Hence $S^\bullet \wedge X$ is homotopically small in $S^\bullet$–mod whenever X is cofibrant and homotopically small. Since these two conditions hold for the generators of $\mathbb{T}\,\mathrm{Sp}$, we see that every element of K_{top} is homotopically small.

Note that the model category $L_{DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F}} DE\mathcal{F}_+$–mod is the same as the model category $L_{\Sigma^* f} DE\mathcal{F}_+$–mod, where $f\colon DE\mathcal{F}_+ \to DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F}$ and $\Sigma^* f$ is the set of all (integer) suspensions and desuspensions of f. This is a similar result to [6, Lemma 4.14], since $\widetilde{E}\mathcal{F}$–localization (in $\mathbb{T}\,\mathrm{Sp}$) is given by smashing with the map of $\mathbb{T}$–spaces $S^0 \to \widetilde{E}\mathcal{F}$.

Proposition 3.2.5 *There is a Quillen equivalence*

$$S^\bullet \wedge -\colon L_{S_\mathbb{Q}}(\mathbb{T}\,\mathrm{Sp}) \rightleftarrows K_{\mathrm{top}}\text{–cell–}S^\bullet\text{–mod :pb.}$$

Proof This follows from the cellularization principle, [13, Proposition 2.7]. It suffices to show that the derived unit is a weak equivalence on the set K of generators for the left-hand side, which are shifts of the objects σ_n for $n > 1$ and $\mathbb{S}$. Each such object is cofibrant and homotopically small, as are the elements of K_{top}.

The derived left adjoint on cofibrant objects (such as the elements of K) is simply the left adjoint. The right derived functor on objects of the form $S^\bullet \wedge k$ for $k \in K$ is weakly equivalent to taking a homotopy pullback of the diagram

$$\begin{tikzcd}
 & \mathbb{S}_\mathbb{Q} \wedge \widetilde{E}\mathcal{F} \wedge k \arrow[d, "\mathrm{Id}\wedge\mathrm{Id}\wedge\lambda\wedge\mathrm{Id}"] \\
\mathbb{S}_\mathbb{Q} \wedge DE\mathcal{F}_+ \wedge k \arrow[r, "\mathrm{Id}\wedge a\wedge\mathrm{Id}\wedge\mathrm{Id}"'] & \mathbb{S}_\mathbb{Q} \wedge \widetilde{E}\mathcal{F} \wedge DE\mathcal{F}_+ \wedge k
\end{tikzcd}$$

where the map $a\colon S^0 \to \widetilde{E}\mathcal{F}$ (of $\mathbb{T}$–spaces) is the map to the cofibre and λ is the unit map. Since homotopy pullbacks commute with smash products, the homotopy pullback of the above is weakly equivalent to the homotopy pullback of

$$\begin{tikzcd} & \widetilde{E}\mathcal{F} \arrow[d] \\ DE\mathcal{F}_+ \arrow[r] & DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F} \end{tikzcd}$$

(in the category $\mathbb{T}\mathrm{Sp}$) smashed with $\mathbb{S}_{\mathbb{Q}} \wedge k$. But the homotopy pullback of the diagram above is $\mathbb{S}$, as discussed after Definition 3.2.2. Hence the derived unit is a weak equivalence (in $L_{\mathbb{S}_{\mathbb{Q}}}(\mathbb{T}\mathrm{Sp})$) on the cells $k \in K$. □

We will show in Proposition 5.1.6 below that this Quillen equivalence is actually a symmetric monoidal Quillen equivalence.

Thus we have separated the homotopical information of $\mathbb{T}\mathrm{Sp}$ into a diagram of three model categories. The advantage of doing so is that we may now remove the equivariance from the model category whilst keeping the correct homotopy category.

3.3 Removing equivariance

Now we are going to remove equivariance using the inflation–fixed points adjunction $(\varepsilon, (-)^{\mathbb{T}})$.

Recall the functor $(-)^{\mathbb{T}}$ of [21, Section 3]. It takes a spectrum indexed on a complete $\mathbb{T}$–universe $\mathcal{U}$ to the $\mathbb{T}$–trivial universe $\mathcal{U}^{\mathbb{T}}$ and then applies the space-level fixed point functor levelwise. We begin by extending this functor to categories of modules over $\mathbb{T}$–equivariant ring spectra.

If A is a commutative ring object in $\mathbb{T}$–equivariant spectra then $A^{\mathbb{T}}$ is a commutative ring object in spectra. We want to compare A–modules in $\mathbb{T}$–equivariant spectra and $A^{\mathbb{T}}$–modules in spectra. Using [14, Section 4] there is a Quillen adjunction

$$A \wedge_{\varepsilon^* A^{\mathbb{T}}} \varepsilon^*(-)\colon A^{\mathbb{T}}\text{–mod} \rightleftarrows A\text{–mod} :(-)^{\mathbb{T}}$$

between right transferred model structures (fibrations and weak equivalences are defined in terms of the underlying categories). To simplify the notation, if $\zeta\colon \varepsilon^* A^{\mathbb{T}} \to A$ is the inclusion of fixed points, we write

$$\zeta_{\#} = A \wedge_{\varepsilon^* A^{\mathbb{T}}} \varepsilon^*(-)$$

for the left adjoint.

We consider several cases of this kind of adjunction and use them to build up an adjunction between $S^\bullet$–mod and a new diagram of model categories $S^\bullet_{\mathrm{top}}$–mod. We then show that this adjunction gives a Quillen equivalence, after cellularizing.

Proposition 3.3.1 *For $\zeta\colon \varepsilon^* DE\mathcal{F}_+^{\mathbb{T}} \to DE\mathcal{F}_+$ the inclusion of fixed points, the adjunction*

$$\zeta_\#\colon L_{\mathbb{S}_{\mathbb{Q}}}(DE\mathcal{F}_+^{\mathbb{T}}\text{–mod}) \rightleftarrows L_{\mathbb{S}_{\mathbb{Q}}}(DE\mathcal{F}_+\text{–mod}) :(-)^{\mathbb{T}}$$

is a symmetric monoidal Quillen equivalence.

Proof We have a Quillen equivalence by [14, Corollaries 8.1 and 9.2]. The left adjoint is strong symmetric monoidal, so the result follows. □

We now left Bousfield localize the model categories in this adjunction. We localize the right-hand side at the set of maps $\Sigma^* f$, where $f\colon DE\mathcal{F}_+ \to DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F}$. Let $(\Sigma^* f)^{\mathbb{T}}$ be the set of maps obtained by applying the derived right adjoint to the maps in $\Sigma^* f$. By [17, Theorem 3.3.20(1)(b)] we obtain the following result:

Proposition 3.3.2 *The adjunction*

$$\zeta_\#\colon L_{(\Sigma^* f)^{\mathbb{T}}} L_{\mathbb{S}_{\mathbb{Q}}}(DE\mathcal{F}_+^{\mathbb{T}}\text{–mod}) \rightleftarrows L_{\Sigma^* f} L_{\mathbb{S}_{\mathbb{Q}}}(DE\mathcal{F}_+\text{–mod}) :(-)^{\mathbb{T}}$$

is a symmetric monoidal Quillen equivalence.

Our final version is where we take A to be the sphere spectrum, so the left adjoint is just ε^*. By [21, Section V, Proposition 3.10] the adjunction

$$\varepsilon^*\colon \mathrm{Sp} \rightleftarrows \mathbb{T}\,\mathrm{Sp} :(-)^{\mathbb{T}}$$

is a symmetric monoidal Quillen adjunction. We localize it to obtain a Quillen equivalence:

Proposition 3.3.3 *The adjunction*

$$\varepsilon^*\colon L_{\mathbb{S}_{\mathbb{Q}}}(\mathrm{Sp}) \rightleftarrows L_{\mathbb{S}_{\mathbb{Q}} \wedge \widetilde{E}\mathcal{F}}(\mathbb{T}\,\mathrm{Sp}) :(-)^{\mathbb{T}}$$

is a symmetric monoidal Quillen equivalence.

Proof Since ε^* is strong monoidal and $\varepsilon^*(\mathbb{S}_{\mathbb{Q}}) = \mathbb{S}_{\mathbb{Q}}$, the above adjunction is a composite of two adjunctions, the second being identity adjunction between $L_{\mathbb{S}_{\mathbb{Q}}}(\mathbb{T}\,\mathrm{Sp})$ and further localization at $\widetilde{E}\mathcal{F}$, namely $L_{\mathbb{S}_{\mathbb{Q}} \wedge \widetilde{E}\mathcal{F}}(\mathbb{T}\,\mathrm{Sp})$.

To verify that this is a Quillen equivalence we will work with the derived unit and the derived counit on generators. The generator for the left-hand side is $\mathbb{S}$. The generators for the right-hand side are $\mathbb{S} = \mathbb{T}/\mathbb{T}_+$ and $(\mathbb{T}/C_n)_+$ for $n \geqslant 1$. But $(\mathbb{T}/C_n)_+$ is weakly equivalent to a point in $L_{\mathbb{S}_\mathbb{Q} \wedge \widetilde{E}\mathcal{F}}(\mathbb{T}\,\mathrm{Sp})$ (that is, $(\mathbb{T}/C_n)_+ \wedge \widetilde{E}\mathcal{F} \simeq *$). So we only need to consider $\mathbb{S}$ for the right-hand side.

The derived functor of $(-)^{\mathbb{T}}$ acts as the geometric $\mathbb{T}$–fixed point functor, because, by definition, for any $H \leq G$, $\phi^H(X) = (X \wedge \widetilde{E}[\not\supseteq H])^H$. With this in mind, it is routine to check that the derived unit and counit are weak equivalences on the generators. It follows that this adjunction is a Quillen equivalence. □

We combine the previous three propositions to compare $S^\bullet$–mod and a new model category of $S^\bullet_{\mathrm{top}}$–mod, where $S^\bullet_{\mathrm{top}}$ is defined by:

Definition 3.3.4 We define $S^\bullet_{\mathrm{top}}$ to be the $\mathscr{P}$–diagram of model categories and adjoint Quillen pairs

$$L_{\mathbb{S}_\mathbb{Q}}(DE\mathcal{F}_+^{\mathbb{T}}\text{–mod}) \underset{\mathrm{Id}}{\overset{\mathrm{Id}}{\rightleftarrows}} L_{\{(\Sigma^* f)^{\mathbb{T}}\}} L_{\mathbb{S}_\mathbb{Q}}(DE\mathcal{F}_+^{\mathbb{T}}\text{–mod}) \underset{U}{\overset{DE\mathcal{F}_+^{\mathbb{T}} \wedge -}{\leftrightarrows}} L_{\mathbb{S}_\mathbb{Q}}(\mathbb{T}\,\mathrm{Sp}),$$

where U denotes the forgetful functor.

By construction, the functor $(-)^{\mathbb{T}}$ induces a functor between $S^\bullet$–mod and $S^\bullet_{\mathrm{top}}$–mod. Since each of the components is a symmetric monoidal Quillen equivalence, we obtain the following from Lemma 3.1.4:

Theorem 3.3.5 *The adjunction*

$$\zeta_\#\colon S^\bullet_{\mathrm{top}}\text{–mod} \rightleftarrows S^\bullet\text{–mod} :(-)^{\mathbb{T}}$$

is a symmetric monoidal Quillen equivalence.

We now extend this Quillen equivalence to a cellularized version. Define $K^{\mathbb{T}}_{\mathrm{top}}$ to be the set of cells given by applying the derived functor of $(-)^{\mathbb{T}}$ to K_{top}. By the cellularization principle of [13, Proposition 2.7], we see that the Quillen equivalence above is preserved by cellularization.

Corollary 3.3.6 *The adjunction below is a Quillen equivalence:*

$$\zeta_\#\colon K^{\mathbb{T}}_{\mathrm{top}}\text{–cell–}S^\bullet_{\mathrm{top}}\text{–mod} \rightleftarrows K_{\mathrm{top}}\text{–cell–}S^\bullet\text{–mod} :(-)^{\mathbb{T}}.$$

As in the previous section, the above Quillen equivalence is symmetric monoidal, but for clarity we postpone the proof of that fact to Section 5.2.

The model category $K^{\mathbb{T}}_{\mathrm{top}}$–cell–$S^\bullet_{\mathrm{top}}$–mod is constructed from model categories of nonequivariant spectra. Hence we have removed the equivariance. The reward for doing so is in the next section, where we can replace our categories based on spectra with categories based on rational chain complexes. Such categories are our first approximation to the algebraic model.

3.4 Passing to algebra

We will replace the model category $K_{\mathrm{top}}^{\mathbb{T}}$–cell–$S_{\mathrm{top}}^{\bullet}$–mod by a Quillen equivalent Ch($\mathbb{Q}$)–model category. The results of [27] and the general theory of diagrams of model categories allow us to do so. To apply the work of [27], we must work with $H\mathbb{Q}$–modules in symmetric spectra. So we give two Quillen equivalences: the first moves us from orthogonal spectra to symmetric spectra, the second from symmetric spectra to $H\mathbb{Q}$–modules.

In more detail, recall $\mathbb{U}$, the forgetful functor from orthogonal spectra (in based topological spaces) to symmetric spectra (in based simplicial sets) and call $\mathbb{P}$ its left adjoint. Define $\mathbb{U}S_{\mathrm{top}}^{\bullet}$ to be the $\mathscr{P}$–diagram of model categories

$$L_{\mathbb{S}_{\mathbb{Q}}}(\mathbb{U}DE\mathcal{F}_+^{\mathbb{T}}\text{–mod}) \underset{\mathrm{Id}}{\overset{\mathrm{Id}}{\rightleftarrows}} L_{\{\mathbb{U}(\Sigma^* f)^{\mathbb{T}}\}} L_{\mathbb{S}_{\mathbb{Q}}}(\mathbb{U}DE\mathcal{F}_+^{\mathbb{T}}\text{–mod}) \underset{U}{\overset{\mathbb{U}DE\mathcal{F}_+^{\mathbb{T}}\wedge -}{\leftrightarrows}} L_{\mathbb{S}_{\mathbb{Q}}}\mathrm{Sp}^{\Sigma}.$$

The functor $\mathbb{U}$ preserves all weak equivalences, so we do not need to apply fibrant replacement when constructing the set $\mathbb{U}(\Sigma^* f)^{\mathbb{T}}$ and the commutative ring spectrum $\mathbb{U}DE\mathcal{F}_+^{\mathbb{T}}$.

Proposition 3.4.1 *The adjunction*

$$\mathbb{U}^{\bullet}\colon S_{\mathrm{top}}^{\bullet}\text{–mod} \leftrightarrows \mathbb{U}S_{\mathrm{top}}^{\bullet}\text{–mod} :\mathbb{P}^{\bullet}$$

is a strong symmetric monoidal Quillen equivalence.

Proof The adjunction $(\mathbb{P},\mathbb{U})$ is a Quillen equivalence between $L_{\mathbb{S}_{\mathbb{Q}}}\mathrm{Sp}$ and $L_{S_{\mathbb{Q}}}\mathrm{Sp}^{\Sigma}$. Furthermore the left adjoint is strong symmetric monoidal, so the result follows by Lemma 3.1.4. □

The second step is to pass from symmetric spectra to $H\mathbb{Q}$–modules using the adjunction $(H\mathbb{Q}\wedge -, U)$. This is a Quillen equivalence between $L_{\mathbb{S}_{\mathbb{Q}}}\mathrm{Sp}^{\Sigma}$ and $H\mathbb{Q}$–mod, and the left adjoint is strong symmetric monoidal. Thus by the same argument as above we get the following:

Proposition 3.4.2 *The adjunction*

$$H\mathbb{Q}\wedge -^{\bullet}\colon \mathbb{U}S_{\mathrm{top}}^{\bullet}\text{–mod} \rightleftarrows H\mathbb{Q}\wedge \mathbb{U}S_{\mathrm{top}}^{\bullet}\text{–mod} :U^{\bullet}$$

is a strong symmetric monoidal Quillen equivalence, where $H\mathbb{Q} \wedge \mathbb{U} S^{\bullet}_{\mathrm{top}}$ denotes the following diagram of model categories:

$$H\mathbb{Q} \wedge \mathbb{U}DE\mathcal{F}_+^{\mathbb{T}}\text{–mod} \underset{\mathrm{Id}}{\overset{\mathrm{Id}}{\rightleftarrows}} L_{\{H\mathbb{Q}\wedge\mathbb{U}(\Sigma^* f)^{\mathbb{T}}\}}(H\mathbb{Q} \wedge \mathbb{U}DE\mathcal{F}_+^{\mathbb{T}}\text{–mod}) \underset{U}{\overset{\mathbb{U}DE\mathcal{F}_+^{\mathbb{T}}\wedge-}{\leftrightarrows}} H\mathbb{Q}\text{–mod}.$$

Here $H\mathbb{Q} \wedge \mathbb{U}\, DE\mathcal{F}_+^{\mathbb{T}}$ denotes first the cofibrant replacement in the model category of commutative ring spectra and then application of $H\mathbb{Q} \wedge -$.

Now we are ready to use the results from [27] to move from topology to algebra on $\mathscr{P}$–diagrams. Let Θ be the derived functor described in [27, Section 2.2]. This functor Θ induces an equivalence between $H\mathbb{Q}$–modules and rational chain complexes.

Definition 3.4.3 By [27, Theorem 1.2] there is a commutative rational differential graded algebra $\widehat{S}_t$, which is naturally weakly equivalent to $\Theta(H\mathbb{Q} \wedge \mathbb{U}\, DE\mathcal{F}_+^{\mathbb{T}})$, such that the model category of $\widehat{S}_t$–modules (in $\mathrm{Ch}(\mathbb{Q})$) is Quillen equivalent to the model category of $H\mathbb{Q}\wedge\mathbb{U}\, DE\mathcal{F}_+^{\mathbb{T}}$–modules (in spectra).

Remark 3.4.4 It is essential for the formality argument in Section 4.1 that the ring spectrum $H\mathbb{Q} \wedge \mathbb{U}\, DE\mathcal{F}_+^{\mathbb{T}}$ is commutative. Without this, one is unable to replace the ring $\widehat{S}_t$ by the simpler ring $\mathcal{O}_{\mathcal{F}}$, nor can one understand the localising set A'' (defined in the next section) in terms of the inclusion $\mathcal{O}_{\mathcal{F}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$.

Let $S^{\bullet}_t$ be the $\mathscr{P}$–diagram of model categories below, where $\Theta(H\mathbb{Q} \wedge \mathbb{U}(\Sigma^* f)^{\mathbb{T}})$ denotes the image of the set of maps $H\mathbb{Q} \wedge \mathbb{U}(\Sigma^* f)^{\mathbb{T}}$ in the category of $\widehat{S}_t$–modules under the derived functor:

$$\widehat{S}_t\text{–mod} \underset{\mathrm{Id}}{\overset{\mathrm{Id}}{\rightleftarrows}} L_{\Theta(H\mathbb{Q}\wedge\mathbb{U}(\Sigma^* f)^{\mathbb{T}})}(\widehat{S}_t\text{–mod}) \underset{U}{\overset{\widehat{S}_t\otimes-}{\leftrightarrows}} \mathrm{Ch}(\mathbb{Q}).$$

Proposition 3.4.5 *There is a zigzag of symmetric monoidal Quillen equivalences*

$$H\mathbb{Q} \wedge \mathbb{U}\, S^{\bullet}_{\mathrm{top}}\text{–mod} \simeq S^{\bullet}_t\text{–mod}.$$

Proof There is a zigzag of symmetric monoidal adjunctions between $H\mathbb{Q}$–modules and $\mathrm{Ch}(\mathbb{Q})$. By [27, Corollary 2.15], this zigzag consists of Quillen equivalences. We can extend this zigzag from $H\mathbb{Q}$–modules to $H\mathbb{Q} \wedge \mathbb{U}\, DE\mathcal{F}_+^{\mathbb{T}}$–modules in a natural way.

We can extend further to diagrams of model categories. Thus we obtain a zigzag of adjunctions between $H\mathbb{Q} \wedge \mathbb{U}\, S^{\bullet}_{\mathrm{top}}$–mod and $S^{\bullet}_t$–mod. At each stage, we have localized the middle category of the diagram at the derived image (ie image under the derived functor) of the set of maps $\{H\mathbb{Q} \wedge \mathbb{U}(\Sigma^* f)^{\mathbb{T}}\}$. We apply Lemma 3.1.4 to see that we have a symmetric monoidal Quillen equivalence, as claimed. □

Corollary 3.4.6 *Denote the derived images (ie images under the derived functor) of the cells* $K^{\mathbb{T}}_{\mathrm{top}}$ *in* $S_t^\bullet$–mod *by* K_t. *Then there is a zigzag of Quillen equivalences*

$$K^{\mathbb{T}}_{\mathrm{top}}\text{–cell–}S^\bullet_{\mathrm{top}}\text{–mod} \simeq K_t\text{–cell–}S^\bullet_t\text{–mod}.$$

Since cellularization is compatible with Quillen equivalences, all Quillen equivalences presented above are still Quillen equivalences after cellularizing at the derived images of the cells from the set $K^{\mathbb{T}}_{\mathrm{top}}$. By the discussion in Sections 5.1 and 5.2, the above zigzag consists of symmetric monoidal Quillen equivalences.

4 Simplifying the algebraic category

We have shown so far that the category of rational $\mathbb{T}$–spectra has an algebraic model of the form K_t–cell–$S_t^\bullet$–mod. However, since this category is not well understood, in this section we perform several steps to obtain a more concrete and easier algebraic model.

4.1 Removing the localization

In this section we have two tasks: replace the commutative dga $\widehat{S}_t$ of Definition 3.4.3 by something simpler and remove the localization of the middle model category, $L_{\Theta(H\mathbb{Q}\wedge\mathbb{U}(\Sigma^* f)^{\mathbb{T}})}(\widehat{S}_t\text{–mod})$.

The main idea is to use a formality argument, similar to the one in [16, Section 10]. However, the important difference lies in adapting the formality argument to one for *modules* over a commutative dga. This is enough to simplify the middle model category in $S_t^\bullet$.

The construction of Θ comes with an isomorphism between $H_*(\Theta X)$ and $\pi_*(X)$ for any $H\mathbb{Q}$–module X. It follows that the homology of $\widehat{S}_t$ is determined by the rational homotopy groups of $DE\mathcal{F}_+^{\mathbb{T}}$. We prove that the homology of $\widehat{S}_t \simeq \Theta(H\mathbb{Q}\wedge\mathbb{U}\,DE\mathcal{F}_+^{\mathbb{T}})$ is so well-structured that $\widehat{S}_t$ is quasi-isomorphic to its homology. We then use this to understand the set of maps $A = \Theta(H\mathbb{Q}\wedge\mathbb{U}(\Sigma^* f)^{\mathbb{T}})$.

Recall that $\mathcal{O}_{\mathcal{F}}$ is the graded ring $\prod_{n\geqslant 1}\mathbb{Q}[c_n]$ with each c_n of degree -2, and $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$ is the colimit over n of $\mathcal{O}_{\mathcal{F}}[c_1^{-1},\dots,c_n^{-1}]$; see Section 2.1.

Lemma 4.1.1 *We have isomorphisms of graded rings*

$$\begin{aligned} H_*(\widehat{S}_t) &\cong H_*\big(\Theta(H\mathbb{Q}\wedge\mathbb{U}(DE\mathcal{F}_+^{\mathbb{T}}))\big) \cong \pi_*(H\mathbb{Q}\wedge\mathbb{U}(DE\mathcal{F}_+^{\mathbb{T}})) \\ &\cong \pi_*(DE\mathcal{F}_+^{\mathbb{T}})\otimes\mathbb{Q} \cong \pi_*^{\mathbb{T}}(DE\mathcal{F}_+)\otimes\mathbb{Q} \cong \mathcal{O}_{\mathcal{F}}, \end{aligned}$$

where the last isomorphism comes from [8].

Note that for the step $\pi_*(DE\mathcal{F}_+^{\mathbb{T}}) \otimes \mathbb{Q} \cong \pi_*^{\mathbb{T}}(DE\mathcal{F}_+) \otimes \mathbb{Q}$ we require $DE\mathcal{F}_+$ to be a (positive) fibrant spectrum.

We want to create a zigzag of quasi-isomorphisms between $\widehat{S}_t$ and $\mathcal{O}_{\mathcal{F}}$. For each $n \geqslant 1$ there is a cycle x_n inside $\widehat{S}_t$ which represents e_n (projection onto factor n) in homology. It follows that the homology of $\widehat{S}_t[(x_n)^{-1}]$ is equal to e_n applied to the homology of $\widehat{S}_t$. Note that for this argument to hold, we need to know that $\widehat{S}_t$ is a commutative dga, which requires that $DE\mathcal{F}_+$ be a commutative ring object in $\mathbb{T}$–spectra.

Define $\widetilde{S}_t = \prod_{n\geqslant 1} \widehat{S}_t[x_n^{-1}]$. There is a canonical map $\alpha\colon \widehat{S}_t \to \widetilde{S}_t$, which is a homology isomorphism. For each $n \geqslant 1$, pick a representative a_n in $\widehat{S}_t[x_n^{-1}]$ for the homology class of c_n. We thus have a map $\mathbb{Q}[c_n] \to \widehat{S}_t[x_n^{-1}]$ which sends c_n to a_n. Define $\beta\colon \mathcal{O}_{\mathcal{F}} \to \widetilde{S}_t$ as the product over n of these maps. We now have our zigzag of quasi-isomorphisms.

Let A' be the image of the set A under (derived) extension of scalars along α. Define a new $\mathscr{P}$–diagram of model categories, $\widetilde{S}_t^{\bullet}$, as

$$\widetilde{S}_t\text{–mod} \underset{\mathrm{Id}}{\overset{\mathrm{Id}}{\rightleftarrows}} L_{A'}(\widetilde{S}_t\text{–mod}) \underset{U}{\overset{\widetilde{S}_t\otimes -}{\leftrightarrows}} \mathrm{Ch}(\mathbb{Q}).$$

Extension and restriction of scalars along $\alpha\colon \widehat{S}_t \to \widetilde{S}_t$ induce a symmetric monoidal Quillen equivalence between $\widehat{S}_t^{\bullet}$–mod and $\widetilde{S}_t^{\bullet}$–mod.

We repeat this construction once more using β. Let A'' be the image of the set A' under restriction of scalars along β. Define a new diagram of model categories, $\widetilde{S}_a^{\bullet}$, as

$$\mathcal{O}_{\mathcal{F}}\text{–mod} \underset{\mathrm{Id}}{\overset{\mathrm{Id}}{\rightleftarrows}} L_{A''}(\mathcal{O}_{\mathcal{F}}\text{–mod}) \underset{U}{\overset{\mathcal{O}_{\mathcal{F}}\otimes -}{\leftrightarrows}} \mathrm{Ch}(\mathbb{Q}).$$

Extension and restriction of scalars along $\beta\colon \mathcal{O}_{\mathcal{F}} \to \widetilde{S}_t$ induce a symmetric monoidal Quillen equivalence between $\widetilde{S}_a^{\bullet}$–mod and $\widetilde{S}_t^{\bullet}$–mod.

We summarize these results in the following:

Proposition 4.1.2 *The adjoint pairs of extension and restriction of scalars along α and β induce symmetric monoidal Quillen equivalences*

$$S_t^{\bullet}\text{–mod} \simeq \widetilde{S}_t^{\bullet}\text{–mod} \simeq \widetilde{S}_a^{\bullet}\text{–mod}.$$

Let $K_{\tilde{t}}$ be the derived images of the cells K_t in $\widetilde{S}_t^{\bullet}$–mod and let $K_{\tilde{a}}$ be the derived images in $\widetilde{S}_a^{\bullet}$–mod. Then we have Quillen equivalences between K_t–cell–$S_t^{\bullet}$–mod, $K_{\tilde{t}}$–cell–$\widetilde{S}_t^{\bullet}$–mod and $K_{\tilde{a}}$–cell–$\widetilde{S}_a^{\bullet}$–mod.

Our next task is to understand the set of maps in A'' so that we can remove the localization in the middle model category in the diagram of model categories $\widetilde{S}^{\bullet}_a$. We show that there is a zigzag of homology isomorphisms between

$$\Theta(\mathbb{U}\,DE\mathcal{F}^{\mathbb{T}}_+) \to \Theta(\mathbb{U}\,(\widetilde{E}\mathcal{F} \wedge DE\mathcal{F}_+)^{\mathbb{T}}) \quad \text{and} \quad j\colon \mathcal{O}_{\mathcal{F}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}.$$

It will follow that we can replace the set A'' by the set of all shifts of j without changing the effect of the localization. That is, we will show that the model categories $L_{A''}(\mathcal{O}_{\mathcal{F}}\text{–mod})$ and $L_{\Sigma^* j}(\mathcal{O}_{\mathcal{F}}\text{–mod})$ are equal.

The zigzag of homology isomorphisms of $\mathcal{O}_{\mathcal{F}}$ modules that we will use is as follows. Factor $\Theta(\mathbb{U}\,DE\mathcal{F}^{\mathbb{T}}_+) \to \Theta(\mathbb{U}\,(\widetilde{E}\mathcal{F} \wedge DE\mathcal{F}_+)^{\mathbb{T}})$ into a cofibration followed by an acyclic fibration (with intermediate term R). Let C be the pushout of the top square below:

$$\begin{tikzcd}
\Theta(\mathbb{U}\,DE\mathcal{F}^{\mathbb{T}}_+) \arrow[r, tail] \arrow[d, "\wr"'] & R \arrow[r, "\simeq"] \arrow[d, "\wr"] & \Theta(\mathbb{U}\,(\widetilde{E}\mathcal{F} \wedge DE\mathcal{F}_+)^{\mathbb{T}}) \\
\widetilde{S}_t \arrow[r, "f"] & C & \\
\mathcal{O}_{\mathcal{F}} \arrow[r, "a"] \arrow[u, "\wr"] \arrow[d, "\mathrm{Id}"'] & C \arrow[u, "\mathrm{Id}"'] \arrow[d] & \\
\mathcal{O}_{\mathcal{F}} \arrow[r] & \mathcal{E}^{-1}C & \\
\mathcal{O}_{\mathcal{F}} \arrow[r, "j"] \arrow[u, "\mathrm{Id}"] & \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \arrow[u, "\mathcal{E}^{-1}a"'] &
\end{tikzcd}$$

Since $\mathcal{O}_{\mathcal{F}}$–mod is left proper it follows that $R \to C$ is a quasi-isomorphism. The functor defined by $M \mapsto \mathcal{E}^{-1}M$ on $\mathcal{O}_{\mathcal{F}}$–modules M is exact. It follows that $C \to \mathcal{E}^{-1}C$ is a homology isomorphism, since $\mathcal{E}^{-1}$ is already inverted on homology. The map f induces a homology isomorphism once $\mathcal{E}$ has been inverted, hence so does a. It follows that $\mathcal{E}^{-1}a$ is a homology isomorphism.

Thus we have shown that model categories $L_{A''}(\mathcal{O}_{\mathcal{F}}\text{–mod})$ and $L_{\Sigma^* j}(\mathcal{O}_{\mathcal{F}}\text{–mod})$ are equal. Now we are ready to remove the localization altogether.

Lemma 4.1.3 *The adjunction induced by the inclusion of rings* $j\colon \mathcal{O}_{\mathcal{F}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$ *induces a symmetric monoidal Quillen equivalence*

$$\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} -\colon L_{\Sigma^* j}(\mathcal{O}_{\mathcal{F}}\text{–mod}) \rightleftarrows \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\text{–mod} : j^*.$$

Proof The cofibrations are unchanged by localization. The weak equivalences of the model category $L_{\Sigma^* j}(\mathcal{O}_{\mathcal{F}}\text{–mod})$ are those maps f such that

$$H_*(\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} f) = \mathcal{E}^{-1}H_*(f)$$

is an isomorphism. The left adjoint preserves (and detects) these new weak equivalences, so we have a symmetric monoidal Quillen adjunction as claimed. The object $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$ is a homotopically small generator for (the homotopy category of) $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$–mod. If we can show that the derived counit of this adjunction is a weak equivalence then it will follow that we have a Quillen equivalence. This follows since the counit map is an isomorphism on the generator,

$$\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}. \qquad \square$$

We use the above result to remove the localization from the middle term in our diagram of model categories. We have a commuting diagram of model categories as below, where U denotes the forgetful functor:

$$\begin{tikzcd}
\mathcal{O}_{\mathcal{F}}\text{–mod} \arrow[r, shift left, "\mathrm{Id}"] \arrow[d, shift right, "\mathrm{Id}"'] & L_{\Sigma^* j}(\mathcal{O}_{\mathcal{F}}\text{–mod}) \arrow[l, shift left, "\mathrm{Id}"] \arrow[r, shift right, "U"'] \arrow[d, shift right, "\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes_{\mathcal{O}_{\mathcal{F}}}-"'] & \mathbb{Q}\text{–mod} \arrow[l, shift right, "\mathcal{O}_{\mathcal{F}}\otimes-"'] \arrow[d, shift right, "\mathrm{Id}"'] \\
\mathcal{O}_{\mathcal{F}}\text{–mod} \arrow[r, shift left, "\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes_{\mathcal{O}_{\mathcal{F}}}-"] \arrow[u, shift right, "\mathrm{Id}"'] & \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\text{–mod} \arrow[l, shift left, "j^*"] \arrow[r, shift right, "U"'] \arrow[u, shift right, "j^*"'] & \mathbb{Q}\text{–mod} \arrow[l, shift right, "\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes-"'] \arrow[u, shift right, "\mathrm{Id}"']
\end{tikzcd}$$

We denote the bottom row by $S_a^\bullet$, the left adjoint from top to bottom by $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} -$, the right adjoint by j^* and we summarize the above in the following:

Proposition 4.1.4 *The adjunction (described above)*

$$\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} -\colon \widetilde{S}_a^\bullet\text{–mod} \rightleftarrows S_a^\bullet\text{–mod} : j^*$$

is a symmetric monoidal Quillen equivalence, and thus the adjunction

$$\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} -\colon K_{\tilde{a}}\text{–cell–}\widetilde{S}_a^\bullet\text{–mod} \rightleftarrows K_a\text{–cell–}S_a^\bullet\text{–mod} : j^*$$

is a Quillen equivalence, where K_a is the derived image of $K_{\tilde{a}}$ under the left adjoint.

Again the adjunction at the level of cellularized categories is a symmetric monoidal Quillen equivalence, by discussion in Section 5.2.

4.2 Removing the cellularization

We now compare K_a–cell–$S_a^\bullet$–mod and the algebraic model $d\mathcal{A}_{\mathrm{dual}}$ of Section 2. The point is to move from a category whose weak equivalences are quite complicated to define to a model category whose weak equivalences are the quasi-isomorphisms. The idea behind this step is similar to one presented in [16, Sections 12 and 13].

We first introduce an adjoint pair relating $S_a^\bullet$–mod and $d\widehat{\mathcal{A}}$. An object

$$\beta\colon M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V$$

of $d\widehat{\mathcal{A}}$ gives an object of $S_a^\bullet$–mod defined by

$$(M,\ \mathcal{E}^{-1}\beta,\ \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V,\ \mathrm{Id},\ V).$$

This functor, which we call $\varkappa$, includes $d\widehat{\mathcal{A}}$ into $S_a^\bullet$–mod. It has a right adjoint Γ_v. Let $(a, \alpha, b, \gamma, c)$ be an object of $S_a^\bullet$–mod. Then we can draw the diagram of $\mathcal{O}_{\mathcal{F}}$–modules

$$a \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} a \to b \leftarrow \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes c.$$

If we take the pullback P of this in the category of $\mathcal{O}_{\mathcal{F}}$–modules in $\mathrm{Ch}(\mathbb{Q})$ we obtain a map $\delta\colon P \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes c$. This map δ is an object of $d\widehat{\mathcal{A}}$. For more details see [11, Section 7]. We call this adjoint pair $(\varkappa, \Gamma_v)$ and we note that it is a strong symmetric monoidal adjunction.

We can compose this adjunction with the adjunction (ι, Γ_h) which relates $d\widehat{\mathcal{A}}$ to $d\mathcal{A}$ (see Section 2.2). We let $\mu = \varkappa \circ \iota$ and $\Gamma = \Gamma_h \circ \Gamma_v$.

Lemma 4.2.1 *The adjunction (μ, Γ) between the categories $d\mathcal{A}$ and $S_a^\bullet$–mod is symmetric monoidal.*

This adjunction is also studied in [4, Section 7], where it is called (inc, Γ) and $S_a^\bullet$ is called $R_a^\bullet$.

Recall that, up to a weak equivalence (and ignoring shifts), the cells K_{top} consist of objects of the form

$$S^\bullet \wedge k = (k \wedge DE\mathcal{F}_+ \to k \wedge DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F} \leftarrow k \wedge \widetilde{E}\mathcal{F}),$$

where $k \in K$, ie $k = \mathbb{S}$ or $k = \sigma_n$ for $n > 1$ (see Section 3.2).

Thus we have to calculate the cells in K_a, ie the derived images of cells from K (or equivalently from K_{top}) in $S_a^\bullet$–mod. Since all required Quillen equivalences are symmetric monoidal (which follows from Section 5), they preserve the unit (up to weak equivalence) and the unit is always cellular. So the derived image of $\mathbb{S} \in L_{\mathbb{S}_{\mathbb{Q}}}\mathbb{T}\mathrm{Sp}$ is the unit in $S_a^\bullet$–mod,

$$\mathcal{O}_{\mathcal{F}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \leftarrow \mathbb{Q}.$$

We will use the simplified notation S^0 for this object. As for the other cells, consider some $S^\bullet \wedge \sigma_n \in K_{\mathrm{top}}$. Let $k_n = (A \to B \leftarrow C)$ be its derived image in $S_a^\bullet$–mod. To recap this process, one takes homotopy $\mathbb{T}$–fixed points of $S^\bullet \wedge \sigma_n$ to get an object of $K_{\mathrm{top}}^{\mathbb{T}}$ and then one applies the derived functor Θ from [27], to get an object of K_t.

Finally, one applies a number of algebraic adjunctions from Section 4.1 to get the object k_n of K_a. All of these adjunctions are constructed by taking Quillen equivalences (which preserve the unit up to weak equivalence) on each of the component categories. It follows that we have isomorphisms

$$H_*(A) = [\mathcal{O}_{\mathcal{F}}, A]_*^{\mathcal{O}_{\mathcal{F}}\text{–mod}} \cong [DE\mathcal{F}_+, DE\mathcal{F}_+ \wedge \sigma_n]_*^{DE\mathcal{F}_+\text{–mod}} \cong [\mathbb{S}, DE\mathcal{F}_+ \wedge \sigma_n]_*^{\mathbb{T}}.$$

Similar isomorphisms also hold for the other two components so, by the calculations of [8, Example 5.8.1], we have

$$\begin{aligned} H_*(A) &= \pi_*^{\mathbb{T}}(DE\mathcal{F}_+ \wedge \sigma_n) = \mathbb{Q}_n\langle 1\rangle, \\ H_*(B) &= \pi_*^{\mathbb{T}}(DE\mathcal{F}_+ \wedge \widetilde{E}\mathcal{F} \wedge \sigma_n) = 0, \\ H_*(C) &= \pi_*^{\mathbb{T}}(\widetilde{E}\mathcal{F} \wedge \sigma_n) = 0, \end{aligned}$$

where $\mathbb{Q}_n\langle 1\rangle$ is the torsion $\mathcal{O}_{\mathcal{F}}$–module consisting of a copy of $\mathbb{Q}$ in factor n and degree 1. It is immediate that there is a homology isomorphism

$$\widetilde{\sigma}_n = (\mathbb{Q}_n\langle 1\rangle \to 0 \leftarrow 0) \to (A \to B \leftarrow C) = k_n$$

given by simply picking a suitable representative cycle for $1 \in \mathbb{Q}_n\langle 1\rangle$. We therefore have the following description of the cells:

Lemma 4.2.2 *The set of cells K_a is given (up to weak equivalence) by all shifts of objects of the form $\widetilde{\sigma}_n$ for $n \geqslant 1$ and all shifts of $S^0 = (\mathcal{O}_{\mathcal{F}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \leftarrow \mathbb{Q})$.*

The above argument on the behaviour of the derived adjunction extends to the following useful result, which tells us that (after applying homology) our derived functors agree with the functor $\pi_*^{\mathcal{A}}$ of [8].

Theorem 4.2.3 *Let X be a rational $\mathbb{T}$–equivariant spectrum. Let ΥX be its derived image in $S_a^\bullet$–mod. Then $H_*(\Upsilon X) \cong \mu\pi_*^{\mathcal{A}}(X)$.*

The adjunction (μ, Γ) is shown to be a symmetric monoidal Quillen equivalence between $d\mathcal{A}$ with the dualizable model structure and a cellularization of $S_a^\bullet$–mod in [4, Theorem 7.6]. The cells for this cellularization are taken to be the "algebraic spheres". An algebraic sphere is an object of the form

$$S^\nu = (\mathcal{O}_{\mathcal{F}}(\nu) \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q} \leftarrow \mathbb{Q}),$$

where $\mathcal{O}_{\mathcal{F}}(\nu)$ is the subset of $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$ consisting of all those x such that $c^\nu x \in \mathcal{O}_{\mathcal{F}}$, for $\nu\colon \mathcal{F} \to \mathbb{Z}_{\geqslant 0}$ of finite support. We also allow negative spheres $S^{-\nu}$ and shifts of such objects. Essentially these are just "partial shifts" of the unit, where we have shifted finitely many factors of $\mathcal{O}_{\mathcal{F}}$ by some varying amount. We let $\{S^\nu\}$ denote the set of such objects.

To show that (μ, Γ) is a Quillen equivalence between $d\mathcal{A}$ with the dualizable model structure and the cellularization of $S_a^\bullet$–mod at the set of cells K_a, we want to use [4, Theorem 7.6], which says that $d\mathcal{A}$ with the dualizable model structure is Quillen equivalent to the cellularization of $S_a^\bullet$–mod at the set of cells $\{S^\nu\}$. Hence, it is enough to show that these two cellularizations agree (that is, produce the same model structure). We will prove that the algebraic spheres can be built via cofibre sequences and coproducts in $S_a^\bullet$–mod from cells in K_a and vice versa. It will follow that the class of K_a–cellular objects equals the class of $\{S^\nu\}$–cellular objects. Hence we will see that the K_a–cellular equivalences and the $\{S^\nu\}$–cellular equivalences agree and that the model categories K_a–cell–$S_a^\bullet$–mod and $\{S^\nu\}$–cell–$S_a^\bullet$–mod are equal.

The unit S^0 (and all its suspensions) is in both sets: in K_a and in the set of "algebraic spheres". So consider the algebraic sphere S^{ν_1} for the function $\nu_1\colon \mathcal{F} \to \mathbb{Z}_{\geqslant 0}$ sending a trivial subgroup to 1 and all other subgroups to 0. There is a cofibre sequence (in $S_a^\bullet$–mod)

$$S^0 \to S^{\nu_1} \to \Sigma\sigma_1,$$

where Σ denotes the suspension. This shows that we can build S^{ν_1} from σ_1 and S^0 and that we can build σ_1 from algebraic spheres. We can also create the negative sphere $S^{-\nu_1}$ using the cofibre sequence

$$S^{-\nu_1} \to S^0 \to \Sigma^{-1}\sigma_1.$$

To build any algebraic sphere we apply the above argument repeatedly. Note that by the definition of an algebraic sphere we need only finitely many steps. Equally we can make all σ_i for $i \geqslant 1$ from the algebraic spheres.

By [4, Theorem 7.6] we have the following:

Proposition 4.2.4 *The pair (μ, Γ) induces a symmetric monoidal Quillen equivalence between the model categories $d\mathcal{A}_{\mathrm{dual}}$ and K_a–cell–$S_a^\bullet$–mod.*

This finishes the proof that $d\mathcal{A}_{\mathrm{dual}}$ provides an algebraic model for the category of rational $\mathbb{T}$–spectra. We leave the consideration that all our Quillen equivalences are in fact symmetric monoidal to the last section.

5 Symmetric monoidal equivalences

All of the adjunctions in the zigzag between $d\mathcal{A}_{\mathrm{dual}}$ and $\mathbb{T}\,\mathrm{Sp}$ have been compatible with the monoidal properties of the categories. By examining the cellularized model structures more clearly we are able to show that each of these model categories is a proper, stable, cellular, monoidal model category that satisfies the monoid axiom. We are thus able to conclude that this zigzag of Quillen equivalences consists of *monoidal* Quillen equivalences. It follows that we also have Quillen equivalences of model categories of ring objects and modules over ring objects.

Our method is to prove a monoidal version of the cellularization principle [13, Proposition 2.7]; see Propositions 5.1.6 and 5.1.7.

5.1 Cellularization of stable model categories

A cellularization of a model category is a right Bousfield localization at a set of objects. Such a localization exists by [17, Theorem 5.1.1] whenever the model category is right proper and cellular. When we are in a stable context the results of [6] can be used.

Those results, which we shall introduce in the next subsection, allow us to understand the sets of generating cofibrations for our cellularized model categories and see that they are all symmetric monoidal and cellular.

In this subsection we recall the notion of cellularization (when $\mathcal{C}$ is stable) and some of basic definitions and results.

Definition 5.1.1 Let $\mathcal{C}$ be a stable model category and K a stable set of objects of $\mathcal{C}$, ie a class of K–cellular objects of $\mathcal{C}$ that is closed under desuspension.[2] We say that a map $f\colon A \to B$ of $\mathcal{C}$ is a *K–cellular equivalence* if the induced map

$$[k, f]^{\mathcal{C}}_*\colon [k, A]^{\mathcal{C}}_* \to [k, B]^{\mathcal{C}}_*$$

is an isomorphism of graded abelian groups for each $k \in K$. An object $Z \in \mathcal{C}$ is said to be *K–cellular* if

$$[Z, f]^{\mathcal{C}}_*\colon [Z, A]^{\mathcal{C}}_* \to [Z, B]^{\mathcal{C}}_*$$

is an isomorphism of graded abelian groups for any K–cellular equivalence f.

Definition 5.1.2 A *right Bousfield localization* or *cellularization* of $\mathcal{C}$ with respect to a set of objects K is a model structure K–cell–$\mathcal{C}$ on $\mathcal{C}$ such that

- the weak equivalences are K cellular equivalences;
- the fibrations of K–cell–$\mathcal{C}$ are the fibrations of $\mathcal{C}$;
- the cofibrations of K–cell–$\mathcal{C}$ are defined via left lifting property.

[2] Note that the class is always closed under suspension.

By [17, Theorem 5.1.1], if $\mathcal{C}$ is a right proper, cellular model category and K a set of objects in $\mathcal{C}$, then the cellularization K–cell–$\mathcal{C}$ of $\mathcal{C}$ with respect to K exists and is a right proper model category. The cofibrant objects of K–cell–$\mathcal{C}$ are called *K–cofibrant* and are precisely the K–cellular and cofibrant objects of $\mathcal{C}$.

We recall some definitions and results from [6] and prove our monoidal version of the cellularization principle. We use $\hat{c}_K$ for a cofibrant replacement functor in K–cell–$\mathcal{C}$.

Definition 5.1.3 Let K be a set of cofibrant objects in a monoidal model category $\mathcal{C}$. We say that K is *monoidal* if the following two conditions hold:

- Any object of the form $k \otimes k'$ for k, $k' \in K$ is K–cellular.
- For $\hat{c}_K S_{\mathcal{C}}$ a K–cofibrant replacement of the unit $S_{\mathcal{C}}$ of $\mathcal{C}$ and any $k \in K$, the map $\hat{c}_K S_{\mathcal{C}} \otimes k \to k$ is a K–cellular equivalence.

The cellularization of a right proper, cellular, stable model category at a stable set of cofibrant objects K is very well behaved (see [6, Theorem 5.9]), in particular it is proper, cellular and stable. Moreover, the second condition of the above definition holds automatically when the unit of $\mathcal{C}$ is K–cellular.

There is another important property we will often want the cells to satisfy, which makes right localization behave in an even more tractable manner; see [6, Section 9]. This property is variously called small, compact or finite. We choose to call it homotopically small to avoid those over-used terms.

Definition 5.1.4 We say that an object X of a stable model category $\mathcal{C}$ is *homotopically small* if, in the homotopy category, $\left[X, \coprod_i Y_i\right]^{\mathcal{C}}$ is canonically isomorphic to $\bigoplus_i [X, Y_i]^{\mathcal{C}}$; see [25, Definition 2.1.2].

Using [25, Lemma 2.2.1] it is routine to check that if K consists of homotopically small objects of $\mathcal{C}$ then K is a set of generators for K–cell–$\mathcal{C}$. Hence we know a set of generators for each of our cellularizations.

Notice that derived functors of both left and right Quillen equivalences preserve homotopically small objects. Now we may turn to monoidal considerations. The following theorem is [6, Theorem 7.2]:

Theorem 5.1.5 *Let $\mathcal{C}$ be a proper, monoidal, cellular, stable model category. Let K be a monoidal and stable set of cofibrant objects of $\mathcal{C}$. Then K–cell–$\mathcal{C}$ is a proper, monoidal, cellular, stable model category. Furthermore, if $\mathcal{C}$ satisfies the monoid axiom then so does K–cell–$\mathcal{C}$.*

The next two results are our upgraded version of the cellularization principle; see [13, Proposition 2.7]. They have slightly different assumptions according to whether the given cells are on the left or right of the adjunction. The first has the cells on the left and behaves as expected. The second starts with cells on the right of the adjunction and here we need to assume that the adjunction is a Quillen equivalence to start with. In both cases we have also assumed that a cofibrant replacement of the unit is in the set of cells (and hence is homotopically small). This simplifies the proofs but is not needed when the adjunction is already a Quillen equivalence.

For the following we let $\widehat{c}$ be the cofibrant replacement functor of $\mathcal{C}$, let $\widehat{c}_K$ be the cofibrant replacement functor of K–cell–$\mathcal{C}$ and let $\widehat{f}$ be the fibrant replacement functor of $\mathcal{D}$.

Proposition 5.1.6 *Consider a symmetric monoidal Quillen adjunction between a pair of proper, cellular, stable, monoidal model categories,*

$$L\colon \mathcal{C} \rightleftarrows \mathcal{D} :R.$$

Let K be a stable and monoidal set of cofibrant objects of $\mathcal{C}$ which contains a cofibrant replacement of the unit. Assume that each element of K and LK is homotopically small and that the unit map $k \to R\widehat{f}Lk$ is a weak equivalence of $\mathcal{C}$ for each $k \in K$. Then LK is a stable monoidal set of cofibrant objects of $\mathcal{D}$ and the unit of $\mathcal{D}$ is in LK (up to weak equivalence). Moreover, we have an induced symmetric monoidal Quillen equivalence

$$L\colon K\text{–cell–}\mathcal{C} \rightleftarrows LK\text{–cell–}\mathcal{D} :R.$$

Proof We apply the cellularization principle [13, Proposition 2.7] to see that (L, R) is a Quillen equivalence on the cellularized categories.

We must show that LK satisfies both parts of the definition of a monoidal set. For the first part, let k and k' be objects of K. Then $Lk \wedge Lk'$ is weakly equivalent to $L(k \wedge k')$, which is LK–cofibrant and hence is LK–cellular. For the second part, the map $L(\widehat{c}S_{\mathcal{C}}) \to S_{\mathcal{D}}$ is a weak equivalence since (L, R) is a monoidal Quillen pair. Hence $S_{\mathcal{D}}$ is in LK (up to weak equivalence) and the second condition holds automatically.

Now we know that LK–cell–$\mathcal{D}$ is a cellular monoidal model category. We must show that (L, R) is a symmetric monoidal Quillen adjunction on the cellularized model categories. We know that the map $L(\widehat{c}S_{\mathcal{C}}) \to S_{\mathcal{D}}$ is a weak equivalence. The comonoidal map $L(X \wedge Y) \to LX \wedge LY$ is also a weak equivalence for any cofibrant X and Y. Hence the proof is complete. □

Proposition 5.1.7 *Consider a symmetric monoidal Quillen equivalence between a pair of proper, cellular, stable, monoidal model categories*

$$L\colon \mathcal{C} \rightleftarrows \mathcal{D} : R.$$

Let H be a stable and monoidal set of cofibrant objects of $\mathcal{D}$ which contains a cofibrant replacement of the unit of $\mathcal{D}$. Assume that every element of H is homotopically small. Then $\hat{c}R\hat{f}H$ is a stable monoidal set of homotopically small cofibrant objects of $\mathcal{C}$ which contains the unit up to weak equivalence. Furthermore we have an induced symmetric monoidal Quillen equivalence

$$L\colon \hat{c}R\hat{f}H\text{–cell–}\mathcal{C} \rightleftarrows H\text{–cell–}\mathcal{D} : R.$$

Proof We apply the cellularization principle [13, Proposition 2.7] to see that (L, R) is a Quillen equivalence on the cellularized categories. We must prove that $K = \hat{c}R\hat{f}H$ is a monoidal set and that the unit of $\mathcal{C}$ is in K (up to weak equivalence).

It is simple to check that L takes K–cellular equivalences between cofibrant objects to H–cellular equivalences. Now consider the pair of maps, for k and k' elements of K,

$$L\hat{c}_K(k \wedge k') \xrightarrow{Lq} L(k \wedge k') \xrightarrow{\upsilon} Lk \wedge Lk'$$

The map υ is the comonoidal map of L and hence is a weak equivalence as (L, R) is monoidal. Since the codomain of υ is H–cellular, so is the domain of υ. The map Lq is L applied to a K–cellular equivalence between cofibrant objects, hence it is a H–cellular equivalence. We have shown that Lq is a H–cellular equivalence between H–cellular objects of $\mathcal{D}$ and thus must be a weak equivalence. Since (L, R) is a Quillen equivalence before cellularization, q must be a weak equivalence of $\mathcal{C}$. Thus $k \wedge k'$ must be K–cellular.

To complete the proof that K is monoidal it will suffice to prove that $S_{\mathcal{C}}$ is K–cellular. Thus we now show that the unit of $\mathcal{C}$ is in K up to weak equivalence. Since (L, R) is a symmetric monoidal Quillen pair, the composite map

$$L\hat{c}S_{\mathcal{C}} \to LS_{\mathcal{C}} \to S_{\mathcal{D}} \to \hat{f}S_{\mathcal{D}}$$

is a weak equivalence. Hence the adjoint $\hat{c}S_{\mathcal{C}} \to R\hat{f}S_{\mathcal{D}}$ is a weak equivalence. Thus we see that $\hat{c}S_{\mathcal{C}}$ is in K up to weak equivalence. We have now shown that the set K is monoidal and that K–cell–$\mathcal{C}$ is a symmetric monoidal model category.

The proof that this adjunction is symmetric monoidal on the cellularized model categories follows the same pattern as the previous case. □

5.2 Application to the classification

We start with the Quillen equivalence of Proposition 3.2.5,

$$S^\bullet \wedge -\colon L_{\mathbb{S}_\mathbb{Q}}(\mathbb{T}\,\mathrm{Sp}) \rightleftarrows K_{\mathrm{top}}\text{–cell–}S^\bullet\text{–mod} :\mathrm{pb}.$$

The set of cells K_{top} is given by $S^\bullet \wedge -$ applied to the set K of generators of $L_{\mathbb{S}_\mathbb{Q}}(\mathbb{T}\,\mathrm{Sp})$. We know that this set is stable and every element is homotopically small and cofibrant. By the proof of Proposition 5.1.6, it also follows that K_{top} is a monoidal set. Thus we may apply Proposition 5.1.6 to see that the adjunction $(S^\bullet \wedge -, \mathrm{pb})$ is symmetric monoidal.

We then have a large number of symmetric monoidal Quillen equivalences relating $S^\bullet$–mod and $S^\bullet_a$–mod. Our initial set of cells K_{top} is monoidal, stable, contains the unit and every element is homotopically small. Hence Propositions 5.1.6 and 5.1.7 tell us that K_{top}–cell–$S^\bullet$–mod and K_a–cell–$S^\bullet_a$–mod are Quillen equivalent via symmetric monoidal Quillen equivalences.

Theorem 5.2.1 *The model category of rational $\mathbb{T}$–spectra, $\mathbb{T}\,\mathrm{Sp}$, is Quillen equivalent to the algebraic model $d\mathcal{A}_{\mathrm{dual}}$. Furthermore, these Quillen equivalences are all symmetric monoidal. Hence the homotopy categories of $\mathbb{T}\,\mathrm{Sp}$ and $d\mathcal{A}_{\mathrm{dual}}$ are equivalent as symmetric monoidal categories.*

Proof This now follows by combining Proposition 3.2.5, Corollaries 3.3.6 and 3.4.6, Section 4.1 and Proposition 4.2.4 with Propositions 5.1.7 and 5.1.6. □

References

[1] **M Ando**, **J P C Greenlees**, *Circle-equivariant classifying spaces and the rational equivariant sigma genus*, Math. Z. 269 (2011) 1021–1104 MR

[2] **D Barnes**, *Rational equivariant spectra*, PhD thesis, University of Sheffield (2008) arXiv

[3] **D Barnes**, *Splitting monoidal stable model categories*, J. Pure Appl. Algebra 213 (2009) 846–856 MR

[4] **D Barnes**, *A monoidal algebraic model for rational* SO(2)*–spectra*, Math. Proc. Cambridge Philos. Soc. 161 (2016) 167–192 MR

[5] **D Barnes**, *Rational $O(2)$–equivariant spectra*, preprint (2016) arXiv To appear in Homology Homotopy Appl.

[6] **D Barnes**, **C Roitzheim**, *Stable left and right Bousfield localisations*, Glasg. Math. J. 56 (2014) 13–42 MR

[7] **J E Bergner**, *Homotopy fiber products of homotopy theories*, Israel J. Math. 185 (2011) 389–411 MR

[8] **J P C Greenlees**, *Rational S^1–equivariant stable homotopy theory*, Mem. Amer. Math. Soc. 661, Amer. Math. Soc., Providence, RI (1999) MR

[9] **J P C Greenlees**, *Rational S^1–equivariant elliptic cohomology*, Topology 44 (2005) 1213–1279 MR

[10] **J P C Greenlees**, *Rational torus-equivariant stable homotopy, I: Calculating groups of stable maps*, J. Pure Appl. Algebra 212 (2008) 72–98 MR

[11] **J P C Greenlees**, *Rational torus-equivariant stable homotopy, II: Algebra of the standard model*, J. Pure Appl. Algebra 216 (2012) 2141–2158 MR

[12] **J P C Greenlees**, **J P May**, *Generalized Tate cohomology*, Mem. Amer. Math. Soc. 543, Amer. Math. Soc., Providence, RI (1995) MR

[13] **J P C Greenlees**, **B Shipley**, *The cellularization principle for Quillen adjunctions*, Homology Homotopy Appl. 15 (2013) 173–184 MR

[14] **J P C Greenlees**, **B Shipley**, *Fixed point adjunctions for equivariant module spectra*, Algebr. Geom. Topol. 14 (2014) 1779–1799 MR

[15] **J P C Greenlees**, **B Shipley**, *Homotopy theory of modules over diagrams of rings*, Proc. Amer. Math. Soc. Ser. B 1 (2014) 89–104 MR

[16] **J P C Greenlees**, **B Shipley**, *An algebraic model for rational torus-equivariant spectra*, preprint (2016) arXiv (69 pages) Submitted for publication

[17] **P S Hirschhorn**, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, Amer. Math. Soc., Providence, RI (2003) MR

[18] **M Hovey**, *Model categories*, Mathematical Surveys and Monographs 63, Amer. Math. Soc., Providence, RI (1999) MR

[19] **M Kędziorek**, *Algebraic models for rational G–spectra*, PhD thesis, University of Sheffield (2014) Available at `http://etheses.whiterose.ac.uk/7699/`

[20] **M Kędziorek**, *An algebraic model for rational* SO(3)*–spectra*, preprint (2016) arXiv

[21] **M A Mandell**, **J P May**, *Equivariant orthogonal spectra and S–modules*, Mem. Amer. Math. Soc. 755, Amer. Math. Soc., Providence, RI (2002) MR

[22] **J E McClure**, *E_∞–ring structures for Tate spectra*, Proc. Amer. Math. Soc. 124 (1996) 1917–1922 MR

[23] **I Patchkoria**, *On the algebraic classification of module spectra*, Algebr. Geom. Topol. 12 (2012) 2329–2388 MR

[24] **S Schwede**, **B Shipley**, *Equivalences of monoidal model categories*, Algebr. Geom. Topol. 3 (2003) 287–334 MR

[25] **S Schwede**, **B Shipley**, *Stable model categories are categories of modules*, Topology 42 (2003) 103–153 MR

[26] **B Shipley**, *An algebraic model for rational S^1–equivariant stable homotopy theory*, Q. J. Math. 53 (2002) 87–110 MR

[27] **B Shipley**, *$H\mathbb{Z}$–algebra spectra are differential graded algebras*, Amer. J. Math. 129 (2007) 351–379 MR

Pure Mathematics Research Centre, Queen's University Belfast
University Road, Belfast, BT7 1NN, United Kingdom

School of Mathematics and Statistics, University of Sheffield
The Hicks Building, Sheffield, S3 7RH, United Kingdom

MATHGEOM, École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland

Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago
508 SEO m/c 249, 851 S. Morgan Street, Chicago, IL 60607-7045, United States

d.barnes@qub.ac.uk, j.greenlees@sheffield.ac.uk,
magdalena.kedziorek@epfl.ch, shipleyb@uic.edu

http://www.qub.ac.uk/puremaths/Staff/David%20Barnes/index.html,
http://www.greenlees.staff.shef.ac.uk/, http://people.epfl.ch/
magdalena.kedziorek, http://homepages.math.uic.edu/~bshipley/

Received: 13 November 2015 Revised: 14 July 2016

Algebraic & Geometric Topology 17 (2017) 1021–1040

On the infinite loop space structure of the cobordism category

HOANG KIM NGUYEN

We show an equivalence of infinite loop spaces between the classifying space of the cobordism category with infinite loop space structure induced by taking disjoint union of manifolds and the infinite loop space associated to the Madsen–Tillmann spectrum.

55P47, 57R56; 57R90

1 Introduction

In this article we show that there is an equivalence of infinite loop spaces between the classifying space of the d–dimensional cobordism category $B\,\mathrm{Cob}_\theta(d)$ and the 0^{th} space of the shifted Madsen–Tillmann spectrum $\mathrm{MT}\theta(d)[1]$. This extends a result by Galatius, Madsen, Tillmann and Weiss [5], who showed an equivalence of topological spaces

$$B\,\mathrm{Cob}_\theta(d) \simeq \mathrm{MT}\theta(d)[1]_0. \tag{1}$$

Note that both spaces in the equivalence above admit infinite loop space structures. The symmetric monoidal structure on the cobordism category, given by disjoint union of manifolds, induces an infinite loop space structure on $B\,\mathrm{Cob}_\theta(d)$, while the infinite loop space structure on $\mathrm{MT}\theta(d)[1]_0$ comes from it being the 0^{th} space of an Ω–spectrum. We will show that the equivalence (1) actually extends to an equivalence of infinite loop spaces with the above mentioned infinite loop space structures.

In more detail, our proof will rely on certain spaces of manifolds introduced by Galatius and Randal-Williams [4], which form an Ω–spectrum denoted here by ψ_θ. Using these spaces, they obtain a new proof of (1), which we record as the following theorem.

Theorem 1.1 *There are weak homotopy equivalences of spaces*

$$B\,\mathrm{Cob}_\theta(d) \simeq \psi_{\theta,0} \simeq \mathrm{MT}\theta(d)[1]_0.$$

In this article, we will show that the equivalences of the above theorem come from equivalences of spectra.

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.1021

Instead of directly constructing an equivalence of spectra, our strategy will be to construct Γ–spaces $\Gamma\,\mathrm{Cob}_\theta(d)$ and $\Gamma\psi_\theta$ with underlying spaces $B\,\mathrm{Cob}_\theta(d)$ and $\psi_{\theta,0}$ respectively, and we show that $\Gamma\psi_\theta$ is a model for the connective cover of the spectrum ψ_θ, denoted by $\psi_{\theta,\geq 0}$. This Γ–structure will be induced by taking disjoint union of manifolds. We then show that their associated spectra have the stable homotopy type of the connective cover of the shifted Madsen–Tillmann spectrum denoted by $\mathrm{MT}\theta(d)[1]_{\geq 0}$, by constructing a Γ–space model for $\mathrm{MT}\theta(d)[1]_{\geq 0}$ and exhibiting an equivalence of Γ–spaces. But more is true; we will see that the equivalences of Theorem 1.1 are the components of this equivalence of Γ–spaces and hence the main result of this article will be the following.

Main Theorem *There are stable equivalences of spectra*

$$\boldsymbol{B}\Gamma\,\mathrm{Cob}_\theta(d) \simeq \psi_{\theta,\geq 0} \simeq \mathrm{MT}\theta(d)[1]_{\geq 0}$$

such that the induced weak equivalences of spaces

$$\Omega^\infty \boldsymbol{B}\Gamma\,\mathrm{Cob}_\theta(d) \simeq \Omega^\infty\psi_\theta \simeq \Omega^\infty \mathrm{MT}\theta(d)[1]$$

are equivalent to the weak equivalences of Theorem 1.1.

Here, $\boldsymbol{B}\Gamma\,\mathrm{Cob}_\theta(d)$ is the spectrum associated to the symmetric monoidal category $\mathrm{Cob}_\theta(d)$. We would like to mention that a similar argument has been given by Madsen and Tillmann in [6] for the case $d=1$.

This article is organized as follows. In the next section we recall some basic notions on spectra and Γ–spaces. This will also serve to fix notation and language. In Section 3 and Section 4 we review the proof of Theorem 1.1 of [4]. In Section 5 we will construct Γ–space models for the spectra ψ_θ and $\mathrm{MT}\theta(d)$, and in Section 6 we will show that these Γ–spaces are equivalent. Finally in Section 7, we will relate these Γ–spaces to the cobordism category with its infinite loop space structure induced by taking disjoint union of manifolds.

Acknowledgements The author would like to thank Ulrich Bunke, John Lind and George Raptis for many helpful conversations and especially Ulrich Bunke for his encouragement to write this article. The author would also like to thank Ulrike Tillmann for her helpful comments, and the anonymous referee, whose suggestions greatly improved the exposition of the paper. Furthermore, the author gratefully acknowledges support from the Deutsche Forschungsgemeinschaft through the SFB 1085 *Higher Invariants*.

2 Conventions on spectra and Γ–spaces

By a *space* we mean a compactly generated weak Hausdorff space. We denote by $\boldsymbol{S}$ the category of spaces and by $\boldsymbol{S}_*$ the category of based spaces. We fix a model for the circle by setting $S^1 := \mathbb{R} \cup \{\infty\}$.

We will work with the Bousfield–Friedlander model of sequential spectra; see Bousfield and Friedlander [2] or Mandell, May, Schwede and Shipley [7]. Recall that a *spectrum* E is a sequence of based spaces $E_n \in \boldsymbol{S}_*$, $n \in \mathbb{N}$ together with structure maps

$$s_n\colon S^1 \wedge E_n \to E_{n+1}.$$

A map of spectra $f\colon E \to F$ is a sequence of maps $f_n\colon E_n \to F_n$ commuting with the structure maps. We denote by **Spt** the category of spectra. A *stable equivalence* is a map of spectra inducing isomorphisms on stable homotopy groups. An Ω*–spectrum* is a spectrum E, where the adjoints of the structure maps $\Sigma E_n \to E_{n+1}$ are weak homotopy equivalences. There is a model structure on **Spt** with weak equivalences the stable equivalences and fibrant objects the Ω–spectra. Moreover, a stable equivalence between Ω–spectra is a levelwise weak homotopy equivalence. We obtain a Quillen adjunction

$$\Sigma^\infty\colon \boldsymbol{S}_* \longleftrightarrow \mathbf{Spt} :\Omega^\infty,$$

where Σ^∞ takes a based space to its suspension spectrum and Ω^∞ assigns to a spectrum its 0^{th} space.

A spectrum E is called *connective* if its negative homotopy groups vanish. The case that E is an Ω–spectrum is equivalent to E_n being $(n-1)$–connected for all $n \in \mathbb{N}$. Note that a map $f\colon E \to F$ between connective Ω–spectra is a stable equivalence if and only if $f_0\colon E_0 \to F_0$ is a weak homotopy equivalence. We denote by $\mathbf{Spt}_{\geq 0}$ the full subcategory of connective spectra. It is a reflective subcategory of **Spt** and we denote the left adjoint of the inclusion by

$$(-)_{\geq 0}\colon \mathbf{Spt} \to \mathbf{Spt}_{\geq 0}.$$

We will need two operations on spectra. The first one is the shift functor

$$(-)[1]\colon \mathbf{Spt} \to \mathbf{Spt}$$

defined on a spectrum E by setting $E[1]_n = E_{n+1}$ and obvious structure maps. The second operation is the loop functor

$$\Omega\colon \mathbf{Spt} \to \mathbf{Spt}$$

defined by $(\Omega E)_n = \Omega(E_n)$ and looping the structure maps.

We recall Segal's infinite loop space machine [9], which provides many examples of connective spectra. We denote by Γ^{op} the skeleton of the category of finite pointed sets and pointed maps, ie its objects are the sets $m_+ := \{*, 1, \dots, m\}$. A Γ*–space* is a functor

$$\Gamma^{\mathrm{op}} \to \boldsymbol{S}_*$$

and we denote by $\Gamma\boldsymbol{S}_*$ the category of Γ–spaces and natural transformations.

There are distinguished maps $\rho_i\colon m_+ \to 1_+$ defined by $\rho_i(k) = *$ if $k \neq i$ and $\rho_i(i) = 1$. Let $A \in \Gamma\boldsymbol{S}_*$. The *Segal map* is the map

$$A(m_+) \xrightarrow{\prod_{i=1}^m \rho_i} \prod_m A(1_+).$$

A Γ–space is called *special* if the Segal map is a weak homotopy equivalence. If $A \in \Gamma\boldsymbol{S}_*$ is special, the set $\pi_0(A(1_+))$ is a monoid with multiplication induced by the span

$$A(1_+) \leftarrow A(2_+) \xrightarrow{\simeq} A(1_+) \times A(1_+),$$

where the left map is the map sending i to 1 for $i = 1, 2$, and the right map is the Segal map. A special Γ–space is called *very special* if this monoid is actually a group.

In [2], Bousfield and Friedlander construct a model structure on $\Gamma\boldsymbol{S}_*$ with fibrant objects the very special Γ–spaces and weak equivalences between fibrant objects levelwise weak equivalences.

There is a functor $\boldsymbol{B}\colon \Gamma\boldsymbol{S}_* \to \mathbf{Spt}$ defined as follows. Denote by $\mathbb{S}\colon \Gamma^{\mathrm{op}} \to \boldsymbol{S}_*$ the inclusion of finite pointed sets into pointed spaces. Given $A \in \Gamma\boldsymbol{S}_*$ we have an (enriched) left Kan extension along $\mathbb{S}$

$$\begin{tikzcd}
\Gamma^{\mathrm{op}} \arrow[r, "A"] \arrow[d, "\mathbb{S}"'] & \boldsymbol{S}_* \\
\boldsymbol{S}_* \arrow[ur, dashed] &
\end{tikzcd}$$

and we denote this left Kan extension by $L_{\mathbb{S}}A$. Now define $\boldsymbol{B}A_n := L_{\mathbb{S}}A(S^n)$. The structure maps are given by the image of the identity morphism $S^1 \wedge S^n \to S^1 \wedge S^n$ under the composite map

$$\begin{aligned}
\boldsymbol{S}_*(S^1 \wedge S^n, S^1 \wedge S^n) &\cong \boldsymbol{S}_*(S^1, \boldsymbol{S}_*(S^n, S^{n+1})) \\
&\to \boldsymbol{S}_*(S^1, \boldsymbol{S}_*(L_{\mathbb{S}}A(S^n), L_{\mathbb{S}}(A(S^{n+1})))) \\
&\cong \boldsymbol{S}_*(S^1 \wedge L_{\mathbb{S}}A(S^n), L_{\mathbb{S}}A(S^{n+1})).
\end{aligned}$$

By the Barratt–Priddy–Quillen theorem $L_{\mathbb{S}}\mathbb{S}$ is the sphere spectrum, hence the notation.

The functor $\boldsymbol{B}$ has a right adjoint $\boldsymbol{A}\colon \mathbf{Spt} \to \Gamma \boldsymbol{S}_*$ given by sending a spectrum $E \in \mathbf{Spt}$ to the Γ–space

$$n_+ \mapsto \mathbf{Spt}(\mathbb{S}^{\times n}, E)$$

using the topological enrichment of spectra. Moreover, the adjoint pair $\boldsymbol{B} \dashv \boldsymbol{A}$ is a Quillen pair which induces an equivalence of categories

$$\mathrm{Ho}(\Gamma \boldsymbol{S}_*) \simeq \mathrm{Ho}(\mathbf{Spt}_{\geq 0}).$$

In view of this equivalence we will say that a *Γ–space A is a model for a connective spectrum E* if there is a stable equivalence $\mathbb{L}\boldsymbol{B}A \simeq E$, where $\mathbb{L}\boldsymbol{B}$ is the left derived functor. The main theorem of Segal [9] states that $\boldsymbol{B}$ sends cofibrant-fibrant Γ–spaces to connective Ω–spectra.

Finally we make the following convention. We will refer to any zigzag of equivalences (of spaces, spectra or Γ–spaces) as simply an *equivalence*.

3 Recollection on spaces of manifolds

We recall the spaces $\Psi_\theta(\mathbb{R}^n)$ of embedded manifolds with tangential structure from Galatius and Randall-Williams [4]. Denote by $\mathrm{Gr}_d(\mathbb{R}^n)$ the Grassmannian manifold of d–dimensional planes in $\mathbb{R}^n$ and denote $BO(d) := \mathrm{colim}_{n \in \mathbb{N}} \mathrm{Gr}_d(\mathbb{R}^n)$ induced by the standard inclusion $\mathbb{R}^n \to \mathbb{R}^{n+1}$. Let $\theta\colon X \to BO(d)$ be a Serre fibration and let $M \subset \mathbb{R}^n$ be a d–dimensional embedded smooth manifold. Then a *tangential θ–structure* on M is a lift

$$\begin{tikzcd} & X \arrow[d, "\theta"] \\ M \arrow[ur, dashed] \arrow[r, "\tau_M"'] & BO(d) \end{tikzcd}$$

where τ_M is the classifying map of the tangent bundle (determined by the embedding). The topological space $\Psi_\theta(\mathbb{R}^n)$ has as underlying set pairs (M, l), where M is a d–dimensional smooth manifold without boundary which is closed as a subset of $\mathbb{R}^n$, and $l\colon M \to X$ is a θ–structure. We refer to [4] and for a description of the topology. We will also in general suppress the tangential structure from the notation.

For $0 \leq k \leq n$, we have the subspaces $\psi_\theta(n, k) \subset \Psi_\theta(\mathbb{R}^n)$ of those manifolds $M \subset \mathbb{R}^n$, satisfying

$$M \subset \mathbb{R}^k \times (-1, 1)^{n-k}.$$

In other words, $\psi_\theta(n,k)$ consists of manifolds with k possibly noncompact and $(n-k)$ compact directions. We denote

$$\Psi_\theta(\mathbb{R}^\infty) := \operatorname{colim}_{n\in\mathbb{N}} \Psi_\theta(\mathbb{R}^n)$$
$$\psi_\theta(\infty,k) := \operatorname{colim}_{n\in\mathbb{N}} \psi_\theta(n,k),$$

where the colimit is again induced by the standard inclusions. In [1] it is shown that the topological spaces $\Psi_\theta(\mathbb{R}^n)$ are metrizable and hence in particular compactly generated weak Hausdorff spaces.

For all $n \in \mathbb{N}$ and $1 \le k \le n-1$ we have a map

$$\begin{aligned} \mathbb{R}\times\psi_\theta(n,k) &\to \psi_\theta(n,k+1), \\ (t,M) &\mapsto M - t\cdot e_{k+1}, \end{aligned}$$

where e_{k+1} denotes the $(k+1)^{\text{st}}$ standard basis vector. This descends to a map $S^1 \wedge \psi_\theta(n,k) \to \psi_\theta(n,k+1)$ when taking as basepoint the empty manifold.

Theorem 3.1 *The adjoint map*

$$\psi_\theta(n,k) \to \Omega\psi_\theta(n,k+1)$$

is a weak homotopy equivalence.

Proof See [4, Theorem 3.20]. □

Definition 3.2 Let ψ_θ be the spectrum with n^{th} space given by

$$(\psi_\theta)_n := \psi_\theta(\infty,n+1)$$

and structure maps given by the adjoints of the translations.

By the above theorem, the spectrum ψ_θ is an Ω–spectrum.

4 The weak homotopy type of $\psi_\theta(\infty,1)$

This section contains a brief review of the main theorem of [5] as proven in [4]. Recall first the construction of the *Madsen–Tillmann spectrum* $\mathrm{MT}\theta(d)$ associated to a Serre fibration $\theta\colon X\to BO(d)$. Denote by $X(\mathbb{R}^n)$ the pullback

$$\begin{tikzcd} X(\mathbb{R}^n) \arrow[r] \arrow[d, "\theta_n"'] & X \arrow[d, "\theta"] \\ \mathrm{Gr}_d(\mathbb{R}^n) \arrow[r] & BO(d) \end{tikzcd}$$

and by $\gamma_{d,n}^{\perp}$ the orthogonal complement of the tautological bundle over $\mathrm{Gr}_d(\mathbb{R}^n)$. Then define the spectrum $T\theta(d)$ to have as n^{th} space the Thom space of the pullback bundle $T\theta(d)_n := \mathrm{Th}(\theta_n^*\gamma_{d,n}^{\perp})$. The structure maps are given by

$$S^1 \wedge \mathrm{Th}(\theta_n^*\gamma_{d,n}^{\perp}) \cong \mathrm{Th}(\theta_n^*\gamma_{d,n}^{\perp} \oplus \varepsilon) \to \mathrm{Th}(\theta_{n+1}^*\gamma_{d,n+1}^{\perp}),$$

where ε denotes the trivial bundle. Then define the Madsen–Tillmann spectrum $\mathrm{MT}\theta(d)$ to be a fibrant replacement of the spectrum $T\theta(d)$. Since the adjoints of the structure maps of $T\theta(d)$ are inclusions, we can give an explicit construction of $\mathrm{MT}\theta(d)$ as

$$\mathrm{MT}\theta(d)_n := \mathrm{colim}_k\, \Omega^k T\theta(d)_{n+k}.$$

Hence we have $\Omega^\infty \mathrm{MT}\theta(d) = \mathrm{colim}_k\, \Omega^k T\theta(d)_k$.

The passage from $\mathrm{MT}\theta(d)$ to our spaces of manifolds is as follows. We have a map

$$\mathrm{Th}(\theta_n^*\gamma_{d,n}^{\perp}) \to \Psi_\theta(\mathbb{R}^n)$$

given by sending an element (V,u,x), where $V \in \mathrm{Gr}_d(\mathbb{R}^n)$, $u \in V^\perp$ and $x \in X$, to the translated plane $V - u \in \Psi_\theta(\mathbb{R}^n)$ with constant θ–structure at x and sending the basepoint to the empty manifold.

Theorem 4.1 [4, Theorem 3.22] *The map* $\mathrm{Th}(\theta_n^*\gamma_{d,n}^{\perp}) \to \Psi_\theta(\mathbb{R}^n)$ *is a weak homotopy equivalence.*

On the other hand, by Theorem 3.1 we also have a weak homotopy equivalence

$$\psi_\theta(n,1) \to \Omega^{n-1}\Psi_\theta(\mathbb{R}^n).$$

Combining the two equivalences, we obtain

$$\Omega^{n-1}\,\mathrm{Th}(\theta_n^*\gamma_{d,n}^{\perp}) \xrightarrow{\simeq} \Omega^{n-1}\Psi_\theta(\mathbb{R}^n) \xleftarrow{\simeq} \psi_\theta(n,1).$$

Now we have a map

$$S^1 \wedge \Psi_\theta(\mathbb{R}^n) \to \Psi_\theta(\mathbb{R}^{n+1}),$$
$$(t, M) \mapsto M \times \{t\},$$

and we obtain the following commutative diagram:

$$\begin{tikzcd}
\Omega^{n-1}\,\mathrm{Th}(\theta_n^*\gamma_{d,n}^{\perp}) \arrow[r, "\simeq"] \arrow[d] & \Omega^{n-1}\Psi_\theta(\mathbb{R}^n) \arrow[d] & \psi_\theta(n,1) \arrow[l, "\simeq"'] \arrow[d] \\
\Omega^{n}\,\mathrm{Th}(\theta_{n+1}^*\gamma_{d,n+1}^{\perp}) \arrow[r, "\simeq"] & \Omega^{n}\Psi_\theta(\mathbb{R}^{n+1}) & \psi_\theta(n+1,1) \arrow[l, "\simeq"']
\end{tikzcd}$$

Finally, letting $n \to \infty$ we can determine the weak homotopy type of $\psi_\theta(\infty,1)$.

Theorem 4.2 *There are weak equivalences of spaces*

$$\Omega^\infty \mathrm{MT}\theta(d)[1] \xrightarrow{\simeq} \operatorname{colim}_{n\in\mathbb{N}} \Omega^{n-1}\Psi_\theta(\mathbb{R}^n) \xleftarrow{\simeq} \psi_\theta(\infty,1).$$

5 Γ–space models for $\mathrm{MT}\theta(d)$ and ψ_θ

In this section we construct Γ–space models for the spectra $\mathrm{MT}\theta(d)$ and ψ_θ. The comparison of these Γ–spaces to the respective spectra relies heavily on results of May and Thomason [8].

We will encounter the following situation.

Definition 5.1 A functor $E\colon \Gamma^{\mathrm{op}} \to \mathbf{Spt}$ is called a Γ*–spectrum*. It is called a *special* Γ–spectrum if the Segal map

$$E(m_+) \to \prod_m E(1_+)$$

is a stable equivalence. Furthermore, we denote by $\Gamma^{(k)}E$ the Γ–space given by evaluating at the k^{th} space, that is,

$$\Gamma^{(k)}E(m_+) := E(m_+)_k.$$

The key proposition for showing that we have constructed the right Γ–spaces will be the following.

Proposition 5.2 *Let $E\colon \Gamma^{\mathrm{op}} \to \mathbf{Spt}$ be projectively fibrant and special. Then the Γ–space $\Gamma^{(k)}E$ is a model for the connective cover of $E(1_+)[k]$.*

Before we can prove the proposition, we will need some lemmas. The first one concerns the behavior of Segal's functor $\boldsymbol{B}$ with respect to the loop functor.

Lemma 5.3 *For $A \in \Gamma\boldsymbol{S}_*$ there is a natural map of spectra*

$$\boldsymbol{B}\Omega A \to \Omega \boldsymbol{B} A$$

which is the identity on 0^{th} spaces.

Proof Since $\mathbb{S}\colon \Gamma^{\mathrm{op}} \to \boldsymbol{S}_*$ is fully faithful, we have a strictly commutative diagram of functors:

$$\begin{tikzcd} \Gamma^{\mathrm{op}} \arrow[r, "\Omega A"] \arrow[d, "\mathbb{S}"'] & \boldsymbol{S}_* \\ \boldsymbol{S}_* \arrow[ur, "L_{\mathbb{S}}\Omega A"'] & \end{tikzcd}$$

The composition of the loop functor with the left Kan extension $\Omega L_{\mathbb{S}} A$ also gives a strictly commutative diagram:

$$\begin{tikzcd} \Gamma^{\mathrm{op}} \arrow[r, "\Omega A"] \arrow[d, "\mathbb{S}"'] & \boldsymbol{S}_* \\ \boldsymbol{S}_* \arrow[ur, "\Omega L_{\mathbb{S}} A"'] & \end{tikzcd}$$

Hence by the universal property of the left Kan extension we get a natural transformation $\gamma\colon L_{\mathbb{S}}\Omega A \Rightarrow \Omega L_{\mathbb{S}} A$. Now the components at the spheres assemble into a map of spectra $\boldsymbol{B}\Omega A \to \Omega \boldsymbol{B} A$, since by naturality we have a commutative diagram:

$$\begin{tikzcd} S^1 \wedge L_{\mathbb{S}}\Omega A(S^n) \arrow[r] \arrow[d, "\mathrm{id}\wedge\gamma"'] & L_{\mathbb{S}}\Omega A(S^{n+1}) \arrow[d, "\gamma"] \\ S^1 \wedge \Omega L_{\mathbb{S}} A(S^n) \arrow[r] & \Omega L_{\mathbb{S}} A(S^{n+1}) \end{tikzcd}$$

Finally, since $S^0 = 1_+ \in \Gamma^{\mathrm{op}}$ the map of spectra is the identity on 0^{th} spaces. □

In general for any $A \in \Gamma\boldsymbol{S}_*$ the spectrum $\boldsymbol{B}A$ might not have the right stable homotopy type, as the functor $\boldsymbol{B}$ only preserves weak equivalences between cofibrant objects. However for very special Γ–spaces, there is a more convenient replacement, which gives the right homotopy type. As a second lemma we record the following fact from [8], which generalizes a construction of [9].

Lemma 5.4 *There is a functor $W\colon \Gamma\boldsymbol{S}_* \to \Gamma\boldsymbol{S}_*$ such that the following hold for all very special $X \in \Gamma\boldsymbol{S}_*$:*

- *The spectrum $\boldsymbol{B}WX$ is a connective Ω–spectrum.*
- *The Γ–space WX is very special and there is a weak equivalence $WX \to X$.*
- *If X, Y are very special and there is a weak equivalence $X \simeq Y$, then $\boldsymbol{B}WX \simeq \boldsymbol{B}WY$.*
- *There is a weak equivalence $W\Omega X \to \Omega WX$.*

Proof See [8, Appendix B]. □

The important thing for us will be that if $X \in \Gamma\boldsymbol{S}_*$ is very special, then $\boldsymbol{B}WX$ has the right stable homotopy type.

Lemma 5.5 (up and across lemma [8; 3]) *Let E^i, $i \in \mathbb{N}$ be a sequence of connective Ω–spectra together with stable equivalences $f^i\colon E^i \to \Omega E^{i+1}$. Let E_0 be the spectrum with $(E_0)_n := E_0^n$ and structure maps given by $f_0^n\colon E_0^n \to \Omega E_0^{n+1}$. Then there is a natural stable equivalence $E^0 \simeq E_0$.*

Note that in particular E_0 is connective. We are now ready to prove our key proposition.

Proof of Proposition 5.2 We prove the proposition for $k = 0$. The argument for higher k is completely analogous.

We first show that the Γ–space $\Gamma^{(0)}E$ is very special. Note that the Γ–spaces $\Gamma^{(k)}E$ are special, since E is projectively fibrant and thus the Segal map is a levelwise equivalence. It remains to show that $\pi_0(\Gamma^{(0)}E(1_+))$ is a group. To this end we compose with the functor $\boldsymbol{A}\colon \mathbf{Spt} \to \Gamma \boldsymbol{S}_*$ to obtain a functor

$$\Gamma^{\mathrm{op}} \xrightarrow{E} \mathbf{Spt} \xrightarrow{\boldsymbol{A}} \Gamma \boldsymbol{S}_*$$

which is equivalently a functor

$$\widehat{A} := \Gamma^{\mathrm{op}} \times \Gamma^{\mathrm{op}} \to \boldsymbol{S}_*.$$

Fixing the first variable gives a Γ–space

$$\widehat{A}(k_+)(-)\colon \Gamma^{\mathrm{op}} \to \boldsymbol{S}_*$$

which is obtained by first evaluating the Γ–spectrum E at k_+ and then applying the functor $\boldsymbol{A}$ to the spectrum $E(k_+)$. In particular, we have

$$\widehat{A}(1_+)(-) = \boldsymbol{A}(E(1_+))\colon \Gamma^{\mathrm{op}} \to \boldsymbol{S}_*,$$

which is very special by construction.

Fixing the second variable gives a Γ–space

$$\widehat{A}(-)(k_+)\colon \Gamma^{\mathrm{op}} \to \boldsymbol{S}_*$$

which is obtained as the composition

$$\Gamma^{\mathrm{op}} \xrightarrow{E} \mathbf{Spt} \xrightarrow{\boldsymbol{A}} \Gamma \boldsymbol{S}_* \xrightarrow{ev_{k_+}} \boldsymbol{S}_*,$$

where the last functor is given by evaluating a Γ–space at the object k_+. In particular, we have

$$\widehat{A}(-)(1_+) = \boldsymbol{A}(E(-))(1_+) = \Gamma^{(0)}E\colon \Gamma^{\mathrm{op}} \to \boldsymbol{S}_*,$$

which is special since $\Gamma^{(0)}E$ is special.

Now we have the following diagram, where the middle square commutes by functoriality:

$$\begin{tikzcd}
 & & \widehat{A}(1_+)(1_+)\times\widehat{A}(1_+)(1_+) \\
 & \widehat{A}(2_+)(2_+) \arrow[r] \arrow[d] & \widehat{A}(2_+)(1_+) \arrow[u, "\simeq"'] \arrow[d] \\
\widehat{A}(1_+)(1_+)\times\widehat{A}(1_+)(1_+) & \widehat{A}(1_+)(2_+) \arrow[l, "\simeq"] \arrow[r] & \widehat{A}(1_+)(1_+)
\end{tikzcd}$$

By the above identification of the Γ–spaces $\widehat{A}(-)(1_+)$ and $\widehat{A}(1_+)(-)$ we see that the right vertical span represents the monoid structure of $\Gamma^{(0)}E$ and the lower horizontal span represents the monoid structure of $\boldsymbol{A}E(1_+)$. In other words, the maps into the products in the lower left and upper right corner are given by the Segal maps while the maps into the lower right corner are the respective multiplications induced by the nontrivial map $2_+\to 1_+$, as are the remaining maps.

Hence we obtain two monoid structures on $\pi_0(\widehat{A}(1_+)(1_+))$ induced by $\boldsymbol{A}E(1_+)$ and $\Gamma^{(0)}E$. The commutativity of the middle square is now precisely the statement that they are compatible, or in other words that one is a homomorphism for the other, thus by the Eckmann–Hilton argument they agree. We now observe that the monoid $\boldsymbol{A}E(1_+)$ is actually a group, since $\pi_0(\boldsymbol{A}E(1_+)(1_+))$ is the 0^{th} stable homotopy group of $E(1_+)$. It follows that $\Gamma^{(0)}E$ is very special.

As a next step, we compose with taking connective covers to obtain a special Γ–spectrum in connective Ω–spectra

$$E_{\geq 0}\colon \Gamma^{\mathrm{op}}\to \mathbf{Spt}_{\geq 0}.$$

Note that $\Gamma^{(0)}E\simeq\Gamma^{(0)}E_{\geq 0}$ and hence $\Gamma^{(0)}E_{\geq 0}$ is very special. For $k\geq 1$, the Γ–spaces $\Gamma^{(k)}E_{\geq 0}$ will automatically be very special since $E_{\geq 0}(1_+)$ is connective and hence $\pi_0(\Gamma^{(k)}E(1_+))\cong\pi_0(E_{\geq 0}(1_+)_k)=0$.

We now consider the spectra associated to the very special Γ–spaces $\Gamma^{(k)}E_{\geq 0}$, ie we apply May and Thomason's replacement followed by Segal's functor to obtain a sequence of connective Ω–spectra

$$\boldsymbol{B}W\Gamma^{(k)}E_{\geq 0}\quad\text{for } k\in\mathbb{N}.$$

Now by Lemma 5.4 we have the following equivalence:

$$\boldsymbol{B}W\Gamma^{(k)}E_{\geq 0}\xrightarrow{\simeq}\boldsymbol{B}W\Omega\Gamma^{(k+1)}E_{\geq 0}\xrightarrow{\simeq}\boldsymbol{B}\Omega W\Gamma^{(k+1)}E_{\geq 0}.$$

By Lemma 5.3 we have a map $\boldsymbol{B}\Omega W\Gamma^{(k)}E\to\Omega\boldsymbol{B}W\Gamma^{(k)}E$ which is the identity on 0^{th} spaces. In particular, since both spectra are Ω–spectra, this map is an equivalence on connective covers. We now observe that since $E(1_+)_{\geq 0}$ is a connective Ω–spectrum, we have

$$\pi_0(\boldsymbol{B}\, W\Gamma^{(k)}E) = \pi_0(E(1^+)_{\geq 0,k}) = 0$$

for $k \geq 1$ and hence $\Omega \boldsymbol{B}\, W\Gamma^{(k)}E$ is connective. Thus we obtain a stable equivalence

$$\boldsymbol{B}\,\Omega W\Gamma^{(k)}E \to \Omega \boldsymbol{B}\, W\Gamma^{(k)}E$$

for $k \geq 1$. Putting all these maps together we obtain a sequence of connective Ω–spectra $\boldsymbol{B}\, W\Gamma^{(k)}E_{\geq 0}$ together with stable equivalences

$$\boldsymbol{B}\, W\Gamma^{(k)}E_{\geq 0} \xrightarrow{\simeq} \boldsymbol{B}\, W\Omega\Gamma^{(k+1)}E_{\geq 0} \xrightarrow{\simeq} \boldsymbol{B}\,\Omega W\Gamma^{(k+1)}E_{\geq 0} \xrightarrow{\simeq} \Omega\boldsymbol{B}\, W\Gamma^{(k+1)}E_{\geq 0},$$

that is, we have $\boldsymbol{B}\, W\Gamma^{(k)}E_{\geq 0} \xrightarrow{\simeq} \Omega\boldsymbol{B}\, W\Gamma^{(k+1)}E_{\geq 0}$. Thus we are in the situation of Lemma 5.5 and conclude that

$$\boldsymbol{B}\, W\Gamma^{(0)}E_{\geq 0} = \boldsymbol{B}\, W\Gamma^{(0)}E \simeq E(1_+)_{\geq 0}. \qquad \square$$

In light of obtaining the right stable homotopy type, we will from now on assume that we replace a Γ–space A by WA before applying the functor $\boldsymbol{B}$, ie in what follows, $\boldsymbol{B}A$ will mean $\boldsymbol{B}\, WA$.

We start with constructing a Γ–space model for the (connective cover of the) spectrum ψ_θ. Recall that ψ_θ has as n^{th} space the space $\psi_\theta(\infty, n+1)$ and structure maps given by translation of manifolds in the $(n+1)^{\text{st}}$ coordinate. The idea is that the spaces $\psi_\theta(\infty, n)$ come with a preferred monoid structure, namely taking disjoint union of manifolds. To make this precise, we introduce the following notation.

Definition 5.6 Let $\theta\colon X \to BO(d)$ be a Serre fibration. We obtain for each $m \in \mathbb{N}$ the Serre fibration

$$\coprod_m \theta\colon \coprod_m X \to BO(d).$$

We denote this Serre fibration by $\theta(m_+)$.

We can now associate to each $m_+ \in \Gamma^{\text{op}}$ the space $\Psi_{\theta(m_+)}(\mathbb{R}^n)$. We think of elements of $\Psi_{\theta(m_+)}(\mathbb{R}^n)$ as manifolds with components labeled by nonbasepoint elements of m_+ together with θ–structures on those labeled components.

Lemma 5.7 *For all $n \in \mathbb{N}$, the spaces $\Psi_{\theta(m_+)}(\mathbb{R}^n)$ assemble into a Γ–space.*

Proof We have to define the induced maps. Let $\sigma\colon m_+ \to k_+$ be a map of based sets. We obtain a map

$$\coprod_{\sigma^{-1}(k_+\backslash\{*\})} X \to \coprod_{k_+\backslash\{*\}} X.$$

Now define the induced map $\Psi_{\theta(m_+)}(\mathbb{R}^n) \to \Psi_{\theta(k_+)}(\mathbb{R}^n)$ as follows. The image of a pair (M, l) is given by the manifold

$$M' := l^{-1}\Bigg(\coprod_{\sigma^{-1}(k_+\setminus\{*\})} X\Bigg)$$

together with $\theta(k_+)$–structure given by the composition

$$M' \xrightarrow{l|_{M'}} \coprod_{\sigma^{-1}(k_+\setminus\{*\})} X \to \coprod_{k_+\setminus\{*\}} X.$$

In other words, we relabel the components of M and forget about those components, which get labeled by the basepoint. Taking the empty manifold as basepoint, it is easy to see that this is functorial in Γ^{op}. □

Note that $\Psi_{\theta(0_+)} \cong *$ since it consists of only the empty manifold and we have $\Psi_{\theta(1_+)}(\mathbb{R}^n) = \Psi_\theta(\mathbb{R}^n)$. Also note that we obtain by restriction for any $k \geq 1$ the Γ–spaces

$$m_+ \mapsto \psi_{\theta(m_+)}(\infty, k).$$

As mentioned above, the Γ–structure can be thought of as taking disjoint union of manifolds. Below we will see that, when stabilizing to $\mathbb{R}^\infty$, taking disjoint union gives a homotopy coherent multiplication on our spaces of manifolds.

Lemma 5.8 *The spectra $\psi_{\theta(m_+)}$ assemble into a projectively fibrant Γ–spectrum.*

Proof By the above lemma we have for each $n \in \mathbb{N}$ and each map of finite pointed sets $\sigma\colon m_+ \to k_+$ a map

$$\sigma_*^n\colon \psi_{\theta(m_+)}(\infty, n+1) \to \psi_{\theta(k_+)}(\infty, n+1)$$

which is functorial in Γ^{op} for fixed n. Thus, we have to show that these maps commute with the structure maps, that is we need to show that the diagram

$$\begin{tikzcd}
S^1 \wedge \psi_{\theta(m_+)}(\infty, n+1) \arrow[r] \arrow[d, "\mathrm{id}\wedge\sigma_*^n"'] & \psi_{\theta(m_+)}(\infty, n+2) \arrow[d, "\sigma_*^{n+1}"] \\
S^1 \wedge \psi_{\theta(k_+)}(\infty, n+1) \arrow[r] & \psi_{\theta(k_+)}(\infty, n+2)
\end{tikzcd}$$

commutes. But this is clear since the structure maps just translate the manifolds in the $(n+1)^{\mathrm{st}}$ coordinate, while the map σ_*^n relabels the components. □

Definition 5.9 We denote by $\Gamma\psi_\theta$ the Γ–spectrum

$$m_+ \mapsto \psi_{\theta(m_+)}.$$

To avoid awkward notation, we will denote the induced Γ–spaces $\Gamma^{(k)}(\Gamma\psi_\theta)$ simply by $\Gamma^{(k)}\psi_\theta$.

Proposition 5.10 *The Γ–space $\Gamma^{(0)}\psi_\theta$ is a model for the connective cover of ψ_θ, ie there is a stable equivalence*

$$\boldsymbol{B}\Gamma^{(0)}\psi_\theta \simeq \psi_{\theta,\geq 0}.$$

Proof We show that $\Gamma\psi_\theta$ is a special Γ–spectrum. The assertion then follows from Proposition 5.2. Since $\psi_{\theta(m_+)}$ is an Ω–spectrum for all $m_+ \in \Gamma^{\mathrm{op}}$, it suffices to show that $\Gamma^{(k)}\psi_\theta$ is a special Γ–space for every k.

We observe that the Segal map for $\Gamma^{(k)}\psi_\theta$

$$\Gamma^{(k)}\psi_\theta(m_+) \to \prod_m \Gamma^{(k)}\psi_\theta(m_+)$$

is an embedding, and we identify its image with a subspace of the product space. This subspace can be characterized as follows. A tuple $(M_1, \dots, M_m)$ lies in this subspace if and only if $M_i \cap M_j = \varnothing \subset \mathbb{R}^\infty$ for all $i \neq j$. We show that this subspace is a weak deformation retract of the product space

$$\prod_m \Gamma^{(k)}\psi_\theta(m_+) = \prod_m \psi_\theta(\infty, k+1).$$

To this end, we need a map that makes manifolds (or more generally any subsets) disjoint inside $\mathbb{R}^\infty$. Consider the maps

$$F\colon \mathbb{R}^\infty \to \mathbb{R}^\infty,$$
$$(x_1, x_2, \dots) \mapsto (0, x_1, x_2, \dots),$$

as well as for any $a \in \mathbb{R}$ the map

$$G_a\colon \mathbb{R}^\infty \to \mathbb{R}^\infty,$$
$$(x_1, x_2, \dots) \mapsto (a + x_1, x_2, \dots).$$

These maps are clearly homotopic to the identity via a straight line homotopy. Choosing $a \in (-1, 1)$, the composition $G_a \circ F\colon \mathbb{R}^\infty \to \mathbb{R}^\infty$ induces a self-map

$$\psi_\theta(\infty, k+1) \to \psi_\theta(\infty, k+1)$$

which is homotopic to the identity. Using for each factor of the product space $\prod_m \psi_\theta(\infty, k+1)$ a different (fixed) real number gives a map

$$\prod_m \psi_\theta(\infty, k+1) \to \prod_m \psi_\theta(\infty, k+1),$$

which is our desired deformation retract; this is also illustrated in Figure 1. $\square$

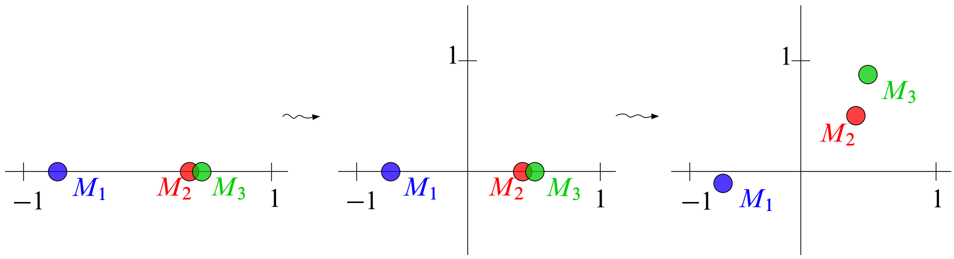


Figure 1: Making manifolds disjoint

Recall from Lemma 5.7 that the association

$$m_+ \mapsto \Psi_{\theta(m_+)}(\mathbb{R}^n)$$

defines a Γ–space for all $n \in \mathbb{N}$.

Definition 5.11 Denote by $\Gamma\Psi_\theta$ the (levelwise) colimit of Γ–spaces

$$\Gamma\Psi_\theta(m_+) := \operatorname{colim}_{n\in\mathbb{N}} \Omega^{n-1}\Psi_{\theta(m_+)}(\mathbb{R}^n).$$

From Theorem 4.2 we obtain for each $m_+ \in \Gamma^{\mathrm{op}}$ equivalences

$$\Gamma^{(0)}\psi_\theta(m_+) = \psi_{\theta(m_+)}(\infty, 1) \xrightarrow{\simeq} \operatorname{colim}_{n\in\mathbb{N}} \Omega^{n-1}\Psi_{\theta(m_+)}(\mathbb{R}^n) = \Gamma\Psi_\theta(m_+)$$

which are clearly functorial in Γ^{op}. Hence we obtain a levelwise equivalence of Γ–spaces $\Gamma^{(0)}\psi_\theta \xrightarrow{\simeq} \Gamma\Psi_\theta$.

Corollary 5.12 *The Γ–space $\Gamma\Psi_\theta$ is a model for the connective cover of the spectrum ψ_θ.*

We now construct a Γ–space model for the Madsen–Tillmann spectrum $\mathrm{MT}\theta(d)$ and we will show in the next section that this Γ–space is equivalent to $\Gamma\Psi_\theta$. As before, we will use the Serre fibrations $\theta(m_+)$. First note that the construction of the Madsen–Tillmann spectrum commutes with coproducts over $BO(d)$, that is we have $\mathrm{MT}\theta(m_+)(d) \cong \bigvee_m \mathrm{MT}\theta(d)$.

Definition 5.13 Define the Γ–spectrum $\Gamma\mathrm{MT}\theta(d)\colon \Gamma^{\mathrm{op}} \to \mathbf{Spt}$ by setting

$$\Gamma\mathrm{MT}\theta(d)(m_+) := \mathrm{MT}\theta(m_+)(d).$$

For any based map $\sigma\colon m_+ \to k_+$, define the induced map to be the fold map

$$\Gamma\mathrm{MT}\theta(d)(m_+) \cong \bigvee_m \mathrm{MT}\theta(d) \to \bigvee_k \mathrm{MT}\theta(d) \cong \Gamma\mathrm{MT}\theta(d)(k_+).$$

As before, we will denote the induced Γ–spaces by $\Gamma^{(k)}\mathrm{MT}\theta(d)$ for all $k \in \mathbb{N}$.

Proposition 5.14 *The Γ–space $\Gamma^{(1)}\mathrm{MT}\theta(d)$ is a model for the connective cover of the spectrum* $\mathrm{MT}\theta(d)[1]$.

Proof Again it suffices to show that $\Gamma\mathrm{MT}\theta(d)$ is special. But this follows easily since in **Spt** we have a stable equivalence

$$\mathrm{MT}\theta(m_+)(d) \cong \bigvee_m \mathrm{MT}\theta(d) \simeq \prod_m \mathrm{MT}\theta(d).$$

Thus by Proposition 5.2 we obtain a stable equivalence

$$\boldsymbol{B}\,\Gamma^{(1)}\mathrm{MT}\theta(d) \simeq \mathrm{MT}\theta(d)[1]_{\geq 0}. \qquad \square$$

6 Equivalence of Γ–space models

In the previous section we have constructed the Γ–space models $\Gamma\Psi_\theta$ for ψ_θ and $\Gamma^{(1)}\mathrm{MT}\theta(d)$ for $\mathrm{MT}\theta(d)[1]_{\geq 0}$. But more is true; by Theorem 4.2 we have for each $m_+ \in \Gamma^{\mathrm{op}}$ a weak equivalence of spaces

$$\Gamma^{(1)}\mathrm{MT}\theta(d)(m_+) = \Omega^\infty \mathrm{MT}\theta(m_+)(d)[1] \xrightarrow{\simeq} \operatorname{colim}_{n\in\mathbb{N}} \Omega^{n-1}\Psi_{\theta(m_+)}(\mathbb{R}^n) = \Gamma\Psi_\theta(m_+).$$

The following lemma shows that these equivalences define a levelwise equivalence of Γ–spaces.

Lemma 6.1 *The weak equivalences of Theorem 4.1*

$$\mathrm{Th}\big(\theta_n^*\gamma_{d,n}^\perp\big) \xrightarrow{\simeq} \Psi_\theta(\mathbb{R}^n)$$

assemble into a map of Γ–spaces. In particular, we obtain a levelwise equivalence

$$\Gamma^{(1)}\mathrm{MT}\theta(d) \xrightarrow{\simeq} \Gamma\Psi_\theta.$$

Proof We need to show that for any map of based sets $\sigma\colon m_+ \to k_+$ the diagram

$$\begin{tikzcd} \mathrm{Th}\big(\theta_n(m_+)^*\gamma_{d,n}^\perp\big) \arrow[r] \arrow[d, "\sigma_*"'] & \Psi_{\theta(m_+)}(\mathbb{R}^n) \arrow[d, "\sigma_*"] \\ \mathrm{Th}\big(\theta_n(k_+)^*\gamma_{d,n}^\perp\big) \arrow[r] & \Psi_{\theta(k_+)}(\mathbb{R}^n) \end{tikzcd}$$

commutes. But this follows easily since the left-hand vertical map is just the fold map. In particular one can view this map as relabeling components of the wedge and mapping components labeled by $*$ to the basepoint. On the other hand this is precisely the description of the right-hand vertical map. $\square$

We can now prove the first part of our main theorem.

Theorem 6.2 *There is an equivalence of spectra*

$$\mathrm{MT}\theta(d)[1]_{\geq 0} \simeq \psi_{\theta,\geq 0}.$$

Proof By Lemma 6.1, we have an equivalence of Γ–spaces

$$\Gamma^{(1)}\mathrm{MT}\theta(d) \xrightarrow{\simeq} \Gamma\Psi_\theta.$$

By Proposition 5.14, $\Gamma^{(1)}\mathrm{MT}\theta(d)$ is a model for the spectrum $\mathrm{MT}\theta(d)[1]_{\geq 0}$, while by Proposition 5.10 and its corollary, the Γ–space $\Gamma\Psi_\theta$ is a model for the connective cover of ψ_θ. Hence we obtain equivalences

$$\mathrm{MT}\theta(d)[1]_{\geq 0} \simeq \boldsymbol{B}\Gamma^{(1)}\mathrm{MT}\theta(d) \simeq \boldsymbol{B}\Gamma\Psi_\theta \simeq \psi_{\theta,\geq 0}. \qquad \square$$

7 The cobordism category

In the previous section we have exhibited an equivalence between the connective covers of the spectra $\mathrm{MT}\theta(d)[1]$ and ψ_θ. It remains to relate these spectra to the (classifying space of the) topological cobordism category.

Classically, the d–dimensional cobordism category has as objects closed $(d-1)$–dimensional manifolds and morphisms given by diffeomorphism classes of cobordisms. It is a symmetric monoidal category with monoidal product given by taking disjoint union of manifolds. We will see that this is also true for the topological variant in a sense we will make precise below. In particular, having a symmetric monoidal structure endows the classifying space of the cobordism category with the structure of an infinite loop space, and we will see that it is equivalent as such to the infinite loop space associated to $\mathrm{MT}\theta(d)[1]$.

Recall that a *topological category* $\mathcal{C}$ has a space of objects $\mathcal{C}_0$ and a space of morphisms $\mathcal{C}_1$ together with source and target maps $s, t\colon \mathcal{C}_1 \to \mathcal{C}_0$, a composition map $c\colon \mathcal{C}_1 \times_{\mathcal{C}_0} \mathcal{C}_1 \to \mathcal{C}_1$, and a unit map $e\colon \mathcal{C}_0 \to \mathcal{C}_1$, which satisfy the usual associativity and unit laws. There have appeared several definitions of the cobordism category as a topological category, which all have equivalent classifying spaces. The relevant model for us will be the topological poset model of [4]. We recall its definition. Define D_θ to be the subspace

$$D_\theta \subset \mathbb{R} \times \psi_\theta(\infty, 1)$$

consisting of pairs (t, M) where $t \in \mathbb{R}$ is a regular value of the projection onto the first coordinate $M \subset \mathbb{R} \times (-1,1)^\infty \to \mathbb{R}$. Order its elements by $(t, M) \leq (t', M')$ if and only if $t \leq t'$ with the usual order on $\mathbb{R}$ and $M = M'$.

Definition 7.1 The *d–dimensional cobordism category* $\mathrm{Cob}_\theta(d)$ is the topological category associated to the topological poset D_θ. That is, its space of objects is given by $\mathrm{ob}(\mathrm{Cob}_\theta(d)) = D_\theta$, and its space of morphisms is given by the subspace $\mathrm{mor}(\mathrm{Cob}_\theta(d)) \subset \mathbb{R}^2 \times \psi_\theta(\infty, 1)$ consisting of triples (t_0, t_1, M), where $t_0 \le t_1$. The source and target maps are simply given by forgetting regular values.

Given a topological category $\mathcal{C}$ we can take its internal nerve yielding a simplicial space

$$N_\bullet\mathcal{C}\colon \Delta^{\mathrm{op}} \to \boldsymbol{S}$$

as follows. The space of 0–simplices and 1–simplices is given by $\mathcal{C}_0$ and $\mathcal{C}_1$ respectively. For $n \ge 2$ the space of n–simplices is given by the n–fold fiber product

$$N_n\mathcal{C} := \mathcal{C}_1 \times_{\mathcal{C}_0} \cdots \times_{\mathcal{C}_0} \mathcal{C}_1.$$

The face and the degeneracy maps are obtained from the structure maps of the topological category. The associativity and unit laws ensure that we indeed obtain a simplicial space. Applying this construction to the cobordism category now yields a simplicial space

$$N_\bullet \mathrm{Cob}_\theta(d)\colon \Delta^{\mathrm{op}} \to \boldsymbol{S}.$$

We will also write $\mathrm{Cob}_\theta(d)$ for the simplicial space obtained from taking the nerve and write $\mathrm{Cob}_\theta(d)_k$ for the space of k–simplices.

Considering $\psi_\theta(\infty, 1)$ as a constant simplicial space, we have a forgetful map of simplicial spaces $\mathrm{Cob}_\theta(d) \to \psi_\theta(\infty, 1)$ defined on k–simplices by

$$\mathrm{Cob}_\theta(d)_k \to \psi_\theta(\infty, 1),$$
$$(\underline{t}, M) \mapsto M.$$

Theorem 7.2 *The forgetful map induces a weak equivalence*

$$B\,\mathrm{Cob}_\theta(d) \xrightarrow{\simeq} \psi_\theta(\infty, 1),$$

where $B\,\mathrm{Cob}_\theta(d)$ is the realization of the simplicial space $\mathrm{Cob}_\theta(d)$.

Proof See [4, Theorem 3.10]. □

We now encode the symmetric monoidal structure of $\mathrm{Cob}_\theta(d)$ in terms of a Γ–structure.

Lemma 7.3 *The simplicial spaces $\mathrm{Cob}_{\theta(m_+)}(d)$ assemble into a Γ–object in simplicial spaces*

$$\mathrm{Cob}_{\theta(-)}(d)\colon \Gamma^{\mathrm{op}} \to \boldsymbol{S}^{\Delta^{\mathrm{op}}}.$$

Proof For $m_+ \in \Gamma^{\mathrm{op}}$ the k–simplices are given as subspaces

$$\mathrm{Cob}_{\theta(m_+)}(d)_k \subset \mathbb{R}^{k+1} \times \psi_{\theta(m_+)}(\infty, 1) = \mathbb{R}^{k+1} \times \Gamma^{(0)}\psi_\theta(m_+).$$

Thus for a map $\sigma\colon m_+ \to n_+$, we define the map

$$\mathrm{Cob}_{\theta(m_+)}(d) \to \mathrm{Cob}_{\theta(n_+)}(d)$$

on k–simplices to be induced by the map

$$\mathrm{id} \times \sigma_*\colon \mathbb{R}^{k+1} \times \Gamma^{(0)}\psi_\theta(m_+) \to \mathbb{R}^{k+1} \times \Gamma^{(0)}\psi_\theta(n_+),$$

where σ_* comes from the functoriality in Γ^{op} of the Γ–space $\Gamma^{(0)}\psi_\theta$. From this description it is clear that the maps just defined are functorial in Δ^{op} and hence define a map of simplicial spaces. □

Definition 7.4 Denote by $\Gamma\,\mathrm{Cob}_\theta(d)$ the Γ–object in simplicial spaces

$$\Gamma\,\mathrm{Cob}_{\theta(m_+)}(d) \to \boldsymbol{S}^{\Delta^{\mathrm{op}}}, \qquad m_+ \mapsto \mathrm{Cob}_{\theta(m_+)}(d).$$

Composing with the realization of simplicial spaces we get a functor

$$B\Gamma\,\mathrm{Cob}_\theta(d)\colon \Gamma^{\mathrm{op}} \to \boldsymbol{S}.$$

We obtain a Γ–space by choosing as basepoints the elements $(\underline{0}, \varnothing) \in \mathrm{Cob}_\theta(d)_k$ for all $k \in \mathbb{N}$.

Lemma 7.5 *The forgetful map induces a levelwise equivalence of Γ–spaces*

$$B\Gamma\,\mathrm{Cob}_\theta(d) \xrightarrow{\simeq} \Gamma^{(0)}\psi_\theta.$$

Proof By construction it is clear that the forgetful maps are functorial in Γ^{op} so that they indeed define a map of Γ–spaces. By Theorem 7.2, these maps are weak equivalences and hence we obtain a levelwise equivalence of Γ–spaces. □

In particular, the Γ–space $B\Gamma\,\mathrm{Cob}_\theta(d)$ is very special, and applying Segal's functor we obtain a connective Ω–spectrum, which we denote by $\boldsymbol{B}\,\Gamma\,\mathrm{Cob}_\theta(d)$ to avoid awkward notation. In conclusion, we obtain an equivalence of spectra

$$\boldsymbol{B}\,\Gamma\,\mathrm{Cob}_\theta(d) \xrightarrow{\simeq} \boldsymbol{B}\,\Gamma^{(0)}\psi_\theta.$$

Combining with Theorem 6.2, we obtain our main theorem.

Main Theorem *There are stable equivalences of spectra*

$$\boldsymbol{B}\,\Gamma\,\mathrm{Cob}_\theta(d) \simeq \boldsymbol{B}\,\Gamma^{(0)}\psi_\theta \simeq \mathrm{MT}\theta(d)[1]_{\geq 0},$$

such that the induced equivalences

$$\Omega^\infty \boldsymbol{B}\,\Gamma\,\mathrm{Cob}_\theta(d) \simeq \Omega^\infty\psi_\theta \simeq \Omega^\infty \mathrm{MT}\theta(d)[1]$$

are equivalent to the equivalences of Theorem 4.2 and Theorem 7.2.

References

[1] **M Bökstedt**, **I Madsen**, *The cobordism category and Waldhausen's K-theory*, from "An alpine expedition through algebraic topology" (C Ausoni, K Hess, B Johnson, W Lück, J Scherer, editors), Contemp. Math. 617, Amer. Math. Soc. (2014) 39–80 MR

[2] **A K Bousfield**, **E M Friedlander**, *Homotopy theory of Γ–spaces, spectra, and bisimplicial sets*, from "Geometric applications of homotopy theory, II" (M G Barratt, M E Mahowald, editors), Lecture Notes in Math. 658, Springer (1978) 80–130 MR

[3] **Z Fiedorowicz**, *A note on the spectra of algebraic K-theory*, Topology 16 (1977) 417–421 MR

[4] **S Galatius**, **O Randal-Williams**, *Monoids of moduli spaces of manifolds*, Geom. Topol. 14 (2010) 1243–1302 MR

[5] **S Galatius**, **U Tillmann**, **I Madsen**, **M Weiss**, *The homotopy type of the cobordism category*, Acta Math. 202 (2009) 195–239 MR

[6] **I Madsen**, **U Tillmann**, *The stable mapping class group and $Q(\mathbb{C}P^\infty_+)$*, Invent. Math. 145 (2001) 509–544 MR

[7] **M A Mandell**, **J P May**, **S Schwede**, **B Shipley**, *Model categories of diagram spectra*, Proc. London Math. Soc. 82 (2001) 441–512 MR

[8] **J P May**, **R Thomason**, *The uniqueness of infinite loop space machines*, Topology 17 (1978) 205–224 MR

[9] **G Segal**, *Categories and cohomology theories*, Topology 13 (1974) 293–312 MR

Fakultät für Mathematik, Universität Regensburg
D-93040 Regensburg, Germany

`hoang-kim.nguyen@mathematik.uni-regensburg.de`

Received: 21 December 2015 Revised: 1 July 2016

msp

Algebraic & Geometric Topology 17 (2017) 1041–1057

Infima of length functions and dual cube complexes

JONAH GASTER

In the presence of certain topological conditions, we provide lower bounds for the infimum of the length function associated to a collection of curves on Teichmüller space that depend on the dual cube complex associated to the collection, a concept due to Sageev. As an application of our bounds, we obtain estimates for the "longest" curve with k self-intersections, complementing work of Basmajian [J. Topol. 6 (2013) 513–524].

51M10; 51M16

Let Σ be an oriented topological surface of finite type. We denote the Teichmüller space of Σ by $\mathcal{T}(\Sigma)$, which we interpret as the deformation space of marked hyperbolic structures on Σ. Given $X \in \mathcal{T}(\Sigma)$ and a free homotopy class (or *closed curve*) γ on Σ, we denote by $\ell(\gamma, X)$ the length of the geodesic representative of γ in the hyperbolic structure determined by X. If $\Gamma = \{\gamma_i\}$ is a collection of closed curves, then we define $\ell(\Gamma, X) = \sum \ell(\gamma_i, X)$.

In this note, we are concerned with translating topological information of Γ into quantitative information about the length function $\ell(\Gamma, \cdot)\colon \mathcal{T}(\Sigma) \to \mathbb{R}$. In particular, we develop tools to estimate the infimum of $\ell(\Gamma, \cdot)$ over $\mathcal{T}(\Sigma)$. This work naturally complements that of Basmajian [3], where such estimates are obtained that depend on the number of self-intersections of Γ. Here we consider a finer topological invariant than the self-intersection number.

A construction of Sageev [7] associates to a curve system Γ an isometric action of $\pi_1\Sigma$ on a finite-dimensional cube complex, or the *dual cube complex* $\mathcal{C}(\Gamma)$ of Γ. In what follows we connect geometric properties of the dual cube complex $\mathcal{C}(\Gamma)$ to the length of the collection of curves Γ on any hyperbolic surface. Indeed, [1, Theorem 3] of Aougab and Gaster suggests that any such information is implicitly contained in the combinatorics of $\mathcal{C}(\Gamma)$. We have:

Theorem A *Suppose that the action of $\pi_1\Sigma$ on $\mathcal{C}(\Gamma)$ has a set of cubes $C_1, \dots, C_m$, of dimensions $n_1, \dots, n_m$, respectively, in distinct $\pi_1\Sigma$–orbits, such that the union of orbits $\bigcup_i \pi_1\Sigma \cdot C_i$ is hyperplane separated. Then*

$$\inf_{X \in \mathcal{T}(\Sigma)} \ell(\Gamma, X) \geq \sum_{i=1}^{m} n_i \log\left(\frac{1+\cos\pi/n_i}{1-\cos\pi/n_i}\right).$$

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.1041

The definition of *hyperplane separated* can be found in Section 1. The main use of this idea is that it allows one to conclude that large chunks of the preimage of the curves Γ in the universal cover embed on the surface under the covering map. The proof of Theorem A proceeds by minimizing the length of these large chunks.

Remark The contribution to the bound above from a cube C_i is useless when C_i is a 2–cube. On the other hand, it still seems reasonable to expect a lower bound on the length function in the presence of many maximal 2–cubes. Note that the presence of m 2–cubes contributes m to the self-intersection number, so that Basmajian's bounds immediately imply a lower bound for the length function that is logarithmic in m. While Basmajian's lower bounds are sharp, the examples that demonstrate sharpness have high-dimensional dual complexes. We expect a positive answer to the following.

Question If $\mathcal{C}(\Gamma)$ contains m maximal 2–cubes, is there a lower bound for the length function $\ell(\Gamma, X)$ that is linear in m?

Remark The bounds in Theorem A are sharp in the following sense: for each $n \in \mathbb{N}$, there exists a set of curves Γ_n on the $(n+1)$–holed sphere $\Sigma_{0,n+1}$, and hyperbolic structures $X_n \in \mathcal{T}(\Sigma_{0,n+1})$ such that

(1) the dual cube complex $\mathcal{C}(\Gamma_n)$ has a hyperplane separated n–cube, and
(2) the hyperbolic length $\ell(\Gamma_n, X_n)$ is asymptotic to $n \log n$.

The problem remains of determining when a collection of curves gives rise to hyperplane separated orbits of cubes in the action of $\pi_1\Sigma$ on the dual cube complex. We offer a sufficient condition below which applies in many cases, toward which we fix some terminology. Recall that a *ribbon graph* is a graph with a cyclic order given to the oriented edges incident to each vertex. A ribbon graph G is *even* if the valence of each vertex is even. When an even ribbon graph G is embedded on a surface Σ, a collection of homotopy classes of curves is determined by G by going straight at each vertex. See Section 6 for a more precise description.

Theorem B *Suppose that $G \hookrightarrow \Sigma$ is an embedding of an even ribbon graph G into Σ with vertices of valence $n_1, \dots, n_m$, such that the complement $\Sigma \setminus G$ contains no monogons, bigons, or triangles. Let Γ indicate the union of the closed curves determined by G. Then G is a minimal position realization of Γ, the self-intersection of Γ is given by $\binom{n_1}{2} + \cdots + \binom{n_m}{2}$, and $\mathcal{C}(\Gamma)$ contains cubes $C_1, \dots, C_m$ of dimensions $n_1, \dots, n_m$, respectively, in distinct $\pi_1\Sigma$–orbits, whose union is hyperplane separated.*

This provides a general method to construct curves with definite self-intersection number and definite hyperplane separated cubes in their dual cube complexes. For example:

Example Consider the curve in Figure 4. Theorem B implies that the curve has six hyperplane separated 3–cubes. The estimate in Theorem A now applies, so that the length of the pictured curve is at least $18\log 3$ in any hyperbolic metric on Σ_6.

Let $\mathfrak{C}_k(\Sigma)$ indicate curves on Σ with self-intersection number k. Basmajian examined the following quantities, showing that they both are asymptotic to $\log k$:

$$m_k(\Sigma) := \min_{\gamma\in\mathfrak{C}_k(\Sigma)} \inf\{\ell(\gamma, X) \mid X \in \mathcal{T}(\Sigma)\},$$
$$M_k := \inf\{m_k(\Sigma) \mid \Sigma \text{ is a finite-type surface with } \chi(\Sigma) < 0\}.$$

Note that, for each k and Σ, there are finitely many mapping class group orbits among $\mathfrak{C}_k(\Sigma)$. This justifies the use of minimum in the definition of $m_k(\Sigma)$ above. One may define analogously

$$\overline{m}_k(\Sigma) := \max_{\gamma\in\mathfrak{C}_k(\Sigma)} \inf\{\ell(\gamma, X) \mid X \in \mathcal{T}(\Sigma)\},$$
$$\overline{M}_k := \sup\{\overline{m}_k(\Sigma) \mid \Sigma \text{ is a finite-type surface with } \chi(\Sigma) < 0\}.$$

The curves that realize the minima $m_k(\Sigma)$ and M_k manage to gain a lot of self-intersection while remaining quite short, which they achieve by winding many times around a very short curve. By constructing explicit families of curves that behave quite differently — namely, they return many times to a fixed small compact set on the surface — we provide a lower bound for $\overline{M}_k$ that grows faster than Basmajian's bounds for the "shortest" curves with k self-intersections.

Theorem C *We have the estimate*

$$\limsup_{k\to\infty} \frac{\overline{M}_k}{k} \geq \frac{\log 3}{3}.$$

Remark It is not hard to observe that

$$\limsup_{k\to\infty} \frac{\overline{m}_k(\Sigma)}{\sqrt{k}} > 0.$$

Indeed, given any k–curve $\gamma \in \mathfrak{C}_k(\Sigma)$, consider the closed curve γ^n given by wrapping n times around γ. The infimum of the length function of γ^n will grow linearly in n, while the self-intersection number will grow quadratically in n. Performing this calculation with a curve with one self-intersection, one finds that

$$\limsup_{k\to\infty} \frac{\overline{m}_k(\Sigma)}{\sqrt{k}} \geq 4\log(1+\sqrt{2}).$$

The problem of sharpness for our examples, namely good upper bounds for $\overline{m}_k(\Sigma)$ and $\overline{M}_k$, seems subtle. In particular, such upper bounds would imply an asymptotically good answer to the following question.

Question Given a curve $\gamma \in \mathfrak{C}_k(\Sigma)$, what is an explicit function $C(k, \Sigma)$ such that there is a point $X \in \mathcal{T}(\Sigma)$ with $\ell(\gamma, X) \le C(k, \Sigma)$?[1]

Note that one could also ask for an upper bound that is independent of Σ, towards which the lower bound in Theorem C is more relevant.

Organization In Section 1 we briefly recall Sageev's construction, and define hyperplane separation. In Section 2 we lay out the necessary tools for the proof of Theorem A, and in Section 3 and Section 4 we prove these tools. Section 5 describes a straightforward method of detecting self-intersection and hyperplane separation, and Section 6 introduces even ribbon graphs and the proof of Theorem B. Finally, Section 7 describes a family of examples to which these tools apply, and contains a proof of Theorem C.

Acknowledgements The author thanks Ara Basmajian, Martin Bridgeman, Spencer Dowdall, and David Dumas for helpful conversations, and Ian Biringer for a correction and reference in regard to Lemma 5.1.

1 Dual cube complexes and hyperplane separation

We recall Sageev's construction. A collection Γ of homotopy classes of curves on Σ gives rise to an isometric action of $\pi_1\Sigma$ on a CAT(0)-cube complex $\mathcal{C}(\Gamma)$. This action is obtained roughly as follows: choose a minimal position realization λ of the curves in Γ, and consider the preimage $\widetilde{\lambda}$ of λ in the universal cover $\widetilde{\Sigma}$. In the language of [9], the set $\widetilde{\lambda}$ decomposes into a union[2] of *elevations*. Each elevation splits $\widetilde{\Sigma}$ into two connected components. A *labeling* of $\widetilde{\lambda}$ is a choice of half-space in the complement of each of the elevations. The one skeleton of the cube complex $\mathcal{C}(\Gamma)$ is built from *admissible* labelings of $\widetilde{\lambda}$, ie choices of half-spaces in the complement of the elevations so that any pair intersect.

[1] While this paper was under review, this question has been given an answer by Aougab et al [2, Theorem 1.4].

[2] An illustrative example is provided by the case that λ is given by the geodesic representatives of Γ relative to a chosen hyperbolic structure, in which case $\widetilde{\lambda} \subset \mathbb{H}^2$ is evidently a union of complete geodesics.

Two such admissible labelings are connected by an edge when they differ on precisely one elevation of $\widetilde{\lambda}$ (in analogy with the dual graph to the set $\widetilde{\lambda}$). Finally, $\mathcal{C}(\Gamma)$ is given by the unique nonpositively curved cube complex with the prescribed 1–skeleton. The action of $\pi_1\Sigma$ on the elevations comprising $\widetilde{\lambda}$ naturally induces a permutation of the labelings, which induces an isometry of $\mathcal{C}(\Gamma)$. See [7; 8; 4] for details.

We collect this information conveniently.

Theorem (Sageev) *The action of $\pi_1\Sigma$ on the* CAT(0) *cube complex $\mathcal{C}(\Gamma)$ is independent of realization. There is a $\pi_1\Sigma$–equivariant incidence-preserving correspondence of the hyperplanes of $\mathcal{C}(\Gamma)$ with the elevations in $\widetilde{\lambda}$, so that maximal n–cubes are in correspondence with maximal collections of n pairwise intersecting elevations of curves in Γ.*

In light of Sageev's theorem we may sometimes identify the elevations in $\widetilde{\lambda}$ with the hyperplanes of the cube complex $\mathcal{C}(\Gamma)$.

Given a cube C in a cube complex $\mathcal{C}$, we denote the set of hyperplanes of $\mathcal{C}$ by $\mathcal{H}(\mathcal{C})$, and the set of hyperplanes of C by $\mathcal{H}(C) \subset \mathcal{H}(\mathcal{C})$.

Definition Suppose C and D are two cubes in a cube complex. We say that C and D are *hyperplane separated* if either $\mathcal{H}(C) \cap \mathcal{H}(D) = \varnothing$, or $|\mathcal{H}(C) \cap \mathcal{H}(D)| = 1$, and, for any $c \in \mathcal{H}(C)$ and $d \in \mathcal{H}(D)$ with $c \neq d$, the hyperplanes c and d are disjoint.

A union of cubes $\bigcup_i C_i$ is hyperplane separated when every pair is hyperplane separated.

2 Proof of Theorem A

Consider an n–cube $C \subset \mathcal{C}(\Gamma)$. The orientation of $\widetilde{\Sigma}$ induces a counterclockwise cyclic ordering of the n elevations of curves from Γ that correspond to the hyperplanes of C. In what follows, we fix a hyperbolic surface $X \in \mathcal{T}(\Sigma)$ and identify the universal cover $\widetilde{\Sigma}$ with $\mathbb{H}^2$. We will work with the Poincaré disk model for $\mathbb{H}^2$, with conformal boundary S^1.

Enumerate the n geodesic representatives $(\gamma_1, \dots, \gamma_n)$ of the elevations of curves that correspond to C, respecting the cyclic order. Each geodesic γ_i has two endpoints $p_i, q_i \in S^1$. Choose these labels so that $p_1, \dots, p_n, q_1, \dots, q_n$ is consistent with the cyclic order of S^1.

Given a trio of elevations $\gamma_{i-1}, \gamma_i, \gamma_{i+1} \subset \mathbb{H}^2$, consider the pair of distinct disjoint geodesics (p_{i-1}, p_{i+1}) and (q_{i-1}, q_{i+1}). We will refer to this pair as the *separators*

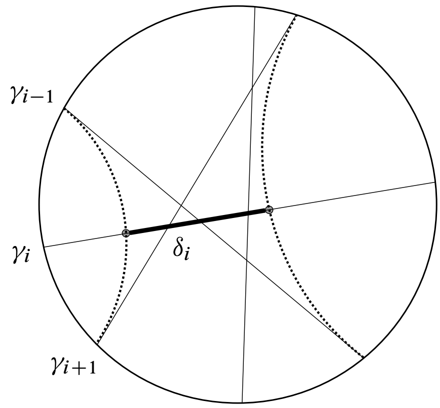


Figure 1: The separators of the hyperplane γ_i and the diagonal δ_i

of the hyperplane γ_i in C, and we will denote the pair by $\mathrm{sep}(\gamma_i, C)$. (When $i = 1$ or $i = n$, the separators of γ_i are the geodesics (p_{i-1}, q_{i+1}) and (q_{i-1}, p_{i+1}), with indices read modulo n.) Let δ_i indicate the portion of γ_i between the separators. We will refer to the arcs $\{\delta_1, \dots, \delta_n\}$ as the *diagonals* of the cube C. See Figure 1 for a schematic picture.

Lemma 2.1 *For each i, we have*

$$\ell(\delta_i) = \log\left|\frac{(p_i - q_{i-1})(p_i - q_{i+1})(q_i - p_{i+1})(q_i - p_{i-1})}{(p_i - p_{i+1})(p_i - p_{i-1})(q_i - q_{i+1})(q_i - q_{i-1})}\right|.$$

Proof The proof is a calculation in $\mathbb{H}^2$. □

Towards Theorem A, we suppose below that $C_1, \dots, C_m$ are cubes of $\mathcal{C}(\Gamma)$ in distinct $\pi_1\Sigma$–orbits. Let $\delta^i_1, \dots, \delta^i_{n_i}$ be the diagonals of the cube C_i, let

$$\mathcal{D}_i := \bigcup_k \delta^i_k$$

indicate the union of the diagonals of C_i, and $\mathcal{D} := \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_m$.

For ease of exposition, we postpone the proof of the following proposition.

Proposition 2.2 *If the union of orbits $\bigcup \pi_1\Sigma \cdot C_i$ is hyperplane separated, then the covering map $\pi\colon \mathbb{H}^2 \to \Sigma$ is injective on the union $\mathcal{D}$ minus a finite set of points.*

Finally, we will need the solution to the following optimization problem, whose proof we also postpone: given $2n$ distinct points $x_1, \dots, x_{2n} \in S^1$, for notational convenience we adopt the natural convention that subscripts should be read modulo $2n$, so that $x_{2n+1} = x_1$ and $x_0 = x_{2n}$. Let $F(x_1, \dots, x_{2n})$ be defined by

$$F(x_1, \dots, x_{2n}) = \log \prod_{j=1}^{2n} \left| \frac{(x_j - x_{j+n+1})(x_j - x_{j+n-1})}{(x_j - x_{j+1})(x_j - x_{j-1})} \right|.$$

Lemma 2.3 *When $(x_1, \dots, x_{2n})$ are cyclically ordered in S^1, we have*

$$F(x_1, \dots, x_{2n}) \geq n \log\left(\frac{1 + \cos \pi/n}{1 - \cos \pi/n} \right).$$

Assuming for now Proposition 2.2 and Lemma 2.3, we are ready to prove Theorem A.

Proof of Theorem A We bound from below the sum of lengths of the curves from Γ in the hyperbolic structure determined by $X \in \mathcal{T}(\Sigma)$. Pull Γ tight to geodesics, and consider the preimage under the covering transformation. As described above, each cube C_i, of dimension n_i, has n_i hyperplanes with n_i corresponding elevations of mutually intersecting geodesics in $\mathbb{H}^2$. These curves determine $2n_i$ cyclically ordered distinct points

$$p_1^i, p_2^i, \dots, p_{n_i}^i, q_1^i, q_2^i, \dots, q_{n_i}^i$$

on S^1, the diagonals $\mathcal{D}_i$, and $\mathcal{D}$, the union of $\mathcal{D}_i$. We estimate

$$\begin{aligned}
\ell(\Gamma, X) &\geq \ell(\mathcal{D}) \\
&= \sum_{i=1}^{m} \ell(\mathcal{D}_i) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n_i} \ell(\delta_j^i) \\
&= \sum_{i=1}^{m} \log \prod_{j=1}^{n_i} \left| \frac{(p_j^i - q_{j-1}^i)(p_j^i - q_{j+1}^i)(q_j^i - p_{j+1}^i)(q_j^i - p_{j-1}^i)}{(p_j^i - p_{j+1}^i)(p_j^i - p_{j-1}^i)(q_j^i - q_{j+1}^i)(q_j^i - q_{j-1}^i)} \right| \\
&\geq \sum_{i=1}^{m} n_i \log\left(\frac{1 + \cos \pi/n_i}{1 - \cos \pi/n_i} \right),
\end{aligned}$$

where the first, fourth and fifth lines follow from Proposition 2.2, Lemma 2.1 and Lemma 2.3, respectively. ◻

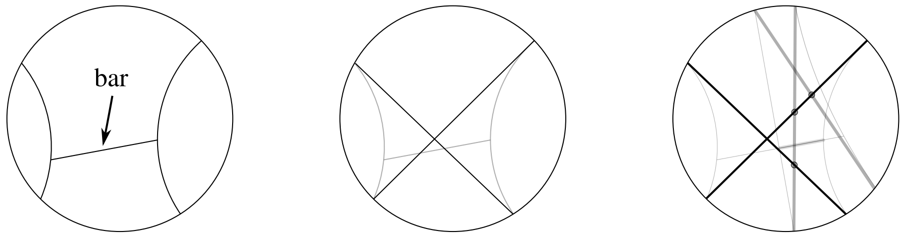


Figure 2: The notion of an H in $\mathbb{H}^2$. From left to right: an H and its bar, the cross of an H, and H's with overlapping bars.

3 Proof of Proposition 2.2

Proposition 2.2 is the sole motivation for the definition of hyperplane separated. We turn to the proof. To aid our exposition, we will say that an H in $\mathbb{H}^2$ is a pair of disjoint geodesics, and a geodesic arc connecting them. Associated to an H is a *cross*, a pair of intersecting geodesics with the same limit points as the H. See Figure 2 left and middle for a schematic. For example, the union of the diagonal δ_i and the separators $\mathrm{sep}(\gamma_i, C)$ form an H, with associated cross $\{\gamma_{i-1}, \gamma_{i+1}\}$.

Lemma 3.1 *Suppose* $\mathrm{H}_1, \mathrm{H}_2 \subset \mathbb{H}^2$ *are distinct* H*'s whose bars overlap in an interval. Then the crosses of* H_1 *and* H_2 *intersect.*

See Figure 2 right for a schematic.

Proof The convex hull of an H is an ideal quadrilateral. By assumption, the convex hulls of H_1 and H_2 intersect. The lemma follows from the following simple observation: if two ideal quadrilaterals intersect, then their crosses intersect. We demonstrate this below. Note that the ideal points of an ideal quadrilateral are cyclically ordered. We say that two such points are *opposite* if they are not neighbors in the cyclic order.

Let P and Q be intersecting ideal quadrilaterals, with cyclically ordered ideal points ∂P and ∂Q in $\partial_\infty \mathbb{H}^2$. As P and Q intersect, there are two points $q, q' \in \partial Q$ lying in distinct components of $\partial_\infty \mathbb{H}^2 \setminus \partial P$. Suppose that q and q' are not opposite vertices. The vertex that follows q' in the cyclic order is either in the same component of $\partial_\infty \mathbb{H}^2 \setminus \partial P$ as q, in which case there are three vertices in the same component as q, or it is in a distinct component from q. Thus if P and Q intersect, there are a pair of opposite vertices of ∂Q in distinct components of $\partial_\infty \mathbb{H}^2 \setminus \partial P$.

Opposite vertices of an ideal quadrilateral are boundary points of the cross of the quadrilateral. Thus there is a geodesic of the cross of Q that runs between distinct components of $\partial_\infty \mathbb{H}^2 \setminus \partial P$, so that the crosses of P and Q intersect. □

Proof of Proposition 2.2 As the union $\mathcal{D}$ is compact, properness of the action of $\pi_1\Sigma$ ensures that there are finitely many elements $g \in \pi_1\Sigma$ such that $g\cdot\mathcal{D}\cap\mathcal{D} \neq \varnothing$. Let $\mathcal{D}'$ indicate the complement in $\mathcal{D}$ of the finitely many points that are transversal intersections of $\mathcal{D}$ with $g\cdot\mathcal{D}$. If π is not injective on $\mathcal{D}'$, then there is an element $1 \neq g \in \pi_1\Sigma$ such that $g\cdot\delta_i^k$ and δ_j^l overlap in an open interval. In particular, note that g sends the hyperplane containing δ_i^k to the hyperplane containing δ_j^l.

If $k=l$ and $g\cdot C_k = C_k$, then $g\cdot\mathcal{D}_k = \mathcal{D}_k$, and by Brouwer's fixed point theorem there would be a fixed point of g, violating freeness of the action of $\pi_1\Sigma$. Since $g\cdot C_k \neq C_l$ for $k \neq l$ (recall that the cubes $\{C_i\}$ are in distinct $\pi_1\Sigma$–orbits), we may thus assume that $g\cdot C_k$ and C_l are distinct cubes that share the common hyperplane containing the diagonals $g\cdot\delta_i^k$ and δ_j^l. Let the hyperplanes of C_k be given by $\{\gamma_1,\dots,\gamma_{n_k}\}$, and those of C_l by $\{\eta_1,\dots,\eta_{n_l}\}$, so that $g\cdot\gamma_i = \eta_j$.

Observe that a trivial consequence of separatedness is that the separators $\operatorname{sep}(\eta_j, g\cdot C_k)$ are not the same pair of geodesics as the separators $\operatorname{sep}(\eta_j, C_l)$: if they were identical, then $g\cdot C_k$ and C_l would be two distinct cubes in the union $\bigcup \pi_1\Sigma\cdot C_i$ that share the hyperplanes corresponding to γ_{j-1}, γ_j, and γ_{j+1}.

Consider then the two H's formed by $g\cdot\delta_i^k$ and $\operatorname{sep}(\eta_j, g\cdot C_k)$ on the one hand, and δ_j^l and $\operatorname{sep}(\eta_j, C_l)$ on the other. By assumption these two H's have overlapping bars, so that by Lemma 3.1 their crosses intersect. Namely, one of $g\cdot\gamma_{i-1}$ and $g\cdot\gamma_{i+1}$ intersects one of η_{j-1} and η_{j+1}. This contradicts separatedness of C. We conclude that π is injective on $\mathcal{D}'$, the union of the diagonals of $C_1,\dots,C_m$ minus finitely many points, as desired. □

4 Proof of Lemma 2.3

We solve the necessary optimization problem.

Proof of Lemma 2.3 Note that F has several useful invariance properties: first it is clear that F is invariant under rotations of S^1. More generally, the conformal automorphisms of the disk $\operatorname{Aut}(\mathbb{D})$ act diagonally on $(S^1)^{2n}$, and for any $\sigma \in \operatorname{Aut}(\mathbb{D})$, $F\circ\sigma = F$. As well, it is immediate from the definition that $F(x_1,x_2,\dots,x_{2n}) = F(x_2,\dots,x_{2n},x_1)$.

For each $j=1,\dots,2n$, let $x_j = e^{i\theta_j}$. Applying a rotation of S^1 if necessary, we assume that $0 \le \theta_1 < \cdots < \theta_{2n} < 2\pi$.

The identity $|e^{i\alpha} - e^{i\beta}| = \sqrt{2-2\cos(\alpha-\beta)}$ implies that

$$\log\left|\frac{e^{i\theta_j} - e^{i\theta_k}}{e^{i\theta_j} - e^{i\theta_l}}\right| = \tfrac{1}{2}\log\frac{1-\cos(\theta_j-\theta_k)}{1-\cos(\theta_j-\theta_l)}.$$

Taking a derivative we find

$$\frac{\partial F}{\partial\theta_j}=\frac{\sin(\theta_j-\theta_{j+n-1})}{1-\cos(\theta_j-\theta_{j+n-1})}+\frac{\sin(\theta_j-\theta_{j+n+1})}{1-\cos(\theta_j-\theta_{j+n+1})}\\ -\frac{\sin(\theta_j-\theta_{j-1})}{1-\cos(\theta_j-\theta_{j-1})}-\frac{\sin(\theta_j-\theta_{j+1})}{1-\cos(\theta_j-\theta_{j+1})}.$$

Since $\sin\theta/(1-\cos\theta)=\cot\frac{\theta}{2}$, we may write the above as

$$\frac{\partial F}{\partial\theta_j}=\cot\frac{\theta_j-\theta_{j+n-1}}{2}+\cot\frac{\theta_j-\theta_{j+n+1}}{2}-\cot\frac{\theta_j-\theta_{j+1}}{2}-\cot\frac{\theta_j-\theta_{j-1}}{2}.$$

Towards candidates for absolute minima of F, we seek solutions to the system of equations $\{\partial F/\partial\theta_j=0\}$. Given the invariance properties of F, any such solution is far from unique, even locally. In order to characterize the unique $\mathrm{Aut}(\mathbb{D})$–orbit of a solution, we pick a j, and fix the choices $\theta_{n+j}-\theta_j=\pi$, and $\theta_{n+j+1}-\theta_{j+1}=\pi$.

With $\theta_j+\pi$ substituted for θ_{n+j}, the equations $\partial F/\partial\theta_j=0$ and $\partial F/\partial\theta_{n+j}=0$ now yield

$$\cot\frac{\theta_j-\theta_{j+n-1}}{2}+\cot\frac{\theta_j-\theta_{j+n+1}}{2}=\cot\frac{\theta_j-\theta_{j+1}}{2}+\cot\frac{\theta_j-\theta_{j-1}}{2},$$
$$\tan\frac{\theta_j-\theta_{j-1}}{2}+\tan\frac{\theta_j-\theta_{j+1}}{2}=\tan\frac{\theta_j-\theta_{j+n-1}}{2}+\tan\frac{\theta_j-\theta_{j+n+1}}{2},$$

respectively. Eliminating $\tan\frac{\theta_j-\theta_{j+1}}{2}$, we find

$$\cot\frac{\theta_j-\theta_{j+n-1}}{2}+\cot\frac{\theta_j-\theta_{j+n+1}}{2}-\cot\frac{\theta_j-\theta_{j-1}}{2}\\ =\left(\tan\frac{\theta_j-\theta_{j+n-1}}{2}+\tan\frac{\theta_j-\theta_{j+n+1}}{2}-\tan\frac{\theta_j-\theta_{j-1}}{2}\right)^{-1}.$$

Recall the remarkable fact that the solutions of the equation

$$\frac{1}{x}+\frac{1}{y}+\frac{1}{z}=\frac{1}{x+y+z}$$

are precisely the equations $x=-y$, $x=-z$, or $y=-z$. As a consequence, we have one of the following:

$$\tan\frac{\theta_j-\theta_{j+n-1}}{2}=-\tan\frac{\theta_j-\theta_{j+n+1}}{2},$$
$$\tan\frac{\theta_j-\theta_{j+n-1}}{2}=\tan\frac{\theta_j-\theta_{j-1}}{2},\quad\text{or}$$
$$\tan\frac{\theta_j-\theta_{j+n+1}}{2}=\tan\frac{\theta_j-\theta_{j-1}}{2}.$$

By assumption,

$$0 < \theta_{j+n-1} - \theta_j < \pi < \theta_{j+n+1} - \theta_j < \theta_{j-1} - \theta_j < 2\pi,$$

so that

$$-\pi < \frac{\theta_j - \theta_{j-1}}{2} < \frac{\theta_j - \theta_{j+n+1}}{2} < -\frac{\pi}{2} < \frac{\theta_j - \theta_{j+n-1}}{2} < 0.$$

The only possibility above is thus the first equation, so that

$$\frac{\theta_{n+j+1} - \theta_j}{2} - \pi = \frac{\theta_j - \theta_{j+n-1}}{2},$$

or $\theta_{j+n+1} + \theta_{j+n-1} = 2\theta_j + 2\pi$. The equation $\partial F/\partial\theta_j = 0$ now implies as well that $\theta_{j-1} + \theta_{j+1} = 2\theta_j + 2\pi$.

On the other hand, we have also assumed that $\theta_{j+n+1} - \theta_{j+1} = \pi$, so

$$\theta_{j+n-1} + \theta_{j+1} = 2\theta_j + 2\pi - \theta_{j+n+1} + \theta_{j+1} = 2\theta_j + \pi.$$

This implies that

$$\begin{aligned} \theta_{j-1} - \theta_{j+n-1} &= \theta_{j-1} - (2\theta_j + \pi - \theta_{j+1}) \\ &= \theta_{j-1} - \theta_2 - (2\theta_j + \pi) \\ &= (2\theta_j + 2\pi) - (2\theta_j + \pi) \\ &= \pi. \end{aligned}$$

We now know that if we make the normalizing assumptions $\theta_{j+n} = \theta_j + \pi$ and $\theta_{j+n+1} = \theta_{j+1} + \pi$, then the equations $\{\partial F/\partial\theta_j = 0, \partial F/\partial\theta_{j+n} = 0\}$ ensure $\theta_{j-1} = \theta_{j+n-1} + \pi$. Using all the equations $\{\partial F/\partial\theta_j = 0\}$, it is now evident that $\theta_{n+k} = \theta_k + \pi$, for each $k = 1, \dots, n$.

We apply this understanding to the equation $\partial F/\partial\theta_j = 0$:

$$\begin{aligned} \cot\frac{\theta_j - \theta_{j-1}}{2} + \cot\frac{\theta_j - \theta_{j+1}}{2} &= \cot\frac{\theta_j - \theta_{j+n-1}}{2} + \cot\frac{\theta_j - \theta_{j+n+1}}{2} \\ &= \cot\frac{\theta_j - \theta_{j-1} - \pi}{2} + \cot\frac{\theta_j - \theta_{j+1} - \pi}{2} \\ &= -\tan\frac{\theta_j - \theta_{j-1}}{2} - \tan\frac{\theta_j - \theta_{j+1}}{2}, \end{aligned}$$

so that

$$-\tan\frac{\theta_j - \theta_{j-1}}{2} - \cot\frac{\theta_j - \theta_{j-1}}{2} = \tan\frac{\theta_j - \theta_{j+1}}{2} + \cot\frac{\theta_j - \theta_{j+1}}{2}.$$

Since $\tan x + \cot x = 2/\sin 2x$, we obtain

$$\sin(\theta_j - \theta_{j-1}) = \sin(\theta_{j+1} - \theta_j).$$

If $\theta_j - \theta_{j-1} = \pi - (\theta_{j+1} - \theta_j)$, then $\theta_{j+1} - \theta_{j-1} = \pi$. However, $\theta_{j+n} = \theta_j + \pi$, and the θ_j are distinct. Thus $\theta_j - \theta_{j-1} = \theta_{j+1} - \theta_j$, for each $j = 1, \ldots, 2n$. Set $\theta_1 = 0$, and we see that $\bigl(1, e^{\pi i/n}, e^{2\pi i/n}, \ldots, e^{(2n-1)\pi i/n}\bigr)$ is the unique $\mathrm{Aut}(\mathbb{D})$–orbit for which the partial derivatives simultaneously vanish. As it is evident that $F(x_1, \ldots, x_{2n})$ goes to $+\infty$ as points x_j and x_{j+1} collide, the absolute minimum of F must occur at a simultaneous zero of its partial derivatives. Evaluating $F\bigl(1, e^{\pi i/n}, e^{2\pi i/n}, \ldots, e^{(2n-1)\pi i/n}\bigr)$ achieves the result. □

5 Bigons and triangles

Towards Theorem B, for the computation of the self-intersection number of a self-intersecting closed curve, we require a slight generalization of the "bigon criterion" of [5]. Recall that a representative λ of a collection of closed curves Γ is in *minimal position* if its intersection points are transverse, and the number of intersections of λ, counted with multiplicity, is minimal among representatives of Γ. A *monogon* is a polygon with one side and a *bigon* is a polygon with two sides.

Definition A representative λ of a collection of closed curves $\Gamma \subset \Sigma$ has an *immersed monogon* if there is an immersion of a monogon whose boundary arc is contained in λ, and it has an *immersed bigon* if there is such an immersion of a bigon.

Lemma 5.1 *If the representative λ of a collection of closed curves on Σ is without immersed monogons and without immersed bigons, then it is in minimal position.*

Remark An error in a previous version of this lemma was pointed out by Ian Biringer, as well as a reference to a very similar statement due to Hass and Scott. The corrected statement is above (see [6, Theorems 3.5 and 4.2]). As they note, a nonprimitive curve demonstrates that the converse is false [6, p. 94].

Proof Suppose λ is without immersed bigons or monogons, and has n transverse self-intersections. Let $G_\lambda \subset \Sigma$ indicate the graph determined by λ, choose a spanning tree T_λ for G_λ, a lift of T_λ to the universal cover, and a representative of each of the n intersection points. At each of these representative intersection points, the preimage of λ in $\widetilde{\Sigma}$ consists of a pair of linked curves: if there was only one curve the covering map would produce an immersed monogon for λ on Σ, and if the pair of curves at this intersection point were not linked the covering map would produce an immersed bigon for λ on Σ. The self-intersection number of Γ is equal to the number of $\pi_1\Sigma$-orbits of linked elevations of curves from Γ in the universal cover $\widetilde{\Sigma}$, so we are done. $\square$

In order to recognize the presence of hyperplane separated cubes in the dual cube complex of Γ, we will employ a lemma.

Lemma 5.2 *Suppose that Γ has a minimal position realization $\lambda \subset \Sigma$ such that λ has k points of transverse self-intersection of orders $n_1, \dots, n_k$, listed with multiplicity. Then $\mathcal{C}(\Gamma)$ has cubes of dimensions $n_1, \dots, n_k$, with multiplicity. Moreover, the $\pi_1\Sigma$–orbit of the union of these cubes is hyperplane separated for the action of $\pi_1\Sigma$ on $\mathcal{C}(\Gamma)$ if and only if the complement $\Sigma \setminus \lambda$ has no triangles.*

Proof Consider the preimage $\widetilde{\lambda} := \pi^{-1}\lambda \subset \widetilde{\Sigma}$, and choose lifts $p_1, \dots, p_k$ of the self-intersection points of λ, where p_i has order n_i. By hypothesis, there are n_i linked elevations from $\widetilde{\lambda}$ through p_i, so that there is an n_i–cube in $\mathcal{C}(\Gamma)$. We denote this n_i–cube corresponding to the choice of lift p_i by C_i.

If the complement $\Sigma \setminus \lambda$ had a triangle, then this triangle would lift to $\widetilde{\Sigma}$, so that the curves corresponding to C_i, for some i, would contain two sides of the lifted triangle. As a consequence, there would be a different lift p' of one of the intersection points, so that p' would also abut this triangle. Let C' indicate the maximal cube corresponding to the lift p'. By construction, C' shares a hyperplane with C_i, while there are two other hyperplanes, one of C_i and one of C', that intersect. As C' is in the same $\pi_1\Sigma$–orbit as C_j, for some j, the union of orbits $\bigcup \pi_1\Sigma \cdot C_i$ is not hyperplane separated.

Finally, suppose the union of orbits is not hyperplane separated. Then there is $g \in \pi_1\Sigma$ such that C_i and $g \cdot C_i$ share a hyperplane γ, and have a pair of other intersecting hyperplanes, say γ_1 and γ_2. In this case, there is a triangle $T \subset \widetilde{\Sigma}$ formed by γ, γ_1 and γ_2. While this triangle may not embed under the covering map, it contains an innermost triangle, namely a triangle in the complement of $\widetilde{\Sigma} \setminus \widetilde{\lambda}$. This triangle must embed under the covering map, so $\Sigma \setminus \lambda$ contains a triangle. $\square$

6 Closed curves from ribbon graphs

We now prove Theorem B, thus obtaining explicit constructions of curves to which Theorem A applies. Recall that a *ribbon graph* is a graph with a cyclic order given to the oriented edges incident to each vertex, and a ribbon graph is *even* if the valence of each edge is even. In what follows, we introduce notation for even ribbon graphs and analyze the closed curves that they determine.

Let $S(n)$ indicate the union of the n line segments

$$\{t\exp(\pi i m/n) \mid t \in [-1,1]\},$$

for $m = 1,\dots,n$, and label the endpoints $\exp(\pi i m/n)$ and $-\exp(\pi i m/n)$ by a_m and a'_m, respectively. Fix a permutation of endpoints μ by $\mu(a_m) = a'_m = \mu^{-1}(a_m)$, for $m = 1,\dots,n$. We refer to $S(n)$ as a *star*, and μ as the *switch* map of the star.

Let $\underline{n}$ be the tuple $(n_1,\dots,n_k)$, and consider the union $S(\underline{n}) := S(n_1) \sqcup \cdots \sqcup S(n_k)$. Let σ be a fixed-point-free, order-two permutation (that is, a pairing) of the set

$$\{a_{j,i}, a'_{j,i} \mid 1 \le j \le n_i, 1 \le i \le k\}.$$

Let $\Gamma(\underline{n},\sigma)$ indicate the graph given by

$$\Gamma(\underline{n},\sigma) := S(\underline{n})/\sim,$$

where $a_{j,i} \sim \sigma(a_{j,i})$. The vertices of $\Gamma(\underline{n},\sigma)$ are in bijection with the stars $S(n_j)$, and the orientation of $\mathbb{C}$ at 0 induces a cyclic order to the vertex contained in each $S(n_j)$. These orientations give $\Gamma(\underline{n},\sigma)$ the structure of an even ribbon graph. Moreover, it is clear that every even ribbon graph can be constructed in this way.

Let $\Sigma(\underline{n},\sigma)$ be the surface with boundary associated to the ribbon graph $\Gamma(\underline{n},\sigma)$. We identify $\Gamma(\underline{n},\sigma)$ as smoothly[3] and incompressibly embedded in $\Sigma(\underline{n},\sigma)$, so that the embedding induces isomorphisms on the level of fundamental groups. By a *closed curve* in $\Gamma(\underline{n},\sigma)$, we mean the free homotopy class of the image of a smooth immersion of S^1 into $\Gamma(\underline{n},\sigma)$.

Lemma 6.1 *Closed curves in $\Gamma(\underline{n},\sigma)$ are in correspondence with fixed cycles of $(\mu\sigma)^l$, for $l > 0$. The closed curves in $\Gamma(\underline{n},\sigma)$ are in minimal position in $\Sigma(\underline{n},\sigma)$, and the total intersection number of these closed curves is given by $\binom{n_1}{2} + \binom{n_2}{2} + \cdots + \binom{n_k}{2}$.*

Proof The first statement is evident. The second follows from Lemma 5.1, since the complement $\Sigma(\underline{n},\sigma) \setminus \Gamma(\underline{n},\sigma)$ contains no disks, and hence no immersed bigons or monogons. □

To obtain closed curves on closed surfaces, one may glue together $\Sigma(\underline{n},\sigma)$ and another (possibly disconnected) surface along its boundary. Some of the components that are glued may be disks, so it is possible that the curves from $\Gamma(\underline{n},\sigma)$ are no longer in minimal position. While Section 5 can be used to build an algorithm that can be applied on a case-by-case basis, a more straightforward control on this phenomenon can be obtained in many cases.

[3]Note that the smooth structure of $\Gamma(\underline{n},\sigma)$ in a neighborhood of its vertices is induced by viewing $S(n_j)$ as an immersed submanifold of $\mathbb{C}$.

Lemma 6.2 *Suppose that $\hat{\Sigma}$ is a surface obtained by a gluing of $\Sigma(\underline{n},\sigma)$, so that there is a natural inclusion $\Sigma(\underline{n},\sigma) \hookrightarrow \hat{\Sigma}$. If the complement $\hat{\Sigma} \setminus \Gamma(\underline{n},\sigma)$ contains no monogons, bigons, or triangles, then the closed curve $\Gamma(\underline{n},\sigma) \subset \hat{\Sigma}$ is in minimal position.*

Proof Suppose that $\Gamma(\underline{n},\sigma) \subset \hat{\Sigma}$ is not in minimal position, so by Lemma 5.1 it has either an immersed monogon or bigon. Suppose that $\phi\colon B \hookrightarrow \hat{\Sigma}$ is an example of the latter. By assumption, the bigon is not embedded. Thus $\phi^{-1}(\Gamma(\underline{n},\sigma))$ consists of the two sides of B, together with some connected arcs that connect opposite sides of the bigon B. It is easy to see by induction on the number of such arcs that the complement in B must contain either a bigon or a triangle. This triangle embeds under ϕ, violating the assumption that $\hat{\Sigma} \setminus \Gamma(\underline{n},\sigma)$ contains no triangles. The case of an immersed monogon is straightforwardly similar. □

Lemmas 6.1, 6.2 and 5.2 imply Theorem B directly.

7 A family of examples and Theorem C

Towards Theorem C, for each k let $\underline{\tau}_k$ indicate the sequence $(3,\dots,3)$ with k terms. The vertices of $S(\underline{\tau}_k)$ are given by

$$\{a_{1,j}, a_{2,j}, a_{3,j}, a'_{1,j}, a'_{2,j}, a'_{3,j} \mid 1 \le j \le k\}.$$

Let σ indicate the following pairing:

$$\begin{aligned} &a_{2,j} \leftrightarrow a'_{1,j}, \\ &a_{3,j} \leftrightarrow a'_{2,j} && \text{for } j = 1,\dots,k, \\ &a_{1,j} \leftrightarrow a'_{3,j+1} && \text{for } j = 1,\dots,k-1, \\ &a_{1,k} \leftrightarrow a'_{3,1}. \end{aligned}$$

See Figure 3 for a schematic picture of $\Gamma(\underline{\tau}_k,\sigma)$, and Figure 4 for a gluing of $\Sigma(\underline{\tau}_6,\sigma)$ to which Lemma 5.2 and Lemma 6.2 apply.

Proof of Theorem C Let Σ indicate the surface $\Sigma(\underline{\tau}_k,\sigma)$, so that Σ contains an embedded minimal position copy of the curve $\Gamma(\underline{\tau}_k,\sigma)$, with the self-intersection number $3k$ by Lemma 6.1.

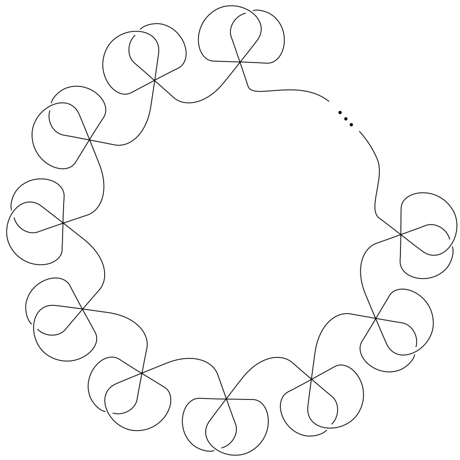


Figure 3: The ribbon graph $\Gamma(\underline{\tau}_k, \sigma)$

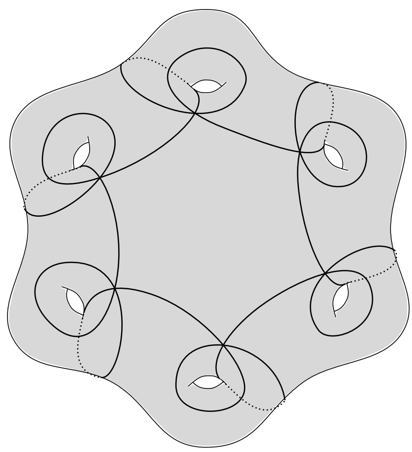


Figure 4: A gluing of $\Sigma(\underline{\tau}_6, \sigma)$ without triangles or bigons in the complement of $\Gamma(\underline{\tau}_6, \sigma)$, such that the given closed curve has six hyperplane separated 3–cubes

By Lemma 5.2, the dual cube complex of $\Gamma(\underline{\tau}_k, \sigma)$ in Σ contains k hyperplane separated 3–cubes. Using Theorem A, we may estimate

$$\begin{aligned}\limsup_{k\to\infty} \frac{\overline{M}_k}{k} &\geq \limsup_{k\to\infty} \frac{\overline{M}_{3k}}{3k} \\ &\geq \limsup_{k\to\infty} \frac{1}{3k} \inf\{\ell(\Gamma(\underline{\tau}_k, \sigma), X) \mid X \in \mathcal{T}(\Sigma)\} \\ &\geq \limsup_{k\to\infty} \frac{1}{3k} k \log\left(\frac{1+\cos\pi/3}{1-\cos\pi/3}\right) = \tfrac{1}{3}\log 3. \qquad \square\end{aligned}$$

Note that in the construction above, it is evident that the genus of $\Sigma(\underline{\tau}_k, \sigma)$ will grow with the self-intersection number of $\Gamma(\underline{\tau}_k, \sigma)$. As a consequence, these lower bounds are not applicable to $\overline{m}_k(\Sigma)$ for a fixed surface Σ. It seems likely[4] that $\overline{m}_k(\Sigma)$ grows as $\sqrt{k}$.

References

[1] **T Aougab**, **J Gaster**, *Curves intersecting exactly once and their dual cube complexes* (2016) arXiv To appear in Groups Geom. Dyn.

[2] **T Aougab**, **J Gaster**, **P Patel**, **J Sapir**, *Building hyperbolic metrics suited to closed curves and applications to lifting simply*, preprint (2016) arXiv

[3] **A Basmajian**, *Universal length bounds for non-simple closed geodesics on hyperbolic surfaces*, J. Topol. 6 (2013) 513–524 MR

[4] **I Chatterji**, **G Niblo**, *From wall spaces to* CAT(0) *cube complexes*, Internat. J. Algebra Comput. 15 (2005) 875–885 MR

[5] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton University Press (2012) MR

[6] **J Hass**, **P Scott**, *Intersections of curves on surfaces*, Israel J. Math. 51 (1985) 90–120 MR

[7] **M Sageev**, *Ends of group pairs and non-positively curved cube complexes*, Proc. London Math. Soc. 71 (1995) 585–617 MR

[8] **M Sageev**, CAT(0) *cube complexes and groups*, from "Geometric group theory" (M Bestvina, M Sageev, K Vogtmann, editors), IAS/Park City Math. Ser. 21, Amer. Math. Soc., Providence, RI (2014) 7–54 MR

[9] **D T Wise**, *Subgroup separability of graphs of free groups with cyclic edge groups*, Q. J. Math. 51 (2000) 107–129 MR

Department of Mathematics, Boston College
140 Commonwealth Avenue, Chestnut Hill, MA 02467, United States

gaster@bc.edu

http://www2.bc.edu/jonah-gaster

Received: 11 January 2016 Revised: 21 June 2016

[4] While this paper was under review, this has been shown in [2, Theorem 1.4].

msp

Algebraic & Geometric Topology 17 (2017) 1059–1104

Two-complete stable motivic stems over finite fields

GLEN MATTHEW WILSON
PAUL ARNE ØSTVÆR

Let ℓ be a prime and $q = p^\nu$, where p is a prime different from ℓ. We show that the ℓ–completion of the n^{th} stable homotopy group of spheres is a summand of the ℓ–completion of the $(n, 0)$ motivic stable homotopy group of spheres over the finite field with q elements, $\mathbb{F}_q$. With this, and assisted by computer calculations, we are able to explicitly compute the two-complete stable motivic stems $\pi_{n,0}(\mathbb{F}_q)^\wedge_2$ for $0 \le n \le 18$ for all finite fields and $\pi_{19,0}(\mathbb{F}_q)^\wedge_2$ and $\pi_{20,0}(\mathbb{F}_q)^\wedge_2$ when $q \equiv 1 \bmod 4$ assuming Morel's connectivity theorem for $\mathbb{F}_q$ holds.

16-04, 14F42, 18G15, 55T15

1 Introduction

The homotopy groups of spheres belong to the most important and puzzling invariants in topology. See Kochman [28] and the more recent works of Isaksen [26] and Wang and Xu [51] for amazing computer-assisted ways of computing these invariants based on the Adams spectral sequence. The Adams spectral sequence of topology is a well-studied method to calculate the stable homotopy groups of spheres; see Adams [2] and Ravenel [40]. With two-primary coefficients, the second page of the Adams spectral sequence has a description in terms of Ext groups over the mod 2 Steenrod algebra

$$E_2^{s,t} = \mathrm{Ext}^{s,t}_{\mathcal{A}^*}(\mathbb{F}_2, \mathbb{F}_2)$$

and converges to the two-complete stable homotopy groups of spheres $(\pi_n^s)^\wedge_2$. Extensive computer calculations of these Ext groups have been carried out by Bruner in [8] and [10]. However, even if one knew completely the answer for the Ext groups in the Adams spectral sequence, one is still not finished with computing the stable homotopy groups of spheres. One needs to know in addition the differentials and all the group extensions hidden in the associated graded groups of the filtration. Only partial results have been obtained in spite of an enormous effort.

Given any field k the stable motivic homotopy category $\mathcal{SH}_k$ over k has the structure of a triangulated category and encodes both topological information and arithmetic

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.1059

information about k. An application of this framework is the proof of Milnor's conjecture on Galois cohomology given by Voevodsky [48]. Just as for the stable homotopy category $\mathcal{SH}$, it is an interesting and deep problem to compute the stable motivic homotopy groups of spheres $\pi_{s,w}(k)$ over k, that is, $\mathcal{SH}_k(\Sigma^{s,w}\mathbb{1},\mathbb{1})$, where $\mathbb{1}$ denotes the motivic sphere spectrum over k. When k has finite mod 2 cohomological dimension and $s \geq w \geq 0$, the motivic Adams spectral sequence (MASS) converges to the two-completion of the stable motivic stems:

$$E_2^{f,(s,w)} = \mathrm{Ext}_{\mathcal{A}^{**}}^{f,(s+f,w)}(H^{**}, H^{**}) \Longrightarrow (\pi_{s,w}\mathbb{1})_2^\wedge.$$

This is a trigraded spectral sequence, where $\mathcal{A}^{**}$ is the bigraded mod 2 motivic Steenrod algebra (see the work of Hoyois, Kelly and Østvær [22] and Voevodsky [48]), and H^{**} is the bigraded mod 2 motivic cohomology ring of k. A construction of the motivic Adams spectral sequence is given in Section 5. The calculational challenges are (1) to identify the motivic Ext groups, (2) to determine the differentials and (3) to reconstruct the abutment from the filtration quotients.

Based on the MASS, Dugger and Isaksen [12] have carried out calculations of the two-complete stable motivic homotopy groups of spheres up to the 34 stem over the complex numbers. Isaksen [25; 26] has extended this work largely up to the 70 stem. We are led to wonder: how do the stable motivic homotopy groups vary for different base fields? Morel [34] has given a complete description of the 0–line $\pi_{n,n}(k)$ in terms of Milnor–Witt K–theory. The 1–line $\pi_{n+1,n}(k)$ is determined by Hermitian and Milnor K–theory groups by the work of Röndigs, Spitzweck and Østvær [41], which generalizes the partial results obtained by Ormsby and Østvær in [39]. Ormsby has investigated the case of related invariants over p–adic fields in [37] and the rationals in Ormsby and Østvær [38], and Dugger and Isaksen [13] have analyzed the case over the real numbers. It is now possible to perform similar calculations over fields of positive characteristic, thanks to work on the motivic Steenrod algebra in positive characteristic by Hoyois, Kelly and Østvær [22]. In this paper we use computer-assisted motivic Ext group calculations in tandem with theoretical arguments to determine stable motivic stems $\pi_{n,0}$ in weight zero over finite fields.

We now state our main results. For a prime ℓ and an abelian group G, we write the ℓ–completion of G by $G_\ell^\wedge$.

Theorem 1.1 *Let $\overline{F}$ be an algebraically closed field of positive characteristic p. For all $s \geq w \geq 0$ or $s < w$, there are isomorphisms $\pi_{s,w}(\overline{F})[p^{-1}] \cong \pi_{s,w}(\mathbb{C})[p^{-1}]$.*

Proof By Proposition 5.14, when $s > w \geq 0$, the groups $\pi_{s,w}(\overline{F})$ and $\pi_{s,w}(\mathbb{C})$ are torsion. The isomorphism $\pi_{s,w}(\overline{F})[p^{-1}] \cong \pi_{s,w}(\mathbb{C})[p^{-1}]$ follows when $s > w \geq 0$

from Theorem 1.3 by summing up the ℓ–primary parts. When $s = w \geq 0$ the result follows by Morel's identification of the 0–line in [34]. If $s < w$ then Morel's connectivity theorem implies that both groups are trivial by Corollary 2.14. □

Let π_n^s denote the n^{th} topological stable stem. Over the complex numbers, Levine [29, Corollary 2] showed there is an isomorphism $\pi_n^s \cong \pi_{n,0}(\mathbb{C})$. We obtain a similar result over any algebraically closed field of positive characteristic p after inverting p.

Corollary 1.2 *Let $\overline{F}$ be an algebraically closed field of positive characteristic p. For all $n \geq 0$ the homomorphism $\mathbb{L}c\colon (\pi_n^s)[p^{-1}] \to \pi_{n,0}(\overline{F})[p^{-1}]$ is an isomorphism.*

We do not expect Levine's theorem to hold over a field which is not algebraically closed. Write $\mathbb{F}_q$ for the finite field with $q = p^\nu$ elements where p is a prime and $\widetilde{\mathbb{F}}_q$ for the union of the field extensions $\mathbb{F}_{q^i}$ over $\mathbb{F}_q$ with i odd. In this paper, we will see how the groups $\pi_{n,0}(\mathbb{F}_q)$ differ from π_n^s using motivic Adams spectral sequence calculations. Corollary 1.2 allows us to identify differentials in the mod 2 motivic Adams spectral sequence over a finite field and identify the two-complete groups $\pi_{n,0}(\mathbb{F}_q)^\wedge_2$ in a range. The analogous calculations with the mod ℓ motivic Adams spectral sequence for ℓ an odd prime are given by Wilson in [52]. The groups take the following form.

Theorem 1.3 *If Morel's connectivity theorem holds for the finite field $\mathbb{F}_q$, then for all $0 \leq n \leq 18$ there is an isomorphism*

$$\pi_{n,0}(\mathbb{F}_q)[p^{-1}] \cong (\pi_n^s \oplus \pi_{n+1}^s)[p^{-1}].$$

In particular, the group $\pi_{4,0}(\mathbb{F}_q)[p^{-1}]$ is trivial.

Proof Propositions 7.15 and 7.18 calculate the two-completion of $\pi_{n,0}(\mathbb{F}_q)$ for n satisfying $0 \leq n \leq 18$. For primes $\ell \neq 2$, the calculations are similar and given by Wilson in [52, Sections 6 and 7]. The ℓ–completions of $\pi_{n,0}(\mathbb{F}_q)$ are shown to agree with the ℓ–primary part of $\pi_{n,0}(\mathbb{F}_q)$ for $n > 0$ in Proposition 5.14. When $n = 0$, the result follows by Morel's identification of $\pi_{0,0}(\mathbb{F}_q)$ with the Grothendieck–Witt ring of $\mathbb{F}_q$, since $\mathrm{GW}(\mathbb{F}_q) \cong \mathbb{Z} \oplus \mathbb{Z}/2$, as shown by Scharlau in [42, Chapter 2, Section 3.3]. □

Remark 1.4 The above theorem depends on Morel's connectivity theorem to prove that the motivic Adams spectral sequence converges to the homotopy groups of the ℓ–completion of the sphere spectrum. The published proof of the theorem by Morel in [34] holds for infinite fields. A private message from Panin gives a new proof of Morel's connectivity theorem which is valid for finite fields. We therefore state our results under the assumption that Morel's connectivity theorem holds for finite fields.

However, our argument for Theorem 1.3 goes through with the field $\mathbb{F}_q$ replaced by $\widetilde{\mathbb{F}}_q$, where Morel's connectivity theorem holds by Proposition 7.22. The uneasy reader may replace $\mathbb{F}_q$ with $\widetilde{\mathbb{F}}_q$ throughout.

In the case of a finite field $\mathbb{F}_q$ with $q \equiv 3 \bmod 4$, we use the ρ–Bockstein spectral sequence to identify the additive structure of the E_2 page of the MASS. Some hidden products in the ρ–Bockstein spectral sequence were identified with the help of computer calculations by Fu and Wilson, which can be found in [16].

It is interesting to note that the pattern $\pi_{n,0}(\mathbb{F}_q)^\wedge_2 \cong (\pi_n^s \oplus \pi_{n+1}^s)^\wedge_2$ obtained in Theorem 1.3 does not hold in general. We show that if $q \equiv 1 \bmod 4$, then

$$\pi_{19,0}(\mathbb{F}_q)^\wedge_2 \cong (\mathbb{Z}/8 \oplus \mathbb{Z}/2) \oplus \mathbb{Z}/4 \quad \text{and} \quad \pi_{20,0}(\mathbb{F}_q)^\wedge_2 \cong \mathbb{Z}/8 \oplus \mathbb{Z}/2.$$

We shall leave open for further investigations the question of whether or not an isomorphism $\pi_{n,0}(\mathbb{F}_q)^\wedge_2 \cong (\pi_n^s \oplus \pi_{n+1}^s)^\wedge_2$ holds when $q \equiv 3 \bmod 4$ and $n = 19, 20$.

Acknowledgements The work on this paper was initiated during the program "A Special Semester in Motivic Homotopy Theory" organized by Marc Levine in 2014; we gratefully acknowledge the hospitality and support from Universität Duisburg-Essen.

We thank the referee for several valuable suggestions and pointing out an error in our calculation of the 19 and 20 stems. Thanks go to Panin for private correspondence on Morel's stable connectivity theorem. Wilson would like to thank Charles Weibel, Dan Isaksen and Kyle Ormsby for helpful conversations regarding this work. This paper constitutes a portion of Wilson's PhD thesis. Østvær gratefully acknowledges support from the RCN projects "Special Geometries", no. 239015 and "Motivic Hopf Equations", no. 250399.

2 The stable motivic homotopy category

We first sketch a construction of the stable motivic homotopy category that will be convenient for our purposes, and in the process, set our notation. Treatments of stable motivic homotopy theory can be found in Voevodsky [47], Jardine [27], Hu [23], Dundas, Röndigs and Østvær [15], Morel [32], Ayoub [4] and the Nordfjordeid lectures [14].

2.1 The unstable motivic homotopy category

A base scheme S is a Noetherian separated scheme of finite Krull dimension. We write Sm/S for the category of smooth schemes of finite type over S. A space over S is a simplicial presheaf on Sm/S. The collection of spaces over S forms the category $\mathrm{Spc}(S)$, where morphisms are natural transformations of functors. We write $\mathrm{Spc}_*(S)$ for the category of pointed spaces.

The first model category structure we endow $\mathrm{Spc}(S)$ with is the projective model structure; see, for example, Blander [5, Theorem 1.4], Dundas, Röndigs and Østvær [15, Theorem 2.7], Hirschhorn [19, Theorem 11.6.1].

Definition 2.1 A map $f\colon X \to Y$ in $\mathrm{Spc}(S)$ is a (global) weak equivalence if for any $U \in \mathrm{Sm}/S$ the map $f(U)\colon X(U) \to Y(U)$ of simplicial sets is a weak equivalence. The projective fibrations are those maps $f\colon X \to Y$ for which $f(U)\colon X(U) \to Y(U)$ is a Kan fibration for any $U \in \mathrm{Sm}/S$. The projective cofibrations are those maps in $\mathrm{Spc}(S)$ which satisfy the left lifting property for trivial projective fibrations. The projective model structure on $\mathrm{Spc}(S)$ consists of the global weak equivalences, the projective fibrations and the projective cofibrations.

The category $\mathrm{Spc}(S)$ equipped with the projective model structure is cellular, proper and simplicial; see Blander [5, Theorem 1.4]. Furthermore, $\mathrm{Spc}(S)$ has the structure of a simplicial monoidal model category, with product $\times$ and internal hom $\underline{\mathrm{Hom}}$.

The constant presheaf functor $c\colon \underline{\mathrm{sSet}} \to \mathrm{Spc}(S)$ associates to a simplicial set A the presheaf cA defined by $cA(U) = A$ for any $U \in \mathrm{Sm}/S$. The functor c is a left Quillen functor when $\mathrm{Spc}(S)$ is equipped with the projective model structure. Its right adjoint $\mathrm{Ev}_S\colon \mathrm{Spc}(S) \to \underline{\mathrm{sSet}}$ satisfies $\mathrm{Ev}_S(X) = X(S)$. One can show that representable presheaves and constant presheaves in $\mathrm{Spc}(S)$ are cofibrant in the projective model structure.

For a smooth scheme X over S, we write h_X for the representable presheaf of simplicial sets. We will occasionally abuse notation and write X for h_X. Although the representable presheaf functor embeds Sm/S into $\mathrm{Spc}(S)$, colimits which exist in Sm/S are not necessarily preserved in $\mathrm{Spc}(S)$. That is, if $X = \operatorname{colim} X_i$ in Sm/S, it need not be true that $h_X = \operatorname{colim} h_{X_i}$, for example, $\operatorname{colim}(h_{\mathbb{A}^1} \leftarrow h_{\mathbb{G}_m} \to h_{\mathbb{A}^1}) \neq h_{\mathbb{P}^1}$, as one can check by applying the Picard group functor. To fix this, one introduces the Nisnevich topology on Sm/S.

Morel and Voevodsky proved in [35, Section 3, Proposition 1.4] that the Nisnevich topology is generated by covers coming from the elementary distinguished squares. Recall that an elementary distinguished square is a pull-back square in Sm/S

$$\begin{tikzcd} V' \arrow[r] \arrow[d] & X' \arrow[d, "f"] \\ V \arrow[r, "j"] & X \end{tikzcd}$$

for which f is an étale map, j is an open embedding and $f^{-1}(X - V) \to X - V$ is an isomorphism, where these subschemes are given the reduced structure. Hence a presheaf of sets F on Sm/S is a Nisnevich sheaf if and only if for any elementary distinguished square the resulting square after applying F is a pull-back square.

Definition 2.2 For a pointed space $\mathcal{X}$ and $n \geq 0$, the n^{th} simplicial homotopy sheaf $\pi_n\mathcal{X}$ of $\mathcal{X}$ is the Nisnevich sheafification of the presheaf $U \mapsto \pi_n(\mathcal{X}(U))$.

Write W_{Nis} for the class of maps $f\colon \mathcal{X} \to \mathcal{Y}$ for which $f_*\colon \pi_n\mathcal{X} \to \pi_n\mathcal{Y}$ is an isomorphism of Nisnevich sheaves for all $n \geq 0$. The Nisnevich local model structure on $\mathrm{Spc}_*(S)$ is the left Bousfield localization of the projective model structure with respect to W_{Nis}.

Definition 2.3 Let $W_{\mathbb{A}^1}$ be the class of maps $\pi_X\colon (X \times \mathbb{A}^1)_+ \to X_+$ for $X \in \mathrm{Sm}/S$. The motivic model structure on $\mathrm{Spc}_*(S)$ is the left Bousfield localization of the projective model structure with respect to $W_{\mathrm{Nis}} \cup W_{\mathbb{A}^1}$. We write $\mathrm{Spc}_*^{\mathbb{A}^1}(S)$ for the category of pointed spaces equipped with the motivic model structure. The homotopy category of $\mathrm{Spc}_*^{\mathbb{A}^1}(S)$ is the pointed motivic homotopy category $\mathcal{H}_*^{\mathbb{A}^1}(S)$.

For pointed spaces $\mathcal{X}$ and $\mathcal{Y}$, we write $[\mathcal{X},\mathcal{Y}]$ for the set of maps $\mathcal{H}_*^{\mathbb{A}^1}(S)(\mathcal{X},\mathcal{Y})$. The n^{th} motivic homotopy sheaf of a pointed space $\mathcal{X}$ over S is the sheaf $\pi_n\mathcal{X}$ associated to the presheaf $U \mapsto [S^n \wedge U_+, \mathcal{X}]$.

There are two circles in the category of pointed spaces: the constant simplicial presheaf S^1 pointed at its 0–simplex and the representable presheaf $\mathbb{G}_m = \mathbb{A}^1 \setminus \{0\}$ pointed at 1. These determine a bigraded family of spheres $S^{i,j} = (S^1)^{\wedge i-j} \wedge \mathbb{G}_m^{\wedge j}$.

Definition 2.4 For a pointed space X over S and natural numbers i and j with $i \geq j$, write $\pi_{i,j}X$ for the set of maps $[S^{i,j}, X]$.

The category of pointed spaces $\mathrm{Spc}_*(S)$ equipped with the induced motivic model category structure has many good properties which make it amenable to Bousfield localization. In particular, $\mathrm{Spc}_*(S)$ is closed symmetric monoidal, pointed simplicial, left proper and cellular.

2.2 The stable Nisnevich local model structure

With the unstable motivic model category in hand, we now construct the stable motivic model category using the general framework laid out by Hovey in [20].

Let T be a cofibrant replacement of $\mathbb{A}^1/\mathbb{A}^1 - \{0\}$. Morel and Voevodsky have shown that T is weakly equivalent to $S^{2,1}$ in $\mathrm{Spc}_*^{\mathbb{A}^1}(S)$ [35, Section 3, Proposition 2.15]. The functor $T \wedge -$ on $\mathrm{Spc}_*^{\mathbb{A}^1}(S)$ is a left Quillen functor, and we may invert it by creating a category of T–spectra.

Definition 2.5 A T–spectrum X is a sequence of spaces $X_n \in \mathrm{Spc}_*^{\mathbb{A}^1}(S)$ equipped with structure maps $\sigma_n\colon T \wedge X_n \to X_{n+1}$. A map of T–spectra $f\colon X \to Y$ is a collection of maps $f_n\colon X_n \to Y_n$ which are compatible with the structure maps. We write $\mathrm{Spt}_T(S)$ for the category of T–spectra of spaces.

To start, the level model category structure on $\mathrm{Spt}_T(S)$ is defined by declaring a map $f\colon X \to Y$ to be a weak equivalence (respectively fibration) if every map $f_n\colon X_n \to Y_n$ is a weak equivalence (respectively fibration) in the motivic model structure on $\mathrm{Spc}_*(S)$. The cofibrations for the level model structure are determined by the left lifting property for trivial level fibrations.

Definition 2.6 Let X be a T–spectrum. For integers i and j, the (i,j) stable homotopy sheaf of X, written as $\pi_{i,j}X$, is the Nisnevich sheafification of the presheaf $U \mapsto \mathrm{colim}_n\, \pi_{i+2n,j+n}X_n(U)$. A map $f\colon X \to Y$ is a stable weak equivalence if for all integers i and j the induced maps $f_*\colon \pi_{i,j}X \to \pi_{i,j}Y$ are isomorphisms.

Definition 2.7 The stable model structure on $\mathrm{Spt}_T(S)$ is the model category where the weak equivalences are the stable weak equivalences and the cofibrations are the cofibrations in the level model structure. The fibrations are those maps with the right lifting property with respect to trivial cofibrations. We write $\mathcal{SH}_S$ for the homotopy category of $\mathrm{Spt}_T(S)$ equipped with the stable model structure.

The stable model structure on $\mathrm{Spt}_T(S)$ can be realized as a left Bousfield localization of the levelwise model structure, as defined by Hovey [20, Definition 3.3].

Just as for the category Spt_{S^1} of simplicial S^1–spectra, there is not a symmetric monoidal category structure on $\mathrm{Spt}_T(S)$ which lifts the smash product $\wedge$ in $\mathcal{SH}_S$. One remedy is to use a category of symmetric T–spectra $\mathrm{Spt}_T^{\Sigma}(S)$. The construction of this category is given by Hovey in [20, Definition 8.7] and Jardine in [27]. It is proven in [20, Theorem 9.1] that there is a zig-zag of Quillen equivalences from $\mathrm{Spt}_T^{\Sigma}(S)$ to $\mathrm{Spt}_T(S)$, hence $\mathcal{SH}_S$ is equivalent to the homotopy category of $\mathrm{Spt}_T^{\Sigma}(S)$ as well. Since Quillen equivalences induce equivalences of homotopy categories, the category $\mathcal{SH}_S$ is a symmetric monoidal triangulated category with shift functor $[1] = S^{1,0} \wedge -$.

Definition 2.8 If E is a T–spectrum over S, write $\pi_{i,j}E$ for $\mathcal{SH}_S(\Sigma^{i,j}\mathbb{1}, E)$. In the case where $E = \mathbb{1}$ and $S = \mathrm{Spec}(R)$ for a ring R, we simply write $\pi_{i,j}(R)$ for $\mathcal{SH}_S(\Sigma^{i,j}\mathbb{1}, \mathbb{1})$.

In addition to the category of T–spectra, we will find it convenient to work with the category of $(\mathbb{G}_m, S^1)$–bispectra; see Jardine [27] or the Nordfjordeid lectures [14].

Definition 2.9 Consider the simplicial circle S^1 as a space over S given by the constant presheaf. An S^1–spectrum over S is a sequence of spaces $X_n \in \mathrm{Spc}_*(S)$

equipped with structure maps $\sigma_n\colon S^1 \wedge X_n \to X_{n+1}$. A map of S^1–spectra over S is a sequence of maps $f_n\colon X_n \to Y_n$ that are compatible with the structure maps. The collection of S^1–spectra over S with compatible maps between them forms a category $\mathrm{Spt}_{S^1}(S)$.

First equip $\mathrm{Spt}_{S^1}(S)$ with the level model structure with respect to the Nisnevich local model structure on $\mathrm{Spc}_*(S)$. The n^{th} stable homotopy sheaf of an S^1–spectrum E over S is the Nisnevich sheaf $\pi_n E = \operatorname{colim} \pi_{n+j} E_j$. A map $f\colon E \to F$ of S^1–spectra over S is a simplicial stable weak equivalence if for all $n \in \mathbb{Z}$ the induced map $f_*\colon \pi_n E \to \pi_n F$ is an isomorphism of sheaves. The stable Nisnevich local model category structure on $\mathrm{Spt}_{S^1}(S)$ is obtained by localizing at the class of simplicial stable equivalences, as in Definition 2.7.

The motivic stable model category structure on $\mathrm{Spt}_{S^1}(S)$ is obtained from the simplicial stable model category structure by left Bousfield localization at the class of maps $W_{\mathbb{A}^1} = \{\Sigma^\infty X_+ \wedge \mathbb{A}^1 \to \Sigma^\infty X_+ \mid X \in \mathrm{Sm}/S\}$. Write $\mathrm{Spt}_{S^1}^{\mathbb{A}^1}(S)$ for the motivic stable model category $L_{W_{\mathbb{A}^1}} \mathrm{Spt}_{S^1}(S)$ and write $\mathcal{SH}_{S^1}^{\mathbb{A}^1}(S)$ for its homotopy category. The n^{th} motivic stable homotopy sheaf of an S^1–spectrum E is the Nisnevich sheaf $\pi_n^{\mathbb{A}^1} E$ associated to the presheaf $U \mapsto \mathcal{SH}_{S^1}^{\mathbb{A}^1}(S^n \wedge \Sigma^\infty U_+, E)$.

Definition 2.10 In the projective model structure on $\mathrm{Spc}_*(S)$, the space $\mathbb{G}_m$ pointed at 1 is not cofibrant. We abuse notation and write $\mathbb{G}_m$ for a cofibrant replacement of $\mathbb{G}_m$. A $(\mathbb{G}_m, S^1)$–bispectrum over S is a $\mathbb{G}_m$–spectrum of S^1–spectra. We write $\mathrm{Spt}_{\mathbb{G}_m, S^1}(S)$ for the category of $(\mathbb{G}_m, S^1)$–bispectra over S. Viewing $\mathrm{Spt}_{\mathbb{G}_m, S^1}(S)$ as the category of $\mathbb{G}_m$–spectra of S^1–spectra, we first equip $\mathrm{Spt}_{\mathbb{G}_m, S^1}(S)$ with the level model category structure with respect to the motivic stable model category structure on $\mathrm{Spt}_{S^1}(S)$. The motivic stable model category structure on $\mathrm{Spt}_{\mathbb{G}_m, S^1}(S)$ is the left Bousfield localization at the class of stable equivalences.

There are left Quillen functors

$$\Sigma^\infty_{S^1}\colon \mathrm{Spc}_*(S) \to \mathrm{Spt}_{S^1}(S) \quad \text{and} \quad \Sigma^\infty_{\mathbb{G}_m}\colon \mathrm{Spt}_{S^1}(S) \to \mathrm{Spt}_{\mathbb{G}_m, S^1}(S).$$

Additionally, the category $\mathrm{Spt}_{\mathbb{G}_m, S^1}(S)$ equipped with the motivic stable model structure is Quillen equivalent to the stable model category structure on $\mathrm{Spt}_T(S)$; see the Nordfjordeid lectures [14, page 216].

Definition 2.11 To any spectrum of simplicial sets $E \in \mathrm{Spt}_{S^1}$ we may associate the constant S^1–spectrum cE over S with value E. That is, cE is the sequence of spaces cE_n with the evident bonding maps. For a simplicial spectrum E, we also write cE for the $(\mathbb{G}_m, S^1)$–bispectrum $\Sigma^\infty_{\mathbb{G}_m} cE$. This defines a left Quillen functor $c\colon \mathrm{Spt}_{S^1} \to \mathrm{Spt}_{\mathbb{G}_m, S^1}(B)$ with right adjoint given by evaluation at S. Compare with Levine [29, Lemma 6.5].

2.3 Base change of stable model categories

Definition 2.12 Let $f\colon R \to S$ be a map of base schemes. Pull-back along f determines a functor $f^{-1}\colon \mathrm{Sm}/S \to \mathrm{Sm}/R$ which induces Quillen adjunctions

$$(f^*, f_*)\colon \mathrm{Spc}_*^{\mathbb{A}^1}(S) \to \mathrm{Spc}_*^{\mathbb{A}^1}(R) \quad \text{and} \quad (f^*, f_*)\colon \mathrm{Spt}_T(S) \to \mathrm{Spt}_T(R).$$

We now discuss some of the properties of base change. A more thorough treatment is given by Morel in [33, Section 5]. The map f_* sends a space $\mathcal{X}$ over R to the space $\mathcal{X} \circ f^{-1}$ over S. The adjoint f^* is given by the formula $(f^*\mathcal{Y})(U) = \mathrm{colim}_{U \to f^{-1}V}\, \mathcal{Y}(V)$. For a smooth scheme X over S, a standard calculation shows $f^*X = f^{-1}X$. Additionally, if cA is a constant simplicial presheaf on Sm/S, it follows that $f^*(cA) = cA$.

The Quillen adjunction (f^*, f_*) extends to both the model category of T–spectra and $(\mathbb{G}_m, S^1)$–bispectra by applying the maps f^*, and respectively f_*, termwise to a given spectrum. In the case of f^* for T–spectra, for instance, the bonding maps of f^*E are given by $T \wedge f^*E_n \cong f^*(T \wedge E_n) \to f^*(E_{n+1})$ as $f^*T = T$. The same reasoning shows that the adjunction (f^*, f_*) extends to $(\mathbb{G}_m, S^1)$–bispectra.

Write Q (respectively R) for the cofibrant (respectively fibrant) replacement functor in $\mathrm{Spt}_T(S)$. The derived functors $\mathbb{L} f^*$ and $\mathbb{R} f_*$ are given by the formulas $\mathbb{L} f^* = f^*Q$ and $\mathbb{R} f_* = f_*R$.

Let $f\colon C \to B$ be a smooth map. The functor $f_\#\colon \mathrm{Sm}/C \to \mathrm{Sm}/B$ sends $\alpha\colon X \to C$ to $f \circ \alpha\colon X \to B$ and, by restricting a presheaf on Sm/B to a presheaf on Sm/C, induces a functor $f_\#\colon \mathrm{Spc}_*^{\mathbb{A}^1}(B) \to \mathrm{Spc}_*^{\mathbb{A}^1}(C)$. The functor f^* is canonically equivalent to $f_\#$ on the level of spaces and spectra.

2.4 The connectivity theorem

Morel establishes the connectivity of the sphere spectrum over fields F by studying the effect of Bousfield localization at $W_{\mathbb{A}^1}$ of the stable Nisnevich local model category structure on $\mathrm{Spt}_{S^1}(F)$ (see Definition 2.9).

An S^1–spectrum E over S is said to be simplicially k–connected if for any $n \le k$, the simplicial stable homotopy sheaves $\pi_n E$ are trivial. An S^1–spectrum E is k–connected if for all $n \le k$ the motivic stable homotopy sheaves $\pi_n^{\mathbb{A}^1} E$ are trivial.

Theorem 2.13 (Morel's connectivity theorem) *If E is a simplicially k–connected S^1–spectrum over an infinite field F, then E is also k–connected.*

Morel's connectivity theorem has been proven when F is an infinite field in [34], but the argument there does not hold for finite fields. Private correspondence with Panin gives a new argument to prove Morel's connectivity theorem for finite fields as well.

The connectivity theorem along with the work of Morel in [32, Section 5] yield the following. This also follows from Voevodsky [47, Theorem 4.14].

Corollary 2.14 *Over a field F where Morel's connectivity theorem holds, the sphere spectrum $\mathbb{1}$ is (-1)–connected. In particular, for all $s-w<0$ the groups $\pi_{s,w}(F)$ are trivial.*

3 Comparison to the stable homotopy category

The following result of Levine is crucial for our calculations [29, Theorem 1].

Theorem 3.1 *If $S=\mathrm{Spec}(\mathbb{C})$, the functor $\mathbb{L}c\colon \mathcal{SH}\to\mathcal{SH}_S$ is fully faithful.*

Proposition 3.2 *Let $f\colon R\to S$ be a map of base schemes. The following diagram of stable homotopy categories commutes:*

$$\begin{tikzcd} & \mathcal{SH} \arrow[dl, "\mathbb{L}c"'] \arrow[dr, "\mathbb{L}c"] & \\ \mathcal{SH}_S \arrow[rr, "\mathbb{L}f^*"] & & \mathcal{SH}_R \end{tikzcd}$$

Proof The result follows by establishing $f^*\circ c=c$ on the level of model categories. For a constant space $cA\in\mathrm{Spc}(S)$, we have $f^*cA=cA$ by the calculation

$$(f^*cA)(U)=\operatorname*{colim}_{U\to f^{-1}V} cA(V)=A,$$

given the formula for f^* in Section 2.3. As the base change map is extended to T–spectra by applying f^* termwise, the claim follows. □

Proposition 3.3 *Let S be a base scheme equipped with a map $\mathrm{Spec}(\mathbb{C})\to S$. Then $\mathbb{L}c\colon\mathcal{SH}\to\mathcal{SH}_S$ is faithful.*

Proof For symmetric spectra X and Y, the map $\mathbb{L}c\colon\mathcal{SH}(X,Y)\to\mathcal{SH}(\mathbb{C})(cX,cY)$ factors through $\mathcal{SH}_S(cX,cY)$ by Proposition 3.2. Theorem 3.1 implies that the map $\mathbb{L}c\colon\mathcal{SH}(X,Y)\to\mathcal{SH}_S(cX,cY)$ must be injective. □

Corollary 3.4 *Write $W(\overline{\mathbb{F}}_p)$ for the ring of Witt vectors of $\overline{\mathbb{F}}_p$ and K for the fraction field of $W(\overline{\mathbb{F}}_p)$ (see Serre* [43, Chapter II, Section 6] *for a definition). Because we have maps $W(\overline{\mathbb{F}}_p)\to K\to\mathbb{C}$, the map $\mathbb{L}c\colon\pi_n^s\to\pi_{n,0}(W(\overline{\mathbb{F}}_p))$ is an injection.*

4 Motivic cohomology

Spitzweck has constructed a spectrum $H\mathbb{Z}$ in $\mathrm{Spt}_T^{\Sigma}(S)$ which represents motivic cohomology $H^{a,b}(X;\mathbb{Z})$ defined using Bloch's cycle complex when S is the Zariski spectrum of a Dedekind domain [45]. Spitzweck establishes enough nice properties of $H\mathbb{Z}$ so that we may construct the motivic Adams spectral sequence over general base schemes and establish comparisons between the motivic Adams spectral sequence over a Hensel local ring in which ℓ is invertible and its residue field.

4.1 Integral motivic cohomology

Definition 4.1 Over the base scheme $\mathrm{Spec}(\mathbb{Z})$, the spectrum $H\mathbb{Z}_{\mathrm{Spec}(\mathbb{Z})}$ is defined by Spitzweck in [45, Definition 4.27]. For a general base scheme S, we define $H\mathbb{Z}_S$ to be $f^*H\mathbb{Z}_{\mathrm{Spec}(\mathbb{Z})}$ where $f\colon S \to \mathrm{Spec}(\mathbb{Z})$ is the unique map.

Let $S = \mathrm{Spec}(D)$ for D a Dedekind domain. For $X \in \mathrm{Sm}/S$, Spitzweck shows there is a canonical isomorphism $\mathcal{SH}_S(\Sigma^\infty X_+, \Sigma^{a,b}H\mathbb{Z}) \cong H^{a,b}(X;\mathbb{Z})$, where $H^{a,b}(-;\mathbb{Z})$ denotes Levine's motivic cohomology defined using Bloch's cycle complex [45, Corollary 7.19]. The isomorphism is functorial with respect to maps in Sm/S. Additionally, if $i\colon \{s\} \to S$ is the inclusion of a closed point with residue field $k(s)$, there is a commutative diagram for $X \in \mathrm{Sm}/S$:

$$\begin{tikzcd}
\mathcal{SH}_S(\Sigma^\infty X_+, \Sigma^{a,b}H\mathbb{Z}) \arrow[r, "\cong"] \arrow[d] & H^{a,b}(X;\mathbb{Z}) \arrow[d] \\
\mathcal{SH}(k(s))(\mathbb{L}i^*\Sigma^\infty X_+, \Sigma^{a,b}H\mathbb{Z}) \arrow[r, "\cong"] & H^{a,b}(\mathbb{L}i^*X;\mathbb{Z})
\end{tikzcd}$$

If the residue field $k(s)$ has positive characteristic, there is a canonical isomorphism of ring spectra $\mathbb{L}i^*H\mathbb{Z}_S \cong H\mathbb{Z}_{k(s)}$ by Spitzweck [45, Theorem 9.16]. For a smooth map of base schemes $f\colon R \to S$, there is an isomorphism $\mathbb{L}f^*H\mathbb{Z}_S \cong H\mathbb{Z}_R$, because when f is smooth we have $\mathbb{L}f^* = f^*$; see Morel [33, page 44]. It is then straightforward to see that $f^*H\mathbb{Z}_S \cong H\mathbb{Z}_R$.

4.2 Motivic cohomology with coefficients $\mathbb{Z}/\ell$

For a prime ℓ, write $H\mathbb{Z}/\ell$ for the cofiber of the map $H\mathbb{Z} \xrightarrow{\ell\cdot} H\mathbb{Z}$ in $\mathcal{SH}_S$. The spectrum $H\mathbb{Z}/\ell$ represents motivic cohomology with $\mathbb{Z}/\ell$ coefficients. For a smooth scheme X over S, we write $H^{**}(X;\mathbb{Z}/\ell)$ for the motivic cohomology of X with $\mathbb{Z}/\ell$ coefficients. When S is the Zariski spectrum of a ring R, we write $H^{**}(R;\mathbb{Z}/\ell)$ for $H^{**}(\mathrm{Spec}(R);\mathbb{Z}/\ell)$. We will frequently omit Spec from our notation when the meaning is clear in other cases as well.

The now resolved Beilinson–Lichtenbaum conjecture allows us to calculate the mod 2 motivic cohomology of a finite field $\mathbb{F}_q$ of odd characteristic. In particular, there is an isomorphism $H^{**}(\mathbb{F}_q;\mathbb{Z}/2) \cong K^M_*(\mathbb{F}_q)/2[\tau]$ where τ has bidegree $(0,1)$ and elements of $K^M_n(\mathbb{F}_q)/2$ have bidegree (n,n). The group $K^M_1(\mathbb{F}_q)/2 \cong \mathbb{F}_q^\times/\mathbb{F}_q^{\times 2}$ is isomorphic to $\mathbb{Z}/2$. We write u for the nontrivial element of $\mathbb{F}_q^\times/\mathbb{F}_q^{\times 2}$ and ρ for the class of -1. It is well known that -1 is a square in $\mathbb{F}_q$ if and only if $q \equiv 1 \bmod 4$. Hence $H^{**}(\mathbb{F}_q;\mathbb{Z}/2) \cong \mathbb{Z}/2[\tau,u]/(u^2)$ and $u=\rho$ if and only if $q \equiv 3 \bmod 4$.

The mod 2 Bockstein homomorphism β is the motivic cohomology operation given by the connecting homomorphism in the long exact sequence of cohomology associated to the short exact sequence of coefficient groups

$$0 \to \mathbb{Z}/2 \to \mathbb{Z}/4 \to \mathbb{Z}/2 \to 0.$$

The Bockstein is a cohomology operation of bidegree $(1,0)$. On the mod 2 motivic cohomology of a finite field $\mathbb{F}_q$, the Bockstein is determined by $\beta(\tau)=\rho$ and $\beta(u)=0$ as it is a derivation. We remark that the Bockstein is trivial on the mod 2 motivic cohomology of a finite field $\mathbb{F}_q$ if and only if $q \equiv 1 \bmod 4$.

Proposition 4.2 *Let D be a Hensel local ring in which ℓ is invertible. Write F for the residue field of D and write $\pi\colon D \to F$ for the quotient map. Then the map $\pi^*\colon H^{**}(D;\mathbb{Z}/\ell) \to H^{**}(F;\mathbb{Z}/\ell)$ is an isomorphism of $\mathbb{Z}/\ell$–algebras. Furthermore, the action of the Bockstein is the same in either case.*

Proof The rigidity theorem for motivic cohomology in Geisser [17, Theorem 1.2(3)] gives the isomorphism. The map $\mathbb{L}\pi^*$ gives comparison maps for the long exact sequences which define the Bockstein over D and F. The rigidity theorem shows the long exact sequences are isomorphic, so the action of the Bockstein is the same in either case. □

4.3 Mod 2 motivic cohomology operations and cooperations

The mod 2 motivic Steenrod algebra over a base scheme S, which we write as $\mathcal{A}^{**}(S)$, is the algebra of bistable mod 2 motivic cohomology operations. A bistable cohomology operation is a family of operations $\theta_{**}\colon H^{**}(-;\mathbb{Z}/2) \to H^{*+a,*+b}(-;\mathbb{Z}/2)$ which are compatible with the suspension isomorphism for both the simplicial circle S^1 and the Tate circle $\mathbb{G}_m$.

When S is the Zariski spectrum of a characteristic 0 field, Voevodsky identified the structure of this algebra in [49; 50]. Voevodsky's calculation was extended to hold where the base is the Zariski spectrum of a field of positive characteristic $p \neq 2$ by Hoyois, Kelly and Østvær in [22]. In particular, the algebra $\mathcal{A}^{**}(S)$ is generated over $\mathbb{F}_2$ by the Steenrod squaring operations Sq^i of bidegree $(i, \lfloor i/2 \rfloor)$ and the operations given by cup products $x \cup -$ where $x \in H^{**}(S; \mathbb{Z}/2)$. The Steenrod squaring operations satisfy motivic Adem relations, which are given by Voevodsky in [49, Section 10] (a minor modification is needed in the case $a + b \equiv 1 \bmod 2$).

We record the structure of the mod 2 dual Steenrod algebra $\mathcal{A}_{**}(\mathbb{F}_q)$ for a finite field $\mathbb{F}_q$ of characteristic different from 2 in the following proposition.

Proposition 4.3 *Let $\mathbb{F}_q$ be a finite field of odd characteristic. The mod 2 dual Steenrod algebra is an associative commutative algebra of the form*

$$\mathcal{A}_{**}(\mathbb{F}_q) \cong H_{**}(\mathbb{F}_q)[\tau_i, \xi_j \mid i \geq 0, j \geq 1]/(\tau_i^2 - \tau\xi_{i+1} - \rho\tau_{i+1} - \rho\tau_0\xi_{i+1}),$$

where τ_i has bidegree $(2^{i+1}-1, 2^i-1)$ and ξ_i has bidegree $(2^{i+1}-2, 2^i-1)$. Note that if $q \equiv 1 \bmod 4$, the relation for τ_i^2 simplifies to $\tau_i^2 = \tau\xi_{i+1}$ as $\rho = 0$.

The structure maps for the Hopf algebroid $(H_{**}(\mathbb{F}_q), \mathcal{A}_{**}(\mathbb{F}_q))$, which we write simply as $(H_{**}, \mathcal{A}_{**})$, are as follows:

(a) *The left unit $\eta_L\colon H_{**} \to \mathcal{A}_{**}$ is given by $\eta_L(x) = x$.*

(b) *The right unit $\eta_R\colon H_{**} \to \mathcal{A}_{**}$ is determined as a map of $\mathbb{Z}/2$–algebras by $\eta_R(\rho) = \rho$ and $\eta_R(\tau) = \tau + \rho\tau_0$. In the case where ρ is trivial, that is, $q \equiv 1 \bmod 4$, the right and left unit agree: $\eta_R = \eta_L$.*

(c) *The augmentation $\epsilon\colon \mathcal{A}_{**} \to H_{**}$ kills τ_i and ξ_i, and for $x \in H_{**}$, it follows that $\epsilon(x) = x$.*

(d) *The coproduct $\Delta\colon \mathcal{A}_{**} \to \mathcal{A}_{**} \otimes_{H_{**}} \mathcal{A}_{**}$ is a map of graded $\mathbb{Z}/2$–algebras determined by*

$$\Delta(x) = x \otimes 1 \quad \text{for } x \in H_{**},$$

$$\Delta(\tau_i) = \tau_i \otimes 1 + 1 \otimes \tau_i + \sum_{j=0}^{i-1} \xi_{i-j}^{2^j} \otimes \tau_j,$$

$$\Delta(\xi_i) = \xi_i \otimes 1 + 1 \otimes \xi_i + \sum_{j=1}^{i-1} \xi_{i-j}^{2^j} \otimes \xi_j.$$

(e) *The antipode c is a map of $\mathbb{Z}/2$–algebras determined by*

$$c(\rho) = \rho, \qquad c(\tau) = \tau + \rho\tau_0,$$

$$c(\tau_i) = \tau_i + \sum_{j=0}^{i-1} \xi_{i-j}^{2^j} c(\tau_j), \qquad c(\xi_i) = \xi_i + \sum_{j=1}^{i-1} \xi_{i-j}^{2^j} c(\xi_j).$$

Proof The calculation can be found in the work of Hoyois, Kelly and Østvær [22] and Voevodsky [49]. □

We now investigate the structure of the Hopf algebroid of mod 2 cohomology cooperations over a Dedekind domain.

Definition 4.4 Let D be a Dedekind domain, and let C denote the set of sequences $(\epsilon_0, r_1, \epsilon_1, r_2, \dots)$ with $\epsilon_i \in \{0, 1\}$, each r_i nonnegative, and only finitely many nonzero terms. The elements $\tau_i \in \mathcal{A}_{2^{i+1}-1, 2^i-1}(D)$ and $\xi_i \in \mathcal{A}_{2^{i+1}-2, 2^i-1}(D)$ are constructed by Spitzweck in [45, Corollary 11.23]. For any sequence $I = (\epsilon_0, r_1, \epsilon_1, r_2, \dots)$ in C, write $\omega(I)$ for the element $\tau_0^{\epsilon_0} \xi_1^{r_1} \cdots$ and $(p(I), q(I))$ for the bidegree of the operation $\omega(I)$.

Spitzweck calculates in [45, Theorem 11.24] that the dual Steenrod algebra is generated by the elements τ_i and ξ_j but does not identify the relations for τ_i^2. We record Spitzweck's calculation in the following proposition.

Proposition 4.5 *Let D be a Dedekind domain. As an $H\mathbb{Z}/2$ module, there is a weak equivalence $\bigvee_{I \in B} \Sigma^{p(I), q(I)} H\mathbb{Z}/2 \to H\mathbb{Z}/2 \wedge H\mathbb{Z}/2$. The map is given by $\omega(I)$ on the factor $\Sigma^{p(I), q(I)} H\mathbb{Z}/2$.*

To obtain the relations for τ_i^2, we find an analog of the result of Voevodsky [49, Theorem 6.10] when D is a Hensel local ring.

Proposition 4.6 *Let D be a Hensel local ring in which 2 is invertible and let F denote the residue field of D. Then the following isomorphism holds:*

$$H^{**}(B\mu_2, \mathbb{Z}/2) \cong H^{**}(D, \mathbb{Z}/2)[\![u, v]\!]/(u^2 = \tau v + \rho u).$$

Here v is the class $v_2 \in H^{2,1}(B\mu_2)$ defined by Spitzweck in [45, page 81] *and $u \in H^{1,1}(B\mu_2; \mathbb{Z}/2)$ is the unique class satisfying $\widetilde{\beta}(u) = v$, where $\widetilde{\beta}$ is the integral Bockstein determined by the coefficient sequence $\mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}/2$.*

Proof The motivic classifying space $B\mu_2$ over D (respectively F) fits into a triangle $B\mu_{2+} \to (\mathcal{O}(-2)_{\mathbb{P}^\infty})_+ \to \mathrm{Th}(\mathcal{O}(-2))$ by [49, (6.2)] and [45, (25)]. From this triangle, we obtain a long exact sequence in mod 2 motivic cohomology [49, (6.3)] and [45, (26)]. The comparison map $\mathbb{L}\pi^*\colon \mathcal{SH}_D \to \mathcal{SH}_F$ induces a homomorphism of these long exact sequences. The rigidity Proposition 4.2 and the 5-lemma then show that the comparison maps are all isomorphisms. As the desired relation holds in the motivic cohomology of $B\mu_2$ over F and the choices of u and v are compatible with base change, the result follows. □

With this result, the relations $\tau_i^2 = \tau\xi_{i+1} + \rho\tau_{i+1} + \rho\tau_0\xi_{i+1}$ in $\mathcal{A}_{**}(D)$ follow when D is a Hensel local ring in which 2 is invertible by the argument given by Voevodsky in [49, Theorem 12.6]. Furthermore, the calculation of Spitzweck in [45, Corollary 11.23] shows that the coproduct Δ is the same as in Proposition 4.3(d). The action of the Steenrod squaring operations $H^{**}(D)$ and $H^{**}(F)$ agree by the naturality of these cohomology operations, since these cohomology groups are isomorphic. This shows that the right unit η_R and the antipode c are given by the formulas in Proposition 4.3(b,e).

Remark 4.7 Let D be a Dedekind domain in which 2 is invertible and consider the map $f\colon \mathbb{Z}\left[\frac{1}{2}\right] \to D$. A key observation of Spitzweck in the proof of Theorem 11.24 in [45] is that the map $\mathbb{L}f^*\colon \mathcal{A}_{**}\left(\mathbb{Z}\left[\frac{1}{2}\right]\right) \to \mathcal{A}_{**}(D)$ satisfies $\mathbb{L}f^*\tau_i = \tau_i$ and $\mathbb{L}f^*\xi_i = \xi_i$ for all i. It follows that for a map $j\colon D \to \widetilde{D}$ of Dedekind domains in which 2 is invertible, $\mathbb{L}j^*\tau_i = \tau_i$ and $\mathbb{L}j^*\xi_i = \xi_i$ for all i.

Proposition 4.8 *Let D be a Hensel local ring in which 2 is invertible and let F denote the residue field of D. Then the comparison map $\pi^*\colon \mathcal{A}_{**}(D) \to \mathcal{A}_{**}(F)$ is an isomorphism of Hopf algebroids.*

Proof Remark 4.7 shows that the map $\pi^*\colon \mathcal{A}_{**}(D) \to \mathcal{A}_{**}(F)$ is an isomorphism of left $H_{**}(F)$ modules. The compatibility of the isomorphism with the coproduct, right unit and antipode was established above. □

The following definition is taken from Dugger and Isaksen [12, Definition 2.11].

Definition 4.9 A set of bigraded objects $X = \{x_{(a,b)}\}$ is said to be motivically finite if for any bigrading (a,b) there are only finitely many objects $y_{(a',b')} \in X$ for which $a \geq a'$ and $2b - a \geq 2b' - a'$. We say a bigraded algebra or module is motivically finite if it has a generating set which is motivically finite.

To motivate the preceding definition, observe that if $H^{**}(X)$ is a motivically finite $H^{**}(F)$ module, then $H^{**}(X)$ is a finite dimensional $\mathbb{F}_\ell$ vector space in each bidegree.

For a Hensel local ring D, the isomorphism $\mathcal{A}_{**}(D) \cong \mathcal{A}_{**}(F)$ of motivically finite algebras gives an isomorphism of their duals $\mathcal{A}^{**}(D) \cong \mathcal{A}^{**}(F)$. See Hoyois, Kelly and Østvær [22, Section 5.2] and Spitzweck [45, Remark 11.25] for the proof that the dual of the Hopf algebroid of cooperations is the Steenrod algebra.

The analogous results of this section hold for mod ℓ motivic cohomology over a base field or a Hensel local ring in which ℓ is invertible for odd primes ℓ. Precise statements can be found in Wilson [52].

5 Motivic Adams spectral sequence

The motivic Adams spectral sequence over a base scheme S may be defined using the appropriate notion of an Adams resolution; see Adams [2], Switzer [46] or Ravenel [40] for treatments in the topological case. We recount the definition for completeness and establish some basic properties of the motivic Adams spectral sequence under base change. We follow Dugger and Isaksen [12, Section 3] for the definition of the motivic Adams spectral sequence. See also the work of Hu, Kriz and Ormsby [24, Section 6].

Let p and ℓ be distinct primes and let $q = p^\nu$ for some integer $\nu \geq 1$. We will be interested in the specific case of the motivic Adams spectral sequence over a field and over a Hensel discrete valuation ring with residue field of characteristic p. We write H for the spectrum $H\mathbb{Z}/\ell$ over the base scheme S and $H^{**}(S)$ for the motivic cohomology of S with $\mathbb{Z}/\ell$ coefficients. The spectrum H is a ring spectrum and is cellular in the sense of Dugger and Isaksen [11] by work of Spitzweck [45, Corollary 11.4].

5.1 Construction of the mod ℓ MASS

Definition 5.1 Consider a spectrum X over the base scheme S and let $\overline{H}$ denote the spectrum in the cofibration sequence $\overline{H} \to \mathbb{1} \to H \to \Sigma\overline{H}$. The standard H–Adams resolution of X is the tower of cofibration sequences $X_{f+1} \to X_f \to W_f$ given by $X_f = \overline{H}^{\wedge f} \wedge X$ and $W_f = H \wedge X_f$:

$$\begin{tikzcd}
X_0 = X \arrow[dr, "j_0"'] & & \overline{H} \wedge X \arrow[ll, "i_1"'] \arrow[dr, "j_1"'] & & \overline{H} \wedge \overline{H} \wedge X \arrow[ll, "i_2"'] & \cdots \arrow[l] \\
& H \wedge X \arrow[ur, "\bullet", "\partial_0"'] & & H \wedge \overline{H} \wedge X \arrow[ur, "\bullet", "\partial_1"'] & &
\end{tikzcd}$$

Compare this with [2, Section 15].

Definition 5.2 Let X be a T–spectrum over S and let $\{X_f, W_f\}$ be the standard H–Adams resolution of X. The motivic Adams spectral sequence for X with respect to H is the spectral sequence determined by the following exact couple:

$$\begin{tikzcd}
\oplus\pi_{**}X_f \arrow[rr, "i_*"] & & \oplus\pi_{**}X_f \arrow[dl, "j_*"] \\
& \oplus\pi_{**}W_f \arrow[ul, "\partial_*"] &
\end{tikzcd}$$

The E_1 term of the motivic Adams spectral sequence is $E_1^{f,(s,w)} = \pi_{s,w}W_f$. The index f is called the Adams filtration, s is the stem and w is the motivic weight. The Adams filtration of $\pi_{**}X$ is given by $F_i\pi_{**}X = \operatorname{im}(\pi_{**}X_i \to \pi_{**}X)$.

Proposition 5.3 *Let $\mathfrak{S}$ denote the category of spectral sequences in the category of abelian groups. The associated spectral sequence to the standard H–Adams resolution defines a functor $\mathfrak{M}\colon \mathcal{SH}_S \to \mathfrak{S}$. Furthermore, the motivic Adams spectral sequence is natural with respect to base change.*

Proof The construction of the standard H–Adams resolution is functorial because $\mathcal{SH}_S$ is symmetric monoidal. Given $X \to X'$ we get induced maps of standard H–Adams resolutions $\{X_f, W_f\} \to \{X'_f, W'_f\}$. As $\pi_{**}(-)$ is a triangulated functor, we get an induced map of the associated exact couples and hence of spectral sequences $\mathfrak{M}(X) \to \mathfrak{M}(X')$.

Let $f\colon R \to S$ be a map of base schemes. The claim is that there is a natural transformation between $\mathfrak{M}\colon \mathcal{SH}_S \to \mathfrak{S}$ and $\mathfrak{M} \circ \mathbb{L} f^*\colon \mathcal{SH}_S \to \mathcal{SH}_R \to \mathfrak{S}$. Let $X \in \mathcal{SH}_S$ and let $\{X_f, W_f\}$ be the standard H_S–Adams resolution of X in $\mathcal{SH}_S$. We may as well assume X is cofibrant, in which case $QX = X$ where Q is the cofibrant replacement functor. Let $\{X'_f, W'_f\}$ denote the standard H_R–Adams resolution of $\mathbb{L} f^* X = f^* X$. Observe that we have $\{f^* X_f, f^* W_f\} = \{X'_f, W'_f\}$, since $f^*\mathbb{1} = \mathbb{1}$, $f^* H_S = H_R$ and $\mathbb{L} f^*$ is a monoidal functor. We therefore have a map $\{\mathbb{L} f^* X_f, \mathbb{L} f^* W_f\} \to \{X'_f, W'_f\}$. Applying $\mathbb{L} f^*\colon \mathcal{SH}_S(\Sigma^{s,w}\mathbb{1}, -) \to \mathcal{SH}_R(\Sigma^{s,w}\mathbb{1}, \mathbb{L} f^* -)$ to $\{X_f, W_f\}$ gives a map of exact couples and therefore a map $\Phi_X\colon \mathfrak{M}_S(X) \to \mathfrak{M}_R(\mathbb{L} f^* X)$. It is straightforward to verify that Φ determines a natural transformation. □

Corollary 5.4 *For a map of base schemes $f\colon R \to S$, there is a map of motivic Adams spectral sequences $\Phi\colon \mathfrak{M}_S(\mathbb{1}) \to \mathfrak{M}_R(\mathbb{1})$. The map Φ is furthermore compatible with the induced map $\pi_{**}(S) \to \pi_{**}(R)$.*

Definition 5.5 A particularly well-behaved family of spectra in $\mathcal{SH}_S$ are the cellular spectra in the sense of Dugger and Isaksen [11, Definition 2.10]. A spectrum $E \in \mathcal{SH}_S$ is cellular if it can be constructed out of the spheres $\Sigma^\infty S^{a,b}$ for any integers a and b by homotopy colimits. A cellular spectrum is of finite type if for some k it has a cell decomposition with no cells $S^{a,b}$ for $a - b < k$ and at most finitely many cells $S^{a,b}$ for any a and b; see Hu, Kriz and Ormsby [24, Section 2].

In the following proposition, Ext is taken in the category of $\mathcal{A}_{**}$–comodules. The homological algebra of comodules is investigated thoroughly in Adams [2], Switzer [46] and Ravenel [40].

Proposition 5.6 *Suppose X is a cellular spectrum over the base scheme S. The motivic Adams spectral sequence for X has E_2 page given by*

$$E_2^{f,(s,w)} \cong \mathrm{Ext}^{f,(s+f,w)}_{\mathcal{A}_{**}(S)}(H_{**}S, H_{**}X),$$

with differentials $d_r\colon E_r^{f,(s,w)} \to E_r^{f+r,(s-1,w)}$ for $r \geq 2$.

Proof Spitzweck proves that H is a cellular spectrum in [45, Corollary 11.4]. The argument given for [12, Proposition 6.10] by Dugger and Isaksen then goes through. The cellularity of X and H is sufficient to ensure that the Künneth theorem holds, which is needed in the argument. □

Corollary 5.7 *If X and X' are cellular spectra over S and $X \to X'$ induces an isomorphism $H_{**}X \to H_{**}X'$, then the induced map $\mathfrak{M}(X) \to \mathfrak{M}(X')$ is an isomorphism of spectral sequences from the E_2 page onwards.*

Corollary 5.8 *Let $f\colon R \to S$ be a map of base schemes and consider a cellular spectrum X over S. Suppose $f^*\colon H_{**}(S) \to H_{**}(R)$, $f^*\colon \mathcal{A}_{**}(S) \to \mathcal{A}_{**}(R)$ and $f^*\colon H_{**}X \to H_{**}(\mathbb{L}f^*X)$ are all isomorphisms. Then $\mathfrak{M}_S(X) \to \mathfrak{M}_R(\mathbb{L}f^*X)$ is an isomorphism of spectral sequences from the E_2 page onwards.*

Corollary 5.9 *Let D be a Hensel local ring in which ℓ is invertible and write F for the residue field of D. Then the comparison map $\mathfrak{M}(D) \to \mathfrak{M}(F)$ is an isomorphism at the E_2 page.*

Proof Propositions 4.2 and 4.8 and Corollary 5.8 give the result when $X = \mathbb{1}$. □

5.2 Convergence of the motivic Adams spectral sequence

To simplify the notation, write $\mathrm{Ext}(R)$ for $\mathrm{Ext}_{\mathcal{A}^{**}(R)}(H^{**}(R), H^{**}(R))$ when working over the base scheme $S = \mathrm{Spec}(R)$. For any abelian group G and any prime ℓ, we write $G_{(\ell)}$ for the ℓ–primary part of G and $G^\wedge_\ell = \varprojlim G/\ell^\nu$ for the ℓ–completion of G. If $\{X_f, W_f\}$ is the standard H–Adams resolution of a spectrum X, the H–nilpotent completion of X is the spectrum $X^\wedge_H = \mathrm{holim}_f\, X/X_f$ defined by Bousfield in [6, Section 5]. The H–nilpotent completion has a tower given by $C_i = \mathrm{holim}_f(X_i/X_f)$.

Proposition 5.10 *Let S be the Zariski spectrum of a field F with characteristic $p \neq \ell$ and let X be a cellular spectrum X over S of finite type* (Definition 5.5). *If either $\ell > 2$ and F has finite mod ℓ cohomological dimension, or $\ell = 2$ and $F[\sqrt{-1}]$ has finite mod 2 cohomological dimension, the motivic Adams spectral sequence converges to the homotopy groups of the H–nilpotent completion of X:*

$$E_2^{f,(s,w)} \Rightarrow \pi_{s,w}(X^\wedge_H).$$

Furthermore, there is a weak equivalence $X^\wedge_H \cong X^\wedge_\ell$.

Proof The argument given by Hu, Kriz and Ormsby in [24], which requires Morel's connectivity theorem for F, carries over to the positive characteristic case from the work of Hoyois, Kelly and Østvær [22]. See Ormsby and Østvær [39, Section 3.1] for the analogous argument for the motivic Adams–Novikov spectral sequence. □

We say a line $s = mf + b$ in the (f,s)–plane is a vanishing line for a bigraded group $G^{f,s}$ if $G^{f,s}$ is zero whenever $0 < s < mf + b$.

Proposition 5.11 *If $\bar{F}$ is an algebraically closed field of characteristic $p \neq \ell$, then a vanishing line for $\mathrm{Ext}^{**}(\bar{F}) \cong \mathrm{Ext}^{**}(W(\bar{F}))$ at the prime ℓ is $s = (2\ell - 3)f$. If $\mathbb{F}_q$ is a finite field of characteristic $p \neq \ell$, then a vanishing line for $\mathrm{Ext}^{**}(\mathbb{F}_q) \cong \mathrm{Ext}^{**}(W(\mathbb{F}_q))$ at the prime ℓ is $s = (2\ell - 3)f - 1$.*

Proof A vanishing line exists for $\mathrm{Ext}(\bar{F}) \cong \mathrm{Ext}(W(\bar{F}))$ when $\bar{F}$ is an algebraically closed fields by comparison with $\mathbb{C}$ and the topological case by work of Dugger and Isaksen [12]. The vanishing line $s = f(2\ell - 3)$ from topology by Adams [1] is therefore a vanishing line for $\mathrm{Ext}(\bar{F}) \cong \mathrm{Ext}(W(\bar{F}))$.

For a finite field $\mathbb{F}_q$, the line $s = f(2\ell - 3) - 1$ is a vanishing line for $\mathrm{Ext}(\mathbb{F}_q) \cong \mathrm{Ext}(W(\mathbb{F}_q))$ by the identification of the E_2 page of the motivic Adams spectral sequence. When $\ell = 2$ this is given in Proposition 7.1 when $q \equiv 1 \bmod 4$ and the calculation of the ρ–BSS when $q \equiv 3 \bmod 4$. For odd ℓ, see Wilson [52]. □

We now discuss the convergence of the motivic Adams spectral sequence over the ring of Witt vectors associated to a finite field or an algebraically closed field. Consult Serre [43, Chapter II, Section 6] for a construction of the ring of Witt vectors associated to a field of positive characteristic.

Proposition 5.12 *Let $W(F)$ be the ring of Witt vectors of a field F that is either a finite field or an algebraically closed field of characteristic p and let ℓ be a prime different from p. The motivic Adams spectral sequence for $\mathbb{1}$ over $W(F)$ converges to $\pi_{**}(\mathbb{1}^{\wedge}_H)$ filtered by the Adams filtration, where $\mathbb{1}^{\wedge}_H$ is the H–nilpotent completion of $\mathbb{1}$.*

Proof The convergence $\mathfrak{M}_{W(F)}(\mathbb{1}) \Rightarrow \pi_{**}(\mathbb{1}^{\wedge}_H)$ follows by the argument given by Dugger and Isaksen [12, Corollary 6.15], given the vanishing line in the motivic Adams spectral sequence from Proposition 5.11. □

Proposition 5.13 *Let R and S be base schemes for which the motivic Adams spectral sequence for $\mathbb{1}$ converges to $\pi_{**}(\mathbb{1}^{\wedge}_H)$; see Propositions 5.10 and 5.12 for examples. A map of base schemes $f\colon R \to S$ yields a comparison map $\mathfrak{M}_S(\mathbb{1}^{\wedge}_H) \to \mathfrak{M}_R(\mathbb{1}^{\wedge}_H)$ which is compatible with the induced map*

$$\pi_{**}(\mathbb{1}^{\wedge}_H(S)) \to \pi_{**}(\mathbb{L}f^*\mathbb{1}^{\wedge}_H(S)) \to \pi_{**}(\mathbb{1}^{\wedge}_H(R)).$$

Proof Let $\{X_f(S), W_f(S)\}$ denote the standard H–Adams resolution of $\mathbb{1}$ over S. We now construct a map $\pi_{**}(\mathbb{1}^\wedge_H(S)) \to \pi_{**}(\mathbb{1}^\wedge_H(R))$. Recall from Proposition 5.3 that $f^* X_f(S) = X_f(R)$. Since $\mathbb{L}f^*$ is a triangulated functor, there are maps $\mathbb{L}f^*(\mathbb{1}/X_f(S)) \to \mathbb{1}/X_f(R)$ and so a map $\mathbb{L}f^*\mathbb{1}^\wedge_H(S) \to \mathbb{1}^\wedge_H(R)$ by the universal property for $\mathbb{1}^\wedge_H(R) = \operatorname{holim} \mathbb{1}/X_f(R)$. Write $C_i(S)$ for the tower of $\mathbb{1}^\wedge_H(S)$ over S defined above (and in Bousfield [6, Section 5]). Similar considerations give a map of towers $\mathbb{L}f^* C_i(S) \to C_i(R)$. Hence $\mathfrak{M}_S(\mathbb{1}^\wedge_H) \to \mathfrak{M}_R(\mathbb{1}^\wedge_H)$ is compatible with the induced map $\pi_{**}(\mathbb{1}^\wedge_H(S)) \to \pi_{**}(\mathbb{1}^\wedge_H(R))$. □

Proposition 5.14 *Let F be a field of characteristic p with finite mod ℓ cohomological dimension for all primes $\ell \neq p$ and suppose $H^{s,w}(F;\mathbb{Z}/\ell)$ is a finite dimensional vector space over $\mathbb{F}_\ell$ for all s and w. Furthermore, assume that the mod ℓ motivic Adams spectral sequence for $\mathbb{1}$ over F has a vanishing line, such as when F is a finite field or an algebraically closed field. Then the ℓ–primary part of $\pi_{s,w}(F)$ is finite whenever $s > w \geq 0$.*

Proof Ananyevsky, Levine and Panin show in [3] that the groups $\pi_{s,w}(F)$ are torsion for $s > w \geq 0$. It follows that the group $\pi_{s,w}(F)$ is the sum of its ℓ–primary subgroups $\pi_{s,w}(F)_{(\ell)}$. We set out to show that $\pi_{s,w}(F)_{(\ell)}$ is finite when $\ell \neq p$.

The motivic Adams spectral sequence converges to $\pi_{**}(\mathbb{1}^\wedge_\ell)$ by Proposition 5.10 (this requires Morel's connectivity theorem). The vanishing line in the motivic Adams spectral sequence shows that the Adams filtration of $\pi_{s,w}(\mathbb{1}^\wedge_\ell)$ has finite length, and as each group $E_2^{f,(s,w)}$ is a finite dimensional $\mathbb{F}_\ell$ vector space, we conclude the groups $\pi_{s,w}(\mathbb{1}^\wedge_\ell)$ are finite. From the long exact sequence of homotopy groups associated to the triangle $\mathbb{1}^\wedge_\ell \to \prod \mathbb{1}/\ell^\nu \to \prod \mathbb{1}/\ell^\nu$ defining $\mathbb{1}^\wedge_\ell$, we extract the short exact sequence of finite groups

$$0 \to \varprojlim{}^1 \pi_{s+1,w}(\mathbb{1}/\ell^\nu) \to \pi_{s,w}(\mathbb{1}^\wedge_\ell) \to \varprojlim \pi_{s,w}(\mathbb{1}/\ell^\nu) \to 0. \tag{5-1}$$

Similarly, from the triangles $\mathbb{1} \xrightarrow{\ell^\nu\cdot} \mathbb{1} \to \mathbb{1}/\ell^\nu$ we extract the short exact sequences

$$0 \to \pi_{s,w}(\mathbb{1})/\ell^\nu \to \pi_{s,w}(\mathbb{1}/\ell^\nu) \to {}_{\ell^\nu}\pi_{s-1,w}(\mathbb{1}) \to 0,$$

which form a short exact sequence of towers. The maps in the tower $\{\pi_{s,w}(\mathbb{1})/\ell^\nu\}$ are given by the reduction maps $\pi_{s,w}(\mathbb{1})/\ell^\nu \to \pi_{s,w}(\mathbb{1})/\ell^{\nu-1}$. Since the tower $\{\pi_{s,w}(\mathbb{1})/\ell^\nu\}$ satisfies the Mittag–Leffler condition, we have $\varprojlim{}^1 \pi_{s,w}(\mathbb{1})/\ell^\nu = 0$. The associated long exact sequence for the inverse limit gives the exact sequence

$$0 \to \pi_{s,w}(\mathbb{1})^\wedge_\ell \to \varprojlim \pi_{s,w}(\mathbb{1}/\ell^\nu) \to \varprojlim {}_{\ell^\nu}\pi_{s-1,w}(\mathbb{1}) \to 0. \tag{5-2}$$

The group $\varprojlim {}_{\ell^\nu}\pi_{s-1,w}(\mathbb{1})$ is the ℓ–adic Tate module of $\pi_{s-1,w}(\mathbb{1})$, which is torsion-free. As $\varprojlim \pi_{s,w}(\mathbb{1}/\ell^\nu)$ is finite by (5-1), the map $\varprojlim \pi_{s,w}(\mathbb{1}/\ell^\nu) \to \varprojlim {}_{\ell^\nu}\pi_{s-1,w}(\mathbb{1})$

is trivial. But since the sequence (5-2) is exact, the group $\varprojlim_{\ell^\nu} \pi_{s-1,w}(\mathbb{1})$ is trivial, $\pi_{s,w}(\mathbb{1})^\wedge_\ell \cong \varprojlim \pi_{s,w}(\mathbb{1}/\ell^\nu)$ and $\pi_{s,w}(\mathbb{1})^\wedge_\ell$ is finite.

Write $K(i)$ for the kernel of the canonical map $\pi_{s,w}(\mathbb{1})^\wedge_\ell \to \pi_{s,w}(\mathbb{1})/\ell^i$. The tower $\cdots \subseteq K(i) \subseteq K(i-1) \subseteq \cdots \subseteq K(1)$ consists of finite groups and so it must stabilize. Hence the tower

$$\cdots \to \pi_{s,w}(\mathbb{1})/\ell^\nu \to \pi_{s,w}(\mathbb{1})/\ell^{\nu-1} \to \cdots \to \pi_{s,w}(\mathbb{1})/\ell$$

must also stabilize. There is then some N for which $\ell^N \pi_{s,w}(\mathbb{1}) = \ell^\nu \pi_{s,w}(\mathbb{1})$ for all $\nu \geq N$, and so $\ell^N \pi_{s,w}(\mathbb{1})$ is ℓ–divisible. From the short exact sequence of towers $\ell^\nu \pi_{s,w}(\mathbb{1}) \to \pi_{s,w}(\mathbb{1}) \to \pi_{s,w}(\mathbb{1})/\ell^\nu$, taking the inverse limit yields the exact sequence

$$0 \to \ell^N \pi_{s,w}(\mathbb{1}) \to \pi_{s,w}(\mathbb{1}) \to \pi_{s,w}(\mathbb{1})^\wedge_\ell \to 0.$$

Since $\pi_{s,w}(\mathbb{1})^\wedge_\ell$ is finite, it is ℓ–primary and there is a short exact sequence

$$0 \to \ell^N \pi_{s,w}(\mathbb{1})_{(\ell)} \to \pi_{s,w}(\mathbb{1})_{(\ell)} \to \pi_{s,w}(\mathbb{1})^\wedge_\ell \to 0.$$

The group $\ell^N \pi_{s,w}(\mathbb{1})_{(\ell)}$ must be zero. Suppose for a contradiction that it is nonzero. Then $\ell^N \pi_{s,w}(\mathbb{1})_{(\ell)}$ must contain $\mathbb{Z}/\ell^\infty$ as a summand, which shows the ℓ–adic Tate module of $\pi_{s,w}(\mathbb{1})$ is nonzero, a contradiction. □

We now identify the groups $\pi_{s,s}(\mathbb{1}^\wedge_\ell)$ for $s \geq 0$.

Proposition 5.15 *Let F be a finite field or an algebraically closed field of characteristic $p \neq \ell$. When $s = w \geq 0$ or $s < w$, the motivic Adams spectral sequence of $\mathbb{1}$ over F converges to the ℓ–completion of $\pi_{s,w}(F)$.*

Proof If $s < w$, the convergence follows from Morel's connectivity theorem. When $s = w \geq 0$, Proposition 5.10 implies that at bidegree (s, w) the motivic Adams spectral sequence converges to the group $\pi_{s,w}(\mathbb{1}^\wedge_\ell)$. Since $\pi_{s-1,s}(\mathbb{1}) = 0$ by Morel's connectivity theorem, the short exact sequence (see, for example, Hu, Kriz and Ormsby [24, (2)])

$$0 \to \mathrm{Ext}(\mathbb{Z}/\ell^\infty, \pi_{s,s}(\mathbb{1})) \to \pi_{s,s}(\mathbb{1}^\wedge_\ell) \to \mathrm{Hom}(\mathbb{Z}/\ell^\infty, \pi_{s-1,s}(\mathbb{1})) \to 0$$

gives an isomorphism $\mathrm{Ext}(\mathbb{Z}/\ell^\infty, \pi_{s,s}(\mathbb{1})) \cong \pi_{s,s}(\mathbb{1}^\wedge_\ell)$. In [34, Corollary 1.25], Morel has calculated $\pi_{0,0}(F) \cong \mathrm{GW}(F)$ and $\pi_{s,s}(F) \cong W(F)$ for $s > 0$ where $W(F)$ is the Witt group of the field F. For the fields under consideration, $\mathrm{GW}(F)$ and $W(F)$ are finitely generated abelian groups. But for any finitely generated abelian group A, there is an isomorphism $\mathrm{Ext}(\mathbb{Z}/\ell^\infty, A) \cong A^\wedge_\ell$, given in Bousfield and Kan [7, Chapter VI, Section 2.1], which concludes the proof. □

6 Stable stems over an algebraically closed field

Let $\overline{F}$ be an algebraically closed field of positive characteristic p. Denote the ring of Witt vectors of $\overline{F}$ by $W = W(\overline{F})$, the field of fractions of W by $K = K(\overline{F})$, and the algebraic closure of K by $\overline{K} = \overline{K}(\overline{F})$. Note that K is a field of characteristic 0. The previous sections have set us up with enough machinery to compare the motivic Adams spectral sequences at a prime $\ell \neq p$ over the associated base schemes $\mathrm{Spec}(\overline{F})$, $\mathrm{Spec}(W)$ and $\mathrm{Spec}(\overline{K})$. We will often write the ring instead of the Zariski spectrum of the ring in our notation. For any Dedekind domain R, we write $\mathrm{Ext}(R)$ for the trigraded ring $\mathrm{Ext}_{\mathcal{A}^{**}(R)}(H^{**}(R), H^{**}(R))$.

Proposition 6.1 *Let $\overline{F}$ be an algebraically closed field of positive characteristic p, and let ℓ be a prime different from p. The E_2 page of the mod ℓ motivic Adams spectral sequence for $\mathbb{1}$ over W, the ring of Witt vectors of $\overline{F}$, is given by*

$$E_2^{f,(s,w)}(W) \cong \mathrm{Ext}^{f,(s+f,w)}(W) \cong \mathrm{Ext}^{f,(s+f,w)}(\overline{F}).$$

Proof Since W is a Hensel local ring with residue field $\overline{F}$, Corollary 5.9 applies. □

Proposition 6.2 *Let $\overline{F}$ be an algebraically closed field of characteristic p. The homomorphism $f\colon W \to \overline{K}$ induces isomorphisms of graded rings*

$$f^*\colon H_{**}(W) \to H_{**}(\overline{K}) \quad \text{and} \quad f^*\colon \mathcal{A}_{**}(W) \to \mathcal{A}_{**}(\overline{K}).$$

Proof Since $H^{**}(S) \cong H_{-*,-*}(S)$, it suffices to establish isomorphisms for motivic cohomology. Because $H^{**}(W) \cong H^{**}(\overline{\mathbb{F}}_p)$, we have $H^{**}(W) \cong \mathbb{F}_\ell[\tau]$ where $\tau \in H^{0,1}(W) \cong \mu_\ell(W)$. We also have that $H^{**}(\overline{K}) \cong \mathbb{F}_\ell[\tau]$. To identify the ring map $f^*\colon H^{**}(W) \to H^{**}(R)$ it suffices to identify the value of $f^*(\tau)$. The homomorphism $f^*\colon H^{0,1}(W) \to H^{0,1}(\overline{K})$ may be identified with $\mu_\ell(W) \to \mu_\ell(\overline{K})$, which is an isomorphism. Hence $f^*\colon H^{**}(W) \to H^{**}(\overline{K})$ is an isomorphism. The argument given for Proposition 4.8 establishes that $f^*\colon \mathcal{A}_{**}(W) \to \mathcal{A}_{**}(\overline{K})$ is an isomorphism. □

Corollary 6.3 *Let $\overline{F}$ be an algebraically closed field of characteristic p. The homomorphisms $W \to \overline{K}$ and $W \to \overline{F}$ induce isomorphisms of motivic Adams spectral sequences for $\mathbb{1}$ from the E_2 page onwards. In particular,* $\mathrm{Ext}(\overline{F}) \cong \mathrm{Ext}(W) \cong \mathrm{Ext}(\overline{K})$.

Lemma 6.4 *Let $f\colon \overline{k} \to \overline{K}$ be an extension of algebraically closed fields of characteristic 0. For all s and $w \geq 0$, base change induces an isomorphism $\pi_{s,w}(\overline{k}) \to \pi_{s,w}(\overline{K})$.*

Proof Let ℓ be prime. The maps $f^*\colon H_{**}(\bar{k}) \to H_{**}(\bar{K})$ and $f^*\colon \mathcal{A}_{**}(\bar{k}) \to \mathcal{A}_{**}(\bar{K})$ are isomorphisms, hence the induced map of cobar complexes $f^*\colon \mathcal{C}^*(\bar{k}) \to \mathcal{C}^*(\bar{K})$ is an isomorphism. It follows that the map $\mathfrak{M}_{\bar{k}}(\mathbb{1}) \to \mathfrak{M}_{\bar{K}}(\mathbb{1})$ is an isomorphism from the E_2 page onwards. The homomorphism $\mathbb{L} f^*\colon \pi_{**}(\mathbb{1}^\wedge_H(\bar{k})) \to \pi_{**}(\mathbb{1}^\wedge_H(\bar{K}))$ is therefore an isomorphism since it is compatible with the map of spectral sequences. Propositions 5.14 and 5.15 identify $\pi_{s,w}(\mathbb{1}^\wedge_H)$ with $\pi_{s,w}(\mathbb{1})^\wedge_\ell$ for all $s \geq w \geq 0$ over both $\bar{k}$ and $\bar{K}$. By the work of Ananyevsky, Levine and Panin [3], the groups $\pi_{s,w}(\bar{k})$ and $\pi_{s,w}(\bar{K})$ are torsion for $s > w \geq 0$ and so they are the sum of their ℓ–primary parts. This establishes the result for $s > w \geq 0$. When $s = w \geq 0$, the result follows by Proposition 5.15 and Morel's identification of the groups $\pi_{n,n}(F)$. If $s < w$, the connectivity theorem applies and gives the isomorphism. $\square$

Corollary 6.5 *Let $\bar{K}$ be an algebraically closed field of characteristic 0. For any $n \geq 0$, the map $\mathbb{L}c\colon \pi^s_n \to \pi_{n,0}(\bar{K})$ is an isomorphism.*

Proof The statement is true when $\bar{K} = \mathbb{C}$ by Levine's theorem. The previous proposition extends the result to an arbitrary algebraically closed field of characteristic 0. $\square$

Theorem 6.6 *Let $\bar{F}$ be an algebraically closed field of characteristic p and let ℓ be a prime different from p. Then there is an isomorphism $\pi_{s,w}(\bar{F})^\wedge_\ell \cong \pi_{s,w}(\mathbb{C})^\wedge_\ell$ for all $s \geq w \geq 0$.*

Proof Consider the homomorphisms $\bar{F} \leftarrow W \to \bar{K}$. The induced maps on the motivic Adams spectral sequence are compatible with the maps of homotopy groups

$$\pi_{**}(\mathbb{1}^\wedge_H(\bar{F})) \leftarrow \pi_{**}(\mathbb{1}^\wedge_H(W)) \to \pi_{**}(\mathbb{1}^\wedge_H(\bar{K})).$$

By Corollary 6.3, the maps $\mathfrak{M}_{\bar{F}}(\mathbb{1}) \leftarrow \mathfrak{M}_W(\mathbb{1}) \to \mathfrak{M}_{\bar{K}}(\mathbb{1})$ are isomorphisms at the E_2 page, and so there are isomorphisms

$$\pi_{**}(\mathbb{1}^\wedge_H(\bar{F})) \cong \pi_{**}(\mathbb{1}^\wedge_H(W)) \cong \pi_{**}(\mathbb{1}^\wedge_H(\bar{K})).$$

For $s \geq w \geq 0$, Propositions 5.14 and 5.15 give isomorphisms

$$\pi_{s,w}(\mathbb{1}^\wedge_H(\bar{F})) \cong \pi_{s,w}(\bar{F})^\wedge_\ell \quad \text{and} \quad \pi_{s,w}(\mathbb{1}^\wedge_H(\bar{K})) \cong \pi_{s,w}(\bar{K})^\wedge_\ell.$$

The result now follows from Lemma 6.4. $\square$

Corollary 6.7 *Let $\bar{F}$ be an algebraically closed field of characteristic p and let ℓ be a prime different from p. The homomorphism $\mathbb{L}c\colon (\pi^s_n)^\wedge_\ell \to \pi_{n,0}(\bar{F})^\wedge_\ell$ is an isomorphism for all $n \geq 0$.*

Proof The previous theorem yields the following diagram for all $n \geq 0$:

$$\begin{tikzcd}
 & (\pi_n^s)^\wedge_\ell \arrow[dl, "\mathbb{L}c"'] \arrow[d, "\mathbb{L}c"] \arrow[dr, "\mathbb{L}c", "\cong"'] & \\
\pi_{n,0}(\bar{F})^\wedge_\ell & \pi_{n,0}(\mathbb{1}^\wedge_H(W)) \arrow[l, "\cong"'] \arrow[r, "\cong"] & \pi_{n,0}(\bar{K})^\wedge_\ell
\end{tikzcd}$$

The map $\mathbb{L}c\colon (\pi_n^s)^\wedge_\ell \to \pi_{n,0}(\bar{K})^\wedge_\ell$ is an isomorphism by Corollary 6.5, and so all of the maps in the above diagram are isomorphisms. □

Corollary 6.8 *For a finite field $\mathbb{F}_q$ with characteristic $p \neq \ell$, the group $(\pi_n^s)^\wedge_\ell$ is a summand of $\pi_{n,0}(\mathbb{F}_q)^\wedge_\ell$ for $n \geq 0$.*

Proof The map $\mathbb{L}c\colon \pi_n^s \to \pi_{n,0}(\bar{\mathbb{F}}_p)$ factors through $\pi_{n,0}(\mathbb{F}_q)$. Passing to the ℓ–completion, Corollary 6.7 implies the composition $(\pi_n^s)^\wedge_\ell \to \pi_{n,0}(\mathbb{F}_q)^\wedge_\ell \to \pi_{n,0}(\bar{\mathbb{F}}_p)^\wedge_\ell$ is an isomorphism. Hence the result. □

7 The motivic Adams spectral sequence for finite fields

We now analyze the two-complete stable stems $\hat{\pi}_{**}(\mathbb{F}_q) = \pi_{**}(\mathbb{F}_q)^\wedge_2$ when q is odd. The results of the previous section allow us to identify the n^{th} topological two-complete stable stem $\hat{\pi}_n^s = (\pi_n^s)^\wedge_2$ as a summand of $\hat{\pi}_{n,0}(\mathbb{F}_q)$. With this, we are able to analyze the MASS for $\mathbb{F}_q$ in a range. We remind the reader that these results assume Morel's connectivity theorem hold for $\mathbb{F}_q$, or the results hold without qualification for the fields $\widetilde{\mathbb{F}}_q$. For the remainder of this section, write H for the mod 2 motivic cohomology spectrum.

7.1 The E_2 page of MASS over $\mathbb{F}_q$ when $q \equiv 1 \bmod 4$

We will make frequent use of the calculation $H^{**}(\mathbb{F}_q;\mathbb{Z}/2) \cong \mathbb{Z}/2[\tau,u]/(u^2)$ which was given in Section 4. Recall τ and u are in bidegree $(0,1)$ and $(1,1)$, respectively.

Proposition 7.1 *The E_2 page of the mod 2 motivic Adams spectral sequence for the sphere spectrum over $\mathbb{F}_q$ with $q \equiv 1 \bmod 4$ is the trigraded algebra*

$$E_2 \cong \mathrm{Ext}(\mathbb{F}_q) \cong \mathbb{F}_2[\tau,u]/(u^2) \otimes_{\mathbb{F}_2[\tau]} \mathrm{Ext}(\bar{\mathbb{F}}_p).$$

We abuse notation and write τ and u for their duals. Hence in the above, τ and u are of bidegree $(0,-1)$ and $(-1,-1)$, respectively.

Proof Consult Dugger and Isaksen [12, Proposition 3.5] for a similar argument. Recall from Proposition 4.3 that we have $\mathcal{A}^{**}(\mathbb{F}_q) \cong \mathcal{A}^{**}(\overline{\mathbb{F}}_p) \otimes_{\mathbb{F}_2[\tau]} \mathbb{F}_2[\tau,u]/(u^2)$ and $H^{**}(\mathbb{F}_q) \cong H^{**}(\overline{\mathbb{F}}_p) \otimes \mathbb{F}_2[\tau,u]/(u^2)$. Since $\mathbb{F}_2[\tau,u]/(u^2)$ is flat as a module over $\mathbb{F}_2[\tau]$, a free resolution $H^{**}(\overline{\mathbb{F}}_p) \leftarrow P^\bullet$ by $\mathcal{A}^{**}(\overline{\mathbb{F}}_p)$ modules determines a free resolution $H^{**}(\mathbb{F}_q) \leftarrow P^\bullet \otimes \mathbb{F}_2[\tau,u]/(u^2)$. It is necessary here that $\mathrm{Sq}^1(\tau) = 0$ for $P^\bullet \otimes \mathbb{F}_2[\tau,u]/(u^2)$ to be a resolution of $\mathcal{A}^{**}(\mathbb{F}_q)$ modules. The canonical map

$$\mathrm{Hom}_{\mathcal{A}^{**}(\overline{\mathbb{F}}_p)}\big(-, H^{**}(\overline{\mathbb{F}}_p)\big) \otimes \mathbb{F}_2[\tau,u]/(u^2) \\ \to \mathrm{Hom}_{\mathcal{A}^{**}(\mathbb{F}_q)}\big(- \otimes \mathbb{F}_2[\tau,u]/(u^2), H^{**}(\mathbb{F}_q)\big)$$

is a natural isomorphism, since a generating set for a module M over $\mathcal{A}^{**}(\overline{\mathbb{F}}_p)$ is also a generating set for $M \otimes \mathbb{F}_2[\tau,u]/(u^2)$ over $\mathcal{A}^{**}(\mathbb{F}_q)$ by Proposition 4.3. We conclude that $\mathrm{Ext}(\overline{\mathbb{F}}_p) \otimes \mathbb{F}_2[\tau,u]/(u^2) \cong \mathrm{Ext}(\mathbb{F}_q)$. □

By the previous proposition, the irreducible elements of $\mathrm{Ext}(\mathbb{C})$ are also irreducible elements of $\mathrm{Ext}(\mathbb{F}_q)$ when $q \equiv 1 \bmod 4$. The only additional irreducible element in $\mathrm{Ext}(\mathbb{F}_q)$ is the class u. The irreducible elements of $\mathrm{Ext}(\mathbb{F}_q)$ up to stem $s = 21$ can be found in Table 1. These were obtained by consulting Isaksen [26, Table 8] and independently verified by computer calculation by Fu and Wilson [16].

element	filtration (f,s,w)	element	filtration (f,s,w)	element	filtration (f,s,w)
u	$(0,-1,-1)$	c_0	$(3,8,5)$	e_0	$(4,17,10)$
τ	$(0,0,-1)$	Ph_1	$(5,9,5)$	P^2h_1	$(9,17,9)$
h_0	$(1,0,0)$	Ph_2	$(5,11,6)$	f_0	$(4,18,10)$
h_1	$(1,1,1)$	d_0	$(4,14,8)$	P^2h_2	$(9,19,10)$
h_2	$(1,3,2)$	h_4	$(1,15,8)$	c_1	$(3,19,11)$
h_3	$(1,7,4)$	Pc_0	$(7,16,9)$	$[\tau g]$	$(4,20,11)$

Table 1: The irreducible elements of $\mathrm{Ext}(\mathbb{F}_q)$ with $q \equiv 1 \bmod 4$ in stem $s \le 21$

We now investigate the motivic May spectral sequence over the finite field $\mathbb{F}_q$ when $q \equiv 1 \bmod 4$. We will find it useful for calculating Massey products in the MASS.

Definition 7.2 Write J for the cokernel of the map $\eta_L\colon H_{**} \to \mathcal{A}_{**}$ in the category of bigraded $\mathbb{F}_2$ vector spaces and consider the increasing filtration of $\mathcal{A}_{**}$ given by

$$F_n\mathcal{A}_{**} = \ker(\mathcal{A}_{**} \xrightarrow{\Delta^n} \mathcal{A}_{**}^{\otimes n+1} \to J^{\otimes n+1}).$$

This filtration on $\mathcal{A}_{**}$ induces a filtration on the cobar complex $(\mathcal{C}, d)$ defined by Ravenel in [40, Definition A1.2.11]. The filtration of the cobar complex is compatible and leads to a spectral sequence [40, Theorem A1.3.9] called the motivic May spectral sequence.

Following the work of Dugger and Isaksen [12, Section 5], we are able to identify the structure of the motivic May spectral sequence over a finite field $\mathbb{F}_q$ when $q \equiv 1 \bmod 4$.

Proposition 7.3 *The associated graded Hopf algebroid $E^0\mathcal{A}_{**}$ to the filtration $F^*\mathcal{A}_{**}$ of the motivic dual Steenrod algebra over a finite field $\mathbb{F}_q$ when $q \equiv 1 \bmod 4$ is the exterior algebra over $H_{**}(\mathbb{F}_q) \cong \mathbb{F}_2[\tau, u]/(u^2)$*

$$E^0\mathcal{A}_{**} \cong E_{H_{**}(\mathbb{F}_q)}(\tau_i, \xi_j^{2^k} \mid i \geq 0, j \geq 1, k \geq 0).$$

If each generator ζ_i of $E^0\mathcal{A}_*^{\mathrm{top}}$ is assigned the weight of τ_{i-1} for $i \geq 1$ and $\zeta_i^{2^j}$ is assigned the weight of $\xi_i^{2^{j-1}}$ for $j \geq 1$, there is an isomorphism of trigraded algebras

$$E^0\mathcal{A}_{**} \cong \mathbb{F}_2[\tau, u]/(u^2) \otimes_{\mathbb{F}_2} E^0\mathcal{A}_*,$$

where $\mathcal{A}_*$ denotes the topological dual Steenrod algebra, which was studied by Milnor in [31].

Proof Since $u \in F^0\mathcal{A}_{**}(\mathbb{F}_q)$ and $\mathcal{A}_{**}(\mathbb{F}_q) \cong \mathbb{F}_2[\tau, u]/(u^2) \otimes_{\mathbb{F}_2[\tau]} \mathcal{A}_{**}(\mathbb{C})$, there are isomorphisms $F^n\mathcal{A}_{**}(\mathbb{F}_q) \cong F^n\mathcal{A}_{**}(\mathbb{C}) \otimes_{\mathbb{F}_2[\tau]} \mathbb{F}_2[\tau, u]/(u^2)$. Over $\mathbb{C}$, there is an isomorphism

$$E^0\mathcal{A}_{**}(\mathbb{C}) \cong \mathbb{F}_2[\tau] \otimes_{\mathbb{F}_2} E^0\mathcal{A}_*,$$

which follows by dualizing the result of Dugger and Isaksen in [12, Proposition 5.2(a)]. The result now follows as $\mathbb{F}_2[\tau] \to \mathbb{F}_2[\tau, u]/(u^2)$ is flat. □

Proposition 7.4 *The E_2 page of the motivic May spectral sequence over a finite field $\mathbb{F}_q$ with $q \equiv 1 \bmod 4$ is given by*

$$\begin{aligned} E_2^{m,f,s,w} &= \mathrm{Ext}^{f,(s+f,w,m)}_{E^0\mathcal{A}_{**}(\mathbb{F}_q)}(H_{**}(\mathbb{F}_q), H_{**}(\mathbb{F}_q)) \\ &\cong \mathbb{F}_2[\tau, u]/(u^2) \otimes_{\mathbb{F}_2[\tau]} \mathrm{Ext}^{f,(s+f,w,m)}_{E^0\mathcal{A}_{**}(\mathbb{C})}(H_{**}(\mathbb{C}), H_{**}(\mathbb{C})), \end{aligned}$$

where f is the Adams filtration (or homological degree), s is the stem, w is the motivic weight and m is the May filtration. The differential d_r changes grading as $d_r\colon E_r^{m,f,s,w} \to E_r^{m+r-1,f+1,s-1,w}$. The motivic May spectral sequence converges to $\mathrm{Ext}_{\mathcal{A}_{**}}(H_{**}, H_{**})$.

To be consistent with the work of Dugger and Isaksen [12; 26], *we write the grading of an element in the May spectral sequence in the form (m, f, s, w).*

Proof The E_2 page of the motivic May spectral sequence is identified by Ravenel in [40, Theorem A1.3.9] in terms of the derived functors of the cotensor product $H_{**}\square_{\mathcal{A}_{**}}-$. In this case, the natural isomorphism $\mathrm{Hom}_{\mathcal{A}_{**}}(H_{**}, -) \cong H_{**}\square_{\mathcal{A}_{**}}-$

identifies the Cotor groups with the Ext groups in the statement of the proposition. The second isomorphism follows formally from the result over $\mathbb{C}$ established by Dugger and Isaksen in [12, Proposition 5.2(b)] by the flatness of $\mathbb{F}_2[\tau,u]/(u^2)$ over $\mathbb{F}_2[\tau]$. $\square$

A description of the motivic May spectral sequence E_2 page over $\mathbb{C}$ is given by Dugger and Isaksen in [12, Section 5] up to the 36 stem, from which one obtains a description of the motivic May spectral sequence E_2 page over $\mathbb{F}_q$ when $q \equiv 1 \bmod 4$ using the previous proposition. One must simply add u to the list of generators of the E_2 page given in [12, Table 1] and the relation $u^2 = 0$.

7.2 The E_2 page of MASS over $\mathbb{F}_q$ when $q \equiv 3 \bmod 4$

For a finite field $\mathbb{F}_q$ with $q \equiv 3 \bmod 4$, the E_2 page of the MASS can be identified in a range using the ρ–Bockstein spectral sequence (ρ–BSS) which was introduced by Hill in [18]. Here $\rho = [-1]$ is the nonzero class in $H^{1,1}(\mathbb{F}_q) \cong \mathbb{F}_q^\times/2$, since -1 is not a square in $\mathbb{F}_q^\times$. We briefly describe the construction of the ρ–BSS and refer the reader to Dugger and Isaksen [13] or Ormsby [38; 37] for more details.

Let $\mathcal{C}$ be the cobar construction corresponding to the Hopf algebroid

$$\big(\mathbb{F}_2[\tau,\rho]/(\rho^2), \mathcal{A}_{**}(\mathbb{F}_q)\big).$$

The filtration of $\mathcal{C}$ given by $0 \subseteq \rho\mathcal{C} \subseteq \mathcal{C}$ determines a spectral sequence, which in this case is just the long exact sequence associated to the short exact sequence of complexes

$$0 \to \rho\mathcal{C} \to \mathcal{C} \to \mathcal{C}/\rho\mathcal{C} \to 0.$$

Note that $\rho\mathcal{C}$ and $\mathcal{C}/\rho\mathcal{C}$ are both isomorphic to the cobar construction over $\mathbb{C}$. Hence we have the following long exact sequence:

$$\cdots \to \rho\mathrm{Ext}^{i,(*,*)}(\mathbb{C}) \to \mathrm{Ext}^{i,(*,*)}(\mathbb{F}_q) \to \mathrm{Ext}^{i,(*,*)}(\mathbb{C}) \xrightarrow{d_1} \rho\mathrm{Ext}^{i+1,(*,*)}(\mathbb{C}) \to \cdots$$

In spectral sequence notation, the E_1 page is given by

$$E_1^{\epsilon,f,(s,w)} \cong \begin{cases} \mathrm{Ext}^{f,(s,w)}(\mathbb{C}) & \text{if } \epsilon = 0, \\ \rho\mathrm{Ext}^{f,(s+1,w+1)}(\mathbb{C}) & \text{if } \epsilon = 1, \\ 0 & \text{otherwise,} \end{cases}$$

with differential $d_1\colon E_1^{\epsilon,f,(s,w)} \to E_1^{\epsilon+1,f+1,(s-1,w)}$. The differential d_1 satisfies the Leibniz rule, so it suffices to identify the differential on irreducible elements. We identify all differentials up to the 20 stem by hand in the following proposition; these calculations have been verified by computer calculations.

Proposition 7.5 *In the ρ–BSS for $\mathbb{F}_q$ with $q \equiv 3 \bmod 4$, every irreducible element x of $\mathrm{Ext}(\mathbb{C})$ in stem $s \leq 19$ other than τ has $d_1(x) = 0$. Also, $d_1(\tau) = \rho h_0$ and $d_1([\tau g]) = \rho h_2 e_0$. Here $[\tau g]$ is the irreducible element of $\mathrm{Ext}(\mathbb{C})$ in stem 20, weight 11 and filtration 4.*

Proof The differential d_1 vanishes on all irreducible classes in $\mathrm{Ext}(\mathbb{C})$ up to stem 20 for degree reasons except for possibly τ, f_0 and $[\tau g]$. The class τ cannot survive the ρ–BSS, since if it did, it would contribute a nonzero element to $\mathrm{Ext}^{0,0,-1}(\mathbb{F}_q) \cong \mathrm{Hom}_{\mathcal{A}}^{0,-1}(H^{**}, H^{**})$, which is trivial. We conclude $d_1(\tau) = \rho h_0$, because this is the only possible nonzero value for $d_1(\tau)$.

The two possibilities for $d_1(f_0)$ are 0 and $\rho h_1 e_0$. Since $h_1 f_0 = 0$ in $\mathrm{Ext}(\mathbb{C})$, we must have $d_1(h_1 f_0) = h_1 d_1(f_0) = 0$; hence $d_1(f_0)$ is annihilated by h_1. But as $\rho h_1 e_0$ is not annihilated by h_1, we must have $d_1(f_0) = 0$.

The only possible nonzero value for $d_1([\tau g])$ is $\rho h_2 e_0$. From the relation $h_0[\tau g] = \tau h_2 e_0$, we calculate $d_1(\tau h_2 e_0) = \rho h_0 h_2 e_0$ and $d_1(h_0[\tau g]) = h_0 d_1([\tau g])$. Hence $h_0 d_1([\tau g]) = h_0 \rho h_2 e_0$, from which the result follows. □

Example 7.6 Since $d_1(h_1) = 0$, we conclude $d_1(\tau h_1) = \rho h_0 h_1 = 0$, as $h_0 h_1$ vanishes in $\mathrm{Ext}(\mathbb{C})$. Hence there is a class $[\tau h_1] \in \mathrm{Ext}^{1,(1,0)}(\mathbb{F}_q)$ which is irreducible.

With this analysis of the ρ–BSS for $\mathbb{F}_q$ with $q \equiv 3 \bmod 4$, the structure of $\mathrm{Ext}(\mathbb{F}_q)$ as a graded abelian group up to stem 21 follows immediately and we may further identify all irreducible elements in this range. The results of this proposition were verified by computer calculation by Fu and Wilson [16].

Proposition 7.7 *When $q \equiv 3 \bmod 4$, the irreducible elements of $\mathrm{Ext}(\mathbb{F}_q)$ up to stem $s = 21$ are given in Table 2.*

Proof The structure of $\mathrm{Ext}(\mathbb{F}_q)$ as an abelian group follows directly from the ρ–BSS and the differentials calculated in Proposition 7.5. We now explain why the tabulated elements comprise all of the irreducible elements in this range. If $y \in H^{**}(\rho\mathcal{C}) \cong \rho\,\mathrm{Ext}(\mathbb{C})$, then we may write $y = \rho \cdot x$ with $x \in H^{**}(\mathcal{C}/\rho\mathcal{C}) \cong \mathrm{Ext}(\mathbb{C})$. So long as $x \neq 1$ and $d_1(x) = 0$, the element y is reducible. By Proposition 7.5 we conclude the only irreducible elements arising from $\rho\,\mathrm{Ext}(\mathbb{C})$ in this range are ρ, $[\rho\tau]$ and $[\rho\tau g]$.

Now consider an element x of $H^{**}(\mathcal{C}/\rho\mathcal{C}) \cong \mathrm{Ext}(\mathbb{C})$ which survives the ρ–BSS, that is, $d_1(x) = 0$. Then x is irreducible in $\mathrm{Ext}(\mathbb{F}_q)$ if and only if for any factorization $x = a \cdot b$ in $\mathrm{Ext}(\mathbb{C})$ with $d_1(a) = d_1(b) = 0$ it follows $a = 1$ or $b = 1$. This observation identifies all of the remaining irreducible elements in $\mathrm{Ext}(\mathbb{F}_q)$ in the range $s \leq 21$. □

element	filtration (f,s,w)	element	filtration (f,s,w)	element	filtration (f,s,w)
ρ	$(0,-1,-1)$	$[\tau c_0]$	$(3,8,4)$	$[\tau P c_0]$	$(7,16,8)$
$[\rho\tau]$	$(0,-1,-2)$	Ph_1	$(5,9,5)$	e_0	$(4,17,10)$
$[\tau^2]$	$(0,0,-2)$	$[\tau P h_1]$	$(5,9,4)$	P^2h_1	$(9,17,9)$
h_0	$(1,0,0)$	Ph_2	$(5,11,6)$	$[\tau P^2 h_1]$	$(9,17,8)$
h_1	$(1,1,1)$	$[\tau h_0 h_3^2]$	$(3,14,7)$	f_0	$(4,18,10)$
$[\tau h_1]$	$(1,1,0)$	d_0	$(4,14,8)$	P^2h_2	$(9,19,10)$
h_2	$(1,3,2)$	$[\tau h_0^2 d_0]$	$(6,14,7)$	c_1	$(3,19,11)$
$[\tau h_2^2]$	$(2,6,3)$	h_4	$(1,15,8)$	$[\tau c_1]$	$(3,19,10)$
h_3	$(1,7,4)$	$[\tau h_0^7 h_4]$	$(8,15,7)$	$[\rho\tau g]$	$(4,19,10)$
$[\tau h_0^3 h_3]$	$(4,7,3)$	Pc_0	$(7,16,9)$	$[\tau^2 g]$	$(4,20,10)$
c_0	$(3,8,5)$				

Table 2: The irreducible elements of $\mathrm{Ext}(\mathbb{F}_q)$ with $q \equiv 3 \bmod 4$ in stem $s \leq 21$

Remark 7.8 Although Proposition 7.7 lists all of the irreducible elements in $\mathrm{Ext}(\mathbb{F}_q)$ when $q \equiv 3 \bmod 4$ in a range, there are hidden products in the ρ–BSS. For example, the product $[\tau h_2^2]\cdot h_1 = \rho c_0$ is hidden in the ρ–BSS. We obtained this product by computer calculation, however the arguments by Dugger and Isaksen in [13, Lemma 6.2] can be used to obtain some products by hand.

7.3 The Adams spectral sequence for $H\mathbb{Z}[p^{-1}]$

We begin with the motivic Adams spectral sequence for $X = H\mathbb{Z}[p^{-1}]$ over a finite field $\mathbb{F}_q$ of characteristic p, as defined in Definition 5.2. In Propositions 7.10 and 7.11 we identify the differentials for $\mathfrak{M}_{\mathbb{F}_q}(H\mathbb{Z}[p^{-1}])$, which converges to $\pi_{**}(H\mathbb{Z}[p^{-1}]^\wedge_2) \cong H_{**}(\mathbb{F}_q;\mathbb{Z})^\wedge_2$. We accomplish this by working backwards from our knowledge of the target group $H^{**}(\mathbb{F}_q;\mathbb{Z})^\wedge_2$, which is isomorphic to $H^*_{\mathrm{et}}(\mathbb{F}_q;\mathbb{Z}_2(*))$ as a consequence of the Beilinson–Lichtenbaum conjecture. Soulé's calculation of $H^*_{\mathrm{et}}(\mathbb{F}_q;\mathbb{Z}_2(*))$ in [44, Paragraphe IV.2] then gives

$$\pi_{s,w}(H\mathbb{Z}[p^{-1}]) \cong \begin{cases} \mathbb{Z}_\ell & \text{if } s = w = 0, \\ \mathbb{Z}/(q^w-1)^\wedge_2 & \text{if } s = -1 \text{ and } w \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Although the spectrum $H\mathbb{Z}[p^{-1}]$ is cellular by the Hopkins–Morel theorem proven by Hoyois [21, Section 8.1], it is unclear if it is of finite type. Instead of relying on Proposition 5.10 for convergence, we establish a weak equivalence of the H–nilpotent completion of $H\mathbb{Z}[p^{-1}]$ with $H\mathbb{Z}^\wedge_2$.

Lemma 7.9 *Let $\mathbb{F}_q$ be a finite field of characteristic $p \neq 2$. The H–nilpotent completion of $H\mathbb{Z}[p^{-1}]$ is weakly equivalent to $H\mathbb{Z}^\wedge_2$.*

Proof We will show that the tower $H\mathbb{Z}/2 \leftarrow H\mathbb{Z}/2^2 \leftarrow H\mathbb{Z}/2^3 \leftarrow \cdots$ under $H\mathbb{Z}[p^{-1}]$ is an H–nilpotent resolution under $H\mathbb{Z}[p^{-1}]$ (as defined by Bousfield in [6, Definition 5.6]). It will then follow that the homotopy limit of this tower is weakly equivalent to the H–nilpotent completion of $H\mathbb{Z}[p^{-1}]$; that is, $H\mathbb{Z}_2^\wedge \cong H\mathbb{Z}[p^{-1}]_H^\wedge$ by the observations of Dugger and Isaksen in [12, Section 7.7], which shows Bousfield's result [6, Proposition 5.8] holds in the motivic stable homotopy category.

The spectrum $H\mathbb{Z}[p^{-1}]$ is the homotopy colimit of the diagram $H\mathbb{Z} \xrightarrow{p\cdot} H\mathbb{Z} \xrightarrow{p\cdot} \cdots$. From the triangle $H\mathbb{Z} \xrightarrow{2^\nu\cdot} H\mathbb{Z} \to H\mathbb{Z}/2^\nu$, we obtain, after inverting p, a triangle $H\mathbb{Z}[p^{-1}] \xrightarrow{2^\nu\cdot} H\mathbb{Z}[p^{-1}] \to H\mathbb{Z}/2^\nu$ since $p \neq 2$ and $H\mathbb{Z}/2^\nu \xrightarrow{p\cdot} H\mathbb{Z}/2^\nu$ is a homotopy equivalence. Consider the following cofibration sequence of towers:

$$\begin{tikzcd}
H\mathbb{Z}[p^{-1}] \arrow[d, "="'] & H\mathbb{Z}[p^{-1}] \arrow[l, "2\cdot"'] \arrow[d, "2\cdot"] & H\mathbb{Z}[p^{-1}] \arrow[l, "2\cdot"'] \arrow[d, "2^2\cdot"] & \cdots \arrow[l] \\
H\mathbb{Z}[p^{-1}] \arrow[d] & H\mathbb{Z}[p^{-1}] \arrow[l, "="'] \arrow[d] & H\mathbb{Z}[p^{-1}] \arrow[l, "="'] \arrow[d] & \cdots \arrow[l] \\
\mathrm{pt} & H\mathbb{Z}/2 \arrow[l] & H\mathbb{Z}/2^2 \arrow[l] & \cdots \arrow[l]
\end{tikzcd}$$

It is clear that $H\mathbb{Z}/2^\nu$ is H–nilpotent for all $\nu \geq 1$. For any H–nilpotent spectrum N we show that the induced map $\operatorname{colim}_\nu \mathcal{SH}_{\mathbb{F}_q}(H\mathbb{Z}/2^\nu, N) \to \mathcal{SH}_{\mathbb{F}_q}(H\mathbb{Z}[p^{-1}], N)$ is an isomorphism following the proof of Bousfield [6, Lemma 5.7]. This isomorphism holds if and only if

$$\operatorname{colim}\bigl\{\mathcal{SH}_{\mathbb{F}_q}\bigl(H\mathbb{Z}\bigl[\tfrac{1}{p}\bigr], N\bigr) \xrightarrow{2\cdot} \mathcal{SH}_{\mathbb{F}_q}\bigl(H\mathbb{Z}\bigl[\tfrac{1}{p}\bigr], N\bigr)\bigr\} \cong \mathcal{SH}_{\mathbb{F}_q}\bigl(H\mathbb{Z}\bigl[\tfrac{1}{p}\bigr], N\bigr)\bigl[\tfrac{1}{2}\bigr]$$

vanishes for all H–nilpotent N. This follows by an inductive proof with the following filtration of the H–nilpotent spectra given in [6, Lemma 3.8]. Take C_0 to be the collection of spectra $H \wedge X$ for X any spectrum, and let C_{m+1} be the collection of the spectra N for which either N is a retract of an element of C_m or there is a triangle $X \to N \to Z$ with X and Z in C_m.

If $N = H \wedge X$, it is clear that $\mathcal{SH}_{\mathbb{F}_q}(H\mathbb{Z}[p^{-1}], N) \xrightarrow{2} \mathcal{SH}_{\mathbb{F}_q}(H\mathbb{Z}[p^{-1}], N)$ is the zero map, which establishes the base case. If the claim holds for N in filtration C_m, the claim holds for N in filtration C_{m+1} by a standard argument. The claim now follows. □

Proposition 7.10 *The mod 2 motivic Adams spectral sequence for $X = H\mathbb{Z}[p^{-1}]$ over $\mathbb{F}_q$ when $q \equiv 1 \bmod 4$ has E_1 page given by*

$$E_1 \cong \mathbb{F}_2[\tau, u, h_0]/(u^2),$$

where $h_0 \in E_1^{1,(0,0)}$.

Write ν_2 for the 2–adic valuation and $\epsilon(q)$ for $\nu_2(q-1)$. For all $r \geq 1$ the differentials d_r vanish on $u\tau^j$ and h_0^j. If $r < \epsilon(q)+\nu_2(j)$ the differentials $d_r\tau^j$ vanish and we have

$$d_{\epsilon(q)+\nu_2(j)}\tau^j = u\tau^{j-1}h_0^{\epsilon(q)+\nu_2(j)}.$$

In particular, the differential d_1 is trivial, so $E_2 \cong E_1$.

Proof We build the following H^{**}–Adams resolution of $H\mathbb{Z}[p^{-1}]$ utilizing the triangles constructed in Lemma 7.9:

$$\begin{tikzcd}
H\mathbb{Z}[p^{-1}] \arrow[dr, "j_0"'] & & H\mathbb{Z}[p^{-1}] \arrow[ll, "2\cdot"'] \arrow[dr, "j_1"'] & & H\mathbb{Z}[p^{-1}] \arrow[ll, "2\cdot"'] & \cdots \arrow[l] \\
& H \arrow[ur, "\partial_0"', "\bullet"] & & H \arrow[ur, "\partial_1"', "\bullet"] & &
\end{tikzcd} \tag{7-1}$$

The spectrum $H\mathbb{Z}[p^{-1}]$ is cellular, and so the motivic Adams spectral sequence for $X = H\mathbb{Z}[p^{-1}]$ converges to $\pi_{**}(H\mathbb{Z}[p^{-1}]^\wedge_H)$ by Proposition 5.10. Lemma 7.9 shows that $\pi_{**}(H\mathbb{Z}[p^{-1}]^\wedge_H) \cong \pi_{**}(H\mathbb{Z}^\wedge_2)$, so the spectral sequence converges:

$$E_2^{f,(s,w)} \Rightarrow H^{-s,-w}(\mathbb{F}_q;\mathbb{Z})^\wedge_2.$$

The groups $H^{s,w}(\mathbb{F}_q;\mathbb{Z})^\wedge_2$ are isomorphic to the groups $H^s_{\mathrm{et}}(\mathbb{F}_q;\mathbb{Z}_2(w))$ which were calculated by Soulé in [44, Paragraphe IV.2]. If $q \equiv 1 \bmod 4$,

$$H^{-s,-w}(\mathbb{F}_q;\mathbb{Z})^\wedge_2 \cong \begin{cases} \mathbb{Z}_\ell & \text{if } s = w = 0, \\ \mathbb{Z}/(q^w-1)^\wedge_2 & \text{if } s = -1 \text{ and } w \geq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7-2}$$

Note that $\nu_2(q^w-1) = \epsilon(q)+\nu_2(w)$ for all natural numbers w. The formulas for the differentials on τ^j are the only choice to give $H^{**}(\mathbb{F}_q;\mathbb{Z})^\wedge_2$ as the E_∞ term. □

Proposition 7.11 *The mod 2 motivic Adams spectral sequence for $X = H\mathbb{Z}[p^{-1}]$ over $\mathbb{F}_q$ when $q \equiv 3 \bmod 4$ has E_1 page given by*

$$E_1 \cong \mathbb{F}_2[\tau,\rho,h_0]/(\rho^2),$$

where $h_0 \in E_1^{1,(0,0)}$.

For all $r \geq 1$ the differentials d_r vanish on $\rho\tau^j$ and h_0^j. For odd natural numbers j, we calculate $d_1(\tau^j) = \rho\tau^{j-1}h_0$. Write $\lambda(q)$ for $\nu_2(q^2-1)$. If $r < \lambda(q)+\nu_2(n)$ the differentials $d_r\tau^{2n}$ vanish and

$$d_{\lambda(q)+\nu_2(n)}\tau^{2n} = \rho\tau^{2n-1}h_0^{\lambda(q)+\nu_2(n)}.$$

Proof The proof of the previous proposition goes through, except the target groups $H^{-s,-w}(\mathbb{F}_q;\mathbb{Z})^\wedge_2$ force different differentials in the spectral sequence when $q \equiv 3 \bmod 4$. Soulé's calculation in (7-2) shows the order of $H^{1,1}(\mathbb{F}_q;\mathbb{Z})^\wedge_2$ is $\nu_2(q-1)=1$, so we conclude $d_1(\tau)=\rho h_0$. As we have $\nu_2(q^{2j}-1)=\lambda(q)+\nu_2(j)$ for all natural numbers j, the claimed formulas for the differentials on τ^{2n} hold. □

Corollary 7.12 *In the MASS of $\mathbb{1}$ over a finite field $\mathbb{F}_q$ with $q \equiv 1 \bmod 4$, the differentials $d_r(\tau^j)$ vanish when $r < \epsilon(q)+\nu_2(j)$ and*

$$d_{\epsilon(q)+\nu_2(j)}\tau^j = u\tau^{j-1}h_0^{\epsilon(q)+\nu_2(j)}.$$

In the MASS of $\mathbb{1}$ over a finite field $\mathbb{F}_q$ with $q \equiv 3 \bmod 4$, the differentials $d_r([\tau^2]^n)$ vanish when $r < \lambda(q)+\nu_2(n)$ and

$$d_{\lambda(q)+\nu_2(n)}[\tau^2]^n = [\rho\tau][\tau^2]^{n-1}h_0^{\lambda(q)+\nu_2(n)}.$$

Proof The unit map $\mathbb{1} \to H\mathbb{Z}[p^{-1}]$ induces a map of motivic Adams spectral sequences $\mathfrak{M}(\mathbb{1}) \to \mathfrak{M}(H\mathbb{Z}[p^{-1}])$. On the E_2 page, observe that when $q \equiv 1 \bmod 4$ the classes τ and u map to τ and u, respectively. When $q \equiv 3 \bmod 4$, the classes $[\tau^2]$, ρ, $[\rho\tau]$ map to $[\tau^2]$, ρ, $[\rho\tau]$, respectively. The identification of the differentials in the MASS for $H\mathbb{Z}[p^{-1}]$ in Propositions 7.10 and 7.11 then force the differentials stated in the corollary. □

Example 7.13 When $q \equiv 3 \bmod 4$, the Massey product $\langle \rho, \rho, h_0\rangle$ in the mod 2 motivic Adams spectral sequence for $H\mathbb{Z}[p^{-1}]$ is $\rho\tau$. Since we have $\rho^2 = 0$ and $d_1(\tau)=\rho h_0$, it follows that $0+\rho\tau$ is in the Massey product. It is straightforward to verify that the indeterminacy is trivial.

7.4 Stable stems over $\mathbb{F}_q$

We now begin an analysis of the differentials in the MASS to identify the two-complete stable stems over $\mathbb{F}_q$. To assist the reader with the computations presented below, Figures 1 and 3 in Section 9 display E_2 page charts of the MASS over $\mathbb{F}_q$. Throughout this section, $\mathbb{F}_q$ is a finite field with q elements where q is odd, and we write $\widehat{G}$ for the two-completion of an abelian group G.

Corollary 6.8 shows that $\widehat{\pi}^s_n$ is a summand of $\widehat{\pi}_{n,0}(\mathbb{F}_q)$ for all $n \geq 0$. We will soon see that for small values of $n \geq 0$ we have $\widehat{\pi}_{n,0}(\mathbb{F}_q) \cong \widehat{\pi}^s_n \oplus \widehat{\pi}^s_{n+1}$. However this pattern fails when $n = 19$ and $q \equiv 1 \bmod 4$.

Lemma 7.14 *For a finite field $\mathbb{F}_q$ with q odd, there is an isomorphism $\pi_{0,0}(\mathbb{F}_q) \cong \pi^s_0 \oplus \pi^s_1$.*

Proof The stem $\pi_{0,0}(\mathbb{F}_q)$ is isomorphic to the Grothendieck–Witt group $\mathrm{GW}(\mathbb{F}_q)$ by Morel [32]. The isomorphism $\mathrm{GW}(\mathbb{F}_q) \cong \mathbb{Z} \oplus \mathbb{Z}/2$ was established by Scharlau in [42, Chapter 2, Section 3.3]. Recall that $\pi_0^s \cong \mathbb{Z}$ and $\pi_1^s \cong \mathbb{Z}/2$. Hence we conclude $\pi_{0,0}(\mathbb{F}_q) \cong \pi_0^s \oplus \pi_1^s$. □

Morel's calculation of $\pi_{0,0}(\mathbb{F}_q)$ shows that $2 = (1-\epsilon) + \rho\eta$, hence multiplication by 2 in $\pi_{**}(\mathbb{F}_q)$ is detected in the mod 2 motivic Adams spectral sequence by the class $h_0 + \rho h_1$ in $\mathrm{Ext}(\mathbb{F}_q)$. This is needed to solve the extension problems when passing from the Adams spectral sequence E_∞ page to the stable stems.

Proposition 7.15 *When* $q \equiv 1 \bmod 4$ *and* $0 \le n \le 18$, *there is an isomorphism* $\widehat{\pi}_{n,0}(\mathbb{F}_q) \cong \widehat{\pi}_n^s \oplus \widehat{\pi}_{n+1}^s$.

Proof Lemma 7.14 takes care of the case when $n = 0$. We now focus on $0 < n \le 18$ where the mod 2 MASS over $\mathbb{F}_q$ converges to the groups $\widehat{\pi}_{n,0}(\mathbb{F}_q)$ by Propositions 5.10 and 5.14.

The irreducible elements of $\mathrm{Ext}(\mathbb{F}_q)$ in this range are given in Table 1. All differentials d_r for $r \ge 2$ vanish on h_0, h_1, h_3, c_0, Ph_1, d_0, Pc_0, P^2h_1 for degree reasons. As $\widehat{\pi}_{3,0}(\mathbb{F}_q)$ must contain $\widehat{\pi}_3^s \cong \mathbb{Z}/8$ as a summand by Corollary 6.8, we conclude $d_2(\tau^2 h_2) = \tau^2 d_2(h_2) = 0$. The only possible nonzero value for $d_2(h_2)$ is uh_1^3. If $d_2(h_2) = uh_1^3$, then $d_2(\tau^2 h_2) = u\tau^2 h_1^3$ would be nonzero by the product structure of $\mathrm{Ext}(\mathbb{F}_q)$ in Proposition 7.1, a contradiction. Hence $d_2(h_2) = 0$.

The nonzero Massey product $Ph_2 = \langle h_3, h_0^4, h_2\rangle$ has no indeterminacy, because $h_3 E_2^{4,(3,2)} + E_2^{4,(7,4)} h_2 = 0$. Since $\widehat{\pi}_{11}^s \cong \mathbb{Z}/8$ is a summand of $\widehat{\pi}_{11,0}$, the differential $d_2(Ph_2)$ must vanish. The nonzero Massey product $P^2h_2 = \langle h_3, h_0^4, h_2\rangle$ has no indeterminacy, because $h_3 E_2^{8,(11,6)} + E_2^{4,(7,4)} Ph_2 = 0$. Since $d_2(Ph_2) = 0$, the topological result of Moss [36, Theorem 1.1(ii)] implies $d_2(P^2h_2) = 0$.

The comparison map $\mathfrak{M}(\mathbb{F}_q) \to \mathfrak{M}(\overline{\mathbb{F}}_p)$ shows that $d_2(h_4)$ and $d_3(h_0h_4)$ must be nonzero, as these differentials are nonzero in $\mathfrak{M}(\overline{\mathbb{F}}_p)$ by Corollary 6.3 and calculations of Isaksen [26, Table 8] over $\mathbb{C}$. The only possible choice for $d_2(h_4)$ is $h_0h_3^2$, but $d_3(h_0h_4)$ is either h_0d_0 or $h_0d_0 + uh_1d_0$. In order to have $\widehat{\pi}_{14}^s \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2$ as a summand of $\widehat{\pi}_{14,0}$, we must have $d_3(h_0h_4) = h_0d_0$. A similar argument establishes $d_2(e_0) = h_1^2 d_0$ and $d_2(f_0) = h_0^2 e_0$. Note that $d_4(h_0^3 h_4) = 0$ for degree reasons.

The elements in weight 0 are all of the form $\tau^j x$ or $u\tau^{j-1}x$ where x is not a multiple of τ and of weight j. The differentials of the elements in weight 0 are now readily identified by using the Leibniz rule from Corollary 7.12. Since $\widehat{\pi}_n^s$ is a summand of $\widehat{\pi}_{n,0}(\mathbb{F}_q)$ for all $n \ge 0$, we see that there are no hidden 2–extensions for $0 < n \le 18$. □

In the proof of the following proposition, we provide some technical details in footnotes for the convenience of the reader. We follow the convention of Dugger and Isaksen [12] and write the grading of an element in the motivic May spectral sequence as (m,s,f,w) where m is the May filtration, s is the stem, f is the Adams filtration and w is the motivic weight. However, we continue to write the grading in the MASS as $(f,(s,w))$.

Proposition 7.16 *When $q \equiv 1 \bmod 4$, there are isomorphisms*

$$\hat{\pi}_{19,0}(\mathbb{F}_q) \cong (\mathbb{Z}/8 \oplus \mathbb{Z}/2) \oplus \mathbb{Z}/4 \quad \text{and} \quad \hat{\pi}_{20,0}(\mathbb{F}_q) \cong \mathbb{Z}/8 \oplus \mathbb{Z}/2.$$

In particular, when $q \equiv 1 \bmod 8$ we find $d_2([\tau g])$ is trivial, and when $q \equiv 5 \bmod 8$ we calculate $d_2([\tau g]) = uh_0h_2e_0$.

Proof When $q \equiv 1 \bmod 4$, it is possible that $d_2([\tau g])$ is $uh_0h_2e_0$. We analyze this differential using Massey products obtained from the May spectral sequence. We show that $\langle \tau, h_1^4, h_4\rangle = \{[\tau g]\}$ in the E_2 page of the MASS using Massey products in the May spectral sequence and the May convergence theorem in Isaksen [26, Theorem 2.2.1].

At the E_4 page of the May spectral sequence we calculate $d_4(b_{21}^2) = h_1^4h_4$ and $d_4(0) = \tau h_1^4$, as $\tau h_1^4 = 0$; hence $[\tau g] = \tau b_{21}^2 \in \langle \tau, h_1^4, h_4\rangle$ in the May spectral sequence. There are no crossing differentials, so the May convergence shows $[\tau g] \in \langle \tau, h_1^4, h_4\rangle$ in the MASS.[1]

The indeterminacy $\tau E_2^{4,(20,12)} + E_2^{3,(5,3)}h_4$ in the MASS is trivial, so we conclude $\langle \tau, h_1^4, h_4\rangle = \{[\tau g]\}$.

We now identify $d_2([\tau g])$ using the following formula of Moss [36, Theorem 1.1(ii)]:

(7-3) $$d_2(\langle \tau, h_1^4, h_4\rangle) \subseteq \langle d_2(\tau), h_1^4, h_4\rangle + \langle \tau, 0, h_4\rangle + \langle \tau, h_1^4, h_0h_3^2\rangle.$$

The Massey product $\langle \tau, 0, h_4\rangle$ contains 0 and has no indeterminacy.[2]

To calculate $\langle \tau, h_1^4, h_0h_3^2\rangle$ we again use the May spectral sequence and the May convergence theorem. We calculate this Massey product at the E_2 page using $d_2(h_2b_{20}) = \tau h_1^4$ and $h_1^4h_0h_3^2 = 0$ and see that $0 \in \langle \tau, h_1^4, h_0h_3^2\rangle$. There are no crossing differentials, so 0 is in this Massey product in the MASS.[3]

[1] In this case, we must check if there are crossing differentials d_t for $t \geq 5$. To see $E_4^{*,5,3,3} = 0$ over $\mathbb{F}_q$, we check $E_4^{*,5,3,3} = 0$ and $E_4^{*,6,3,4} = 0$ over $\mathbb{C}$ using the chart in [12, Appendix C]. All that is in $(*,5,3,3)$ is h_1b_{20}, but this does not survive to E_4. And nothing is in $(*,6,3,4)$ even at the E_2 page.

To see $E_5^{*,20,4,12}$ is trivial over $\mathbb{F}_q$, observe that all that is in $E_4^{*,20,4,12}$ over $\mathbb{C}$ is b_{21}^2, which does not survive to the E_5 page. The group $E_4^{*,21,4,13}$ over $\mathbb{C}$ is trivial. A potential contribution from $h_0h_3^3$ or $h_0h_2^2h_4$ is ruled out by weight reasons, and because they do not survive to the E_4 page from the differentials $d_2(h_0(1))$ and $d_2(h_0b_{22})$.

[2] Here $0 = d_2(h_1^4)$ is in grading $E_2^{6,(3,4)}$, so the indeterminacy is $\tau E_3^{6,(19,12)} + E_3^{5,(4,3)}h_4$. The degree of $h_1^2e_0$ is $6,(19,12)$, but it does not survive to the E_3 page. The group $E_3^{5,(4,3)}$ is trivial by checking the E_2 page.

[3] Note that $a_{01} = h_1b_{20}$ is in degree $(5,5,3,3)$ and a_{12} is in degree $(8,19,6,12)$. Then for a_{01} crossing differentials occur in $(?,5,3,3)$, which is trivial from the fourth page on. For a_{12} crossing

The indeterminacy for $\langle \tau, h_1^4, h_0h_3^2\rangle$ in the MASS is $\tau E_2^{6,(19,12)} + E_2^{3,(5,3)}h_0h_3^2$, which is trivial. The group $E_2^{6,(19,12)}$ is generated by $h_1^2 e_0$, which is annihilated by τ, while $E_2^{3,(5,3)}$ is trivial.

We now handle the Massey product $\langle d_2(\tau), h_1^4, h_4\rangle$, which depends on the base field. Let us suppose that $q \equiv 1 \bmod 8$ so that $d_2(\tau) = 0$ by Corollary 7.12. If a_{12} is in the E_1 page of the MASS with $d_1(a_{12}) = h_1^4 h_4$, then the Massey product contains $0 \cdot h_4 + 0 \cdot a_{12} = 0$. It is straightforward to check that the indeterminacy $0 \cdot E_2^{4,(20,12)} + E_2^{5,(4,3)} \cdot h_4$ is trivial. We conclude $d_2([\tau g]) = 0$ when $q \equiv 1 \bmod 8$.

When $q \equiv 5 \bmod 8$, Corollary 7.12 establishes $d_2(\tau) = uh_0^2$. We identify the Massey product $\langle uh_0^2, h_1^4, h_4\rangle$ using the May spectral sequence and the May convergence theorem. At the E_4 page of the May spectral sequence we have $d_4(b_{21}^2) = h_1^4 h_4$ and $uh_0^2 h_1^4 = 0$. Hence $uh_0^2 b_{21}^2 + 0h_4 = uh_0 b_{21} h_2 h_0(1) = uh_0 h_2 e_0$ is in the Massey product under consideration. It is straightforward to verify that there are no crossing differentials in this case.[4]

The indeterminacy of $\langle uh_0^2, h_1^4, h_4\rangle$ in the MASS is $uh_0^2 E_2^{4,(20,12)} + E_2^{5,(4,3)} h_4$, which is trivial. Thus the May convergence theorem shows the Massey product is exactly $\{uh_0 h_2 e_0\}$ and we conclude $d_2([\tau g]) = uh_0 h_2 e_0$ if $q \equiv 5 \bmod 8$.

We now analyze the differentials in the MASS in the 19 and 20 stems. Since $[\tau g]$ has weight 11, the class $\tau^{11}[\tau g]$ is in $E^{4,(20,0)}$. If $q \equiv 1 \bmod 8$, we calculate $d_2(\tau^{11}[\tau g]) = \tau^{11} uh_0 h_2 e_0 = u\tau^{10} h_0^2 [\tau g]$. If $q \equiv 5 \bmod 8$, then $d_2(\tau^{11}[\tau g]) = u\tau^{10} h_0^2 [\tau g]$. This resolves all differentials in the 19 and 20 stems, so the calculation of the 19 stem follows.

As $\hat{\pi}_{20}^s \cong \mathbb{Z}/8$ must be a summand of $\hat{\pi}_{20,0}(\mathbb{F}_q)$, we conclude there is a hidden extension from $u\tau^{11} h_2^2 h_4 = u\tau^{11} h_3^3$ to $\tau^{12} h_2 e_0$. The calculation of the 20 stem now follows. $\square$

---

differentials occur in degree $(m', 19, 6, 12)$ with $m' \geq 8$. The only thing in this filtration, stem and weight is $h_1^2 e_0$, which has May filtration 10. But note that both h_1^2 and e_0 are permanent cycles, so that $h_1^2 e_0$ is as well. So there are no crossing differentials in this case.

[4] As $a_{01} = 0$ in $E_4^{9,4,5,3}$ and $a_{12} = b_{12}^2$, we must check two conditions: (1) whenever $m' \geq 9$ and $m' - 5 < t$ that d_t is trivial on $E_t^{m',4,5,3}$ and (2) whenever $m' \geq 8$ and $m' - 4 < t$ that d_t is trivial on $E_t^{m',20,4,12}$. Condition (1) is easily verified as $E_4^{*,4,5,3} = 0$ over $\mathbb{C}$ and $E_4^{*,5,5,4} = 0$ over $\mathbb{C}$ as well. We conclude $E_4^{*,4,5,3} = 0$ over $\mathbb{F}_q$ as only these two groups can contribute to this graded piece. We remark that uh_1^5 does not contribute any terms, since to get the weight correct one needs to multiply by τ which annihilates the element. For condition (2), we will check that for all $t \geq 6$ the differentials vanish on $E_t^{(*,20,4,12)}$. This graded piece contains b_{21}^2 at the E_4 page, but it does not survive to $E_5 = E_6$. The only other possible elements in this group arise from elements in $E_t^{*,21,4,13}$ over $\mathbb{C}$ which we have seen is trivial at the E_4 page. This verifies the hypotheses of May's convergence theorem.

Remark 7.17 Note that over $\mathbb{F}_q$ with $q \equiv 5 \bmod 8$ the map $\mathbb{L}c\{g\}$ is detected by $u\tau^{11}h_3^3$, which is in Adams filtration 3. But over $\overline{\mathbb{F}}_q$, the map $\mathbb{L}c\{g\}$ is in Adams filtration 4.

Proposition 7.18 *When $q \equiv 3 \bmod 4$ and $0 \le n \le 18$, there is an isomorphism $\hat{\pi}_{n,0}(\mathbb{F}_q) \cong \hat{\pi}^s_n \oplus \hat{\pi}^s_{n+1}$.*

Proof The case $n = 0$ is resolved by Lemma 7.14, so we now consider $0 < n \le 18$, where we may use the motivic Adams spectral sequence as in Proposition 7.15.

The differentials d_r for $r \ge 2$ vanish on the following generators for degree reasons:

$$[\rho\tau],\ \ \rho,\ \ h_0,\ \ h_1,\ \ h_3,\ \ [\tau h_2^2],\ \ [\tau c_0],\ \ [\tau P h_1],\ \ d_0,\ \ [\tau P c_0],\ \ [\tau P^2 h_1].$$

Since $\hat{\pi}^s_1 \cong \mathbb{Z}/2$ is a summand of $\hat{\pi}_{1,0}(\mathbb{F}_q)$, we must have $d_r([\tau h_1]) = 0$ for all $r \ge 2$. Since $\hat{\pi}^s_3 \cong \mathbb{Z}/8$ is a summand of $\hat{\pi}_{3,0}(\mathbb{F}_q)$, we must have $d_2(h_2) = 0$. An argument similar to that given for Proposition 7.15 establishes

$$d_2(h_4) = h_0 h_3^2, \quad d_2(e_0) = h_1^2 d_0, \quad d_2(f_0) = h_0^2 e_0$$

by comparison to $\mathfrak{M}(\overline{\mathbb{F}}_q)$. Also, we determine $d_r([\tau c_1]) = 0$ for $r \ge 2$ by comparing with $\mathfrak{M}(\overline{\mathbb{F}}_q)$, as the class $[\tau c_1]$ must be a permanent cycle.

The one exceptional case is $d_3(h_0 h_4)$. Here we must have $d_3(h_0 h_4) = h_0 d_0 + \rho h_1 d_0$ in order for $\hat{\pi}^s_{14} = \mathbb{Z}/2 \oplus \mathbb{Z}/2$ to be a summand of $\hat{\pi}_{14,0}(\mathbb{F}_q)$.

The elements in weight 0 are all of the form $[\tau^2]^i x$ or $[\rho\tau][\tau^2]^{i-1} x$ where x is not a multiple of τ^2 and weight $2i$, or of the form $\rho[\tau^2]^i x$ if x is not a multiple of τ^2 and of weight $2i+1$. The differentials of the elements in weight 0 are now determined by using the Leibniz rule. Since $\lambda(q) = \nu_2(q^2-1) \ge 3$, we have $d_2(\tau^2) = 0$. This is sufficient to ensure that for elements x in stems $s \le 19$ there are no nontrivial differentials of the form $d_r([\tau^2]^i x) = \rho\tau^{2i-1} h_0^r x$ when $[\tau^2]^i x$ has weight 0. This resolves all differentials in weight 0 for stems $s \le 19$ and there are no hidden 2–extensions in this range. Hence for $0 < n \le 18$ there is an isomorphism $\hat{\pi}_{n,0}(\mathbb{F}_q) \cong \hat{\pi}^s_n \oplus \hat{\pi}^s_{n+1}$. □

Remark 7.19 When $q \equiv 3 \bmod 4$, it is unclear whether $d_2([\tau^2 g]) = [\rho\tau g]$ or $d_2([\tau^2 g]) = 0$. This is all that obstructs the identification of the stems $\hat{\pi}_{19,0}(\mathbb{F}_q)$ and $\hat{\pi}_{20,0}(\mathbb{F}_q)$ in this case.

7.5 Base change for finite fields

Proposition 7.20 *Let $q = p^\nu$, where p is an odd prime. For a field extension $f\colon \mathbb{F}_q \to \mathbb{F}_{q^i}$ with i odd, the induced maps*

$$\mathbb{L}f^*\colon H^{**}(\mathbb{F}_q) \to H^{**}(\mathbb{F}_{q^i}) \quad \text{and} \quad \mathbb{L}f^*\colon \mathcal{A}^{**}(\mathbb{F}_q) \to \mathcal{A}^{**}(\mathbb{F}_{q^i})$$

are isomorphisms.

Proof The claim follows by checking on étale cohomology. The map on cohomology is determined by $H^1_{\mathrm{et}}(\mathbb{F}_q;\mu_2) \to H^1_{\mathrm{et}}(\mathbb{F}_{q^i};\mu_2)$, which is just the induced map $\mathbb{F}_q^\times/2 \to \mathbb{F}_{q^i}^\times/2$. So long as i is odd, this map is an isomorphism. □

Corollary 7.21 *For $q = p^\nu$ with p an odd prime, the induced map $\mathfrak{M}(\mathbb{F}_q) \to \mathfrak{M}(\mathbb{F}_{q^i})$ is an isomorphism of spectral sequences whenever i is odd.*

Proposition 7.22 *Let $q = p^\nu$ with p an odd prime. Let $\widetilde{\mathbb{F}}_q$ denote the union of the field extensions $\mathbb{F}_{q^i}$ over $\mathbb{F}_q$ with i odd. The field extension $f\colon \mathbb{F}_q \to \widetilde{\mathbb{F}}_q$ induces isomorphisms $\mathbb{L}f^*\colon H^{**}(\mathbb{F}_q) \to H^{**}(\widetilde{\mathbb{F}}_q)$ and $\mathbb{L}f^*\colon \mathcal{A}^{**}(\mathbb{F}_q) \to \mathcal{A}^{**}(\widetilde{\mathbb{F}}_q)$. Hence the map $\mathfrak{M}(\mathbb{F}_q) \to \mathfrak{M}(\widetilde{\mathbb{F}}_q)$ is an isomorphism of spectral sequences.*

Proof This follows by a colimit argument using Proposition 7.20. □

Corollary 7.23 *For any integers s and $w \geq 0$, there is an isomorphism $\hat{\pi}_{s,w}(\mathbb{F}_q) \cong \hat{\pi}_{s,w}(\widetilde{\mathbb{F}}_q)$.*

Proposition 7.24 *Let $q = p^\nu$, where p is an odd prime. For a field extension $f\colon \mathbb{F}_q \to \mathbb{F}_{q^i}$ with i even, the map $f^*\colon H^{1,*}(\mathbb{F}_q) \to H^{1,*}(\mathbb{F}_{q^j})$ is trivial, and the map $f^*\colon H^{0,*}(\mathbb{F}_q) \to H^{0,*}(\mathbb{F}_{q^j})$ is injective.*

Proof The map is determined by $\mathbb{L}f^*\colon H^{1,1}(\mathbb{F}_q) \to H^{1,1}(\mathbb{F}_{q^i})$, which is just the map $\mathbb{F}_q^\times/2 \to \mathbb{F}_{q^i}^\times/2$. However, any nonsquare $x \in \mathbb{F}_q^\times$ will be a square in $\mathbb{F}_{q^i}^\times$ when i is even. □

Corollary 7.25 *Let $q = p^\nu$ with p an odd prime. For a field extension $f\colon \mathbb{F}_q \to \mathbb{F}_{q^i}$ with i even, the induced map $\mathfrak{M}(\mathbb{F}_q) \to \mathfrak{M}(\mathbb{F}_{q^i})$ kills the class u (respectively ρ and $[\rho\tau]$) and all of their multiples at the E_2 page.*

Proof The induced map of cobar complexes is determined from Proposition 7.24 and shows the class u (respectively ρ and $[\rho\tau]$) is killed under base change. □

8 Implementation of motivic Ext group calculations

The computer calculations used in this paper were done with the program available from Fu and Wilson [16]. The program is written in Python and calculates $\mathrm{Ext}(F)$ when F is $\mathbb{C}$, $\mathbb{R}$ or $\mathbb{F}_q$ by producing a minimal resolution of $H^{**}(F)$ by $\mathcal{A}^{**}(F)$ modules in a range. With this complex in hand, the program then produces its dual and calculates cohomology in each degree.

To calculate a free resolution of $H^{**}(F)$ of $\mathcal{A}^{**}(F)$ modules, we first need the program to efficiently perform calculations in $\mathcal{A}^{**}(F)$. The mod 2 motivic Steenrod algebra is generated by the squaring operations Sq^i and the cup products $\alpha \cup -$ for $\alpha \in H^{**}(F)$. These generators satisfy Adem relations, which are recorded in [22, Section 5.1] by Hoyois, Kelly and Østvær and in [49, Theorem 10.2] by Voevodsky. Additionally, one needs the commutation relations $\mathrm{Sq}^{2i}\tau = \tau\,\mathrm{Sq}^{2i} + \tau\rho\,\mathrm{Sq}^{2i-1}$ for $i > 0$ and $\mathrm{Sq}^{2i+1}\tau = \tau\,\mathrm{Sq}^{2i+1} + \rho\,\mathrm{Sq}^{2i} + \rho^2\,\mathrm{Sq}^{2i-1}$ for $i \geq 0$ which are obtained from the Cartan formula. With these relations, the program can calculate the canonical form of any element of $\mathcal{A}^{**}$, that is, as a sum of monomials $\alpha \cdot \mathrm{Sq}^I$ where $\alpha \in H^{**}(F)$ and I is an admissible sequence.

With the algebra of $\mathcal{A}^{**}(F)$ available to the program, it then proceeds to calculate a minimal resolution of $H^{**}(F)$ by $\mathcal{A}^{**}(F)$ modules. This is where a great deal of computational effort is spent. To clarify what a minimal resolution is in practice, let $\prec$ denote the order on $\mathbb{Z} \times \mathbb{Z}$ given by $(m_1, n_1) \prec (m_2, n_2)$ if and only if $m_1 + n_1 < m_2 + n_2$, or $m_1 + n_1 = m_2 + n_2$ and $n_1 < n_2$. The reader is encouraged to compare this definition with the definition of McCleary in [30, Definition 9.3] and consult Bruner [9] for detailed calculations of a minimal resolution for the Adams spectral sequence of topology.

Definition 8.1 A resolution of $H^{**}(F)$ by $\mathcal{A}^{**}(F)$ modules $H^{**}(F) \leftarrow P^\bullet$ is a minimal resolution if the following conditions are satisfied:

(1) Each module P^i is equipped with ordered basis $\{h_i(j)\}$ such that if $j \leq k$ then $\deg h_i(j) \preceq \deg h_i(k)$.

(2) $\mathrm{im}(h_i(k)) \notin \mathrm{im}(\langle h_i(j) \mid j < k\rangle)$.

(3) $\deg h_i(k)$ is minimal with respect to degree in the order $\prec$ over all elements in $P^{i-1} \setminus \mathrm{im}(\langle h_i(j) \mid j < k\rangle)$.

The computer program calculates the first n maps and modules in a minimal resolution up to bidegree $(2n, n)$. With this, it then calculates the dual of the resolution by applying the functor $\mathrm{Hom}_{\mathcal{A}^{**}(F)}(-, H^{**}(F))$ to the resolution $P^\bullet$. With the cochain complex $\mathrm{Hom}_{\mathcal{A}^{**}(F)}(P^\bullet, H^{**}(F))$ in hand, the program calculates cohomology in each degree, that is, $\mathrm{Ext}^{f,(s+f,w)}(\mathbb{F}_q)$.

Because the program calculates an explicit resolution of $H^{**}(F)$, the products of elements in $\mathrm{Ext}(F)$ can be obtained from the composition product; see McCleary [30, Theorem 9.5].

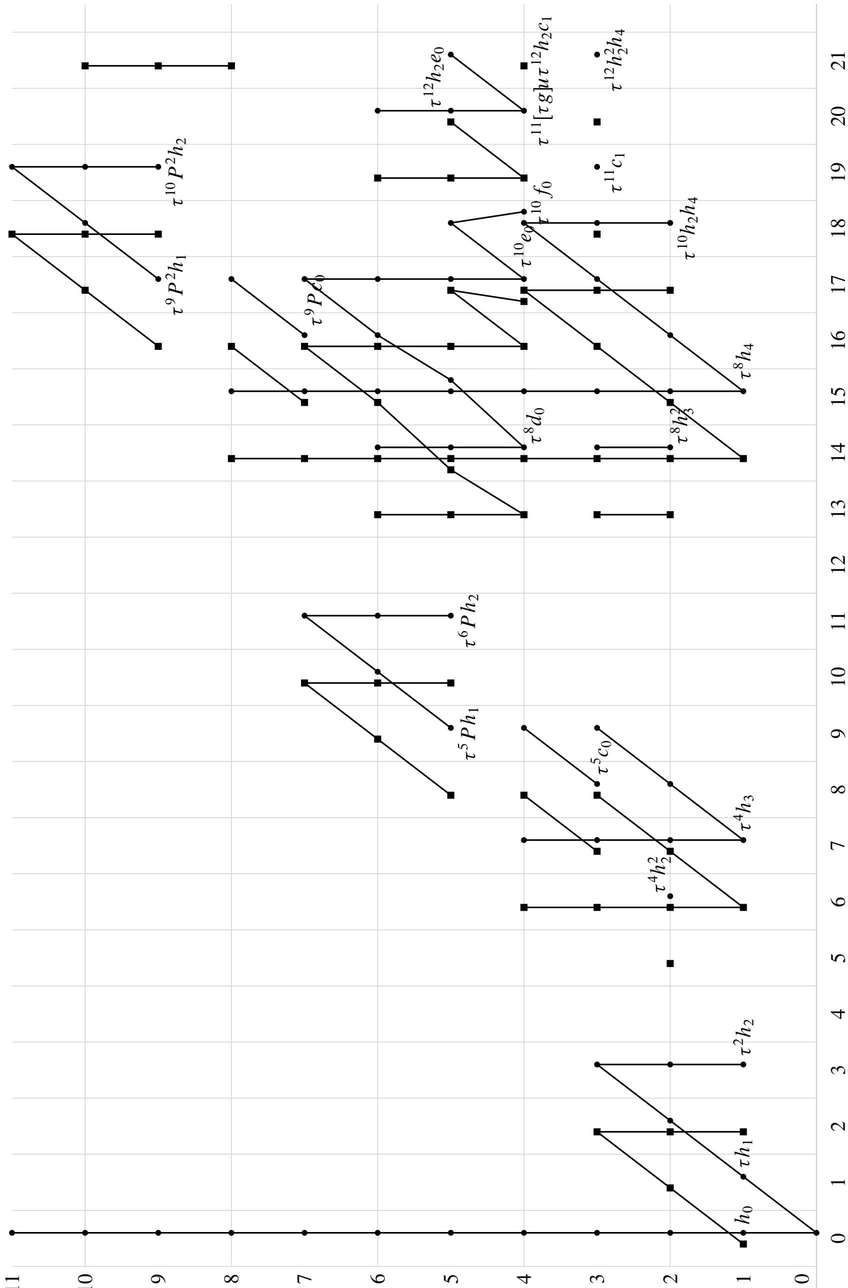


Figure 1: E_2 page of MASS for $\mathbb{F}_q$ with $q \equiv 1 \bmod 4$, weight 0

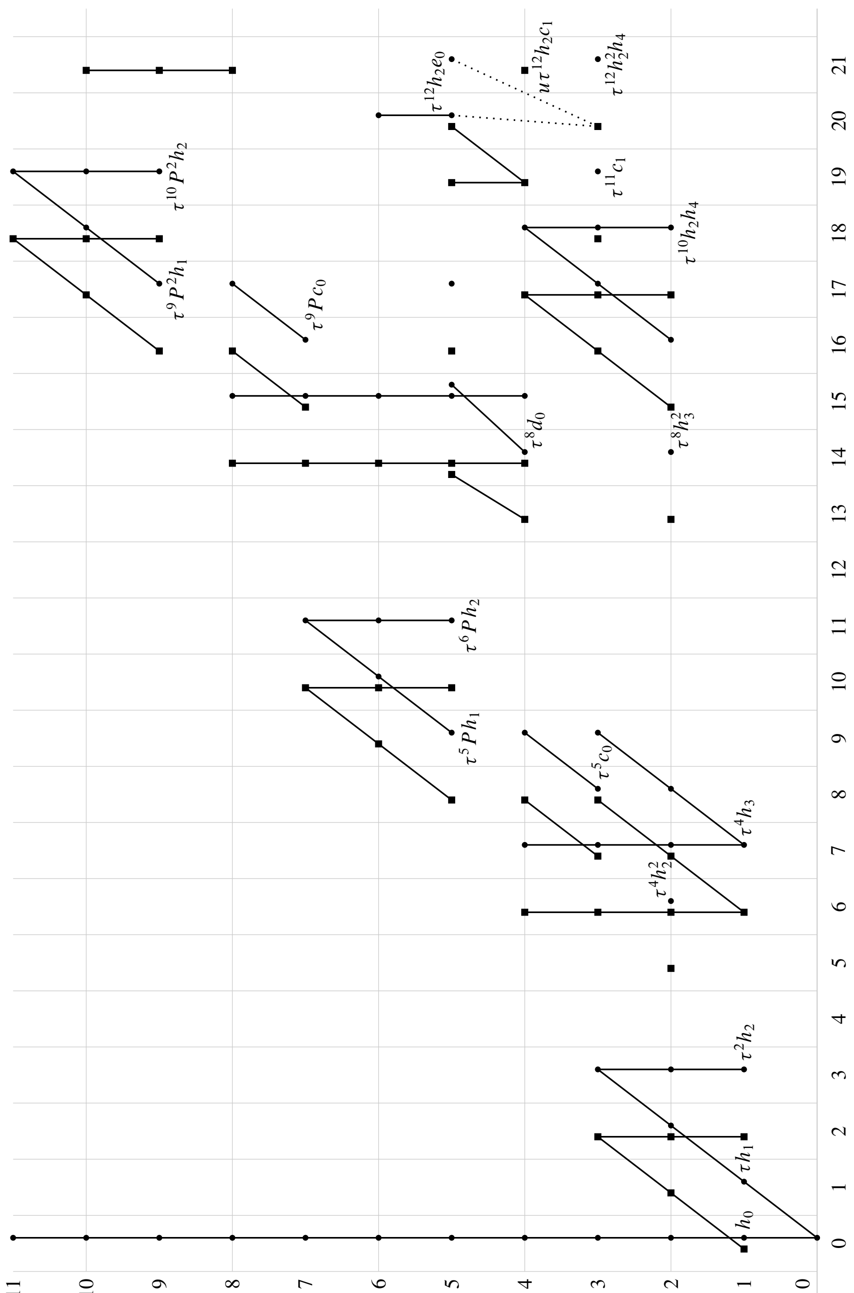


Figure 2: E_∞ page of MASS for $\mathbb{F}_q$ with $q \equiv 1 \bmod 4$, weight 0

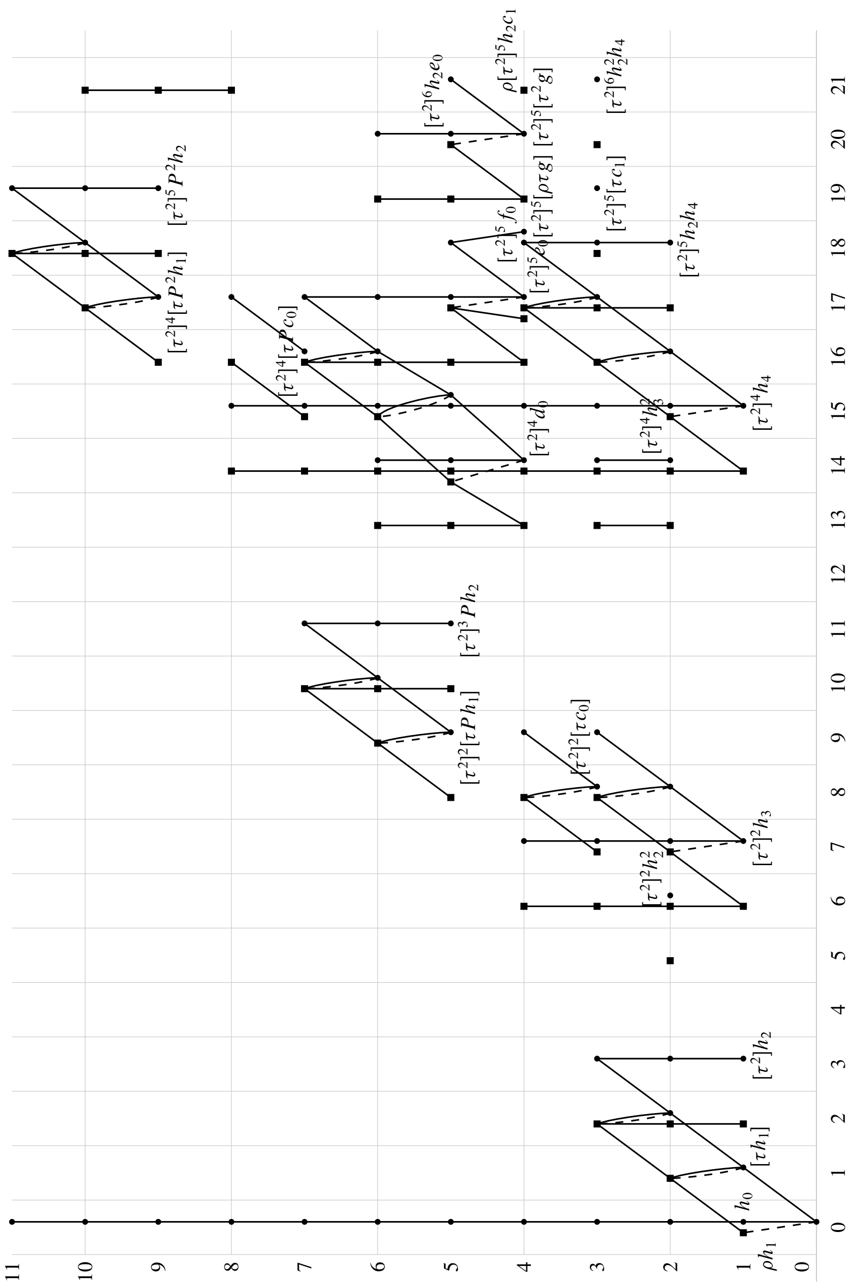


Figure 3: E_2 page of MASS for $\mathbb{F}_q$ with $q \equiv 3 \bmod 4$, weight 0

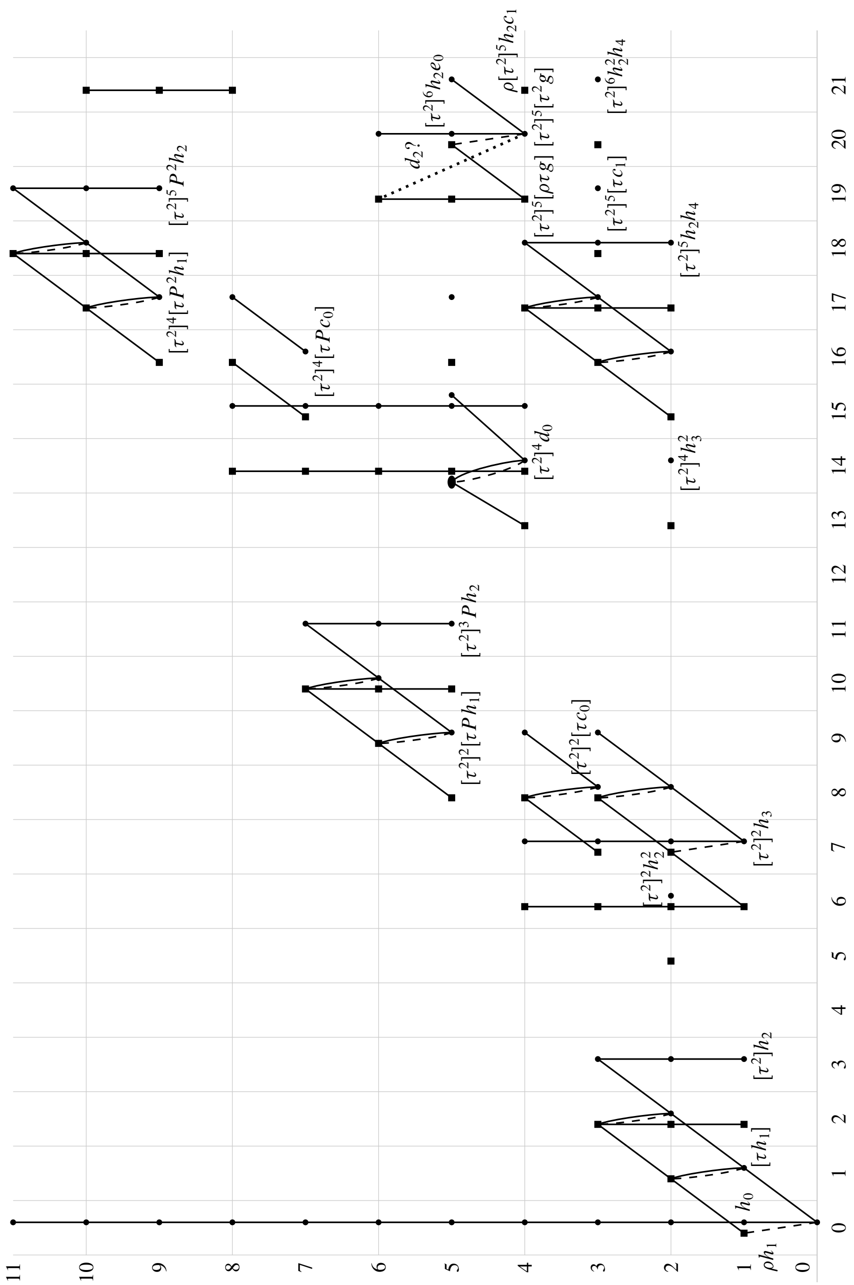


Figure 4: E_∞ page of MASS for $\mathbb{F}_q$ with $q \equiv 3 \bmod 4$, weight 0

9 Charts

The weight 0 part of the E_2 page of the mod 2 MASS over $\mathbb{F}_q$ is depicted in Figures 1 and 3 according to the case $q \equiv 1 \bmod 4$ or $q \equiv 3 \bmod 4$. The weight 0 part of the E_∞ page of the mod 2 MASS over $\mathbb{F}_q$ can be found in Figures 2 and 4.

In each chart, a circular or square dot in grading (s, f) represents a generator of the $\mathbb{F}_2$ vector space in the graded piece of the spectral sequence. The square dots are used to indicate that the given element is divisible by u, ρ or $\rho\tau$, depending on the case. Circular dots denote elements which are not divisible by u, ρ or $\rho\tau$. In Figure 4, there is an oval dot which corresponds to the class with representative $\tau^8\rho h_1 d_0 \equiv \tau^8 h_0 d_0$, as the class $\rho h_1 d_0 + h_0 d_0$ is killed.

We indicate the product of a given class by h_0 with a solid, vertical line. In the case $q \equiv 3 \bmod 4$, multiplication by ρh_1 plays an important role, so nonzero products by ρh_1 are indicated by dashed vertical lines. In particular, when $q \equiv 3 \bmod 4$, multiplication by 2 in $\hat{\pi}_{**}(\mathbb{F}_q)$ is detected by multiplication by $h_0 + \rho h_1$. The lines of slope 1 indicate multiplication by τh_1 or $[\tau h_1]$ depending on the case. We caution the reader that the product structure displayed in this chart was obtained by computer calculation and not all products were established by hand in this paper. For example, the products in the 8 stem by h_0 are hidden in the May spectral sequence.

Dotted lines are used in two separate instances in these charts. The first use is in Figure 2, where dotted lines indicate hidden extensions by h_0 and τh_1. The other instance is in Figure 4 to indicate an unknown d_2 differential.

Additional charts obtained from the program of Fu and Wilson [16] may be found at the website `http://math.rutgers.edu/~wilson47/image_viewer/`.

References

[1] **J F Adams**, *A finiteness theorem in homological algebra*, Proc. Cambridge Philos. Soc. 57 (1961) 31–36 MR

[2] **J F Adams**, *Stable homotopy and generalised homology*, University of Chicago Press (1974) MR

[3] **A Ananyevskiy**, **M Levine**, **I Panin**, *Witt sheaves and the η–inverted sphere spectrum*, preprint (2015) arXiv

[4] **J Ayoub**, *Les six opérations de Grothendieck et le formalisme des cycles évanescents dans le monde motivique, I*, Astérisque 314, Société Mathématique de France, Paris (2007) MR

[5] **B A Blander**, *Local projective model structures on simplicial presheaves*, K–Theory 24 (2001) 283–301 MR

[6] **A K Bousfield**, *The localization of spectra with respect to homology*, Topology 18 (1979) 257–281 MR

[7] **A K Bousfield**, **D M Kan**, *Homotopy limits, completions and localizations*, Lecture Notes in Mathematics 304, Springer (1972) MR

[8] **R R Bruner**, Ext *in the nineties*, from "Algebraic topology" (M C Tangora, editor), Contemp. Math. 146, Amer. Math. Soc., Providence, RI (1993) 71–90 MR

[9] **R R Bruner**, *An Adams spectral sequence primer*, notes (2009) Available at `http://math.wayne.edu/~rrb/papers/adams.pdf`

[10] **R R Bruner**, *The cohomology of the mod* 2 *Steenrod algebra*, unpublished manuscript (2016) Available at `http://math.wayne.edu/~rrb/papers/cohom.pdf`

[11] **D Dugger**, **D C Isaksen**, *Motivic cell structures*, Algebr. Geom. Topol. 5 (2005) 615–652 MR

[12] **D Dugger**, **D C Isaksen**, *The motivic Adams spectral sequence*, Geom. Topol. 14 (2010) 967–1014 MR

[13] **D Dugger**, **D Isaksen**, *Low-dimensional Milnor–Witt stems over* $\mathbb{R}$, Ann. K-Theory 2 (2017) 175–210 MR

[14] **B I Dundas**, **M Levine**, **P A Østvær**, **O Röndigs**, **V Voevodsky**, *Motivic homotopy theory*, Springer (2007) MR

[15] **B I Dundas**, **O Röndigs**, **P A Østvær**, *Motivic functors*, Doc. Math. 8 (2003) 489–525 MR

[16] **K Fu**, **G M Wilson**, *Motivic Adams spectral sequence program*, Python code (2015) Available at `https://github.com/glenwilson/MassProg`

[17] **T Geisser**, *Motivic cohomology over Dedekind rings*, Math. Z. 248 (2004) 773–794 MR

[18] **M A Hill**, *Ext and the motivic Steenrod algebra over* $\mathbb{R}$, J. Pure Appl. Algebra 215 (2011) 715–727 MR

[19] **P S Hirschhorn**, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, Amer. Math. Soc., Providence, RI (2003) MR

[20] **M Hovey**, *Spectra and symmetric spectra in general model categories*, J. Pure Appl. Algebra 165 (2001) 63–127 MR

[21] **M Hoyois**, *From algebraic cobordism to motivic cohomology*, J. Reine Angew. Math. 702 (2015) 173–226 MR

[22] **M Hoyois**, **S Kelly**, **P A Østvær**, *The motivic Steenrod algebra in positive characteristic*, preprint (2013) arXiv To appear in J. Eur. Math. Soc.

[23] **P Hu**, *S–modules in the category of schemes*, Mem. Amer. Math. Soc. 767, Amer. Math. Soc., Providence, RI (2003) MR

[24] **P Hu**, **I Kriz**, **K Ormsby**, *Remarks on motivic homotopy theory over algebraically closed fields*, J. K-Theory 7 (2011) 55–89 MR

[25] **D C Isaksen**, *Classical and motivic Adams charts*, preprint (2014) arXiv

[26] **D C Isaksen**, *Stable stems*, preprint (2014) arXiv

[27] **J F Jardine**, *Motivic symmetric spectra*, Doc. Math. 5 (2000) 445–553 MR

[28] **S O Kochman**, *Stable homotopy groups of spheres: a computer-assisted approach*, Lecture Notes in Mathematics 1423, Springer (1990) MR

[29] **M Levine**, *A comparison of motivic and classical stable homotopy theories*, J. Topol. 7 (2014) 327–362 MR

[30] **J McCleary**, *A user's guide to spectral sequences*, 2nd edition, Cambridge Studies in Advanced Mathematics 58, Cambridge University Press (2001) MR

[31] **J Milnor**, *The Steenrod algebra and its dual*, Ann. of Math. 67 (1958) 150–171 MR

[32] **F Morel**, *An introduction to $\mathbb{A}^1$–homotopy theory*, from "Contemporary developments in algebraic K–theory" (M Karoubi, A O Kuku, C Pedrini, editors), ICTP Lect. Notes XV, Abdus Salam Int. Cent. Theoret. Phys., Trieste (2004) 357–441 MR

[33] **F Morel**, *The stable $\mathbb{A}^1$–connectivity theorems*, K–Theory 35 (2005) 1–68 MR

[34] **F Morel**, *$\mathbb{A}^1$–algebraic topology over a field*, Lecture Notes in Mathematics 2052, Springer (2012) MR

[35] **F Morel**, **V Voevodsky**, *$\mathbf{A}^1$–homotopy theory of schemes*, Inst. Hautes Études Sci. Publ. Math. 90 (1999) 45–143 MR

[36] **R M F Moss**, *Secondary compositions and the Adams spectral sequence*, Math. Z. 115 (1970) 283–310 MR

[37] **K M Ormsby**, *Motivic invariants of p–adic fields*, J. K-Theory 7 (2011) 597–618 MR

[38] **K M Ormsby**, **P A Østvær**, *Motivic Brown–Peterson invariants of the rationals*, Geom. Topol. 17 (2013) 1671–1706 MR

[39] **K M Ormsby**, **P A Østvær**, *Stable motivic π_1 of low-dimensional fields*, Adv. Math. 265 (2014) 97–131 MR

[40] **D C Ravenel**, *Complex cobordism and stable homotopy groups of spheres*, Pure and Applied Mathematics 121, Academic Press, Orlando, FL (1986) MR

[41] **O Röndigs**, **M Spitzweck**, **P A Østvær**, *The first stable homotopy groups of motivic spheres*, preprint (2016) arXiv

[42] **W Scharlau**, *Quadratic and Hermitian forms*, Grundl. Math. Wissen. 270, Springer (1985) MR

[43] **J-P Serre**, *Local fields*, Graduate Texts in Mathematics 67, Springer (1979) MR

[44] **C Soulé**, *K–théorie des anneaux d'entiers de corps de nombres et cohomologie étale*, Invent. Math. 55 (1979) 251–295 MR

[45] **M Spitzweck**, *A commutative $\mathbb{P}^1$–spectrum representing motivic cohomology over Dedekind domains v3*, preprint (2013) arXiv

[46] **R M Switzer**, *Algebraic topology—homotopy and homology*, Grundl. Math. Wissen. 212, Springer (1975) MR

[47] **V Voevodsky**, *$\mathbf{A}^1$–homotopy theory*, from "Proceedings of the International Congress of Mathematicians, I", Documenta Mathematica, Bielefeld (1998) 579–604 MR

[48] **V Voevodsky**, *Motivic cohomology with $\mathbf{Z}/2$–coefficients*, Publ. Math. Inst. Hautes Études Sci. 98 (2003) 59–104 MR

[49] **V Voevodsky**, *Reduced power operations in motivic cohomology*, Publ. Math. Inst. Hautes Études Sci. 98 (2003) 1–57 MR

[50] **V Voevodsky**, *Motivic Eilenberg–Maclane spaces*, Publ. Math. Inst. Hautes Études Sci. 112 (2010) 1–99 MR

[51] **G Wang**, **Z Xu**, *The algebraic Atiyah–Hirzebruch spectral sequence of real projective spectra*, preprint (2016) arXiv

[52] **G M Wilson**, *Motivic stable stems over finite fields*, PhD thesis, Rutgers University–New Brunswick (2016) MR Available at `http://search.proquest.com/docview/1844381634`

Department of Mathematics, University of Oslo
PO Box 1053, 0316 Oslo, Norway

`glenw@math.uio.no`, `paularne@math.uio.no`

Received: 2 February 2016 Revised: 3 October 2016

Algebraic & Geometric Topology 17 (2017) 1105–1130

Operad bimodules and composition products on André–Quillen filtrations of algebras

NICHOLAS J KUHN
LUÍS ALEXANDRE PEREIRA

If $\mathcal{O}$ is a reduced operad in a symmetric monoidal category of spectra (S–modules), an $\mathcal{O}$–algebra I can be viewed as analogous to the augmentation ideal of an augmented algebra. From the literature on topological André–Quillen homology, one can see that such an I admits a canonical (and homotopically meaningful) decreasing $\mathcal{O}$–algebra filtration $I \xleftarrow{\sim} I^1 \leftarrow I^2 \leftarrow I^3 \leftarrow \cdots$ satisfying various nice properties analogous to powers of an ideal in a ring.

We more fully develop such constructions in a manner allowing for more flexibility and revealing new structure. With R a commutative S–algebra, an $\mathcal{O}$–bimodule M defines an endofunctor of the category of $\mathcal{O}$–algebras in R–modules by sending such an $\mathcal{O}$–algebra I to $M \circ_{\mathcal{O}} I$. We explore the use of the bar construction as a derived version of this. Letting M run through a decreasing $\mathcal{O}$–bimodule filtration of $\mathcal{O}$ itself then yields the augmentation ideal filtration as above. The composition structure of the operad then induces pairings among these bimodules, which in turn induce natural transformations $(I^i)^j \to I^{ij}$, fitting nicely with previously studied structure.

As a formal consequence, an $\mathcal{O}$–algebra map $I \to J^d$ induces compatible maps $I^n \to J^{dn}$ for all n. This is an essential tool in the first author's study of Hurewicz maps for infinite loop spaces, and its utility is illustrated here with a lifting theorem.

55P43; 18D50

1 Introduction

Let S–mod be the category of symmetric spectra (see Hovey, Shipley and Smith [7]), one of the standard symmetric monoidal models for the category of spectra. Let S denote the sphere spectrum, and let $\mathcal{O}$ be a reduced operad in S–mod. If R is a commutative S–algebra, we let $\mathsf{Alg}_{\mathcal{O}}(R)$ denote the category of $\mathcal{O}$–algebras in R–modules.

The starting point of this paper is the observation that if M is a reduced $\mathcal{O}$–bimodule and $I \in \mathsf{Alg}_{\mathcal{O}}(R)$ then $M \circ_{\mathcal{O}} I$ is again in $\mathsf{Alg}_{\mathcal{O}}(R)$, and that many interesting constructions on $\mathcal{O}$–algebras are derived versions of functors of I of this form.

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.1105

Our first goal, presented in Section 2, is to develop the basic properties of a suitable derived version of $M \circ_{\mathcal{O}} I$, the bar construction $B(M, \mathcal{O}, I)$, noting particularly how structure on the category of $\mathcal{O}$–bimodules is reflected in the category of endofunctors of $\mathcal{O}$–algebras.

In Section 2.1 and Section 2.2, we begin by introducing the setting in which we wish to work. This includes the model structure on $\mathsf{Alg}_{\mathcal{O}}(R)$ developed by Harper [3], which piggybacks off of the "positive" model structure on S–mod first exploited by Shipley [15].

Theorem 2.11 lays out the basic properties of $B(M, \mathcal{O}, I)$ needed for homotopical analysis. For example, a levelwise homotopy fibration sequence in the bimodule variable M induces a homotopy fibration sequence in $\mathsf{Alg}_{\mathcal{O}}(R)$.

In Section 2.5 and Section 2.6, we describe $B(M, \mathcal{O}, I)$ in the case where M is concentrated in one level, in terms of the topological André–Quillen spectrum $\mathrm{TQ}(I) = B(\mathcal{O}(1), \mathcal{O}, I)$. This allows us to easily identify, in Section 2.7, the Goodwillie tower of $F_M(I) = B(M, \mathcal{O}, I)$, viewed as an endofunctor of $\mathsf{Alg}_{\mathcal{O}}(R)$. In particular, one learns that $\partial_* \mathrm{Id} = \mathcal{O}$, $\partial_* F_M = M$, and one gets the expected chain rule: $\partial_*(F_M \circ F_N) \simeq M \circ_{\mathcal{O}} N$. See Pereira [13] for more about Goodwillie calculus in this setting.

Remark 1.1 $\mathrm{TQ}(I)$ can be informally viewed as I/I^2: its study goes back to Basterra [1]. The results in Section 2, and their proofs, clearly have much in common with Harper and Hess [5, Section 4], and our definition of $\mathrm{TQ}(I)$ agrees with that in Harper [4]. However those authors use only one special family of bimodules in the M variable, whereas for applications in this paper, and in ongoing work, greater generality is essential. In particular, we try to make clear that, on the one hand, our constructions connect nicely to $\mathrm{TQ}(I)$ and, on the other, they are well suited to iteration using the monoidal properties of $\circ$.

Another new aspect of our work, also crucial to applications (see, for example, Kuhn [9]), is that throughout we also have change-of-rings statements allowing for passage from $\mathsf{Alg}_{\mathcal{O}}(R)$ to $\mathsf{Alg}_{\mathcal{O}}(R')$, given a map $R \to R'$ of commutative S–algebras.

An $\mathcal{O}$–algebra I can be viewed as similar to the augmentation ideal in an augmented ring. In Section 3, we apply our bar construction to a natural decreasing $\mathcal{O}$–bimodule filtration of $\mathcal{O}$ itself, defining, for $I \in \mathsf{Alg}_{\mathcal{O}}(R)$, a homotopically meaningful natural augmentation ideal filtration

$$I \xleftarrow{\sim} I^1 \leftarrow I^2 \leftarrow I^3 \leftarrow \cdots .$$

The results of the Section 2 show that I^n/I^{n+1} is determined by $\mathcal{O}(n)$ and $\mathrm{TQ}(I)$. Our model makes it easy to analyze connectivity properties: if R and $\mathcal{O}$ are connective and I is $(c-1)$–connected, then I^n will be $(nc-1)$–connected.

The construction of such a filtration, or more precisely, the associated tower under I,

$$I^1/I^2 \leftarrow I/I^3 \leftarrow I/I^4 \leftarrow \cdots,$$

goes back to Minasian [12] and Kuhn [8] when $\mathcal{O} = \mathrm{Com}$. Harper and Hess [5] construct this tower in exactly the same way we do. However, we now add new structure by taking advantage of the observation that a pairing of bimodules

$$L \circ_{\mathcal{O}} M \to N$$

will induce a natural transformation of functors of $I \in \mathsf{Alg}_{\mathcal{O}}(R)$,

$$F_L(F_M(I)) \to F_N(I).$$

The multiplication $\mathcal{O} \circ \mathcal{O} \to \mathcal{O}$ induces pairings among our $\mathcal{O}$–bimodule filtration of $\mathcal{O}$, and these in turn induce natural pairings

$$(I^n)^m \to I^{mn}$$

satisfying expected properties. As a formal consequence, an $\mathcal{O}$–algebra map $I \to J^d$ induces compatible maps $I^n \to J^{dn}$ for all n.

This seems to be fundamental structure which has not previously appeared in the literature. The following result is a consequence illustrating its utility.

Theorem 1.2 *Let $f\colon I \to J$ be a map in the homotopy category* $\mathrm{ho}\,\mathsf{Alg}_{\mathcal{O}}(R)$*. If f factors as $f = f_s \circ \cdots \circ f_1$ such that* $\mathrm{TQ}(f_i)$ *is null for all i, then there is a lifting in* $\mathrm{ho}\,\mathsf{Alg}_{\mathcal{O}}(R)$*:*

$$\begin{tikzcd} & J^{2^s} \arrow[d] \\ I \arrow[ur, dotted, "\tilde{f}"] \arrow[r, "f"] & J \end{tikzcd}$$

We restate this, with slightly different notation, as Theorem 3.11.

Further applications in this spirit can be seen in the work of the first author on Hurewicz maps of infinite loop spaces [9], the project whose needs motivated this paper.

The deeper proofs from Section 2 are deferred to Section 4, which itself is supported by the Appendix. Much of the technical work consists of generalizing results and arguments from Pereira [14] from S–mod to R–mod for a general R.

2 General results about derived composition products

2.1 Our categories of modules and algebras

In this paper, the category of S–modules will mean the category of symmetric spectra as defined in [7]: $X \in S$–mod consists of a sequence $X_0, X_1, X_2, \dots$ of simplicial sets equipped with extra structure.

With the smash product as product and sphere spectrum S as unit, S–mod is a closed symmetric monoidal category. There is a notion of weak equivalence, and various model structures on S–mod compatible with these, such that the resulting quotient category models the standard stable homotopy category.

Recall that a symmetric sequence in S–mod then consists of a sequence

$$X(0), X(1), X(2), \dots,$$

where $X(n)$ is a symmetric spectrum equipped with an action of the n^{th} symmetric group Σ_n. The category of such symmetric sequences in S–mod, $\mathsf{Sym}(S)$, admits a composition product $\circ$ defined by

$$(X \circ Y)(s) = \bigvee_r X(r) \wedge_{\Sigma_r} \Big(\bigvee_{\phi\colon \mathbf{s} \to \mathbf{r}} Y(s_1(\phi)) \wedge \cdots \wedge Y(s_r(\phi)) \Big), \tag{1}$$

where $\mathbf{s} = \{1, \dots, s\}$ and $s_k(\phi)$ is the cardinality of $\phi^{-1}(k)$. With this product, $(\mathsf{Sym}(S), \circ, S(1))$ is monoidal, where $S(1) = (*, S, *, *, \dots)$.

An operad $\mathcal{O}$ is then a monoid in this category, and one makes sense of left $\mathcal{O}$–modules, right $\mathcal{O}$–modules, and $\mathcal{O}$–bimodules in the usual way. Furthermore, if X is a right $\mathcal{O}$–module, and Y is a left $\mathcal{O}$–module, the symmetric sequence $X \circ_{\mathcal{O}} Y$ can be defined as the coequalizer in $\mathsf{Sym}(S)$ of the two evident maps

$$X \circ \mathcal{O} \circ Y \rightrightarrows X \circ Y.$$

Extra structure on X or Y can then induce evident extra structure on $X \circ_{\mathcal{O}} Y$.

For the purposes of this paper, it is natural to require that our operads $\mathcal{O}$ and bimodules M be reduced: $\mathcal{O}(0) = * = M(0)$. By contrast, an $\mathcal{O}$–algebra is a left $\mathcal{O}$–module I concentrated in level 0: $I(n) = *$ for all $n > 0$.

If R is a commutative S–algebra, these definitions and constructions extend to the category of R–modules. Furthermore, one can mix and match. For example, if X is a symmetric sequence in S–mod and Y is a symmetric sequence in R–mod, $X \circ Y$ will be the symmetric sequence in R–mod with

$$(X \circ Y)(s) = \bigvee_r X(r) \wedge_{\Sigma_r} \Big(\bigvee_{\phi\colon \mathbf{s} \to \mathbf{r}} Y(s_1(\phi)) \wedge_R \cdots \wedge_R Y(s_r(\phi)) \Big).$$

We denote by $\mathsf{Sym}(R)$ the category of symmetric sequences in R–mod, $\mathsf{Alg}_{\mathcal{O}}(R)$ the category of $\mathcal{O}$–algebras in R–mod and $\mathsf{Mod}^l_{\mathcal{O}}(R)$ the category of left $\mathcal{O}$–modules in $\mathsf{Sym}(R)$.

Remark 2.1 If $\mathcal{O}$ is an operad in S–mod, then the symmetric sequence $R \wedge \mathcal{O}$, defined as $(R \wedge \mathcal{O})(n) = R \wedge \mathcal{O}(n)$ is naturally an operad when viewed either in $\mathsf{Sym}(R)$ or $\mathsf{Sym}(S)$. It is easily checked that the category $\mathsf{Mod}^l_{\mathcal{O}}(R)$ is isomorphic to $\mathsf{Mod}^l_{R\wedge\mathcal{O}}(S)$, and that both of these are isomorphic to $\mathsf{Mod}^l_{R\wedge\mathcal{O}}$, the category of left $R \wedge \mathcal{O}$–modules in $\mathsf{Sym}(R)$. A similar remark holds for the three corresponding categories of algebras.

2.2 Model structures

We specify model structures on the various categories just described.

We accept as given the S–model structure on symmetric spectra (called S–modules in this paper) as defined in [7, Definition 5.3.6] and [15, Theorem 2.4]. This is monoidal with respect to the smash product [7, Corollary 5.3.8].

We then give $\mathsf{Sym}(S)$ its associated *injective* model structure: weak equivalences and cofibrations are those morphisms which are levelwise weak equivalences and cofibrations in S–mod. That this structure exists was checked in [14, Theorem 3.8 and Section 5.3].[1]

As in [11, Section 15], [15], [5] and [14, Section 5.3], we need positive variants of these model structures. Weak equivalences will be as before, but there are fewer cofibrations: for $X \to Y$ in S–mod to be a positive cofibration, we now insist that $X_0 \to Y_0$ also be an isomorphism, and for $M \to N$ in $\mathsf{Sym}(S)$ to be a positive cofibration, we now insist that $M(0)_0 \to N(0)_0$ also be an isomorphism.[2] It is worth noting that if $M \in \mathsf{Sym}(S)$ is reduced, then it is positive cofibrant exactly when each $M(n)$ is cofibrant, when viewed in S–mod.

Given a commutative S–algebra R, the positive R–model structure on R–modules is then defined to be the projective structure induced from that on S–mod with its positive structure: weak equivalences and fibrations in R–mod are the maps which are weak equivalences and positive fibrations in S–mod. Similarly, we define the positive structure on $\mathsf{Sym}(R)$, the category of symmetric sequences in R–mod, to be the projective structure induced from that on $\mathsf{Sym}(S)$ with its positive structure: weak equivalences and fibrations in $\mathsf{Sym}(R)$ are the maps which are weak equivalences and positive fibrations in $\mathsf{Sym}(S)$.

[1] This structure is different from the associated projective structure used in [3; 4; 5].

[2] On $\mathsf{Sym}(S)$, this agrees with [14] but is different from [5], where it is required that $M(n)_0 \to N(n)_0$ be an isomorphism for all n.

Thanks to Remark 2.1, the following theorem is an immediate consequence of [14, Theorem 1.4]; see also [3]. Special cases go back to [15].

Theorem 2.2 *$\mathsf{Alg}_{\mathcal{O}}(R)$ has a projective model structure induced from the positive structure on R–mod: $f\colon I \to J$ is a weak equivalence if it is one in R–mod (and thus in S–mod), and a fibration if it is a positive fibration in R–mod (and thus in S–mod). Similarly, $\mathsf{Mod}^{l}_{\mathcal{O}}(R)$ has a projective model structure induced from the positive structure on $\mathsf{Sym}(R)$: $f\colon M \to N$ is a weak equivalence if it is one in $\mathsf{Sym}(R)$ (and thus in $\mathsf{Sym}(S)$), and a fibration if it is a positive fibration in $\mathsf{Sym}(R)$ (and thus in $\mathsf{Sym}(S)$).*

The next lemma says that the model structure on $\mathsf{Alg}_{\mathcal{O}}(R)$ is really the same as the model structure on $\mathsf{Mod}^{l}_{\mathcal{O}}(R)$, restricted to the subcategory of modules concentrated in degree 0.

Lemma 2.3 *An algebra map $I \to J$ is a cofibration in $\mathsf{Alg}_{\mathcal{O}}(R)$ if and only if it is a cofibration in $\mathsf{Mod}^{l}_{\mathcal{O}}(R)$, when I and J are regarded as objects in $\mathsf{Sym}(R)$ concentrated in level 0.*

Proof The inclusion $\mathsf{Alg}_{\mathcal{O}}(R) \hookrightarrow \mathsf{Mod}^{l}_{\mathcal{O}}(R)$ has right adjoint given by $M \mapsto M(0)$. This is a Quillen pair, as it is easily checked that this right adjoint preserves weak equivalences and fibrations. □

2.3 Cofibrancy assumption on $\mathcal{O}$ and first consequences

Unless stated otherwise, we make the following standing cofibrancy assumption about our operad $\mathcal{O}$.

Assumption 2.4 *The map $S(1) \to \mathcal{O}$ is assumed to be a positive cofibration in $\mathsf{Sym}(S)$.*

As $\mathcal{O}(0) = *$ has been assumed earlier, equivalently this means that, in S–mod, $S \to \mathcal{O}(1)$ is a cofibration, and $\mathcal{O}(n)$ is cofibrant for all n.

Notation 2.5 Let $\mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$ be the full subcategory of $\mathsf{Alg}_{\mathcal{O}}(R)$ consisting of $\mathcal{O}$–algebras in R–mod which are cofibrant when just viewed as R–modules.

A key advantage of our particular model structure on $\mathsf{Alg}_{\mathcal{O}}(R)$ is that the following property holds.

Proposition 2.6 *The forgetful functor $\mathsf{Alg}_{\mathcal{O}}(R) \to R$–mod preserves cofibrations between cofibrant objects. In particular, if I is cofibrant in $\mathsf{Alg}_{\mathcal{O}}(R)$, then $I \in \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$.*

When $R = S$, this is [14, Theorem 1.5]. We defer the proof of the general case to Section 4.

It follows that a functorial cofibrant replacement functor on $\mathsf{Alg}_{\mathcal{O}}(R)$ takes values in $\mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$. More elementary but also useful is that $\mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$ is well behaved under change of rings.

Lemma 2.7 *Let $R \to R'$ be a map of commutative S–algebras. Then*

$$R' \wedge_R \_\_ \colon \mathsf{Alg}_{\mathcal{O}}(R) \to \mathsf{Alg}_{\mathcal{O}}(R')$$

restricts to a functor

$$R' \wedge_R \_\_ \colon \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}} \to \mathsf{Alg}_{\mathcal{O}}(R')^{\mathrm{c}}$$

which preserves weak equivalences.

Proof This is immediate since $R' \wedge_R \_\_$ is left adjoint to a forgetful functor that is easily seen to be right Quillen. □

2.4 General properties of the bar construction

We will make much use of the bar construction. Given an $\mathcal{O}$–bimodule M and $I \in \mathsf{Alg}_{\mathcal{O}}(R)$, $B(M, \mathcal{O}, I) \in \mathsf{Alg}_{\mathcal{O}}(R)$ is defined as the geometric realization of the simplicial object $B_\bullet(M, \mathcal{O}, I)$ in R–mod defined by

$$B_n(M, \mathcal{O}, I) = M \circ \overbrace{\mathcal{O} \circ \cdots \circ \mathcal{O}}^{n} \circ I.$$

Similarly if M and N are $\mathcal{O}$–bimodules, then $B(M, \mathcal{O}, N)$ is again an $\mathcal{O}$–bimodule.

The theme of the next set of results is that this construction is well behaved when the $\mathcal{O}$–bimodules are positive cofibrant in $\mathsf{Sym}(S)$, and $I \in \mathsf{Alg}_{\mathcal{O}}(R)$ is cofibrant in R–mod. (We recall that a reduced $M \in \mathsf{Sym}(S)$ is positively cofibrant exactly when it is levelwise cofibrant.)

Proposition 2.8 *Let M, N be levelwise cofibrant $\mathcal{O}$–bimodules. Then $B(M, \mathcal{O}, N)$ is again levelwise cofibrant. Similarly, if M is levelwise cofibrant and I is in $\mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$, then $B(M, \mathcal{O}, I) \in \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$.*

The first statement is immediately implied by [14, Theorem 1.6] which says that $B_\bullet(M, \mathcal{O}, N)$ is Reedy cofibrant in the category of simplicial objects in $\mathsf{Sym}(S)$. We defer the proof of the second statement for general R to Section 4.

We also record the following, which shows that the bar construction can be usefully used as a derived circle product. This will also be proved in Section 4.

Proposition 2.9 *Let M be a levelwise cofibrant right $\mathcal{O}$–module. If I is cofibrant in $\mathsf{Alg}_{\mathcal{O}}(R)$, the natural map $B(M, \mathcal{O}, I) \to M \circ_{\mathcal{O}} I$ is a weak equivalence. Similarly, if N is cofibrant in $\mathsf{Mod}^{l}_{\mathcal{O}}(S)$, then $B(M, \mathcal{O}, N) \to M \circ_{\mathcal{O}} N$ is a weak equivalence.*

To emphasize the functors defined by levelwise cofibrant bimodules, we change notation.

Definition 2.10 If M is a levelwise cofibrant $\mathcal{O}$–bimodule, define

$$F_M^R\colon \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}} \to \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$$

by the formula $F_M^R(I) = B(M, \mathcal{O}, I)$.

Theorem 2.11 *The F_M^R construction satisfies the following properties:*

(a) *The functor sending (M, I) to $F_M^R(I)$ takes weak equivalences in either the M or I variable to weak equivalences in $\mathsf{Alg}_{\mathcal{O}}(R)$.*

(b) *A levelwise homotopy fibration*[3] *sequence of levelwise cofibrant $\mathcal{O}$–bimodules*

$$L \to M \to N$$

induces a homotopy fibration sequence in $\mathsf{Alg}_{\mathcal{O}}(R)$

$$F_L^R(I) \to F_M^R(I) \to F_N^R(I).$$

(c) *There is a natural isomorphism of functors*

$$F_M^R \circ F_N^R \simeq F_{B(M,\mathcal{O},N)}^R.$$

(d) *Let $R \to R'$ be a map of commutative S–algebras. There is a natural isomorphism in $\mathsf{Alg}_{\mathcal{O}}(R')$*

$$F_M^{R'}(R' \wedge_R I) \simeq R' \wedge_R F_M^R(I).$$

Parts (a) and (b) will be proved in Section 4. By contrast, parts (c) and (d) are straightforward. Part (c) follows from the natural isomorphism

$$B(M, \mathcal{O}, B(N, \mathcal{O}, I)) \simeq B(B(M, \mathcal{O}, N), \mathcal{O}, I),$$

while part (d) follows from the natural isomorphism

$$R' \wedge_R B(M, \mathcal{O}, I) \simeq B(M, \mathcal{O}, R' \wedge_R I).$$

Remark 2.12 As there is a natural map $B(M, \mathcal{O}, N) \to M \circ_{\mathcal{O}} N$, it follows that a bimodule pairing

$$\mu\colon M \circ_{\mathcal{O}} N \to L$$

[3] Equivalently, we could say cofibration, as levelwise homotopy fibration sequences agree with levelwise homotopy cofibration sequences.

induces a natural transformation

$$\mu\colon F^R_M \circ F^R_N \to F^R_L$$

defined as the composite

$$F^R_M \circ F^R_N \simeq F^R_{B(M,\mathcal{O},N)} \to F^R_{M\circ_{\mathcal{O}}N} \to F^R_L.$$

See Section 3 for examples of this.

2.5 Topological André–Quillen homology

In the next two subsections, we consider our constructions when M is concentrated in just one level, ie there exists an n such that $M(m) = *$ for all $m \neq n$. We show that then $F^R_M(I)$ is determined by $M(n)$ together with the topological André–Quillen homology of I.

We first need to define this last term in our context. The S–module $\mathcal{O}(1)$ will be an associative S–algebra, and can be viewed as an operad concentrated in level 1. From this point of view, the evident maps $\mathcal{O}(1) \to \mathcal{O}$ and $\mathcal{O} \to \mathcal{O}(1)$ are both maps of operads, and the second of these gives $\mathcal{O}(1)$ the structure of an $\mathcal{O}$–bimodule concentrated in level 1.

Let $R\mathcal{O}(1)$–mod be the category of $R \wedge \mathcal{O}(1)$–modules. It is illuminating to note that this category is also $\mathsf{Alg}_{\mathcal{O}(1)}(R)$, when one views $\mathcal{O}(1)$ as an operad. The map $\mathcal{O} \to \mathcal{O}(1)$ induces a functor

$$z\colon R\mathcal{O}(1)\text{–mod} \to \mathsf{Alg}_{\mathcal{O}}(R)$$

with left adjoint

$$Q = \mathcal{O}(1) \circ_{\mathcal{O}} \_\colon \mathsf{Alg}_{\mathcal{O}}(R) \to R\mathcal{O}(1)\text{–mod}$$

making the pair of functors into a Quillen pair.

Definition 2.13 Define $\mathrm{TQ}\colon \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}} \to R\mathcal{O}(1)$–mod by the formula $\mathrm{TQ}(I) = B(\mathcal{O}(1), \mathcal{O}, I)$.

The next proposition is a special case of Proposition 2.9.

Proposition 2.14 *If I is cofibrant in $\mathsf{Alg}_{\mathcal{O}}(R)$, the natural map $\mathrm{TQ}(I) \to Q(I)$ is an equivalence.*

As TQ is thus equivalent to the left derived functor of the left Quillen functor Q, one has the next two consequences. To state the first, we let $[I, J]_{\mathsf{Alg}}$ denote morphisms between I and J in the homotopy category of $\mathsf{Alg}_{\mathcal{O}}(R)$, and we similarly let $[M, N]_{\mathsf{Mod}}$ denote morphisms between M and N in the homotopy category of $R\mathcal{O}(1)$–mod.

Corollary 2.15 *There is an adjunction in the associated homotopy categories*

$$[\mathrm{TQ}(I), N]_{\mathsf{Mod}} \simeq [I, z(N)]_{\mathsf{Alg}}.$$

Corollary 2.16 *If $I \to J \to K$ is a homotopy cofibration sequence in $\mathsf{Alg}_{\mathcal{O}}(R)$, then*

$$\mathrm{TQ}(I) \to \mathrm{TQ}(J) \to \mathrm{TQ}(K)$$

is a homotopy cofibration sequence in $R\mathcal{O}(1)$–mod.

The next result is a particular instance of Theorem 2.11(d).

Proposition 2.17 *Let $R \to R'$ be a map of commutative S–algebras. There is a natural isomorphism*

$$\mathrm{TQ}(R' \wedge_R I) \simeq R' \wedge_R \mathrm{TQ}(I).$$

The first TQ here is with respect to the S–algebra R'.

2.6 $\mathcal{O}$–bimodules with one term

As before, we can view $\mathcal{O}(1)$ as either a commutative S–algebra or an operad concentrated in level 1.

Suppose $M \in \mathsf{Sym}(S)$ is a right $\mathcal{O}(1)$–module with the operad interpretation, ie one has $M \circ \mathcal{O}(1) \to M$ making appropriate diagrams commute. Unraveling the definitions, one sees that this structure map amounts to Σ_n–equivariant maps

$$M(n) \wedge \mathcal{O}(1)^{\wedge n} \to M(n)$$

exhibiting $M(n)$ as an $\mathcal{O}(1)^{\wedge n}$–module. Equivalently, each $M(n)$ will be a right $\Sigma_n \wr \mathcal{O}(1)$–module, where $\Sigma_n \wr \mathcal{O}(1)$ is the associative algebra with underlying S–module $\bigvee_{\sigma \in \Sigma_n} \mathcal{O}(1)^{\wedge n}$, and evident "twisted" multiplication.

From this, it is easy to see that if $J \in \mathsf{Alg}_{\mathcal{O}(1)}(R) = R\mathcal{O}(1)$–mod, then

$$M \circ_{\mathcal{O}(1)} J = \bigvee_n M(n) \wedge_{\Sigma_n \wr \mathcal{O}(1)} J^{\wedge_R n}.$$

Now suppose, given $M(n)$, an $(\mathcal{O}(1), \Sigma_n \wr \mathcal{O}(1))$–bimodule. Abusing notation, we will also write $M(n)$ for the symmetric sequence concentrated at level n:

$$M(n) = (*, \dots, *, M(n), *, \dots).$$

From this point of view, $M(n)$ is precisely an $\mathcal{O}(1)$–bimodule, where $\mathcal{O}(1)$ is viewed as an operad. Furthermore, an $\mathcal{O}$–bimodule structure on $M(n)$ will necessarily be an $\mathcal{O}(1)$–bimodule structure pulled back via the map of operads $\mathcal{O} \to \mathcal{O}(1)$.

Theorem 2.18 *Suppose $M(n)$ is also a cofibrant S–module. Then, for $I \in \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$, there is a natural isomorphism*

$$F^R_{M(n)}(I) = z(M(n) \wedge_{\Sigma_n \wr \mathcal{O}(1)} \mathrm{TQ}(I)^{\wedge_R n}),$$

and a natural equivalence

$$z(B(M(n), \mathcal{O}(1), \mathrm{TQ}(I))) \xrightarrow{\sim} F^R_{M(n)}(I).$$

Proof We suppress some applications of z, the pullback along $\mathcal{O} \to \mathcal{O}(1)$. Firstly, one has natural isomorphisms

$$\begin{aligned} M(n) \wedge_{\Sigma_n \wr \mathcal{O}(1)} \mathrm{TQ}(I)^{\wedge_R n} &= M(n) \wedge_{\Sigma_n \wr \mathcal{O}(1)} B(\mathcal{O}(1), \mathcal{O}, I)^{\wedge_R n} \\ &= M(n) \circ_{\mathcal{O}(1)} B(\mathcal{O}(1), \mathcal{O}, I) \\ &= B(M(n), \mathcal{O}, I) \\ &= F^R_{M(n)}(I). \end{aligned}$$

Secondly, the equivalence $B(M(n), \mathcal{O}(1), \mathcal{O}(1)) \xrightarrow{\sim} M(n)$ induces the equivalence

$$\begin{aligned} B(M(n), \mathcal{O}(1), \mathrm{TQ}(I)) &= B(M(n), \mathcal{O}(1), B(\mathcal{O}(1), \mathcal{O}, I)) \\ &= B(B(M(n), \mathcal{O}(1), \mathcal{O}(1)), \mathcal{O}, I) \\ &\xrightarrow{\sim} B(M(n), \mathcal{O}, I) \\ &= F^R_{M(n)}(I). \end{aligned}$$ □

Corollary 2.19 *Let $f\colon I \to J$ be a morphism in $\mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$. With $M(n)$ as in the theorem, if $\mathrm{TQ}(f)$ is a weak equivalence, so is $F^R_{M(n)}(f)$.*

2.7 The Goodwillie tower of F^R_M

The second author has studied Goodwillie calculus on the category $\mathsf{Alg}_{\mathcal{O}}(R)$ [13]. Here we sketch how our results above lead to an understanding of the Goodwillie tower of the functor F^R_M.

Given a levelwise cofibrant $\mathcal{O}$–bimodule M, let $M^{\leq n}$ denote the $\mathcal{O}$–bimodule with

$$M^{\leq n}(k) = \begin{cases} M(k) & \text{if } k \leq n, \\ * & \text{if } k > n. \end{cases}$$

Definition 2.20 Let $P_n F_M^R = F_{M^{\leq n}}^R \colon \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}} \to \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$.

Theorem 2.21 *The Goodwillie tower of the functor F_M^R identifies with*

$$P_1 F_M^R \leftarrow P_2 F_M^R \leftarrow P_3 F_M^R \leftarrow \cdots,$$

and its n^{th} derivative $\partial_n F_M^R$ identifies with $M(n)$.

Sketch of proof The sequence of $\mathcal{O}$–bimodules

$$M(n) \to M^{\leq n} \to M^{\leq (n-1)}$$

satisfies the hypothesis of Theorem 2.11(b). Thus the homotopy fiber of the map

$$P_n F_M^R(I) \to P_{n-1} F_M^R(I)$$

identifies as $F_{M(n)}^R(I)$, which Theorem 2.18 tells us is

$$z(M(n) \wedge_{\Sigma_n \wr \mathcal{O}(1)} \mathrm{TQ}(I)^{\wedge_R n}).$$

This is a homogeneous n–excisive functor; note that Corollary 2.16 first tells us that TQ is a homogeneous linear functor. See [13, Theorem 3.2] for more detail.

It follows that $P_n F_M^R$ is n–excisive. With a bit more care, one can now check that the natural transformation $F_M^R \to P_n F_M^R$ identifies with the map from F_M^R to its n–excisive quotient: the proof of [13, Theorem 4.3] generalizes immediately to our setting. □

Under connectivity hypotheses, one gets very concrete convergence estimates. Say that $X \in \mathsf{Sym}(S)$ is connective if each $X(n) \in S$–mod is connective, ie -1–connected.

Proposition 2.22 *If R, M, and $\mathcal{O}$ are connective, and I is $(c-1)$–connected, then the map $F_M^R(I) \to P_n F_M^R(I)$ is $(n+1)c$–connected.*

Proof We need to show that the homotopy fiber is $((n+1)c-1)$–connected. By Theorem 2.11(b), this homotopy fiber identifies with $B(M^{>n}, \mathcal{O}, I)$, where

$$M^{>n}(k) = \begin{cases} M(k) & \text{if } k > n, \\ * & \text{if } k \leq n. \end{cases}$$

This fiber then is the homotopy colimit (in R–modules) of a diagram of R–modules of the form

$$M(r) \wedge \mathcal{O}(s_1) \wedge \cdots \wedge \mathcal{O}(s_k) \wedge I^{\wedge_R t},$$

with $t \geq r > n$. In particular, it is a homotopy colimit of a diagram of $((n+1)c-1)$–connected R–modules, and so is itself $((n+1)c-1)$–connected. □

These results also show the following, when combined with Corollary 2.19.

Theorem 2.23 *Let $f\colon I \to J$ be a morphism in $\mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$. If $\mathrm{TQ}(f)$ is a weak equivalence, so is $P_n F^R_M(f)$ for any n and any levelwise cofibrant $\mathcal{O}$–bimodule M. Furthermore, if R, M and $\mathcal{O}$ are connective, and I and J are 0–connected, then $F^R_M(f)$ is itself a weak equivalence.*

Special cases of this theorem appear in [8] and [5].

3 Application to the augmentation ideal filtration

In our constructions, when the $\mathcal{O}$–bimodule is $\mathcal{O}$ itself, the resulting functor, sending an $\mathcal{O}$–algebra I to $F^R_{\mathcal{O}}(I) = B(\mathcal{O}, \mathcal{O}, I)$, is naturally weakly equivalent to the identity. In this section we study structure on the augmentation ideal filtration of I arising from using the levelwise bimodule filtration of $\mathcal{O}$ in conjunction with the operad structure $\mathcal{O} \circ \mathcal{O} \to \mathcal{O}$.

3.1 Construction and basic properties of the filtration

Definitions 3.1 Let $1 \leq i < m \leq \infty$.

(a) Let $\mathcal{O}^m_i$ denote the $\mathcal{O}$–bimodule with $\mathcal{O}^m_i(k) = \begin{cases} \mathcal{O}(k) & \text{if } i \leq k < m, \\ * & \text{otherwise.} \end{cases}$

(b) For $I \in \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$, let $I^i_m = F^R_{\mathcal{O}^m_i}(I) = B(\mathcal{O}^m_i, \mathcal{O}, I)$.

Note that there is a natural weak equivalence $I^1_\infty \to I$. We sometimes write I^i for I^i_∞, and readers are encouraged to view I^i_m as I^i/I^m; see Theorem 3.4(b) below.

For $j \leq i$ and $n \leq m$, it is not hard to see that the evident map

$$\mathcal{O}^m_i \to \mathcal{O}^n_j$$

is a map of $\mathcal{O}$–bimodules, and thus induces natural maps

$$I^i_m \to I^j_n$$

for all $I \in \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$.

Special cases of these are illustrated in the next examples.

Example 3.2 $I \in \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$ has a natural augmentation ideal filtration

$$I \xleftarrow{\sim} I^1 \leftarrow I^2 \leftarrow I^3 \leftarrow \cdots .$$

Example 3.3 $I_n^1 = P_{n-1} F_{\mathcal{O}}^R(I)$ in the notation of the last section, so the tower

$$I_2^1 \leftarrow I_3^1 \leftarrow I_4^1 \leftarrow \cdots$$

identifies with the Goodwillie tower of the identity functor on $\mathsf{Alg}_{\mathcal{O}}(R)$. This tower, defined as we do here, is the subject of study in [5].

These examples are related: the filtration of the first example appears as the homotopy fibers of the maps from I to the tower in the second example. More precisely, there are homotopy fiber sequences

$$I^n \to I^1 \to I_n^1.$$

This is a special case of property (b) in the next theorem.

Theorem 3.4 *The functors sending I to I_n^i satisfy the following properties:*

(a) *They preserve weak equivalences in the variable $I \in \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$.*

(b) *For $1 \le i < m < l \le \infty$, the sequence $I_l^m \to I_l^i \to I_m^i$ is a homotopy fiber sequence. In particular, $I^m \to I^i \to I_m^i$ is a homotopy fiber sequence.*

(c) *There are natural isomorphisms $I_2^1 = z(\mathrm{TQ}(I))$, and more generally, $I_{k+1}^k = z(\mathcal{O}(k) \wedge_{\Sigma_k \wr \mathcal{O}(1)} \mathrm{TQ}(I)^{\wedge_R k})$.*

(d) *Let $R \to R'$ be a map of commutative S–algebras. There is a natural isomorphism $R' \wedge_R I_n^i \simeq (R' \wedge_R I)_n^i$.*

All of these properties follow immediately from the more general results of Section 2. For example, part (b) follows from Theorem 2.11(b) applied to the sequence of $\mathcal{O}$–bimodules

$$\mathcal{O}_m^l \to \mathcal{O}_i^l \to \mathcal{O}_i^m.$$

Our connectivity estimates of Section 2.7 give the following.

Proposition 3.5 *Suppose R and $\mathcal{O}$ are connective. If I is $(c-1)$–connected, then I^n is $(nc-1)$–connected.*

3.2 Composition properties of the filtration

Now we look at composition structure. As will be shown in the proof of the next proposition, it is not hard to see that the operad composition

$$\mu\colon \mathcal{O} \circ \mathcal{O} \to \mathcal{O}$$

induces maps of $\mathcal{O}$–bimodules

$$\mu\colon \mathcal{O}_i^\infty \circ_{\mathcal{O}} \mathcal{O}_j^\infty \to \mathcal{O}_{ij}^\infty,$$

and these pairings, in turn, define natural maps

$$\mu\colon (I^j)^i \to I^{ij}$$

for all $I \in \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$.

With a little more care, one can check the following.

Proposition 3.6 *Given $i < m$, $j < n$, and $ij < N \leq \min(ij + (n-j), mj)$, the operad structure map*

$$\mu\colon \mathcal{O} \circ \mathcal{O} \to \mathcal{O}$$

induces maps of $\mathcal{O}$–bimodules

$$\mu\colon \mathcal{O}_i^m \circ_{\mathcal{O}} \mathcal{O}_j^n \to \mathcal{O}_{ij}^N$$

making the following diagram commute:

$$\begin{tikzcd}
\mathcal{O} \circ_{\mathcal{O}} \mathcal{O} \arrow[d, "\mu", "\wr"'] & \mathcal{O}_i^\infty \circ_{\mathcal{O}} \mathcal{O}_j^\infty \arrow[l, hook'] \arrow[r, two heads] \arrow[d, "\mu"] & \mathcal{O}_i^m \circ_{\mathcal{O}} \mathcal{O}_j^n \arrow[d, "\mu"] \\
\mathcal{O} & \mathcal{O}_{ij}^\infty \arrow[l, hook'] \arrow[r, two heads] & \mathcal{O}_{ij}^N
\end{tikzcd}$$

These thus induce natural maps $\mu\colon (I_n^j)_m^i \to I_N^{ij}$ making the following diagram commute:

$$\begin{tikzcd}
I \arrow[d, equal] & (I^j)^i \arrow[l] \arrow[r] \arrow[d, "\mu"] & (I_n^j)_m^i \arrow[d, "\mu"] \\
I & I^{ij} \arrow[l] \arrow[r] & I_N^{ij}
\end{tikzcd}$$

Proof We begin by observing that $(\mathcal{O}_i^\infty \circ \mathcal{O}_j^\infty)(s)$ equals a wedge of S–modules of the form

$$\mathcal{O}(r) \wedge \mathcal{O}(s_1) \wedge \cdots \wedge \mathcal{O}(s_r)$$

such that $s = s_1 + \cdots + s_r$, $r \geq i$, and $s_k \geq j$ for all k.

These conditions force $s \geq ij$, and thus the dotted arrow exists in the following diagram:

$$\begin{tikzcd} \mathcal{O} \circ \mathcal{O} \arrow[r, "\mu"] & \mathcal{O} \\ \mathcal{O}_i^\infty \circ \mathcal{O}_j^\infty \arrow[u, hook] \arrow[r, dotted, "\mu"] & \mathcal{O}_{ij}^\infty \arrow[u, hook] \end{tikzcd}$$

Similarly, if $ij < N \leq \min(ij + (n-j), mj)$, then the dotted arrow exists in this diagram:

$$\begin{tikzcd} \mathcal{O}_i^\infty \circ \mathcal{O}_j^\infty \arrow[r, "\mu"] \arrow[d, two heads] & \mathcal{O}_{ij}^\infty \arrow[d, two heads] \\ \mathcal{O}_i^m \circ \mathcal{O}_j^n \arrow[r, dotted, "\mu"] & \mathcal{O}_{ij}^N \end{tikzcd}$$

To see this, note that a wedge summand as above maps to $*$ under the quotient $\mathcal{O}_i^\infty \circ \mathcal{O}_j^\infty \twoheadrightarrow \mathcal{O}_i^m \circ \mathcal{O}_j^n$ (the left map of the diagram) exactly when either $r \geq m$ or $s_k \geq n$ for at least one k. In the first case, it follows that $s \geq mj$. In the second case, it follows that $s \geq (r-1)j + n \geq (i-1)j + n = ij + (n-j)$. We conclude that if $N \leq \min(ij + (n-j), mj)$, then $s \geq N$, so this summand also maps to $*$ under the composite $\mathcal{O}_i^\infty \circ \mathcal{O}_j^\infty \xrightarrow{\mu} \mathcal{O}_{ij}^\infty \twoheadrightarrow \mathcal{O}_{ij}^N$. Thus the dotted arrow in the diagram exists.

Thus, the bimodule map $\mathcal{O}_i^m \circ \mathcal{O}_j^n \to \mathcal{O}_{ij}^{\min(ij+(n-j),mj)}$ induces an $\mathcal{O}$–bimodule map $\mathcal{O}_i^m \circ_{\mathcal{O}} \mathcal{O}_j^n \to \mathcal{O}_{ij}^{\min(ij+(n-j),mj)}$. This follows formally from the fact that each of the maps $\mathcal{O} \hookrightarrow \mathcal{O}_i^\infty \twoheadrightarrow \mathcal{O}_i^m$ are maps of $\mathcal{O}$–bimodules, combined with the evident fact that the operad pairing $\mathcal{O} \circ \mathcal{O} \to \mathcal{O}$ induces a map $\mathcal{O} \circ_{\mathcal{O}} \mathcal{O} \to \mathcal{O}$. □

Addendum 3.7 *The construction shows a bit more compatibility than listed above: given $i \leq i'$, $m \leq m'$, $j \leq j'$, $n \leq n'$, $N \leq N'$, with $i < m$, $i' < m'$, $j < n$, $j' < m'$, $ij < N \leq \min(ij + (n-j), mj)$, and $i'j' < N' \leq \min(i'j' + (n'-j'), m'j')$, there is a commutative diagram of $\mathcal{O}$–bimodules*

$$\begin{tikzcd} \mathcal{O}_{i'}^{m'} \circ_{\mathcal{O}} \mathcal{O}_{j'}^{n'} \arrow[r, two heads] \arrow[d, "\mu"] & \mathcal{O}_i^m \circ_{\mathcal{O}} \mathcal{O}_j^n \arrow[d, "\mu"] \\ \mathcal{O}_{i'j'}^{N'} \arrow[r, two heads] & \mathcal{O}_{ij}^N \end{tikzcd}$$

and thus a commutative diagram

$$\begin{tikzcd} (I_{n'}^{j'})_{m'}^{i'} \arrow[r] \arrow[d, "\mu"] & (I_n^j)_m^i \arrow[d, "\mu"] \\ I_{N'}^{i'j'} \arrow[r] & I_N^{ij} \end{tikzcd}$$

Example 3.8 For simplicity, let $D_i(M) = \mathcal{O}(i) \wedge_{\Sigma_i \wr \mathcal{O}(1)} M^{\wedge_R i}$, for $M \in R\mathcal{O}(1)$–mod, and let $T = \mathrm{TQ}$. With this notation, Theorem 2.18 tells us that there is an isomorphism $I^i_{i+1} \simeq zD_i T(I)$. Then there is a commutative diagram

$$\begin{tikzcd}
(I^j_{j+1})^i_{i+1} \arrow[rr, "\mu"] \arrow[d, equal] & & I^{ij}_{ij+1} \arrow[d, equal] \\
zD_i T(zD_j T(I)) \arrow[r] & zD_i D_j T(I) \arrow[r] & zD_{ij} T(I)
\end{tikzcd}$$

where the lower left map is induced by the counit $TzM \to M$ (which one can check is a projection onto a wedge summand), and the lower right map is induced by the operad structure map $\mathcal{O}(i) \wedge \mathcal{O}(j)^{\wedge i} \to \mathcal{O}(ij)$. Diagrams like this suggest that our composition structure should be useful in doing calculations in spectral sequences associated to the augmentation ideal filtration. There are hints of how this might go in [10, Theorem 1.6].

3.3 Application to lifting filtrations

Theorem 3.9 *Let $I, J \in \mathsf{Alg}_{\mathcal{O}}(R)^{\mathrm{c}}$, and let $f\colon I \to J^d$ be a morphism in $\mathsf{Alg}_{\mathcal{O}}(R)$. Then f induces $\mathcal{O}$–algebra maps $f_n\colon I^n \to J^{dn}$ for all n, and the assignment sending f to f_n is both functorial and preserves weak equivalences. Furthermore, the maps f_n are compatible as n varies: for $m < n$, the following diagram commutes:*

$$\begin{tikzcd}
I_n \arrow[r, "f_n"] \arrow[d] & J^{dn} \arrow[d] \\
I^m \arrow[r, "f_m"] & J^{dm}
\end{tikzcd}$$

Proof Let f_n be the composite $I^n \xrightarrow{f^n} (J^d)^n \xrightarrow{\mu} J^{dn}$. □

Definition 3.10 Say that a map $f \in [I, J]_{\mathsf{Alg}}$ has AQ–filtration[4] at least s if f factors in $\mathrm{ho}(\mathsf{Alg}_{\mathcal{O}}(R))$ as the composition of s maps

$$I = I(0) \xrightarrow{f(1)} I(1) \xrightarrow{f(2)} I(2) \to \cdots \to I(s-1) \xrightarrow{f(s)} I(s) = J$$

such that $\mathrm{TQ}(f(i))$ is null for each i.

[4]The reader can decide if AQ stands for André–Quillen or Adams–Quillen.

Theorem 3.11 *Let $f \in [I,J]_{\mathsf{Alg}}$ have AQ–filtration at least s. Then there exists $\tilde{f} \in [I, J^{2^s}]_{\mathsf{Alg}}$ such that*

$$\begin{tikzcd} & J^{2^s} \arrow[d] \\ I \arrow[ur, "\tilde{f}"] \arrow[r, "f"'] & J \end{tikzcd}$$

commutes in $\mathrm{ho}(\mathsf{Alg}_{\mathcal{O}}(R))$.

Proof We work in $\mathrm{ho}(\mathsf{Alg}_{\mathcal{O}}(R))$.

Let $f = f(s) \circ \cdots \circ f(1)$ as in Definition 3.10.

For each i between 1 and s, there is an exact sequence of pointed sets

$$[I(i-1), I(i)^2]_{\mathsf{Alg}} \to [I(i-1), I(i)]_{\mathsf{Alg}} \to [I(i-1), I(i)^1_2]_{\mathsf{Alg}},$$

and there are identifications

$$[I(i-1), I(i)^1_2]_{\mathsf{Alg}} \simeq [I(i-1), z(\mathrm{TQ}(I(i)))]_{\mathsf{Alg}} \simeq [\mathrm{TQ}(I(i-1)), \mathrm{TQ}(I(i))]_{\mathsf{Mod}}.$$

It follows that since $\mathrm{TQ}(f(i))$ is null, $f(i)$ lifts to $\tilde{f}(i)\colon I(i-1) \to I(i)^2$. Then Theorem 3.9 gives maps

$$\tilde{f}(i)_{2^{i-1}}\colon I(i-1)^{2^{i-1}} \to I(i)^{2^i}.$$

Now let $\tilde{f}$ be the composite of these s maps: $\tilde{f} = \tilde{f}(s)_{2^{s-1}} \circ \cdots \circ \tilde{f}(1)$. □

The theorem, combined with Proposition 3.5, has the following corollary.

Corollary 3.12 *Suppose that R and $\mathcal{O}$ are connective and $J \in \mathsf{Alg}_{\mathcal{O}}(R)$ is $(c-1)$–connected. If $f\colon I \to J$ has AQ–filtration s, then $f_*\colon \pi_*(I) \to \pi_*(J)$ will be zero for $* < 2^s c$.*

For more results in this spirit, see [9].

4 Deferred proofs

In this section we prove Propositions 2.6, 2.8, and 2.9 and Theorem 2.11. When $R = S$, so that our algebras just have the underlying structure of an S–module, these results can be deduced from the second author's work, specifically [14, Theorem 1.1]. The case of a general R requires a suitable generalization of that result, which we state as Theorem 4.4.

4.1 The homotopical behavior of the composition product

Fixing a commutative S–algebra R, it is useful to generalize the context slightly.

Notation 4.1 Let $\mathcal{P}$ be an operad in R–mod, ie a monoid object for the monoidal structure $\circ_R$ in $\mathsf{Sym}(R)$ defined just as in (1) but with $\wedge$ replaced by $\wedge_R$. We then denote by $\mathsf{Mod}^r_{\mathcal{P}}$, $\mathsf{Mod}^l_{\mathcal{P}}$, and $\mathsf{Alg}_{\mathcal{P}}$ the associated categories of left modules, right modules, and algebras over $\mathcal{P}$ in $\mathsf{Sym}(R)$. We endow $\mathsf{Mod}^l_{\mathcal{P}}$, and $\mathsf{Alg}_{\mathcal{P}}$ with the model structure as in Theorem 2.2.[5]

Remark 4.2 As noted in Remark 2.1, there are identifications

$$\mathsf{Alg}_{R\wedge\mathcal{O}} \simeq \mathsf{Alg}_{\mathcal{O}}(R) \simeq \mathsf{Alg}_{R\wedge\mathcal{O}}(S)$$

and

$$\mathsf{Mod}^l_{R\wedge\mathcal{O}} \simeq \mathsf{Mod}^l_{\mathcal{O}}(R) \simeq \mathsf{Mod}^l_{R\wedge\mathcal{O}}(S).$$

By contrast, one only has a proper inclusion of categories

$$\mathsf{Mod}^r_{R\wedge\mathcal{O}} \subset \mathsf{Mod}^r_{R\wedge\mathcal{O}}(S),$$

where $\mathsf{Mod}^r_{R\wedge\mathcal{O}}$ is the category of right $R\wedge\mathcal{O}$–modules in $\mathsf{Sym}(R)$, and $\mathsf{Mod}^r_{R\wedge\mathcal{O}}(S)$ is the category of right $R\wedge\mathcal{O}$–modules in $\mathsf{Sym}(S)$.

To see the reason for this, assume for simplicity that $\mathcal{O}(1) = S$. Then if $N \in \mathsf{Mod}^r_{R\wedge\mathcal{O}}(S)$, $N(n)$ will be a right $\Sigma_n \wr R$–module. But unwinding definitions reveals that, for any $N \in \mathsf{Mod}^r_{R\wedge\mathcal{O}}$, this $\Sigma_n \wr R$–module structure on $M(n)$ must be one pulled back along the canonical ring map $\Sigma_n \wr R \to \Sigma_n \times R$.

To state our main technical theorem, we need the following construction.

Definition 4.3 Suppose given a map $f_1\colon M \to N$ in $\mathsf{Mod}^r_{\mathcal{P}}$ and a map $f_2\colon A \to B$ in $\mathsf{Mod}^l_{\mathcal{P}}$. Let $(M \circ_{\mathcal{P}} B) \vee_{M\circ_{\mathcal{P}} A} (N \circ_{\mathcal{P}} A)$ be the pushout of the diagram

$$\begin{tikzcd} M \circ_{\mathcal{P}} A \arrow[r, "f_1\circ_{\mathcal{P}} A"] \arrow[d, "M\circ_{\mathcal{P}} f_2"'] & N \circ_{\mathcal{P}} A \\ M \circ_{\mathcal{P}} B & \end{tikzcd}$$

in $\mathsf{Sym}(R)$, and then define the *pushout circle product* of f_1 and f_2, to be the natural map

$$f_1 \square^{\circ_{\mathcal{P}}} f_2\colon (M \circ_{\mathcal{P}} B) \vee_{M\circ_{\mathcal{P}} A} (N \circ_{\mathcal{P}} A) \to N \circ_{\mathcal{P}} B.$$

[5] Note that we do not need to equip $\mathsf{Mod}^r_{\mathcal{P}}$ with a model structure.

Theorem 4.4 *Suppose $f_2\colon A \to B$ is a cofibration between cofibrant objects in $\mathsf{Mod}^l_{\mathcal{P}}$. If a $f_1\colon M \to N$ in $\mathsf{Mod}^r_{\mathcal{P}}$ is an underlying positive cofibration in $\mathsf{Sym}(R)$, then so is*

$$f_1 \square^{\circ_{\mathcal{P}}} f_2\colon (M \circ_{\mathcal{P}} B) \vee_{M \circ_{\mathcal{P}} A} (N \circ_{\mathcal{P}} A) \to N \circ_{\mathcal{P}} B.$$

Furthermore, this map will be a weak equivalence if either f_1 or f_2 is a weak equivalence.

When $R = S$, this theorem nearly coincides with [14, Theorem 1.1], and we defer the proof in the general case to the Appendix. For the purpose of proving results stated in Section 2, we will just need the following corollary.

Corollary 4.5 *Let $\mathcal{O}$ be an operad in S–mod. Suppose $f_2\colon I \to J$ is a cofibration between cofibrant objects in $\mathsf{Alg}_{\mathcal{O}}(R)$. If a map $f_1\colon M \to N$ in $\mathsf{Mod}^r_{\mathcal{O}}(S)$ is an underlying positive cofibration in $\mathsf{Sym}(S)$, then*

$$f_1 \square^{\circ_{\mathcal{O}}} f_2\colon (M \circ_{\mathcal{O}} J) \vee_{M \circ_{\mathcal{O}} I} (N \circ_{\mathcal{O}} I) \to N \circ_{\mathcal{O}} J$$

will be a positive cofibration in R–mod.

Furthermore, this map will be a weak equivalence if either f_1 or f_2 is a weak equivalence.

Proof Since the functor $R \wedge \_\colon \mathsf{Sym}(S) \to \mathsf{Sym}(R)$ sends positive cofibrations and trivial cofibrations in $\mathsf{Sym}(S)$ respectively to positive cofibrations and trivial cofibrations in $\mathsf{Sym}(R)$, the result follows immediately from Theorem 4.4 applied to $\mathcal{P} = R \wedge \mathcal{O}$, $R \wedge f_1$ and f_2. Note that the positive model structure on $\mathsf{Sym}(R)$ restricts on level 0 to the positive model structure on R–mod. □

4.2 Proofs of results from Section 2

Proof of Proposition 2.6 If f_1 is the map $* \to \mathcal{O}$, and $f_2\colon I \to J$ is map in $\mathsf{Alg}_{\mathcal{O}}(R)$, then $f_1 \square^{\circ_{\mathcal{O}}} f_2$ is just the map $f_2\colon I \to J$, now viewed as a map in R–mod.

If I is cofibrant in $\mathsf{Alg}_{\mathcal{O}}(R)$, then applying Corollary 4.5 to the map $f_2\colon * \to I$, shows that I will be cofibrant in R–mod.

Similarly, if $f_2\colon I \to J$ is a cofibration between cofibrant objects in $\mathsf{Alg}_{\mathcal{O}}(R)$, we learn that $f_2\colon I \to J$ is a cofibration in R–mod. □

Proof of Proposition 2.8 For the first statement, we note that $B(M, \mathcal{O}, N)$ is the realization of the simplicial object $B_\bullet(M, \mathcal{O}, N)$, and thus will be cofibrant in $\mathsf{Sym}(S)$ if $B_\bullet(M, \mathcal{O}, N)$ is Reedy cofibrant in $\mathsf{Sym}(S)^{\Delta^{\mathrm{op}}}$. That this is true, under our hypotheses on M and N, is precisely the conclusion of [14, Theorem 1.6].

Proving the second statement is similar: one sees that $B_\bullet(M,\mathcal{O},I)$ is Reedy cofibrant in R–mod$^{\Delta^{\mathrm{op}}}$ by noting that the proof of [14, Theorem 1.6] (and in particular that of the auxiliary [14, Lemma 5.47]) goes through if one simply replaces the very last application of [14, Theorem 1.1] by an application of Corollary 4.5. □

Proof of Proposition 2.9 First note that by Corollary 4.5 the functor

$$M \circ_{\mathcal{O}} \_\colon \mathsf{Alg}_{\mathcal{O}}(R) \to R\text{–mod}$$

sends trivial cofibrations between cofibrant algebras to weak equivalences, and hence, by Ken Brown's lemma [6, Corollary 7.7.2], also preserves all weak equivalences between cofibrant algebras.

Hence, rewriting the map

$$B(M,\mathcal{O},I) \to M \circ_{\mathcal{O}} I$$

as

$$M \circ_{\mathcal{O}} (B(\mathcal{O},\mathcal{O},I) \to I)$$

one sees that it suffices to show that $B(\mathcal{O},\mathcal{O},I)$ is cofibrant in $\mathsf{Alg}_{\mathcal{O}}(R)$.

$B(\mathcal{O},\mathcal{O},I)$ is the realization of the simplicial algebra $B_\bullet(\mathcal{O},\mathcal{O},I)$, viewed as a simplicial object in R–mod. By [5, Proposition 6.11], this agrees with the realization of $B_\bullet(\mathcal{O},\mathcal{O},I)$, viewed as a simplicial object in $\mathsf{Alg}_{\mathcal{O}}(R)$. Thus it suffices to show that $B_\bullet(\mathcal{O},\mathcal{O},I)$ is Reedy cofibrant in $\mathsf{Alg}_{\mathcal{O}}(R)^{\Delta^{\mathrm{op}}}$.

Checking this involves showing that the latching maps for $B_\bullet(\mathcal{O},\mathcal{O},I)$ are cofibrations in $\mathsf{Alg}_{\mathcal{O}}(R)$. These depend only on $B_\bullet(\mathcal{O},\mathcal{O},I)$ together with its degeneracies, ie face maps can be ignored. From this perspective

$$B_\bullet(\mathcal{O},\mathcal{O},I) \simeq \mathcal{O} \circ B_\bullet(S(1),\mathcal{O},I),$$

where $S(1)$ is our notation for the unit symmetric sequence $(*,S,*,*,\dots)$ under $\circ$.

Hence, letting $\ell_n^{\mathcal{O}}$ and ℓ_n respectively denote the n^{th} latching map construction on $\mathbb{N}$–graded objects with degeneracies in $\mathsf{Alg}_{\mathcal{O}}(R)$ and R–mod, one has

$$\ell_n^{\mathcal{O}}(B_\bullet(\mathcal{O},\mathcal{O},I)) \simeq \mathcal{O} \circ \ell_n(B_\bullet(S(1),\mathcal{O},I)).$$

Since $\mathcal{O} \circ \_\colon R\text{–mod} \to \mathsf{Alg}_{\mathcal{O}}(R)$ is a left Quillen functor, $\ell_n^{\mathcal{O}}(B_\bullet(\mathcal{O},\mathcal{O},I))$ will be a cofibration in $\mathsf{Alg}_{\mathcal{O}}(R)$ if $\ell_n(B_\bullet(S(1),\mathcal{O},I))$ is a cofibration in R–mod. But the latter map *is* a cofibration, since it is a special case of the latching maps shown to be cofibrations in the proof of Proposition 2.8. □

Proof of Theorem 2.11(a) and (b) In this proof we focus on the identification $\mathrm{Alg}_{\mathcal{O}}(R) \simeq \mathrm{Alg}_{R\wedge\mathcal{O}}(S)$ so as to be able to directly apply [14, Theorem 1.1].

For part (a), note first that

$$F_M^R(I) = M \circ_{\mathcal{O}} B(\mathcal{O}, \mathcal{O}, I).$$

That $F_M^R(I)$ preserves weak equivalences in the I variable then follows from the proof of Proposition 2.9, where it was shown both that $B(\mathcal{O}, \mathcal{O}, I)$ is a cofibrant algebra and that $M \circ_{\mathcal{O}} \_\_$ preserves weak equivalences between cofibrant algebras.

To see that weak equivalences are also preserved in the M variable, one uses a similar argument: using the identifications of Remark 4.2 to change perspective to S–mod, one applies [14, Theorem 1.1] to any trivial cofibration $f_1\colon M \to N$ in $\mathrm{Mod}^r_{R\wedge\mathcal{O}}(S)$ and the map $f_2 = * \to B(\mathcal{O}, \mathcal{O}, I)$. One concludes that the functor sending M to $F_M^R(I)$ sends trivial cofibrations to weak equivalences. One now again uses Ken Brown's lemma.

The intuition behind part (b) comes from the observation that (1), the formula for the composition product $X \circ Y$ of symmetric sequences, is linear in the variable X. Our official proof goes as follows. Note that homotopy fibration sequences in $\mathrm{Alg}_{\mathcal{O}}(R)$ are detected by considering them as sequences in S–mod. Again using the identifications of Remark 4.2 to change perspectives, one immediately reduces to [14, Theorem 1.8] applied to the operad $R \wedge \mathcal{O}$ in S–modules. □

Appendix: Proof of Theorem 4.4

We now turn to the task of proving Theorem 4.4. If one just tries to redo all the work in [14] with a general commutative S–algebra R replacing occurrences of S, one finds that most of results generalize, with the key exception being the characterization of S cofibrations in [14, Proposition 3.9], which fails for general R (and, in particular, cofibrations in $\mathrm{Sym}(R)$ can not be detected by first forgetting the Σ_n–actions at each level). Here we take a somewhat blended approach: we use a string of arguments from [14] to ultimately reduce ourselves to a situation covered by [14, Theorem 1.1].

We begin by noting the following elementary lemma, a consequence of the fact that the positive model structure on $\mathrm{Sym}(R)$ is the projective structure induced from the positive model structure on $\mathrm{Sym}(S)$.

Lemma A.1 *A set of generating cofibrations for* $\mathrm{Sym}(R)$ *can be obtained by applying* $R \wedge \_\_$ *to a set of generating cofibrations for* $\mathrm{Sym}(S)$.

Let us remind ourselves of our goal. Given $f_1\colon M \to N$ in $\mathsf{Mod}^r_{\mathcal{P}}$ and $f_2\colon A \to B$ in $\mathsf{Mod}^l_{\mathcal{P}}$, we are considering the pushout corner map, in $\mathsf{Sym}(R)$, of the following commutative square:

$$
\begin{tikzcd}
M \circ_{\mathcal{P}} A \arrow[r, "M \circ_{\mathcal{P}} f_2"] \arrow[d, "f_1 \circ_{\mathcal{P}} A"] & M \circ_{\mathcal{P}} B \arrow[d, "f_1 \circ_{\mathcal{P}} B"] \\
N \circ_{\mathcal{P}} A \arrow[r, "N \circ_{\mathcal{P}} f_2"] & N \circ_{\mathcal{P}} B
\end{tikzcd}
\tag{2}
$$

By this we mean the map from the pushout of the upper left corner of the square to the lower right term.

We wish to show that if f_2 is a cofibration between cofibrant objects[6] in $\mathsf{Mod}^l_{\mathcal{P}}$, then if f_1 is a positive cofibration in $\mathsf{Sym}(R)$, so is the pushout corner map. Furthermore, in this situation, if either f_1 or f_2 is a weak equivalence, so is the pushout corner map.

We will address this last statement at the end of the Appendix, and focus on the first statement. For this, we try to follow the proof of [14, Theorem 1.1], which is the case when $R = S$. The majority of the arguments in that proof are agnostic as to the category or model structure used — in particular, the filtrations of [14, Proposition 5.20] cover R–mod — with the exception of the two instances where [14, Theorems 1.2, 1.3] are used.

As in [14], we first assume that f_2 is a map between algebras, rather than more general left $\mathcal{P}$–modules. In this case, arguing as in [14, Section 5.4], one reduces to the case where $f_2\colon A \to B$ is the pushout of a generating cofibration. Using Lemma A.1, this means that f_2 is the lower horizontal map of a pushout in $\mathsf{Alg}_{\mathcal{P}}$ of the form

$$
\begin{tikzcd}
\mathcal{P} \circ_R (R \wedge X) \arrow[r] \arrow[d] & \mathcal{P} \circ_R (R \wedge Y) \arrow[d] \\
A \arrow[r, "f_2"] & B
\end{tikzcd}
$$

with $X \to Y$ a generating positive cofibration in S–mod.

The key is now to use the infinite filtration of the horizontal maps in (2) given by [14, Proposition 5.20]. (This key filtration is a generalization of similar filtrations appearing in [5, Proposition 5.10] and [2, proofs of Theorems 1.4 and 12.5].) Arguing as in [14, Section 5.4], one is reduced to studying the pushout corner maps of the following

[6]This is a bit redundant: if A is cofibrant, and f_2 is a cofibration, then B is necessarily cofibrant.

squares, for which we will shortly explain our notation:

$$\begin{tikzcd} M_A(r) \wedge_{R\times\Sigma_r} (R \wedge Q^r_{r-1}) \arrow[r] \arrow[d] & M_A(r) \wedge_{R\times\Sigma_r} (R \wedge Y^{\wedge r}) \arrow[d] \\ N_A(r) \wedge_{R\times\Sigma_r} (R \wedge Q^r_{r-1}) \arrow[r] & N_A(r) \wedge_{R\times\Sigma_r} (R \wedge Y^{\wedge r}) \end{tikzcd} \tag{3}$$

Firstly, if we view $X \to Y$ as a functor $\{\mathbf{0} \to \mathbf{1}\} \to S$–mod, we can smash this functor with itself r times, obtaining a cubical diagram $\{\mathbf{0} \to \mathbf{1}\}^{\times r} \to S$–mod. We let Q^r_{r-1} denote the colimit of this cube with the terminal object $\mathbf{1}^r$ removed; this comes with an evident map $Q^r_{r-1} \to Y^{\wedge r}$.

Secondly, as in [14, Definition 5.15], M_A denotes the $M \circ_{\mathcal{P}} \big(\mathcal{P} \coprod A\big)$, where the coproduct is taken in $\mathsf{Mod}^l_{\mathcal{P}}$.

We wish to show that the pushout corner map of (3) is a positive cofibration in R–mod. Since $X \to Y$ is a positive cofibration in S–mod, [14, Theorem 1.2] tells us that $Q^r_{r-1} \to Y^{\wedge r}$ is appropriately cofibrant in the category of S–modules with a Σ_r action.

If the map $M_A \to N_A$ were a generating positive cofibration in $\mathsf{Sym}(R)$, one would be able to pull a $R \wedge (-)$ factor out of the pushout corner map (by Lemma A.1), reducing to the S case, which in turn follows by applying [14, Theorems 1.2, 1.3] as in the proof of [14, Theorem 1.1].

Hence, by standard arguments, it suffices to show that $M_A \to N_A$ is a positive cofibration in $\mathsf{Sym}(R)$. This would follow from the special case of our theorem when f_2 has the form $i\colon \mathcal{P} \to \mathcal{P} \coprod A$, which would say that the pushout corner map of the middle square of the diagram

$$\begin{tikzcd} M \arrow[r, equal] \arrow[d, "f_1"] & M \circ_{\mathcal{P}} \mathcal{P} \arrow[r, "M\circ_{\mathcal{P}} i"] \arrow[d, "f_1\circ_{\mathcal{P}}\mathcal{P}"] & M \circ_{\mathcal{P}} \big(\mathcal{P} \coprod A\big) \arrow[r, equal] \arrow[d, "f_1\circ_{\mathcal{P}}(\mathcal{P}\coprod A)"] & M_A \arrow[d] \\ N \arrow[r, equal] & N \circ_{\mathcal{P}} \mathcal{P} \arrow[r, "N\circ_{\mathcal{P}} i"] & N \circ_{\mathcal{P}} \big(\mathcal{P} \coprod A\big) \arrow[r, equal] & N_A \end{tikzcd} \tag{4}$$

is a positive cofibration in $\mathsf{Sym}(R)$.

Now we use our assumption that A is cofibrant in $\mathsf{Alg}_{\mathcal{P}}$, and basically proceed as before. The map i can be assumed to be an infinite composition of maps of the form $\mathcal{P} \coprod A_\beta \xrightarrow{\mathcal{P}\coprod i_\beta} \mathcal{P} \coprod A_{\beta+1}$, where i_β is the lower horizontal map of a pushout in $\mathsf{Alg}_{\mathcal{P}}$

of the form

$$\begin{tikzcd}
\mathcal{P} \circ_R (R \wedge X_\beta) \arrow[r] \arrow[d] & \mathcal{P} \circ_R (R \wedge Y_\beta) \arrow[d] \\
A_\beta \arrow[r] & A_{\beta+1}
\end{tikzcd}$$

with $X_\beta \to Y_\beta$ a generating positive cofibration in S–mod.

It suffices to show by induction on β that $N_{A_\beta} \coprod_{M_{A_\beta}} M_{A_{\beta+1}} \to N_{A_{\beta+1}}$ is a positive cofibration. Note that the induction hypothesis then implies $M_{A_\beta} \to N_{A_\beta}$ is a positive cofibration.

After a filtration argument as before, one is left needing to show that the pushout corner map in

$$\begin{tikzcd}
M_{(\mathcal{P} \coprod A_\beta)}(r) \check{\wedge}_{R \times \Sigma_r} (R \wedge Q^r_{r-1}) \arrow[r] \arrow[d] & M_{(\mathcal{P} \coprod A_\beta)}(r) \check{\wedge}_{R \times \Sigma_r} (R \wedge Y_\beta^{\check{\wedge} r}) \arrow[d] \\
N_{(\mathcal{P} \coprod A_\beta)}(r) \check{\wedge}_{R \times \Sigma_r} (R \wedge Q^r_{r-1}) \arrow[r] & N_{(\mathcal{P} \coprod A_\beta)}(r) \check{\wedge}_{R \times \Sigma_r} (R \wedge Y_\beta^{\check{\wedge} r})
\end{tikzcd}$$

is a positive cofibration in $\mathsf{Sym}(R)$, where $\check{\wedge}$ denotes the smash product in $\mathsf{Sym}(R)$.

Using the obvious analogue of Lemma A.1 for R bisymmetric sequences and [14, Propositions 5.43, 5.44] (the analogues of [14, Theorems 1.2, 1.3] for $\mathsf{Sym}(S)$) just as in the argument following (3), one further reduces to just needing to show that $M_{\mathcal{P} \coprod A_\beta} \to N_{\mathcal{P} \coprod A_\beta}$ is a positive cofibration in $\mathsf{biSym}(R)$, the category of bisymmetric sequences of R–modules. (The notion of cofibration is defined by analogy with $\mathsf{Sym}(R)$.) But since [14, Proposition 5.19] identifies the (r,s) level of this map with $M_{A_\beta}(r+s) \to N_{A_\beta}(r+s)$, the result follows by our induction hypothesis.

To deal with the case of f_2 a general map of left modules one repeats the argument in the last paragraph of the proof of [14, Theorem 1.1].

Finally, the case where either f_1 or f_2 are additionally weak equivalences follows by using the identifications of Remark 4.2 to reduce the question to the S–mod level and then applying the monomorphism part of [14, Theorem 1.1].

References

[1] **M Basterra**, *André–Quillen cohomology of commutative S–algebras*, J. Pure Appl. Algebra 144 (1999) 111–143 MR

[2] **A D Elmendorf**, **M A Mandell**, *Rings, modules, and algebras in infinite loop space theory*, Adv. Math. 205 (2006) 163–228 MR

[3] **J E Harper**, *Homotopy theory of modules over operads in symmetric spectra*, Algebr. Geom. Topol. 9 (2009) 1637–1680 MR

[4] **J E Harper**, *Bar constructions and Quillen homology of modules over operads*, Algebr. Geom. Topol. 10 (2010) 87–136 MR

[5] **J E Harper**, **K Hess**, *Homotopy completion and topological Quillen homology of structured ring spectra*, Geom. Topol. 17 (2013) 1325–1416 MR

[6] **P S Hirschhorn**, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, Amer. Math. Soc., Providence, RI (2003) MR

[7] **M Hovey**, **B Shipley**, **J Smith**, *Symmetric spectra*, J. Amer. Math. Soc. 13 (2000) 149–208 MR

[8] **N J Kuhn**, *Localization of André–Quillen–Goodwillie towers, and the periodic homology of infinite loopspaces*, Adv. Math. 201 (2006) 318–378 MR

[9] **N J Kuhn**, *Adams filtration and generalized Hurewicz maps for infinite loopspaces*, preprint (2014) arXiv

[10] **N Kuhn**, **J McCarty**, *The mod* 2 *homology of infinite loopspaces*, Algebr. Geom. Topol. 13 (2013) 687–745 MR

[11] **M A Mandell**, **J P May**, **S Schwede**, **B Shipley**, *Model categories of diagram spectra*, Proc. London Math. Soc. 82 (2001) 441–512 MR

[12] **V Minasian**, *André–Quillen spectral sequence for* THH, Topology Appl. 129 (2003) 273–280 MR

[13] **L A Pereira**, *Goodwillie calculus in the category of operads over a spectral operad*, preprint (2015) Available at `http://www.faculty.virginia.edu/luisalex/Papers/GoodwillieCalculusAlgO.pdf`

[14] **L A Pereira**, *Cofibrancy of operadic constructions in positive symmetric spectra*, Homology Homotopy Appl. 18 (2016) 133–168 MR

[15] **B Shipley**, *A convenient model category for commutative ring spectra*, 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) 473–483 MR

Department of Mathematics, University of Virginia
Charlottesville, VA 22904, United States

`njk4x@virginia.edu`, `lp2h@virginia.edu`

`http://pi.math.virginia.edu/Faculty/Kuhn/`

Received: 21 February 2016 Revised: 20 June 2016

Algebraic & Geometric Topology 17 (2017) 1131–1150

Odd primary homotopy types of SU(n)–gauge groups

STEPHEN THERIAULT

Let $\mathcal{G}_k(\mathrm{SU}(n))$ be the gauge group of the principal $\mathrm{SU}(n)$–bundle with second Chern class k. If p is an odd prime and $n \le (p-1)^2+1$, we classify the p–local homotopy types of $\mathcal{G}_k(\mathrm{SU}(n))$.

55P15; 54C35

1 Introduction

Let G be a topological group, B be a space and $P \to B$ be a principal G–bundle over B. The *gauge group* $\mathcal{G}(P)$ is the group of G–equivariant automorphisms of P that fix B. Crabb and Sutherland [2] showed that, even though there may be infinitely many inequivalent principal G–bundles over B, their gauge groups have only finitely many distinct homotopy types. However, their argument did not give an explicit enumeration of the homotopy types. Using different methods, Kono [14] gave an explicit enumeration in the case of gauge groups of principal $\mathrm{SU}(2)$–bundles over S^4. He then asked whether this can be done more generally.

Since then there has been considerable effort to classify the homotopy types of gauge groups in specific cases. Let G be a simply connected, simple compact Lie group and let BG be its classifying space. The number of equivalence classes of principal G–bundles over S^4 is in one-to-one correspondence with homotopy classes of maps $[S^4, BG] \cong \mathbb{Z}$, and the correspondence in the case of $G = \mathrm{SU}(n)$ is given by the value of the second Chern class. Let $P_k \to S^4$ be the principal G–bundle corresponding to $k \in \mathbb{Z}$ and let $\mathcal{G}_k(G)$ be its gauge group. For integers a and b, let (a,b) be their greatest common divisor. Then:

- $\mathcal{G}_k(\mathrm{SU}(2)) \simeq \mathcal{G}_{k'}(\mathrm{SU}(2))$ if and only if $(12,k) = (12,k')$ (Kono [14]);
- $\mathcal{G}_k(\mathrm{SU}(3)) \simeq \mathcal{G}_{k'}(\mathrm{SU}(3))$ if and only if $(24,k) = (24,k')$ (Hamanaka and Kono [6]);
- $\mathcal{G}_k(\mathrm{Sp}(2)) \simeq_{(p)} \mathcal{G}_{k'}(\mathrm{Sp}(2))$ if and only if $(40,k) = (40,k')$ (Theriault [21]);
- $\mathcal{G}_k(\mathrm{SU}(5)) \simeq_{(p)} \mathcal{G}_{k'}(\mathrm{SU}(5))$ if and only if $(120,k) = (120,k')$ (Theriault [23]);

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.1131

where the homotopy equivalence in the third and fourth cases is p–local for any prime p or rational (using $p=0$ to indicate rational localization). Bounds, but not a classification, were obtained in the case of $\mathcal{G}_k(G_2)$ in Kishimoto, Theriault and Tsutaya [13], and classifications involving spheres of different dimensions or non-simply connected Lie groups can be found in Hamanaka and Kono [7], Kamiyama, Kishimoto, Kono and Tsukuda [10] and Kishimoto, Kono and Tsutaya [12]. In all cases the fixed number in the greatest common divisor is the order of the Samelson product $S^3 \wedge G \xrightarrow{\langle i,1\rangle} G$, where i is the inclusion of the bottom cell and 1 is the identity map on G.

Here we consider $\mathcal{G}_k(\mathrm{SU}(n))$ for all n. There is a canonical map $j_n\colon \Sigma\mathbb{C}\mathrm{P}^{n-1} \to \mathrm{SU}(n)$ that induces a projection onto the generating set in cohomology. In what follows, while spaces will be localized at a prime p, it is more illuminating to write the order of a map as an integer m rather than the p–component of m. We prove the following:

Theorem 1.1 *Localize at an odd prime p. Then:*

(a) *If $n \geq 2$, the composite $S^3 \wedge \Sigma\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{1\wedge j_n} S^3 \wedge \mathrm{SU}(n) \xrightarrow{\langle i,1\rangle} \mathrm{SU}(n)$ has order at most $n(n^2-1)$.*

(b) *If $n \leq (p-1)^2+1$, the composite $S^3 \wedge \Sigma\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{1\wedge j_n} S^3 \wedge \mathrm{SU}(n) \xrightarrow{\langle i,1\rangle} \mathrm{SU}(n)$ has order exactly $n(n^2-1)$.*

(c) *If $n \leq (p-1)^2+1$, the map $S^3 \wedge \mathrm{SU}(n) \xrightarrow{\langle i,1\rangle} \mathrm{SU}(n)$ has order $n(n^2-1)$.*

(d) *If $n \leq (p-1)^2+1$, there is a homotopy equivalence $\mathcal{G}_k(\mathrm{SU}(n)) \simeq \mathcal{G}_{k'}(\mathrm{SU}(n))$ if and only if $(n(n^2-1),k) = (n(n^2-1),k')$.*

Theorem 1.1(d) significantly improves on the known classifications of the homotopy types of gauge groups. It is the first general result; all the previous results held for specific Lie groups G and involved proofs that used particular properties of that Lie group. It recovers the known odd primary information for $\mathrm{SU}(2)$, $\mathrm{SU}(3)$ and $\mathrm{SU}(5)$ and gives exact information in a large range of previously unknown cases. For example, $\mathcal{G}_k(\mathrm{SU}(4)) \simeq_{(p)} \mathcal{G}_{k'}(\mathrm{SU}(4))$ at odd primes if and only if $(60,k) = (60,k')$.

Hamanaka and Kono [6], refining a result of Sutherland [19], showed that for any $n \geq 2$ the (integral) order of $\langle i,1\rangle$ is at least $n(n^2-1)$ by considering certain homotopy sets $[X,\mathrm{SU}(n)]$ where X is a sphere or a suspension of $\mathbb{C}\mathrm{P}^2$. Theorem 1.1(a)–(b) are much stronger reformulations of their result at odd primes. We obtain part (a) by closely examining a map constructed by Toda [24] to give a topological proof of Bott periodicity. The restriction to $n \leq (p-1)^2+1$ in parts (b) and (c) arises from the fact that for these n the space $\Sigma\mathbb{C}\mathrm{P}^{n-1}$ "generates" $\mathrm{SU}(n)$ in a sense that will be made precise in Section 5. Part (d) follows as a consequence of part (c) and other general results, to be discussed in Section 6.

Acknowledgement The author would like to thank the referee for making many suggestions that have improved the exposition of the paper.

2 Gauge groups and function spaces

In this section we discuss some general results that translate the study of gauge groups into that of function spaces, which is more suited to topological methods. This holds for any simply connected, simple compact Lie group G, so it is stated that way. Since $[S^4, BG] \cong \mathbb{Z}$, principal G–bundles over S^4 are classified by their second Chern class, which can take any integer value. Let $P_k \to S^4$ be the principal G–bundle corresponding to $k \in \mathbb{Z}$, and let $\mathcal{G}_k(G)$ be its gauge group. As this is a group it has a classifying space $B\mathcal{G}_k(G)$.

Let $\mathrm{Map}(S^4, BG)$ and $\mathrm{Map}^*(S^4, BG)$ respectively be the spaces of freely continuous and pointed continuous maps between S^4 and BG. The components of each space are in one-to-one correspondence with the integers, where the integer is determined by the degree of a map $S^4 \to BG$. By [3] or [1], there is a homotopy equivalence $B\mathcal{G}_k(G) \simeq \mathrm{Map}_k(S^4, BG)$ between $B\mathcal{G}_k(G)$ and the component of $\mathrm{Map}(S^4, BG)$ consisting of maps of degree k. Evaluating a map at the basepoint of S^4, we obtain a map ev: $B\mathcal{G}_k(G) \xrightarrow{\mathrm{ev}} BG$ whose fibre is homotopy equivalent to $\mathrm{Map}^*_k(S^4, BG)$. It is well known that each component of $\mathrm{Map}^*(S^4, BG)$ is homotopy equivalent to $\Omega^3_0 G$, the component of $\Omega^3 G$ containing the basepoint. Putting all this together, for each $k \in \mathbb{Z}$ there is a homotopy fibration sequence

$$G \xrightarrow{\partial_k} \Omega^3_0 G \to B\mathcal{G}_k(G) \xrightarrow{\mathrm{ev}} BG, \tag{1}$$

where ∂_k is the fibration connecting map. In particular, the gauge group $\mathcal{G}_k(G)$ is the homotopy fibre of ∂_k, and it is by understanding the map ∂_k that information will be deduced about $\mathcal{G}_k(G)$.

Note that, while the components of $\mathrm{Map}^*(S^4, BG)$ are all homotopy equivalent, the same need not be true for the components of $\mathrm{Map}(S^4, BG)$. For example, in [25; 26] it was shown that there is a homotopy equivalence

$$\mathrm{Map}_k(S^4, B\mathrm{SU}(2)) \simeq \mathrm{Map}_{k'}(S^4, B\mathrm{SU}(2))$$

if and only if $k = \pm k'$. However, many components become homotopy equivalent after looping. In the $\mathrm{SU}(2)$ example, Kono [14] showed that $\Omega\mathrm{Map}_k(S^4, B\mathrm{SU}(2)) \simeq \Omega\mathrm{Map}_{k'}(S^4, B\mathrm{SU}(2))$ (ie $\mathcal{G}_k(\mathrm{SU}(2)) \simeq \mathcal{G}_{k'}(\mathrm{SU}(2))$) if and only if $(12, k) = (12, k')$. This example also shows that Theorem 1.1(d) likely cannot be upgraded to a statement about the classifying spaces of the relevant gauge groups.

The triple adjoint of ∂_k was described in [15, Theorem 2.7]. More precisely, the homotopy class of a homotopy fibration connecting map is determined only up to self-homotopy equivalences of its domain and range; in [15, Theorem 2.7] it was shown that choices of self-homotopy equivalences could be made which allow for the triple adjoint of ∂_k to be described in terms of Samelson products. In fact, in [15] a four-fold adjoint is taken using the fact that $G \simeq \Omega BG$, and this four-fold adjoint is described in terms of Whitehead products. We choose for ease of presentation to use only a three-fold adjoint, which is described in terms of a Samelson product.

Let $i\colon S^3 \to G$ be the inclusion of the bottom cell and let $1\colon G \to G$ be the identity map. In general, for an H–space Y, let $k\colon Y \to Y$ be the k^{th}–power map.

Lemma 2.1 *The triple adjoint of the map $G \xrightarrow{\partial_k} \Omega_0^3 G$ is homotopic to the Samelson product $S^3 \wedge G \xrightarrow{\langle k\circ i,1\rangle} G$.* □

The linearity of the Samelson product implies that $\langle k\circ i, 1\rangle \simeq k\circ\langle i,1\rangle$. Taking adjoints therefore implies the following:

Corollary 2.2 *There is a homotopy $\partial_k \simeq k\circ\partial_1$.* □

In what follows we will prove results about the order of ∂_1 by proving them about the order of $\langle i,1\rangle$.

3 Properties of Toda's map

Take cohomology with $\mathbb{Z}$–coefficients. Recall that $H^*(\mathbb{C}\mathrm{P}^\infty) \cong \mathbb{Z}[x]$, where x has degree 2. Write σx^i for the image of x^i under the suspension isomorphism $H^{2i}(\mathbb{C}\mathrm{P}^\infty) \xrightarrow{\cong} H^{2i+1}(\Sigma\mathbb{C}\mathrm{P}^\infty)$. In his topological proof of Bott periodicity, Toda [24] constructed a map

$$f\colon \Sigma^3\mathbb{C}\mathrm{P}^\infty \to \Sigma\mathbb{C}\mathrm{P}^\infty$$

with the property that $f^*(\sigma x^m) = m\sigma^3 x^{m-1}$ for $m \geq 2$. Composing, we obtain a composite

$$g\colon \Sigma^5\mathbb{C}\mathrm{P}^\infty \xrightarrow{\Sigma^2 f} \Sigma^3\mathbb{C}\mathrm{P}^\infty \xrightarrow{f} \Sigma\mathbb{C}\mathrm{P}^\infty$$

with the property that $g^*(\sigma x^m) = m(m-1)\sigma^5 x^{m-2}$ for $m \geq 3$.

Let

$$g_{2n+1}\colon \Sigma^5\mathbb{C}\mathrm{P}^{n-2} \to \Sigma\mathbb{C}\mathrm{P}^n$$

be the restriction of g to $(2n+1)$–skeletons. Then skeletal inclusions give a commutative square:

$$\begin{tikzcd} \Sigma^5\mathbb{C}P^{n-3} \arrow[r, "g_{2n-1}"] \arrow[d] & \Sigma\mathbb{C}P^{n-1} \arrow[d] \\ \Sigma^5\mathbb{C}P^{n-1} \arrow[r, "g_{2n+3}"] & \Sigma\mathbb{C}P^{n+1} \end{tikzcd}$$

Let $X^{n+1} = \mathbb{C}P^{n+1}/\mathbb{C}P^{n-1}$ be the stunted complex projective space. The commutativity of the preceding square implies that there is a homotopy cofibration diagram

(2)
$$\begin{tikzcd} \Sigma^5\mathbb{C}P^{n-3} \arrow[r, "g_{2n-1}"] \arrow[d] & \Sigma\mathbb{C}P^{n-1} \arrow[d] \\ \Sigma^5\mathbb{C}P^{n-1} \arrow[r, "g_{2n+3}"] \arrow[d] & \Sigma\mathbb{C}P^{n+1} \arrow[d] \\ \Sigma^5 X^{n-1} \arrow[r, "\overline{g}_{2n+3}"] \arrow[d] & \Sigma X^{n+1} \arrow[d] \\ \Sigma^6\mathbb{C}P^{n-3} \arrow[r, "\Sigma g_{2n-1}"] & \Sigma^2\mathbb{C}P^{n-1} \end{tikzcd}$$

for some map $\overline{g}_{2n+3}$.

We describe some properties of X^{n+1} and $\overline{g}_{2n+3}$. The quotient map $\mathbb{C}P^{n+1} \to X^{n+1} = \mathbb{C}P^{n+1}/\mathbb{C}P^{n-1}$ induces a map $H^*(X^{n+1}) \to H^*(\mathbb{C}P^{n+1})$. A straightforward long exact sequence argument immediately shows that $H^*(X^{n+1}) \cong \mathbb{Z}\{y_n, y_{n+1}\}$, where y_n and y_{n+1} are the images of x^n and x^{n+1}, respectively. The homotopy commutativity of the middle square in (2) and the description of g^* immediately imply the following:

Lemma 3.1 *We have:*

(a) $(\overline{g}_{2n+3})^*(\sigma y_n) = n(n-1)\sigma^5 y_{n-2}$.

(b) $(\overline{g}_{2n+3})^*(\sigma y_{n+1}) = (n+1)n\,\sigma^5 y_{n-1}$. □

It is useful to identify the homotopy type of X^{n+1}. Observe that X^{n+1} has a CW–structure consisting of two cells, one each in dimensions $2n$ and $2n+2$. The structure of the Steenrod algebra on $\mathbb{C}P^{n+1}$ (see [27, Chapter VIII, Theorem 9.2], for example) implies that there is a Sq^2 connecting the two generators in $H^*(X^{n+1}; \mathbb{Z}/2\mathbb{Z})$ if and only if n is odd. Also, Sq^2 detects the stable generator η_m of $\pi_{m+1}(S^m) \cong \mathbb{Z}/2\mathbb{Z}$ (see [27, Chapter VIII, Corollary 8.8], for example). As the cofibre of η_m is $\Sigma^{m-2}\mathbb{C}P^2$, we obtain the following:

Lemma 3.2 *We have:*

(a) *If n is odd then $X^{n+1} \simeq \Sigma^{2n-2}\mathbb{C}P^2$.*

(b) *If n is even then $X^{n+1} \simeq S^{2n} \vee S^{2n+2}$.* □

At this point we localize at an odd prime p; the explanation as to why this is done will be deferred to Remark 4.5. At odd primes, the map η_m generating the stable group $\pi_{m+1}(S^m) \cong \mathbb{Z}/2\mathbb{Z}$ is null homotopic. Thus $\Sigma^{m-2}\mathbb{C}P^2 \simeq S^m \vee S^{m+2}$ for $m \geq 3$. Consequently, Lemma 3.2 implies the following:

Corollary 3.3 *Localize at an odd prime p. Then $X^{n+1} \simeq S^{2n} \vee S^{2n+2}$.* □

By Corollary 3.3, the map $\Sigma^5 X^{n-1} \xrightarrow{\overline{g}_{2n+3}} \Sigma X^{n+1}$ is a self-map of $S^{2n+1} \vee S^{2n+3}$. This lets us determine the homotopy class of $\overline{g}_{2n+3}$. In general, suppose that there is a map $h\colon S^m \vee S^{m+2} \to S^m \vee S^{m+2}$, where $m \geq 3$. Let h_1 and h_2 be the restrictions of h to S^m and S^{m+2}, respectively. The map $h_1\colon S^m \to S^m \vee S^{m+2}$ is determined by pinching to S^m and S^{m+2}. The pinch map to S^m is a map of some degree, say d_1, and the pinch to S^{m+2} is null homotopic for dimensional reasons. Thus $h_1 \simeq d_1 + *$. Since $m \geq 2$, the Hilton–Milnor theorem implies that $\pi_{m+2}(S^m \vee S^{m+2}) \cong \pi_{m+2}(S^m) \oplus \pi_{m+2}(S^{m+2})$. Therefore, the map $h_2\colon S^{m+2} \to S^m \vee S^{m+2}$ is also determined by pinching it to S^m and S^{m+2}. The pinch map to S^m is an element of $\pi_{m+2}(S^m) \cong 0$ (at odd primes) and the pinch map to S^{m+2} is a map of some degree, say d_2. Therefore $h_2 \simeq * + d_2$. Thus $h \simeq d_1 \vee d_2$. In particular, the homotopy class of h is determined by the map it induces in cohomology. Hence, from Lemma 3.1 we immediately obtain the following:

Lemma 3.4 *Localize at an odd prime p. The map $\Sigma^5 X^{n-1} \xrightarrow{\overline{g}_{2n+3}} \Sigma X^{n+1}$ is homotopic to the wedge of degree maps $d_1 \vee d_2$, where $d_1 = n(n-1)$ and $d_2 = (n+1)n$.* □

4 An upper bound for the order of ∂_1 in the unitary case

We wish to estimate the order of the map $SU(n) \xrightarrow{\partial_1} \Omega_0^3 SU(n)$. By Lemma 2.1, it is equivalent to calculate the order of the Samelson product $S^3 \wedge SU(n) \xrightarrow{\langle i,1\rangle} SU(n)$. Let $SU(\infty)$ be the infinite special unitary group and let t be the standard inclusion $t\colon SU(n) \to SU(\infty)$. There is a homotopy fibration sequence

$$\Omega(SU(\infty)/SU(n)) \to SU(n) \xrightarrow{t} SU(\infty) \to SU(\infty)/SU(n).$$

Since t is a group homomorphism, it is an H–map, and so it commutes with Samelson products. That is, $t \circ \langle i, 1\rangle \simeq \langle t \circ i, t\rangle$. Since $SU(\infty)$ is an infinite loop space, it is homotopy commutative. Therefore the Samelson product $\langle t \circ i, t\rangle$ is null homotopic, implying that there is a lift

$$\begin{tikzcd} & \Omega(\mathrm{SU}(\infty)/\mathrm{SU}(n)) \arrow[d] \\ S^3 \wedge \mathrm{SU}(n) \arrow[ur, "\lambda"] \arrow[r, "\langle i,1\rangle"] & \mathrm{SU}(n) \end{tikzcd}$$

for some map λ.

There is a canonical map

$$j_n\colon \Sigma\mathbb{C}\mathrm{P}^{n-1} \to \mathrm{SU}(n),$$

which induces an epimorphism in cohomology (see [27, Chapter VII, Section 4.6], for example). Let

$$q_n\colon \mathbb{C}\mathrm{P}^{n-1} \to X^{n-1} = \mathbb{C}\mathrm{P}^{n-1}/\mathbb{C}\mathrm{P}^{n-3}$$

be the quotient map. Observe that $\Omega(\mathrm{SU}(\infty)/\mathrm{SU}(n))$ is $(2n-1)$–connected. Therefore, the restriction of the composite $S^3\wedge\Sigma\mathbb{C}\mathrm{P}^{n-1}\xrightarrow{1\wedge j_n}S^3\wedge\mathrm{SU}(n)\xrightarrow{\lambda}\Omega(\mathrm{SU}(\infty)/\mathrm{SU}(n))$ to $S^3\wedge\Sigma\mathbb{C}\mathrm{P}^{n-3}$ is null homotopic. This implies that there is a homotopy commutative diagram

$$\begin{tikzcd} & S^3\wedge\Sigma X^{n-1} \arrow[rr, "\nu"] & & \Omega(\mathrm{SU}(\infty)/\mathrm{SU}(n)) \arrow[d] \\ S^3\wedge\Sigma\mathbb{C}\mathrm{P}^{n-1} \arrow[ur, "\Sigma^4 q_n"] \arrow[r, "1\wedge j_n"] & S^3\wedge\mathrm{SU}(n) \arrow[urr, "\lambda"] \arrow[rr, "\langle i,1\rangle"] & & \mathrm{SU}(n) \end{tikzcd} \tag{3}$$

for some map ν. Hamanaka and Kono [5, Proposition 5.2] showed that λ can be chosen so that $\lambda^*\circ(1\wedge j_n)*$ is a degree 1 isomorphism in dimensions $2n$ and $2n+2$. The homotopy commutativity of the left square in (3) therefore implies that ν^* is a degree 1 isomorphism in dimensions $2n$ and $2n+2$.

Let Y be the $(2n+2)$–skeleton of $\Omega(\mathrm{SU}(\infty)/\mathrm{SU}(n))$ and let $\overline{\nu}\colon \Sigma^4X^{n-1}\to Y$ be the factorization of ν through the $(2n+2)$–skeleton. An integral homology Serre spectral sequence calculation for the homotopy fibration $\Omega\mathrm{SU}(\infty)\to\Omega(\mathrm{SU}(\infty)/\mathrm{SU}(n))\to\mathrm{SU}(n)$ immediately shows that Y can be given a CW–structure with two cells, one each in dimensions $2n$ and $2n+2$. Localized at an odd prime this implies that $Y\simeq S^{2n}\vee S^{2n+2}$. Thus $\overline{\nu}$ is a self-map of $S^{2n}\vee S^{2n+2}$, so arguing as for Lemma 3.4 shows that $\overline{\nu}$ is homotopic to a wedge of degree maps. Since ν^* is a degree 1 isomorphism in dimensions $2n$ and $2n+2$. Hence, including Y into $\Omega(\mathrm{SU}(\infty)/\mathrm{SU}(n))$ we obtain the following:

Lemma 4.1 *The map $\Sigma^4X^{n-1}\xrightarrow{\nu}\Omega(\mathrm{SU}(\infty)/\mathrm{SU}(n))$ induces a degree 1 isomorphism in cohomology in dimensions $2n$ and $2n+2$.* □

We now assemble several pieces of information, aimed at establishing the homotopy commutativity of (10). The naturality of the Samelson product implies that the composite $S^3 \wedge \Sigma\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{1\wedge j_n} S^3 \wedge \mathrm{SU}(n) \xrightarrow{\langle i,1\rangle} \mathrm{SU}(n)$ is the Samelson product $\langle i, j_n\rangle$. Notice that the map $\Omega(\mathrm{SU}(\infty)/\mathrm{SU}(n)) \to \mathrm{SU}(n)$ is a loop map. So we can take the adjoint of diagram (3) to obtain a homotopy commutative diagram

$$\begin{tikzcd}
\Sigma^5 X^{n-1} \arrow[r, "\widetilde{\nu}"] & \mathrm{SU}(\infty)/\mathrm{SU}(n) \arrow[d] \\
\Sigma^5\mathbb{C}\mathrm{P}^{n-1} \arrow[u, "\Sigma^5 q_n"] \arrow[r, "\widetilde{\langle i,j_n\rangle}"] & B\mathrm{SU}(n)
\end{tikzcd} \tag{4}$$

where $\widetilde{\langle i, j_n\rangle}$ and $\widetilde{\nu}$ are the adjoints of $\langle i, j_n\rangle$ and ν, respectively.

On the other hand, by Corollary 3.3, $\Sigma^5 X^{n-1} \simeq \Sigma X^{n+1} \simeq S^{2n+1} \vee S^{2n+3}$ and, by Lemma 4.1, $\widetilde{\nu}$ induces a degree 1 isomorphism in cohomology in dimensions $2n+1$ and $2n+3$. So we may choose a homotopy equivalence $\Sigma^5 X^{n-1} \simeq \Sigma X^{n+1}$ so that there is a homotopy commutative diagram

$$\begin{tikzcd}
\Sigma^5\mathbb{C}\mathrm{P}^{n-1} \arrow[r] & \Sigma^5 X^{n-1} \arrow[d, "\wr"] \arrow[dr, "\widetilde{\nu}"] & \\
 & \Sigma X^{n+1} \arrow[r, "\iota"] & \mathrm{SU}(\infty)/\mathrm{SU}(n)
\end{tikzcd} \tag{5}$$

where ι is the inclusion.

Next, the map $\Sigma\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{j_n} \mathrm{SU}(n)$ is natural with respect to the usual inclusion of $\mathrm{SU}(n)$ into $\mathrm{SU}(n+1)$. Let j be the composite $\Sigma\mathbb{C}\mathrm{P}^{n+1} \xrightarrow{j_{n+2}} \mathrm{SU}(n+2) \to \mathrm{SU}(\infty)$. Consider the diagram:

$$\begin{tikzcd}
\Sigma\mathbb{C}\mathrm{P}^{n-1} \arrow[r] \arrow[d, "j_n"] & \Sigma\mathbb{C}\mathrm{P}^{n+1} \arrow[r, "\Sigma q_{n+2}"] \arrow[d, "j"] & \Sigma X^{n+1} \arrow[r] \arrow[d, "a"] & \Sigma^2\mathbb{C}\mathrm{P}^{n-1} \arrow[d, "b"] \\
\mathrm{SU}(n) \arrow[r] & \mathrm{SU}(\infty) \arrow[r] & \mathrm{SU}(\infty)/\mathrm{SU}(n) \arrow[r] & B\mathrm{SU}(n)
\end{tikzcd}$$

The left square commutes by the naturality of j_n. As the top row is a homotopy cofibration and the bottom row is a homotopy fibration, the homotopy commutativity of the left square induces the middle square and right square for some maps a and b. It is standard that the CW–structure of $\mathrm{SU}(\infty)/\mathrm{SU}(n)$ is taken so that a is the inclusion of the $(2n+3)$–skeleton. The Peterson–Stein formulas (see [8, Formula 3.4.2]) imply that the map b may be taken to be the adjoint $\tilde{j}_n$ of j_n. Therefore, reorienting the right square in the previous diagram, we obtain a homotopy commutative square:

$$\begin{tikzcd} \Sigma X^{n+1} \arrow[r, "a"] \arrow[d] & \mathrm{SU}(\infty)/\mathrm{SU}(n) \arrow[d] \\ \Sigma^2\mathbb{C}\mathrm{P}^{n-1} \arrow[r, "\tilde{j}_n"] & B\mathrm{SU}(n) \end{tikzcd} \tag{6}$$

The last preparatory step is to manipulate degree maps on $\Sigma^5\mathbb{C}\mathrm{P}^{n-1}$. By [17], there is a p–local homotopy equivalence

$$\Sigma\mathbb{C}\mathrm{P}^{n-1} \simeq \bigvee_{i=1}^{p-1} A_i,$$

where $H^*(A_i;\mathbb{Z}/p\mathbb{Z})$ consists of those elements in $H^*(\Sigma\mathbb{C}\mathrm{P}^{n-1};\mathbb{Z}/p\mathbb{Z})$ that are in degrees of the form $2i+2(p-1)k+1$ for some $k\geq 0$. Select indices i_0 and i_1 such that $2i_0+2(p-1)k_0+1=2n-3$ and $2i_1+2(p-1)k_1+1=2n-1$ for some integers k_0 and k_1. Then A_{i_0} and A_{i_1} inherit the mod-p cohomology generators of $\Sigma\mathbb{C}\mathrm{P}^{n-1}$ in degrees $2n-3$ and $2n-1$, respectively. For dimension and connectivity reasons, there is a homotopy commutative diagram

$$\begin{tikzcd} \Sigma\mathbb{C}\mathrm{P}^{n-1} \arrow[rr, "\Sigma q_n"] \arrow[d, "\wr"] & & \Sigma X^{n-1} \arrow[d, "\wr"] \\ \bigvee_{i=1}^{p-1} A_i \arrow[r, "Q"] & A_{i_0}\vee A_{i_1} \arrow[r, "q_{i_0}\vee q_{i_1}"] & S^{2n-3}\vee S^{2n-1} \end{tikzcd} \tag{7}$$

where Q is the pinch map onto the two designated summands and q_{i_0} and q_{i_1} are the pinch maps onto the top cells of A_{i_0} and A_{i_1}, respectively. Let $\overline{\gamma}$ be the composite

$$\overline{\gamma}\colon \Sigma^5\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{\simeq} \bigvee_{i=1}^{p-1}\Sigma^4 A_i \xrightarrow{\Sigma^4 Q} \Sigma^4 A_{i_0}\vee\Sigma^4 A_{i_1} \xrightarrow{d_1\vee d_2} \Sigma^4 A_{i_0}\vee\Sigma^4 A_{i_1} \hookrightarrow \bigvee_{i=1}^{p-1}\Sigma^4 A_i \xrightarrow{\simeq} \Sigma^5\mathbb{C}\mathrm{P}^{n-1},$$

where $d_1=n(n-1)$ and $d_2=(n+1)n$ are the degree maps. Notice that all the maps in the composite defining $\overline{\gamma}$ are suspensions and so they commute with degree maps. Therefore, the definition of $\overline{\gamma}$ and (7) imply that there is a homotopy commutative diagram:

(8)
$$\begin{tikzcd}
\Sigma^5\mathbb{C}\mathrm{P}^{n-1} \arrow[r, "\overline{\gamma}"] \arrow[d, "\Sigma^5 q_n"] & \Sigma^5\mathbb{C}\mathrm{P}^{n-1} \arrow[d, "\Sigma^5 q_n"] \\
\Sigma^5 X_{n-1} \arrow[d, "\wr"] & \Sigma^5 X_{n-1} \arrow[d, "\wr"] \\
S^{2n+1} \vee S^{2n+3} \arrow[r, "d_1 \vee d_2"] & S^{2n+1} \vee S^{2n+3}
\end{tikzcd}$$

It is also useful at this point to define the map γ by the composite

$$\gamma\colon \Sigma^5\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{\simeq} \bigvee_{i=1}^{p-1} \Sigma^4 A_i \xrightarrow{\Sigma^4 Q} \Sigma^4 A_{i_0} \vee \Sigma^4 A_{i_1} \xrightarrow{d \vee d} \Sigma^4 A_{i_0} \vee \Sigma^4 A_{i_1}$$
$$\hookrightarrow \bigvee_{i=1}^{p-1} \Sigma^4 A_i \xrightarrow{\simeq} \Sigma^5\mathbb{C}\mathrm{P}^{n-1},$$

where $d = n(n^2-1)$. As before, the definition of γ and (7) imply that there is a homotopy commutative diagram:

(9)
$$\begin{tikzcd}
\Sigma^5\mathbb{C}\mathrm{P}^{n-1} \arrow[r, "\gamma"] \arrow[d, "\Sigma^5 q_n"] & \Sigma^5\mathbb{C}\mathrm{P}^{n-1} \arrow[d, "\Sigma^5 q_n"] \\
\Sigma^5 X_{n-1} \arrow[r, "d"] & \Sigma^5 X_{n-1}
\end{tikzcd}$$

Now we put things together. Consider the diagram:

(10)
$$\begin{tikzcd}
\Sigma^5\mathbb{C}\mathrm{P}^{n-1} \arrow[r, "\overline{\gamma}"] \arrow[dd, "\Sigma^5 q_n"] & \Sigma^5\mathbb{C}\mathrm{P}^{n-1} \arrow[d, "\Sigma^5 q_n"] & \\
& \Sigma^5 X^{n-1} \arrow[d, "\wr"] \arrow[dr, "\widetilde{\nu}"] & \\
\Sigma^5 X^{n-1} \arrow[r, "\overline{g}_{2n+3}"] \arrow[d] & \Sigma X^{n+1} \arrow[r, "a"] \arrow[d] & \mathrm{SU}(\infty)/\mathrm{SU}(n) \arrow[d] \\
\Sigma^6\mathbb{C}\mathrm{P}^{n-3} \arrow[r, "\Sigma g_{2n-1}"] & \Sigma^2\mathbb{C}\mathrm{P}^{n-1} \arrow[r, "\widetilde{j}_n"] & B\mathrm{SU}(n)
\end{tikzcd}$$

The upper left rectangle homotopy commutes by (8) and the description of $\overline{g}_{2n+3}$ in Lemma 3.4; the upper right triangle homotopy commutes by (5); the lower left square homotopy commutes by (2); and the lower right square homotopy commutes by (6). Thus the entire diagram homotopy commutes. In the upper direction around the diagram, by (4) the composite

$$\Sigma^5\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{\Sigma^5 q_n} \Sigma^5 X^{n-1} \xrightarrow{\widetilde{\nu}} \mathrm{SU}(\infty)/\mathrm{SU}(n) \to B\mathrm{SU}(n)$$

is homotopic to $\widetilde{\langle i, j_n\rangle}$, so the upper direction around the diagram is homotopic to $\widetilde{\langle i, j_n\rangle} \circ \overline{\gamma}$. On the other hand, in the lower direction around the diagram the left column is null homotopic since it is two consecutive maps in a homotopy cofibration. Therefore $\widetilde{\langle i, j_n\rangle} \circ \overline{\gamma}$ is null homotopic.

Lemma 4.2 *Localize at an odd prime p. The map $\Sigma^5\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{\widetilde{\langle i,j_n\rangle}} B\mathrm{SU}(n)$ has order dividing $n(n^2-1)$.*

Proof We wish to show that the composite $\Sigma^5\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{d} \Sigma^5\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{\widetilde{\langle i,j_n\rangle}} B\mathrm{SU}(n)$ is null homotopic, where $d = n(n^2-1)$. For convenience, label the map from $\mathrm{SU}(\infty)/\mathrm{SU}(n)$ to $B\mathrm{SU}(n)$ by δ. Consider the string of homotopies

$$\widetilde{\langle i, j_n\rangle} \circ d \simeq \delta \circ \widetilde{\nu} \circ \Sigma^5 q_n \circ d \simeq \delta \circ \widetilde{\nu} \circ d \circ \Sigma^5 q_n \simeq \delta \circ \widetilde{\nu} \circ \Sigma^5 q_n \circ \gamma \simeq \widetilde{\langle i, j_n\rangle} \circ \gamma.$$

From left to right, the first homotopy holds by (4), the second holds since maps which are suspensions commute with degree maps, the third holds by (9), and the fourth holds by (4). Observe that, by its definition, γ factors through $\overline{\gamma}$. Therefore, as $\widetilde{\langle i, j_n\rangle} \circ \overline{\gamma}$ is null homotopic, so is $\widetilde{\langle i, j_n\rangle} \circ \gamma$. Hence $\widetilde{\langle i, j_n\rangle} \circ d$ is null homotopic, as asserted. □

Recall that, by definition, $\widetilde{\langle i, j_n\rangle}$ is the adjoint of the Samelson product $\langle i, j_n\rangle$, which is homotopic to the composite $S^3 \wedge \Sigma\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{1\wedge j_n} S^3 \wedge \mathrm{SU}(n) \xrightarrow{\langle i,1\rangle} \mathrm{SU}(n)$. Therefore Lemma 4.2 immediately implies the following:

Corollary 4.3 *Localize at an odd prime p. The composite $S^3 \wedge \Sigma\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{1\wedge j_n} S^3 \wedge \mathrm{SU}(n) \xrightarrow{\langle i,1\rangle} \mathrm{SU}(n)$ has order at most $n(n^2-1)$.* □

Also, the map $\langle i, 1\rangle$ is the triple adjoint of the map $\mathrm{SU}(n) \xrightarrow{\partial_1} \Omega^3_0\mathrm{SU}(n)$. Therefore, Corollary 4.3 implies the following:

Proposition 4.4 *Localize at an odd prime p. The composite $\Sigma\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{j_n} \mathrm{SU}(n) \xrightarrow{\partial_1} \Omega^3_0\mathrm{SU}(n)$ has order at most $n(n^2-1)$.* □

Remark 4.5 It is unclear whether an analogue of Proposition 4.4 holds at the prime 2. At odd p the splitting $\Sigma\mathbb{C}\mathrm{P}^{n-1} \simeq \bigvee_{i=1}^{p-1} A_i$ lets us work with the two different degree maps $d_1 = n(n-1)$ and $d_2 = (n+1)n$ on S^{2n+1} and S^{2n+3} separately. However, at 2 there is no such splitting of $\Sigma\mathbb{C}\mathrm{P}^{n-1}$, so it is unclear how a map $\overline{\gamma}$ can be defined so as to obtain a diagram as in (8).

5 Retractile Lie groups and an upper bound for the order of ∂_1

This section is aimed at proving parts (b) and (c) of Theorem 1.1. To do so we consider those n for which $\mathrm{SU}(n)$ has the special property of being retractile (a term defined in a moment). The results in this section hold in more generality than just the $\mathrm{SU}(n)$ case, so they are stated this way. Throughout this section spaces and maps have been localized at an odd prime p and homology is taken with mod-p coefficients.

Definition 5.1 An H–space B is *retractile* if there is a space A and a map $i\colon A \to B$ with the following properties:

(i) There is an algebra isomorphism $H_*(B) \cong \Lambda(\widetilde{H}_*(A))$.

(ii) i_* is the inclusion of the generating set in homology.

(iii) The map $\Sigma A \xrightarrow{\Sigma i} \Sigma B$ has a left homotopy inverse.

Many simply connected, simple compact Lie groups are retractile. In [20] it was shown that the retractile cases are:

$$\begin{array}{llll} \mathrm{SU}(n), & n \le (p-1)^2+1; & G_2, F_4, E_6, & p \ge 3; \\ \mathrm{Sp}(n), & 2n \le (p-1)^2+1; & E_7, & p \ge 5; \\ \mathrm{Spin}(2n+1), & 2n \le (p-1)^2+1; & E_8, & p \ge 7; \\ \mathrm{Spin}(2n), & 2(n-1) \le (p-1)^2+1; & & \end{array}$$

further, in each of these cases the space A is a co-H–space.

Let G be a simply connected, simple compact Lie group which is retractile. Recall the map $G \xrightarrow{\partial_1} \Omega^3_0 G$ from Section 2. Define the map $\overline{\partial}_1$ by the composite

$$\overline{\partial}_1\colon A \xrightarrow{i} G \xrightarrow{\partial_1} \Omega^3_0 G.$$

We will prove the following, which relates the order of ∂_1 to that of $\overline{\partial}_1$:

Proposition 5.2 *Let G be a retractile, simply connected, simple compact Lie group. If $p^r \circ \overline{\partial}_1$ is null homotopic then so is $p^r \circ \partial_1$.*

Proposition 5.2 is very useful, in practise it tends to be much easier to prove facts about $\overline{\partial}_1$ rather than ∂_1. The proposition says that this is good enough. Granting Proposition 5.2, we obtain the following corollary, which lets us go on to prove Theorem 1.1(a)–(c).

Corollary 5.3 *Let G be a retractile simply connected, simple compact Lie group. Then ∂_1 has order p^r if and only if $\overline{\partial}_1$ has order p^r.*

Proof If ∂_1 has order p^r then, as $\overline{\partial}_1$ factors through ∂_1, the order of $\overline{\partial}_1$ is at most p^r. Conversely, if $\overline{\partial}_1$ has order p^r then by Proposition 5.2 ∂_1 has order at most p^r. Hence, if a and b are the orders of ∂_1 and $\overline{\partial}_1$, respectively, then $a \le b \le a$, implying that $a = b$. □

Proof of Theorem 1.1(a)–(c) Part (a) is the statement of Corollary 4.3. For part (c), $\mathrm{SU}(n)$ is retractile if $n \le (p-1)^2+1$, and in this case the space A is $\Sigma\mathbb{C}\mathrm{P}^{n-1}$ and the map i is the map $\Sigma\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{j_n} \mathrm{SU}(n)$. Therefore the composite $\Sigma\mathbb{C}\mathrm{P}^{n-1} \xrightarrow{j_n} \mathrm{SU}(n) \xrightarrow{\partial_1} \Omega^3_0\mathrm{SU}(3)$ is $\overline{\partial}_1$. Let p^r be the p–component of $n(n^2-1)$. By Proposition 4.4, $p^r \circ \overline{\partial}_1$ is null homotopic, so Proposition 5.2 implies that $p^r \circ \partial_1$ is null homotopic. By [6, Theorem 2.4 and Lemma 2.5], ∂_1 has order at least p^r. Therefore the order of ∂_1 equals p^r. This proves part (c), and part (b) now follows by Corollary 5.3. □

It remains to prove Proposition 5.2. To do so we describe a homotopy decomposition of ΣG which is designed to behave well with respect to $p^r \circ \partial_1$. In [4] it was shown that if G is retractile then $\Sigma G \simeq \Sigma A \vee C$, where C can be chosen so that it factors through the Hopf construction on G. We need a refinement of this, so the construction will be described in more detail.

In general, let X and Y be path-connected, pointed spaces and let I be the unit interval. The *reduced join* of X and Y is the quotient space $X * Y = (X \times Y \times I)/{\sim}$, where $(x,y,0) \sim (x,y',0)$, $(x,y,1) \sim (x',y,1)$ and $(*,*,t) \sim (*,*,0)$ for all $x,\ x' \in X$, $y,\ y' \in Y$ and $t \in I$. The reduced suspension $\Sigma(X \times Y)$ is also a quotient of $X \times Y \times I$, given by $\Sigma(X \times Y) = (X \times Y \times I)/{\sim'}$, where $(x,y,0) \sim' (x',y',0)$, $(x,y,1) \sim' (x',y',1)$ and $(*,*,t) \sim' (*,*,0)$. The relations imply that the quotient map from $X \times Y \times I$ to $\Sigma(X \times Y)$ factors through $X * Y$. Thus there is a map $X * Y \to \Sigma(X \times Y)$. Note that this map is natural in both variables. Note also that the composites $X * Y \to \Sigma(X \times Y) \xrightarrow{\Sigma\pi_1} \Sigma X$ and $X * Y \to \Sigma(X \times Y) \xrightarrow{\Sigma\pi_2} \Sigma Y$ are null homotopic, where π_1 and π_2 are the projections onto the first and second factors, respectively. There is also a natural homotopy equivalence $\Sigma X \wedge Y \simeq X * Y$ (see [18, Proposition 7.7.1], for example), so we obtain a natural map

$$\Sigma X \wedge Y \to \Sigma(X \times Y), \tag{11}$$

which composes trivially with $\Sigma\pi_1$ and $\Sigma\pi_2$. Iterating, if $X_1, \dots, X_k$ are path-connected, pointed spaces then there is a map

$$\Sigma X_1 \wedge \cdots \wedge X_k \to \Sigma(X_1 \times \cdots \times X_k) \tag{12}$$

which is natural in all k variables. In particular, if $X^{\times k}$ is the Cartesian product of k copies of X with itself and $X^{\wedge k}$ is the k–fold smash product of X with itself, then there is a natural map $\Sigma X^{\wedge k} \to \Sigma X^{\times k}$.

Suppose now that X is an H–space with multiplication $\mu\colon X \times X \to X$. Let μ^* be the composite
$$\mu^*\colon \Sigma X \wedge X \to \Sigma(X \times X) \xrightarrow{\Sigma\mu} \Sigma X.$$

This composite is known as the *Hopf construction* on X. In particular, if G is a simply connected, simple compact Lie group then there is a Hopf construction $\Sigma G \wedge G \xrightarrow{\mu^*} \Sigma G$.

Now we bring in the space A. As G is homotopy associative, the James construction [9] implies that the map $A \xrightarrow{i} G$ extends to an H–map $j\colon \Omega\Sigma A \to G$. By the Bott–Samelson theorem, there is an algebra isomorphism $H_*(\Omega\Sigma A) \cong T(\widetilde{H}_*(A))$, where $T(\,\cdot\,)$ is the free tensor algebra functor. Since G is retractile, there is an algebra isomorphism $H_*(G) \cong \Lambda(\widetilde{H}_*(A))$; observe also that $\Lambda(\widetilde{H}_*(A)) \cong S(\widetilde{H}_*(A))$, where $S(\,\cdot\,)$ is the free symmetric algebra functor. Since j is an H–map, j_* is an algebra map and so is determined by its restriction to the generating set $\widetilde{H}_*(A)$ of $T(\widetilde{H}_*(A))$. Since j is an extension of i and i_* is the inclusion of the generating set into $\Lambda(\widetilde{H}_*(A)) \cong S(\widetilde{H}_*(A))$, the map j_* is therefore the abelianization of the tensor algebra. Let m be the number of generators in $\widetilde{H}_*(A)$. For $1 \le k \le m$, let M_k be the submodule of $H_*(G) \cong \Lambda(\widetilde{H}_*(A))$ consisting of monomials of length k. Observe that there is a module isomorphism $H_*(G) \cong \bigoplus_{k=1}^m M_k$.

Let $E\colon A \to \Omega\Sigma A$ be the suspension map. Let E_k be the composite
$$E_k\colon A^{\times k} \xrightarrow{E^{\times k}} (\Omega\Sigma A)^{\times k} \to \Omega\Sigma A,$$
where the left map is the iterated loop space multiplication. In [4] it was shown that if $1 \le k \le m$ then there is a retract $S_k(A)$ of $\Sigma A^{\wedge k}$ such that $S_1(A) = \Sigma A$, $H_*(S_k(A)) \cong \Sigma M_k$ and the composite
$$\phi_k\colon S_k(A) \to \Sigma A^{\wedge k} \to \Sigma A^{\times k} \xrightarrow{\Sigma E_k} \Sigma\Omega\Sigma A \xrightarrow{\Sigma j} \Sigma G$$
induces an isomorphism onto the submodule ΣM_k of $H_*(\Sigma G)$. Taking the wedge sum over $1 \le k \le m$ gives a map
$$\phi\colon \bigvee_{k=1}^{m} S_k(A) \to \Sigma G$$
which induces an isomorphism in homology and so is a homotopy equivalence. Notice that, when $k = 1$, we have $S_1(A) = \Sigma A$ and $\phi_1 \simeq \Sigma i$. Thus if we let $C = \bigvee_{k=2}^m S_k(A)$

then there is a homotopy equivalence

$$\Sigma A \vee C \xrightarrow{\phi} \Sigma B,$$

where ϕ restricted to ΣA is $\phi_1 \simeq \Sigma i$. The refinement on ϕ we need is to show that its restriction to C factors not just through the Hopf construction $\Sigma G \wedge G \xrightarrow{\mu^*} \Sigma G$, as stated in [4], but through the composite

$$\bar{\mu}^*\colon \Sigma G \wedge A \xrightarrow{\Sigma 1 \wedge i} \Sigma G \wedge G \xrightarrow{\mu^*} \Sigma G.$$

Lemma 5.4 *The restriction of the homotopy equivalence $\Sigma A \vee C \xrightarrow{\phi} G$ to C factors through $\bar{\mu}^*$.*

Proof Consider the diagram

$$\begin{tikzcd}
S_k(A) \arrow[r] & \Sigma A^{\wedge k} \arrow[r, "\Sigma E^{\wedge k}"] \arrow[d] & \Sigma(\Omega\Sigma A)^{\wedge k} \arrow[r, "\Sigma j^{\wedge k}"] \arrow[d] & \Sigma G^{\wedge k} \arrow[d] \\
& \Sigma A^{\times k} \arrow[r, "\Sigma E^{\times k}"] \arrow[dr, "\Sigma E_k"'] & \Sigma(\Omega\Sigma A)^{\times k} \arrow[r, "\Sigma j^{\times k}"] \arrow[d, "\Sigma\mu_k"] & \Sigma G^{\times k} \arrow[d, "\Sigma\mu_k"] \\
& & \Sigma\Omega\Sigma A \arrow[r, "\Sigma j"] & \Sigma G
\end{tikzcd} \tag{13}$$

where μ_k is the iterated multiplication on $\Omega\Sigma A$ and G (both of which are homotopy associative, so the order in which the multiplication is taken is irrelevant). The two upper squares homotopy commute by the naturality of (12), the left lower triangle homotopy commutes by the definition of E_k, and the lower right square homotopy commutes since j is an H–map. By the definition of ϕ_k, the lower direction around the diagram is $\Sigma\phi_k$. On the other hand, since $j \circ E \simeq i$, if $k \geq 2$ then the upper direction around the diagram can be rewritten as the composite

$$S_k(A) \to \Sigma A^{\wedge k} \xrightarrow{i^{\wedge(k-1)} \wedge 1} \Sigma G^{\wedge(k-1)} \wedge A \to \Sigma G^{\times(k-1)} \wedge A \xrightarrow{\Sigma\mu_{k-1} \wedge 1} \Sigma G \wedge A \xrightarrow{\Sigma 1 \wedge i} \Sigma G \wedge G \xrightarrow{\mu^*} \Sigma G.$$

Notice that the last two maps in this composite define $\bar{\mu}^*$, so the homotopy commutativity of (13) shows that ϕ_k factors through $\bar{\mu}^*$. □

Now we relate the homotopy equivalence for ΣG to the order of the boundary map $G \xrightarrow{\partial_1} \Omega_0^3 G$. In general, if $\Omega B \xrightarrow{\partial} F \to E \to B$ is a homotopy fibration sequence then there is a homotopy action

$$\theta\colon \Omega B \times F \to F$$

such that the restriction of θ to ΩB is ∂, the restriction of θ to F is the identity map, and there is a homotopy commutative diagram:

$$\begin{tikzcd} \Omega B \times \Omega B \arrow[r, "\mu"] \arrow[d, "1\times\partial"] & \Omega B \arrow[d, "\partial"] \\ \Omega B \times F \arrow[r, "\theta"] & F \end{tikzcd} \tag{14}$$

In our case, since $G \xrightarrow{\partial_1} \Omega_0^3 G$ is a homotopy fibration connecting map there is a homotopy action $\theta\colon G \times \Omega_0^3 G \to \Omega_0^3 G$.

Let ev: $\Sigma\Omega_0^3 G \to \Omega^2 G$ be the canonical evaluation map. For an H–space X, recall that $p^r\colon X \to X$ is the p^r–power map. For a co-H–space Y, let $\underline{p}^r\colon Y \to Y$ be the map of degree p^r. Recall that $\overline{\partial}_1$ is the composite

$$\overline{\partial}_1\colon A \xrightarrow{i} G \xrightarrow{\partial_1} \Omega_0^3 G.$$

Let

$$\phi_C\colon C \to \Sigma G$$

be the restriction of the homotopy equivalence $\Sigma A \vee C \xrightarrow{\phi} \Sigma G$ to C.

Lemma 5.5 *Let G be a retractile, simply connected, simple compact Lie group. Suppose that $p^r \circ \overline{\partial}_1$ is null homotopic. Then for $2 \le k \le m$ the composite*

$$C \xrightarrow{\phi_C} \Sigma G \xrightarrow{\Sigma\partial_1} \Sigma\Omega_0^3 G \xrightarrow{\mathrm{ev}} \Omega^2 G \xrightarrow{p^r} \Omega^2 G$$

is null homotopic.

Proof Consider the diagram

$$\begin{tikzcd} G \times A \arrow[r, "\pi_1"] \arrow[d, "1\times\underline{p}^r"] & G \arrow[dd, "i_1"] \\ G \times A \arrow[d, "1\times i"] & \\ G \times G \arrow[r, "1\times\partial_1"] \arrow[d, "\mu"] & G \times \Omega_0^3 G \arrow[d, "\theta"] \\ G \arrow[r, "\partial_1"] & \Omega_0^3 G \end{tikzcd}$$

where π_1 is the projection onto the first factor and i_1 is the inclusion of the first factor. Since $\overline{\partial}_1 = \partial_1 \circ i$, the hypothesis that $p^r \circ \overline{\partial}_1$ is null homotopic implies that the upper rectangle homotopy commutes. The lower square homotopy commutes by (14). Suspending and using the naturality of (11) and the definition of $\bar{\mu}^*$, we obtain a homotopy commutative diagram:

$$\begin{tikzcd}
\Sigma G \wedge A \arrow[r] \arrow[d, "\Sigma(1\wedge \underline{p}^r)"] & \Sigma(G\times A) \arrow[r, "\Sigma\pi_1"] \arrow[d, "\Sigma(1\times \underline{p}^r)"] & \Sigma G \arrow[dd, "\Sigma i_1"] & \\
\Sigma G \wedge A \arrow[r] \arrow[ddr, "\bar{\mu}^*"'] & \Sigma(G\times A) \arrow[d, "\Sigma(1\times i)"] & & \\
& \Sigma(G\times G) \arrow[r, "\Sigma(1\times\partial_1)"] \arrow[d, "\Sigma\mu"] & \Sigma(G\times \Omega^3_0 G) \arrow[d, "\Sigma\theta"] & \\
& \Sigma G \arrow[r, "\Sigma\partial_1"] & \Sigma\Omega^3_0 G \arrow[r, "\mathrm{ev}"] & \Omega^2 G
\end{tikzcd}$$

By (11), the composite along the top row is null homotopic. Thus the lower direction around the diagram is null homotopic. Observe that the map $\Sigma 1\wedge \underline{p}^r$ is homotopic to the degree p^r map on $\Sigma G\wedge A$, so we obtain a null homotopy for $\mathrm{ev}\circ\Sigma\partial_1\circ\bar{\mu}^*\circ\underline{p}^r$. In $[\Sigma X, \Omega Y]$ the group structure induced by the comultiplication equals that induced by the multiplication, so we obtain a null homotopy for the composite $\Sigma G\wedge A \xrightarrow{\bar{\mu}^*} \Sigma G \xrightarrow{\Sigma\partial_1} \Sigma\Omega^3_0 G \xrightarrow{\mathrm{ev}} \Omega^2 G \xrightarrow{p^r} \Omega^2 G$. By Lemma 5.4, the restriction of the homotopy equivalence $\Sigma A\vee C \xrightarrow{\phi} \Sigma G$ to C factors through $\bar{\mu}^*$. Thus the composite

$$C \xrightarrow{\phi_C} \Sigma G \xrightarrow{\Sigma\partial_1} \Sigma\Omega^3_0 G \xrightarrow{\mathrm{ev}} \Omega^2 G \xrightarrow{p^r} \Omega^2 G$$

is null homotopic. $\square$

Putting all this together, we prove Proposition 5.2:

Proof of Proposition 5.2 Consider the composite

$$\psi\colon \Sigma A\vee C \xrightarrow{\phi} \Sigma G \xrightarrow{\Sigma\partial_1} \Omega^3_0 G \xrightarrow{\mathrm{ev}} \Omega^2 G \xrightarrow{p^r} \Omega^2 G.$$

The restriction of ϕ to ΣA is homotopic to Σi, and by definition $\overline{\partial}_1 = \partial_1\circ i$. So the restriction of ψ to ΣA is homotopic to $p^r\circ\mathrm{ev}\circ\Sigma\overline{\partial}_1$, which is null homotopic since $p^r\circ\overline{\partial}_1$ is null homotopic. By Lemma 5.5, the hypothesis that $p^r\circ\overline{\partial}_1$ is null homotopic implies that the restriction of ψ to C is null homotopic. Thus ψ is null homotopic. As ϕ is a homotopy equivalence, this implies that $p^r\circ\mathrm{ev}\circ\Sigma\partial_1$ is null homotopic. Taking adjoints, we obtain that $p^r\circ\partial_1$ is null homotopic. $\square$

6 Applications to gauge groups

The first application is to prove Theorem 1.1(d) by counting the number of distinct homotopy types of gauge groups. This requires two known results. Hamanaka and Kono [6] proved the following:

Theorem 6.1 *If $\mathcal{G}_k(\mathrm{SU}(n)) \simeq \mathcal{G}_{k'}(\mathrm{SU}(n))$ then $(n(n^2-1,k) = (n(n^2-1),k')$.* □

Theorem 6.1 improved an earlier result of Sutherland [19] by a factor of 2 in the case where n is odd. On the other hand, the author [21] proved the following. For positive integers a and b, let (a,b) be their greatest common divisor.

Theorem 6.2 *Let X be a space and Y be an H–space with a homotopy inverse. Suppose there is a map $X \xrightarrow{f} Y$ of order m, where m is finite. For a positive integer $k \geq 0$, let F_k be the homotopy fibre of $k \circ f$. If $(m,k) = (m,k')$ then F_k and $F_{k'}$ are homotopy equivalent when localized rationally or at any prime.* □

Proof of Theorem 1.1(d) Consider the map $\mathrm{SU}(n) \xrightarrow{\partial_1} \Omega_0^3\mathrm{SU}(n)$. By Lemma 2.1, $k \circ \partial_1 \simeq \partial_k$, and recall that the homotopy fibre of ∂_k is $\mathcal{G}_k(\mathrm{SU}(n))$. Localize at an odd prime p. If $n \leq (p-1)^2+1$ then Theorem 1.1(c) says that ∂_1 has order $n(n^2-1)$. So, by Theorem 6.2, if $(n(n^2-1),k) = (n(n^2-1),k')$ then $\mathcal{G}_k(\mathrm{SU}(n)) \simeq \mathcal{G}_{k'}(\mathrm{SU}(n))$ when localized at any odd prime. The converse is Theorem 6.1. □

The next application is to a p–local homotopy decomposition of $\mathcal{G}_k(\mathrm{SU}(n))$. Assume from now on that all spaces and maps are localized at p. In [16] it was shown that there is a homotopy equivalence $\mathrm{SU}(n) \simeq \prod_{i=1}^{p-1} B_i$, where the generators of $H^*(B_i;\mathbb{Z}/p\mathbb{Z})$ are those of $H^*(\mathrm{SU}(n);\mathbb{Z}/p\mathbb{Z})$ which occur in dimensions of the form $2i+2(p-1)t+1$ for some $t \geq 0$. Let i_0 and i_1 be such that $2n-3 = 2i_0+2(p-1)t_0+1$ and $2n-1 = 2i_1+2(p-1)t_1+1$ for some integers t_0 and t_1. In what follows, when we write B_{i+2} and $i = p-2$ or $p-1$ we mean B_1 or B_2, respectively. In [22] in the retractile case and in [11] in the general case it was shown that there is a homotopy equivalence

$$\mathcal{G}_k(\mathrm{SU}(n)) \simeq \Bigg(\prod_{\substack{i=1 \\ i\neq i_0,i_1}}^{p-1} B_i \times \Omega_0^4 B_{i+2}\Bigg) \times X_{i_0} \times X_{i_1}, \tag{15}$$

where for $j \in \{i_0,i_1\}$ there are homotopy fibrations $\Omega_0^4 B_{j+2} \to X_j \to B_j$.

The extra information we can now add is when the spaces X_{i_0} and X_{i_1} decompose as products. If $\mathrm{SU}(n)$ is retractile and k is a multiple of $n(n^2-1)$ then, by Theorem 1.1(c) and Lemma 2.1, $\partial_k \simeq k \circ \partial_1$ is null homotopic. Therefore, $\mathcal{G}_k(\mathrm{SU}(n)) \simeq \mathrm{SU}(n) \times \Omega_0^4\mathrm{SU}(n)$, implying that there are homotopy equivalences $X_{i_0} \simeq B_{i_0} \times \Omega^4 B_{i_0+2}$ and $X_{i_1} \simeq B_{i_1} \times \Omega_0^4 B_{i_1+2}$. On the other hand, by Theorem 1.1(d), there is a homotopy equivalence $\mathcal{G}_k(\mathrm{SU}(n)) \simeq \mathrm{SU}(n) \times \Omega_0^4\mathrm{SU}(n))$ if and only if k is a multiple of $n(n^2-1)$. Hence if k is not a multiple of $n(n^2-1)$ then there cannot be simultaneous homotopy equivalences $X_{i_0} \simeq B_{i_0} \times \Omega^4 B_{i_0+2}$ and $X_{i_1} \simeq B_{i_1} \times \Omega_0^4 B_{i_1+2}$. Thus we obtain the following:

Theorem 6.3 *Localize at an odd prime p and suppose $n \leq (p-1)^2+1$. Then in* (15) *there are homotopy equivalences $X_{i_0} \simeq B_{i_0} \times \Omega^4 B_{i_0+2}$ and $X_{i_1} \simeq B_{i_1} \times \Omega_0^4 B_{i_1+2}$ if and only if k is a multiple of $n(n^2-1)$.* □

References

[1] **M F Atiyah**, **R Bott**, *The Yang–Mills equations over Riemann surfaces*, Philos. Trans. Roy. Soc. London Ser. A 308 (1983) 523–615 MR

[2] **M C Crabb**, **W A Sutherland**, *Counting homotopy types of gauge groups*, Proc. London Math. Soc. 81 (2000) 747–768 MR

[3] **D H Gottlieb**, *Applications of bundle map theory*, Trans. Amer. Math. Soc. 171 (1972) 23–50 MR

[4] **J Grbić**, **S Theriault**, *Self-maps of low rank Lie groups at odd primes*, Canad. J. Math. 62 (2010) 284–304 MR

[5] **H Hamanaka**, **A Kono**, *On* $[X, \mathrm{U}(n)]$ *when* $\dim X$ *is* $2n$, J. Math. Kyoto Univ. 43 (2003) 333–348 MR

[6] **H Hamanaka**, **A Kono**, *Unstable* K^1*–group and homotopy type of certain gauge groups*, Proc. Roy. Soc. Edinburgh Sect. A 136 (2006) 149–155 MR

[7] **H Hamanaka**, **A Kono**, *Homotopy type of gauge groups of* SU(3)*–bundles over* S^6, Topology Appl. 154 (2007) 1377–1380 MR

[8] **J R Harper**, *Secondary cohomology operations*, Graduate Studies in Mathematics 49, Amer. Math. Soc., Providence, RI (2002) MR

[9] **I M James**, *Reduced product spaces*, Ann. of Math. 62 (1955) 170–197 MR

[10] **Y Kamiyama**, **D Kishimoto**, **A Kono**, **S Tsukuda**, *Samelson products of* SO(3) *and applications*, Glasg. Math. J. 49 (2007) 405–409 MR

[11] **D Kishimoto**, **A Kono**, **S Theriault**, *Refined gauge group decompositions*, Kyoto J. Math. 54 (2014) 679–691 MR

[12] **D Kishimoto**, **A Kono**, **M Tsutaya**, *On* p*–local homotopy types of gauge groups*, Proc. Roy. Soc. Edinburgh Sect. A 144 (2014) 149–160 MR

[13] **D Kishimoto**, **S Theriault**, **M Tsutaya**, *The homotopy types of G_2–gauge groups*, preprint

[14] **A Kono**, *A note on the homotopy type of certain gauge groups*, Proc. Roy. Soc. Edinburgh Sect. A 117 (1991) 295–297 MR

[15] **G E Lang, Jr**, *The evaluation map and* EHP *sequences*, Pacific J. Math. 44 (1973) 201–210 MR

[16] **M Mimura**, **G Nishida**, **H Toda**, Mod *p decomposition of compact Lie groups*, Publ. Res. Inst. Math. Sci. 13 (1977) 627–680 MR

[17] **M Mimura**, **H Toda**, *Cohomology operations and homotopy of compact Lie groups, I*, Topology 9 (1970) 317–336 MR

[18] **P Selick**, *Introduction to homotopy theory*, Fields Institute Monographs 9, Amer. Math. Soc., Providence, RI (1997) MR

[19] **W A Sutherland**, *Function spaces related to gauge groups*, Proc. Roy. Soc. Edinburgh Sect. A 121 (1992) 185–190 MR

[20] **S D Theriault**, *The odd primary H–structure of low rank Lie groups and its application to exponents*, Trans. Amer. Math. Soc. 359 (2007) 4511–4535 MR

[21] **S D Theriault**, *The homotopy types of* Sp(2)*–gauge groups*, Kyoto J. Math. 50 (2010) 591–605 MR

[22] **S D Theriault**, *Odd primary homotopy decompositions of gauge groups*, Algebr. Geom. Topol. 10 (2010) 535–564 MR

[23] **S Theriault**, *The homotopy types of* SU(5)*–gauge groups*, Osaka J. Math. 52 (2015) 15–29 MR

[24] **H Toda**, *A topological proof of theorems of Bott and Borel–Hirzebruch for homotopy groups of unitary groups*, Mem. Coll. Sci. Univ. Kyoto. Ser. A. Math. 32 (1959) 103–119 MR

[25] **S Tsukuda**, *A remark on the homotopy type of the classifying space of certain gauge groups*, J. Math. Kyoto Univ. 36 (1996) 123–128 MR

[26] **M Tsutaya**, *A note on homotopy types of connected components of* $\mathrm{Map}(S^4, BSU(2))$, J. Pure Appl. Algebra 216 (2012) 826–832 MR

[27] **G W Whitehead**, *Elements of homotopy theory*, Graduate Texts in Mathematics 61, Springer, New York (1978) MR

Mathematical Sciences, University of Southampton
Southampton, SO17 1BJ, United Kingdom

s.d.theriault@soton.ac.uk

Received: 23 March 2016 Revised: 6 July 2016

Algebraic & Geometric Topology 17 (2017) 1151–1162

Relative Thom spectra via operadic Kan extensions

JONATHAN BEARDSLEY

We show that a large number of Thom spectra, that is, colimits of morphisms $BG \to B\mathrm{GL}_1(\mathbb{S})$, can be obtained as iterated Thom spectra, that is, colimits of morphisms $BG \to B\mathrm{GL}_1(Mf)$ for some Thom spectrum Mf. This leads to a number of new relative Thom isomorphisms, for example $M\mathrm{U}[6,\infty) \wedge_{M\mathrm{String}} M\mathrm{U}[6,\infty) \simeq M\mathrm{U}[6,\infty) \wedge \mathbb{S}[B^3\mathrm{Spin}]$. As an example of interest to chromatic homotopy theorists, we also show that Ravenel's $X(n)$ filtration of $M\mathrm{U}$ is a tower of intermediate Thom spectra determined by a natural filtration of $B\mathrm{U}$ by subbialagebras.

An errata was posted on 26 May 2017 in an online supplement.

55N22, 55P42

1 Introduction

We prove several new results about Thom spectra which are $\mathbb{E}_n$–ring spectra. The most immediately accessible results are relative Thom isomorphisms like the following from Section 3:

- $M\mathrm{Spin} \wedge_{M\mathrm{String}} M\mathrm{Spin} \simeq M\mathrm{Spin} \wedge \mathbb{S}[K(\mathbb{Z},4)]$.
- $M\mathrm{SO} \wedge_{M\mathrm{U}} M\mathrm{SO} \simeq M\mathrm{SO} \wedge \mathbb{S}[\mathrm{Spin}]$.
- $M\mathrm{U} \wedge_{M\mathrm{Sp}} M\mathrm{U} \simeq M\mathrm{U} \wedge \mathbb{S}[\mathrm{SO}/\mathrm{U}]$.
- $M\mathrm{U}[6,\infty) \wedge_{M\mathrm{String}} M\mathrm{U}[6,\infty) \simeq M\mathrm{U}[6,\infty) \wedge \mathbb{S}[B^3\mathrm{Spin}]$.
- $H\mathbb{Z}/2 \wedge_{H\mathbb{Z}} H\mathbb{Z}/2 \simeq H\mathbb{Z}/2 \wedge \mathbb{S}[S^1]$.

However, in consideration of the fact that there are now a number of different methods for defining such objects, we will take a moment to clarify precisely which models we use for the remainder (though we do not expect that the choice of model is relevant to the veracity of the statements). By the category of spectra, which we denote by $\mathcal{S}$, we will always mean Lurie's symmetric monoidal quasicategory of spectra defined in [13, Section 1.4.3]. In general, except when we explicitly state otherwise, we will always be working with quasicategories and all of our constructions will be homotopy invariant. For example, all of our tensor products are derived (as they must be when working internally to a quasicategory), all of our limits and colimits are the quasicategorical analogs of homotopy colimits and limits, and our functors are actually morphisms

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.1151

of simplicial sets. We also make use of Lurie's notion of ∞–operads, defined and described in [13, Chapter 2]. We will not review the theory of ∞–operads here except to say that they are a natural generalization of the notion of multicategories and the categories of operators of May and Thomason [17]. We are especially interested in the $\mathbb{E}_n$ ∞–operads of [13, Chapter 5] and will refer to them frequently in this paper. These ∞–operads capture the same structure as Boardman and Vogt's little n–cubes operads, which use embeddings of n–dimensional cubes to parametrize multiplicative structure (see May [15, Chapter 4]). It is nontrivial to show that these quasicategorical constructions behave identically to their model category theoretic analogs, and that results obtained thereby are compatible with results obtained using model categories. The interested reader is invited to refer to Lurie [12; 13] for proofs that these conditions are met. We recognize, of course, that these references are expansive in their own right, so will endeavor to give more specific citations throughout the paper.

Thom spectra, the main objects of investigation here, are classically constructed by considering spaces associated to stable spherical bundles on topological spaces (see eg Sullivan [22]). However, after work of May and Sigurdsson [16] and later work of Ando, Blumberg, Gepner, Hopkins and Rezk [2], it became clear that there was an alternative way to think of Thom spectra: as quotients of ring spectra by group actions. In general, given an $\mathbb{E}_n$–ring spectrum R, there is an n–fold loop space of units, $\mathrm{GL}_1(R)$. Thus a morphism of n–fold loop spaces $X \to \mathrm{GL}_1(R)$ gives an action of X on R, and induces a morphism of $(n-1)$–fold loop spaces $BX \to B\mathrm{GL}_1(R)$. By working quasicategorically, we can see that there is in fact a functor (thinking of BX and $B\mathrm{GL}_1(R)$ as quasicategories) $B\mathrm{GL}_1(R) \hookrightarrow L\mathrm{Mod}_R$ which is fully faithful. Thus we have a morphism of simplicial sets $BX \to L\mathrm{Mod}_R$ (the quasicategory of left R–modules, where R is considered as an $\mathbb{E}_1$–ring spectrum) which is picking out an action of X on R by R–module equivalences. This determines a diagram in $L\mathrm{Mod}_R$ whose colimit, as the thing on which every point of X, including the identity, acts in the same way, must be exactly R/X. When $R = \mathbb{S}$, the sphere spectrum $B\mathrm{GL}_1(\mathbb{S})$ is precisely the classifying space of stable spherical fibrations, and taking the colimit of a morphism $BX \to B\mathrm{GL}_1(\mathbb{S}) \hookrightarrow L\mathrm{Mod}_{\mathbb{S}}$ produces a spectrum which is equivalent to the one produced classically by taking a sequence of Thom spaces over BX (see [2, Proposition 3.23]). Thus, for instance, $M\mathrm{U}$ is just $\mathbb{S}/\mathrm{U}$ and $M\mathrm{O}$ is just $\mathbb{S}/\mathrm{O}$, and so on and so forth.

Similarly, if we have an action of a Lie group G on a smooth manifold X, we can take its homogeneous space X/G. If we happen to have an inclusion of a normal subgroup $H \hookrightarrow G$, then we obtain an action of H on X and can also take the homogeneous space X/H. It is a classical fact then that X/H admits an action of G/H and moreover that the iterated homogeneous space $(X/H)/(G/H)$ is diffeomorphic to X/G (see Bourbaki [5, Section 1.6, Proposition 13]). It stands to reason then that something similar should be true for actions of n–fold loop spaces on a ring spectrum R, and that is one of the main theorems of this document (see Theorem 1) if we allow ourselves to replace the condition "H is a normal subgroup of G" with "there is a fiber sequence of n–fold loop spaces $H \to G \to G/H$". Specifically, if we have a G–action on an $\mathbb{E}_n$–ring spectrum R, then we obtain a G/H–action on R/H and a sequence of ring spectra $R \to R/H \to (R/H)/(G/H) \simeq R/G$. In other words, R/G can be produced as the Thom spectrum associated to an action map $G/H \to B\mathrm{GL}_1(R/H)$.

In Section 2 we show that a number of classical Thom spectra over $\mathbb{S}$ can in fact be constructed as Thom spectra over *other* Thom spectra. This statement is made rigorous by the following theorem:

Theorem 1 *Suppose $Y \xrightarrow{i} X \xrightarrow{q} B$ is a fiber sequence of reduced $\mathbb{E}_n$–monoidal Kan complexes for $n > 1$ with i and q both maps of $\mathbb{E}_n$–algebras. Let $f\colon X \to B\mathrm{GL}_1(\mathbb{S})$ be a morphism of $\mathbb{E}_n$–monoidal Kan complexes for $n > 1$. Then there is a morphism of $\mathbb{E}_{n-1}$–algebras $B \to B\mathrm{GL}_1(M(f \circ i))$ whose associated Thom spectrum is equivalent to Mf.*

By constructing Mf as a Thom spectrum over an intermediate Thom spectrum, we get a relative Thom isomorphism:

Corollary *There is a morphism of $\mathbb{E}_{n-1}$–ring spectra $R \to M(f \circ i) \to Mf$ and a relative Thom isomorphism $Mf \wedge_{M(f\circ i)} Mf \simeq Mf \wedge_R R[B]$, where $R[B] = R \wedge_{\mathbb{S}} \Sigma^\infty_+ B$.*

The proof requires certain technical details and constructions from Lurie [13], so we separate the relevant lemmas into their own subsection, Section 2.1, and refer to them as needed. In Section 3 we give a number of examples of constructions of intermediate Thom spectra which are $\mathbb{E}_n$–rings. The last example we present is a new construction of $M\mathrm{U}$, which bears some resemblance to Lazard's construction of the Lazard ring in [10]. This construction is unrelated to recent work regarding $M\mathrm{U}$ and complex orientations by McKeown [18]. This paper comprises work contained in the author's doctoral thesis.

Let us fix some notation for the remainder of the paper: the quasicategory of spectra will be denoted by $\mathcal{S}$ and the quasicategory of Kan complexes, sometimes called spaces, will be denoted by $\mathcal{T}$; the quasicategory of small quasicategories will be denoted by $q\mathcal{C}at$ (to avoid set-theoretic issues we assume the existence of inaccessible cardinals as necessary, as in Lurie [12, Section 1.2.15]); $\mathcal{O}^\otimes$ or $\mathcal{O}$ will always refer to an ∞–operad; $\mathbb{E}_n$ will refer to the little n–cubes ∞–operad, but sometimes when considering the $\mathbb{E}_\infty$–operad in its role as the terminal ∞–operad we will denote it by $\mathcal{F}in_*$, to indicate that it is equivalent to the nerve of the category of finite pointed sets; for an $\mathbb{E}_n$–ring spectrum R, we denote by $L\mathrm{Mod}_R$ the $\mathbb{E}_{n-1}$–monoidal quasicategory of left R–modules over R as an $\mathbb{E}_1$–ring spectrum; $B\mathrm{GL}_1(R)$ will be the Kan complex defined in [2], ie the delooping of the Kan complex of homotopy automorphisms of R in $L\mathrm{Mod}_R$.

2 Intermediate Thom spectra

The following theorem describes our general method for producing intermediate Thom spectra:

Theorem 1 *Suppose $Y \xrightarrow{i} X \xrightarrow{q} B$ is a fiber sequence of reduced $\mathbb{E}_n$–monoidal Kan complexes for $n > 1$ with i and q both maps of $\mathbb{E}_n$–algebras. Let $f\colon X \to B\mathrm{GL}_1(R)$ be a morphism of $\mathbb{E}_n$–monoidal Kan complexes for $n > 1$. Then there is a morphism of $\mathbb{E}_{n-1}$–algebras $B \to B\mathrm{GL}_1(M(f \circ i))$ whose associated Thom spectrum is equivalent to Mf.*

The following two corollaries follow immediately from Theorem 1:

Corollary 2 *Given the assumptions of Theorem 1, there is an equivalence of $\mathbb{E}_{n-1}$–R–algebras $Mf \simeq M(f \circ i) \wedge_{R[\Omega B]} R$, where R is equipped with the trivial $R[\Omega B]$–module structure.*

Proof For a fiber sequence $Y \to X \to B$ of $\mathbb{E}_n$–spaces we have a fiber sequence $\Omega B \to Y \to X$ such that X is equivalent to a bar construction $\mathrm{Bar}_\bullet(Y, \Omega B, *)$. Thus, since the Thom spectrum functor is symmetric monoidal and preserves colimits (see [1, Corollary 8.1] or Lewis' slightly weaker result in [11]), the Thom spectrum of $X \to B\mathrm{GL}_1(R)$ is equivalent to the bar construction in $\mathbb{E}_n$–R–algebras, and so in general only admits the structure of an $\mathbb{E}_{n-1}$–algebra. □

Remark 3 Constructing Thom spectra as bar constructions is not a new idea, and should be compared to the bar construction definition of generalized Thom spectra given in [16, Sections 23.4 and 23.5].

Corollary 4 *Given the assumptions of Theorem 1, there is a morphism of $\mathbb{E}_{n-1}$–R–algebra spectra $R \to M(f \circ i) \to Mf$ and a relative Thom isomorphism $Mf \wedge_{M(f\circ i)} Mf \simeq Mf \wedge_R R[B]$, where $R[B] = R \wedge_{\mathbb{S}} \Sigma^\infty_+ B$.*

Proof The fact that the equivalence exists and is an equivalence of $\mathbb{E}_{n-1}$–R–algebras follows from [1, Corollary 1.8]. In particular, we know that the equivalence is given by a morphism $Mf \wedge_{M(f\circ i)} Mf \to Mf \wedge_{M(f\circ i)} Mf \wedge_R R[B] \to Mf \wedge_R R[B]$, where the first map is the Thom diagonal and the second map is the $M(f\circ i)$–algebra structure map of Mf. □

We now give a proof of Theorem 1, though it relies on lemmas which we defer to Section 2.1. It also makes crucial use of the notion of an *operadic left Kan extension*, as described in [13, Section 3.1.2].

Proof Note that $M(f\circ i)$ is an $\mathbb{E}_n$–algebra, so $B\mathrm{GL}_1(M(f\circ i))$ is an $(n-1)$–fold loop space, so we cannot hope for the desired map to be more structured than this. By Lemmas 5 and 6 the $\mathbb{E}_{n-1}$–monoidal left Kan extension of $X \xrightarrow{f} B\mathrm{GL}_1(\mathbb{S}) \hookrightarrow \mathcal{S}$ along $q\colon X \to B$ exists and takes the unique 0–simplex of B to $M(f\circ i)$. By Proposition 8, this Kan extension factors as a morphism of $\mathbb{E}_{n-1}$–monoidal Kan complexes through $B\mathrm{GL}_1(M(f\circ i))$. Taking the Thom spectrum of the induced morphism $B \to B\mathrm{GL}_1(M(f\circ i))$ produces $M(f\circ i)/(\Omega B)$ as a Thom spectrum over $M(f\circ i)$. Moreover, taking the colimit of the functor $B \to B\mathrm{GL}_1(M(f\circ i))$ is equivalent to forming the left operadic Kan extension along the map $B \to *$. By Lemma 7 and [13, Corollary 3.1.4.2] we have that the left operadic Kan extension along $X \to B$ followed by the left operadic Kan extension along $B \to *$ is equivalent to the left operadic Kan extension along $X \to *$ (ie Kan extensions compose). Thus the iterated Kan extension which produces $M(f\circ i) = \mathbb{S}/\Omega Y$ and then quotients it by the action of ΩB is equivalent to the one-step Kan extension producing $\mathbb{S}/\Omega X \simeq Mf$ with an “action” of the trivial $\mathbb{E}_{n-1}$–space. Hence Mf is produced as a Thom spectrum over $M(f\circ i)$. □

2.1 The lemmas

Lemma 5 *Let X be a Kan complex and $f\colon X \to \mathcal{C}$ an $\mathbb{E}_n$–monoidal morphism of quasicategories, where $\mathcal{C}$ is a cocomplete quasicategory. Then, for any morphism of $\mathbb{E}_n$–monoidal Kan complexes $p\colon X \to B$, the operadic Kan extension of f along p exists.*

Proof Since X and B are Kan complexes, hence essentially small, and $\mathcal{C}$ is cocomplete, the result follows from [13, Corollary 3.1.3.5]. □

Lemma 6 *Let $Y \xrightarrow{i} X \xrightarrow{q} B$ be a fiber sequence of $\mathbb{E}_n$–monoidal Kan complexes. The $\mathbb{E}_{n-1}$–monoidal left Kan extension of an $\mathbb{E}_n$–monoidal morphism*

$$f\colon X \to B\mathrm{GL}_1(\mathbb{S}) \to L\mathrm{Mod}(M(f\circ i))$$

along $q\colon X \to B$ is computed by taking the colimit of the composition

$$\mathrm{fib}(X \to B) \simeq Y \to X \to B\mathrm{GL}_1(\mathbb{S}) \to L\mathrm{Mod}(M(f \circ i)).$$

Proof Following the notation given in Definition 3.1.2.2 and the construction in Remark 3.1.3.15 of [13], we have a correspondence of ∞–operads given by

$$\mathcal{M}^{\otimes} \simeq (X^{\otimes} \times \Delta^1) \amalg \coprod_{X^{\otimes} \times \{1\}} B^{\otimes} \to \mathcal{F}in_* \times \Delta^1.$$

In other words, there is a family of ∞–operads indexed by Δ^1 which looks like $X^{\otimes}$ (the ∞–operad associated to X as an $\mathbb{E}_n$–monoidal Kan complex) at one end and $B^{\otimes}$ at the other end. Formula $(*)$ of [13, Definition 3.1.2.2] states that the value of the desired Kan extension at a 0–simplex $\sigma \in B$ is given by the colimit diagram

$$((\mathcal{M}^{\otimes}_{\mathrm{act}})_{/\sigma} \times_{\mathcal{M}^{\otimes}} X^{\otimes})^{\triangleright} \to (\mathcal{M}^{\otimes})^{\triangleright}_{/\sigma} \to \mathcal{M}^{\otimes} \to \mathcal{T},$$

where the morphism $(\mathcal{M}^{\otimes})^{\triangleright}_{/\sigma} \to \mathcal{M}^{\otimes}$ takes the cone point to σ. In other words, the value of the Kan extension at σ is computed by taking the colimit over the diagram in $\mathcal{M}^{\otimes}$ of objects (and active morphisms) living over σ. As the simplicial set $\mathcal{M}^{\otimes}$ is nothing more than the mapping cylinder of the morphism of $\mathbb{E}_n$–monoidal Kan complexes $X^{\otimes} \to B^{\otimes}$, we have the result. □

Lemma 7 *There is a Δ^2–family of ∞–operads induced by the morphisms of $\mathbb{E}_{n-1}$–monoidal Kan complexes $X \to B$ and $B \to *$, denoted by $\mathcal{M}^{\otimes} \to \Delta^2 \times \mathcal{F}in_*$, and the induced projection $\mathcal{M}^{\otimes} \to \Delta^2$ is a flat categorical fibration.*

Proof The equivalence of morphisms $(X \to B \to *) \simeq (X \to *)$ is given by a 2–simplex in the quasicategory of $\mathbb{E}_{n-1}$–monoidal quasicategories, hence by a morphism of simplicial sets in $\mathrm{Hom}(\Delta^2, \mathrm{Hom}(\mathbb{E}^{\otimes}_{n-1}, q\mathcal{C}at)) \simeq \mathrm{Hom}(\Delta^2 \times \mathbb{E}^{\otimes}_{n-1}, q\mathcal{C}at)$. By the quasicategorical Grothendieck construction of [12], we obtain a cocartesian fibration of simplicial sets $p\colon \mathcal{M}^{\otimes} \to \Delta^2 \times \mathbb{E}^{\otimes}_{n-1}$ such that $p^{-1}(0) \simeq X^{\otimes}$, $p^{-1}(1) \simeq B^{\otimes}$ and $p^{-1}(2) \simeq *^{\otimes}$, where $X^{\otimes}$, $B^{\otimes}$ and $*^{\otimes}$ are the ∞–operads witnessing the $\mathbb{E}_{n-1}$–monoidal structure on X, B and $*$. The projection map induces a family of ∞–operads $\mathcal{M}^{\otimes} \to \Delta^2$. This projection is a flat fibration as it satisfies the requirements of [13, Example B.3.4], ie there are cocartesian lifts of every edge in $\Delta^2 \simeq \Delta^2 \times * \subset \Delta^2 \times \mathcal{F}in_*$. □

Proposition 8 *Let $Y \xrightarrow{i} X \xrightarrow{q} B$ be a fiber sequence of reduced, connected $\mathbb{E}_n$–monoidal Kan complexes. The left operadic Kan extension of an $\mathbb{E}_n$–morphism $f\colon X \to B\mathrm{GL}_1(\mathbb{S}) \to L\mathrm{Mod}(M(f \circ i))$ along the $\mathbb{E}_n$–morphism $q\colon X \to B$ factors as a morphism of $\mathbb{E}_{n-1}$–monoidal Kan complexes through $B\mathrm{GL}_1(M(f \circ i))$.*

Proof Note that the left operadic Kan extension along q takes the unique zero simplex of B to $M(f \circ i)$ by Lemma 6. Since B is a Kan complex it must be that this Kan extension factors through $B\mathrm{GL}_1(M(f \circ i))$. This morphism is only $\mathbb{E}_{n-1}$ since $B\mathrm{GL}_1(M(f \circ i))$ is only an $(n-1)$–fold loop space. □

3 Examples

A large number of morphisms of $\mathbb{E}_n$–monoidal Kan complexes fit into the framework described in the introduction and Theorem 1. In the following we repeatedly use the fact from [4, Examples 6.39] that there is a sequence of infinite loop maps $\mathrm{U} \to \mathrm{O} \to \mathrm{GL}_1(\mathbb{S})$ (where they write F for $\mathrm{GL}_1(\mathbb{S})$). The delooped (infinite loop) map $B\mathrm{O} \to B\mathrm{GL}_1(\mathbb{S})$ is called the j–homomorphism and the composition $B\mathrm{U} \to B\mathrm{O} \to B\mathrm{GL}_1(\mathbb{S})$ is called the complex j–homomorphism. We also use the fact that deloopings and connective covers (modeled by a bar construction and based loops on a bar construction, respectively) take $\mathbb{E}_n$–spaces to $\mathbb{E}_{n-1}$–spaces and $\mathbb{E}_n$–spaces to $\mathbb{E}_n$–spaces, respectively.

(1) $B\mathrm{SU} \to B\mathrm{U} \to \mathbb{C}\mathrm{P}^\infty$ is a fiber sequence of infinite loop spaces. The complex j–homomorphism $B\mathrm{U} \to B\mathrm{GL}_1(\mathbb{S})$ is a morphism of infinite loop spaces.

(2) $B\mathrm{String} \to B\mathrm{Spin} \to K(\mathbb{Z}, 4)$ is a fiber sequence of infinite loop spaces. Using the covering map $B\mathrm{Spin} \to B\mathrm{O}$ composed with the j–homomorphism, we obtain a map of infinite loop spaces $B\mathrm{Spin} \to B\mathrm{GL}_1(\mathbb{S})$.

(3) $B\mathrm{U} \to B\mathrm{SO} \to \mathrm{Spin}$ is a fiber sequence of infinite loop spaces as a result of [9, Table 2.1.1], and the map $B\mathrm{SO} \to B\mathrm{GL}_1(\mathbb{S})$ comes from the classical j–homomorphism, as above.

(4) That $B\mathrm{Sp} \to B\mathrm{U} \to \mathrm{SO}/\mathrm{U}$ is a fiber sequence of infinite loop spaces also follows from [9].

(5) $B\mathrm{String} \to B\mathrm{U}[6, \infty) \to B^3\mathrm{Spin}$ is a fiber sequence of infinite loop spaces, again from [9]. The map $B\mathrm{U}[6, \infty) \to B\mathrm{GL}_1(\mathbb{S})$ is the obvious one.

(6) $B\mathrm{Spin} \to B\mathrm{SO} \to B(\mathrm{SO}/\mathrm{Spin})$ is clearly a fiber sequence of infinite loop spaces, and the map $B\mathrm{SO} \to B\mathrm{GL}_1(\mathbb{S})$ is clear.

(7) $\Omega\mathrm{SU}(n) \to \Omega\mathrm{SU}(n+1) \to \Omega S^{2n+1}$ is a fiber sequence of $\mathbb{E}_2$–spaces, as shown in [20, Diagram 9.1.2]. Since, by Bott periodicity, $\Omega\mathrm{SU} \simeq B\mathrm{U}$, there is a morphism of $\mathbb{E}_2$–spaces $\Omega\mathrm{SU}(n+1) \to \Omega\mathrm{SU} \simeq B\mathrm{U} \to B\mathrm{GL}_1(\mathbb{S})$.

(8) $B\mathrm{SO} \to B\mathrm{O} \to \mathbb{Z}/2$ is the usual fiber sequence of infinite loop spaces giving the 1–connected cover.

(9) $\Omega^2 S^3[3, \infty) \to \Omega^2 S^3 \to S^1$ is a fiber sequence of $\mathbb{E}_2$–spaces, after [14], and the morphism $\Omega^2 S^3 \to B\mathrm{GL}_1(\mathbb{S})$ is also the one given there.

Thus from Corollaries 2 and 4 we obtain the following equivalences (with respect to the numbering given above):

(1) $MU \simeq MSU \wedge_{\mathbb{S}[S^1]} \mathbb{S}$ and $MU \wedge_{MSU} MU \simeq MU \wedge \mathbb{S}[\mathbb{C}P^\infty]$.

(2) $M\mathrm{Spin} \simeq M\mathrm{String} \wedge_{K(\mathbb{Z},3)} \mathbb{S}$ and
$M\mathrm{Spin} \wedge_{M\mathrm{String}} M\mathrm{Spin} \simeq M\mathrm{Spin} \wedge \mathbb{S}[K(\mathbb{Z},4)]$.

(3) $MSO \simeq MU \wedge_{SO/U} \mathbb{S}$ and $MSO \wedge_{MU} MSO \simeq MSO \wedge \mathbb{S}[\mathrm{Spin}]$.

(4) $MU \simeq MSp \wedge_{\mathbb{S}[U/Sp]} \mathbb{S}$ and $MU \wedge_{MSp} MU \simeq MU \wedge \mathbb{S}[SO/U]$.

(5) $MU[6,\infty) \simeq M\mathrm{String} \wedge_{BB\mathrm{Spin}} \mathbb{S}$ and
$MU[6,\infty) \wedge_{M\mathrm{String}} MU[6,\infty) \simeq MU[6,\infty) \wedge \mathbb{S}[B^3\mathrm{Spin}]$.

(6) $MSO \simeq M\mathrm{Spin} \wedge_{\mathbb{S}[SO/\mathrm{Spin}]} \mathbb{S}$ and
$MSO \wedge_{M\mathrm{Spin}} MSO \simeq MSO \wedge \mathbb{S}[B(SO/\mathrm{Spin})]$.

(7) $X(n+1) \simeq X(n) \wedge_{\Omega^2 SU(n)} \mathbb{S}$ and
$X(n+1) \wedge_{X(n)} X(n+1) \simeq X(n+1) \wedge \mathbb{S}[\Omega S^{2n+1}]$.

(8) $MO \simeq MSO \wedge_{\mathbb{S}[\mathbb{Z}/2]} \mathbb{S}$ and $MO \wedge_{MSO} MO \simeq MO \wedge \mathbb{S}[\mathbb{R}P^\infty]$.

(9) $H\mathbb{Z}/2 \simeq H\mathbb{Z} \wedge_{\mathbb{S}[\mathbb{Z}]} \mathbb{S}$ and $H\mathbb{Z}/2 \wedge_{H\mathbb{Z}} H\mathbb{Z}/2 \simeq H\mathbb{Z}/2 \wedge \mathbb{S}[S^1]$.

Remark 9 Some of the examples given above can be verified by computations using the spectral sequence found in [8, Theorem 6.4],

$$\mathrm{Tor}_{p,q}^{E_*(R)}(E_*(M), E_*(N)) \Rightarrow E_{p+q}(M \wedge_R N).$$

For instance, for $E = H\mathbb{Z}$ we can relatively easily check that

$$H_*(X(n+1) \wedge_{X(n)} X(n+1); \mathbb{Z}) \cong H_*(X(n+1); \mathbb{Z}) \otimes_{\mathbb{Z}} H_*(\Omega S^{2n+1}; \mathbb{Z}).$$

Similar computations can be made for MU over MSU as well as for the fiber sequences appearing in Bott periodicity. Much of the relevant algebra for the latter has in fact already been determined in [6]. It is the author's hope that the above equivalences will be of use to homotopy theorists doing the much harder computations related to various connective covers of BO.

Remark 10 The relative Thom isomorphisms described above can be interpreted as torsor conditions for modules over spectral algebraic group schemes. In particular, if X is a Kan complex (and thus a coalgebra by the diagonal map) then we may think of an equivalence $Mf \wedge_{M(f\circ i)} Mf \simeq Mf \wedge \mathbb{S}[X]$ as giving $\mathrm{Spec}(Mf)$ the structure of a $\mathrm{Spec}(\mathbb{S}[X])$–torsor over $\mathrm{Spec}(M(f \circ i))$. Indeed, in terminology familiar to noncommutative geometers, many of the above examples are *Hopf–Galois extensions* in the sense of Rognes [21]. We delay an investigation of this structure to future work.

3.1 A new construction of $M\mathrm{U}$

The Lazard ring, which classifies formal group laws over discrete rings, is constructed iteratively by obstruction theory, one polynomial generator at a time (see [10]). The spectrum $M\mathrm{U}$, which classifies complex oriented ring spectra, is given in [19] as the colimit of the sequence of spectra $X(n)$ described in the previous section. Moreover, the spectra $X(n)$ are strongly related to rings used to construct the Lazard ring. This naturally leads to the question of whether or not the $X(n)$ spectra, and thus $M\mathrm{U}$, can also be constructed by some form of obstruction or deformation theory. Theorem 1 and its corollaries indicate that $X(n+1)$ is a "torsor" over $X(n)$ for the coalgebra $\mathbb{S}[\Omega S^{2n+1}]$. The stable splitting of ΩS^{2n+1} then further implies that $X(n+1)$ can be thought of a twisted polynomial extension of $X(n)$ (by a polynomial algebra with a single generator in degree $2n$).

What we show in this section is that even more is true. By invoking [3, Theorem 4.10], we can deduce that $X(n+1)$ is in fact a so-called *versal* $\mathbb{E}_1$–$X(n)$*–algebra of characteristic* χ_n, where χ_n is a class in $\pi_{2n-1}(X(n))$. This terminology, introduced in [23], indicates that $X(n+1)$ can be thought of as a highly structured ($\mathbb{E}_1$, to be specific) homotopy quotient of $X(n)$ along χ_n. It is never equivalent to the simpler process of "coning off" that class. What is true, however, is that $X(n+1)$–module structure on a spectrum (where we are thinking of $X(n+1)$ as an $\mathbb{E}_1$–algebra) is equivalent to an $X(n)$–module structure on that spectrum and a null-homotopy of multiplication by χ_n. Moreover, it is a result of the nilpotence theorem of [7] that each χ_n is nilpotent for all n. Recalling that π_{2n-1} is the first homotopy degree of $X(n)$ which is not either polynomial or empty, we see then that our construction of $M\mathrm{U}$ is given by iteratively attaching $\mathbb{E}_1$–cells along nilpotent elements, which is exactly what one might expect to do if one wished to construct the universal nilpotence detecting ring spectrum (which $M\mathrm{U}$ is).

Definition 11 Given $\alpha \in \pi_k(R)$ for R an $\mathbb{E}_n$–ring spectrum, we define the versal $\mathbb{E}_n$–R–algebra of characteristic α to be the pushout in $\mathbb{E}_n$–R–algebras

$$\begin{tikzcd}
\mathrm{Fr}_{\mathbb{E}_n}(\Sigma^k R) \arrow[r, "\mathrm{adj}(0)"] \arrow[d, "\mathrm{adj}(\alpha)"] & R \arrow[d, dashed] \\
R \arrow[r, dashed] & R/\!/\alpha
\end{tikzcd}$$

where $\mathrm{Fr}_{\mathbb{E}_n}$ is the free $\mathbb{E}_n$–algebra functor and the maps $\mathrm{adj}(\alpha)$ and $\mathrm{adj}(0)$ are the adjoints of the associated maps of R–modules $\alpha\colon \Sigma^k R \to R$ and $0\colon \Sigma^k R \to R$.

Corollary 12 *Let $X(n)$ be the Thom spectrum associated to the morphism of $\mathbb{E}_2$–monoidal Kan complexes $\Omega \mathrm{SU}(n) \to B\mathrm{U} \to B\mathrm{GL}_1(\mathbb{S})$. Then $X(n+1)$ is a versal $\mathbb{E}_1$–algebra over $X(n)$ of characteristic χ_n where χ_n is a canonical class in $\pi_{2n-1}(X(n))$.*

Proof Given the fiber sequence $\Omega\mathrm{SU}(n) \to \Omega\mathrm{SU}(n+1) \to \Omega S^{2n+1}$ and an application of Theorem 1 above, we can identify $X(n+1)$ as the $\mathbb{E}_1$–monoidal Thom spectrum given by the $\mathbb{E}_1$–monoidal left Kan extension $\Omega S^{2n+1} \to B\mathrm{GL}_1(X(n))$. By application of standard adjunctions, the map of $\mathbb{E}_1$–monoidal Kan complexes $\widetilde{\widetilde{\chi}}_n \in \mathrm{Map}_{\mathbb{E}_1}(\Omega S^{2n+1}, B\mathrm{GL}_1(X(n)))$ induces a map of Kan complexes

$$\widetilde{\chi}_n \in \mathrm{Map}_{\mathcal{T}}\big(S^{2n-1}, \mathrm{GL}_1(X(n))\big).$$

Note $\widetilde{\chi}_n$ must have image contained in a connected component $u \in \pi_0(\mathrm{GL}_1(X(n))) \simeq \mathbb{Z}/2$ which induces a translation $\tau_u\colon \Omega^\infty X(n) \to \Omega^\infty X(n)$. The composition

$$\tau_u \circ \widetilde{\chi}_n\colon S^{2n-1} \to \Omega^\infty X(n)$$

lifts to a morphism of spectra $\chi_n\colon \mathbb{S}^{2n-1} \to X(n)$. An application of [3, Theorem 4.10] gives that $X(n+1)$ is the versal $\mathbb{E}_1$–algebra of characteristic χ_n on $X(n)$. In other words, $X(n+1)$ in the following diagram is a pushout:

$$\begin{tikzcd} F_{\mathbb{E}_1}(\Sigma^{2n-1}X(n)) \arrow[r, "\mathrm{adj}(0)"] \arrow[d, "\mathrm{adj}(\chi_n)"] & X(n) \arrow[d] \\ X(n) \arrow[r] & X(n+1) \end{tikzcd}$$

This concludes the proof. □

Remark 13 The content of [3] allows us to consider $X(n+1)$ as the $\mathbb{E}_1$–spectrum obtained by attaching an $\mathbb{E}_1$–$X(n)$–cell to $X(n)$ along the map χ_n described above. Note that χ_1, as a nonzero element of $\pi_1(\mathbb{S})$, must be equivalent to η, the Hopf element. The Hopf element is, of course, the first nilpotent element in the stable homotopy groups of spheres and so, again, it stands to reason that it would be the first element eliminated in an effort to construct the maximal nilpotence detecting ring spectrum.

The following result is included since it follows immediately from [3]:

Corollary 14 *The $\mathbb{E}_1$–cotangent complex of the $\mathbb{E}_1$–algebra $X(n+1)$ in $X(n)$–modules is equivalent to $\Sigma^{2n} F_{\mathbb{E}_1}(X(n)) \wedge_{X(n)} X(n+1)$.*

Proof Compare [3, Proposition 5.4]. □

References

[1] **M Ando**, **A J Blumberg**, **D Gepner**, *Parametrized spectra, multiplicative Thom spectra, and the twisted Umkehr map*, preprint (2011) arXiv

[2] **M Ando**, **A J Blumberg**, **D Gepner**, **M J Hopkins**, **C Rezk**, *An ∞–categorical approach to R–line bundles, R–module Thom spectra, and twisted R–homology*, J. Topol. 7 (2014) 869–893 MR

[3] **O Antolín-Camarena**, **T Barthel**, *A simple universal property of Thom ring spectra*, preprint (2014) arXiv

[4] **J M Boardman**, **R M Vogt**, *Homotopy invariant algebraic structures on topological spaces*, Lecture Notes in Mathematics 347, Springer, New York (1973) MR

[5] **N Bourbaki**, *Lie groups and Lie algebras: Chapters 1–3*, Hermann, Paris (1975) MR Reprinted, Springer, Berlin (1998)

[6] **H Cartan**, *Démonstrations homologique des théorèmes de périodicité de Bott, II: Homologie et cohomologie des groupes classiques et de leurs espaces homogènes*, from "Périodicité des groupes d'homotopie stables des groupes classiques, d'après Bott", Séminaire Henri Cartan (1959/60) 12, Secrétariat Mathématique, Paris (1961) Exposé 17

[7] **E S Devinatz**, **M J Hopkins**, **J H Smith**, *Nilpotence and stable homotopy theory, I*, Ann. of Math. 128 (1988) 207–241 MR

[8] **A D Elmendorf**, **I Kriz**, **M A Mandell**, **J P May**, *Rings, modules, and algebras in stable homotopy theory*, Mathematical Surveys and Monographs 47, Amer. Math. Soc., Providence, RI (1997) MR

[9] **N Kitchloo**, **G Laures**, **W S Wilson**, *The Morava K–theory of spaces related to BO*, Adv. Math. 189 (2004) 192–236 MR

[10] **M Lazard**, *Commutative formal groups*, Lecture Notes in Mathematics 443, Springer, Berlin (1975) MR

[11] **L G Lewis, Jr**, *The stable category and generalized Thom spectra*, PhD thesis, The University of Chicago (1978) MR Available at `http://search.proquest.com/docview/251798616`

[12] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton University Press (2009) MR

[13] **J Lurie**, *Higher algebra*, preprint (2014) Available at `http://math.harvard.edu/~lurie/`

[14] **M Mahowald**, *Ring spectra which are Thom complexes*, Duke Math. J. 46 (1979) 549–559 MR

[15] **J P May**, *The geometry of iterated loop spaces*, Lectures Notes in Mathematics 271, Springer, Berlin (1972) MR

[16] **J P May**, **J Sigurdsson**, *Parametrized homotopy theory*, Mathematical Surveys and Monographs 132, Amer. Math. Soc., Providence, RI (2006) MR

[17] **J P May**, **R Thomason**, *The uniqueness of infinite loop space machines*, Topology 17 (1978) 205–224 MR

[18] **J C McKeown**, *Complex cobordism vs representing formal group laws*, preprint (2016) arXiv

[19] **D C Ravenel**, *Complex cobordism and stable homotopy groups of spheres*, Pure and Applied Mathematics 121, Academic Press, Orlando (1986) MR

[20] **D C Ravenel**, *Nilpotence and periodicity in stable homotopy theory*, Annals of Mathematics Studies 128, Princeton University Press (1992) MR

[21] **J Rognes**, *Galois extensions of structured ring spectra: stably dualizable groups*, Mem. Amer. Math. Soc. 898, Amer. Math. Soc., Providence, RI (2008) MR

[22] **D Sullivan**, *René Thom's work on geometric homology and bordism*, Bull. Amer. Math. Soc. 41 (2004) 341–350 MR

[23] **M Szymik**, *Commutative $\mathbb{S}$–algebras of prime characteristics and applications to unoriented bordism*, Algebr. Geom. Topol. 14 (2014) 3717–3743 MR

Department of Mathematics, University of Washington
C-138 Padelford, Seattle, WA 98195, United States

jbeards1@u.washington.edu

http://www.math.washington.edu/~jbeards1

Received: 31 March 2016 Revised: 21 August 2016

Algebraic & Geometric Topology 17 (2017) 1163–1188

On the geometry and topology of partial configuration spaces of Riemann surfaces

BARBU BERCEANU
DANIELA ANCA MĂCINIC
ŞTEFAN PAPADIMA
CLEMENT RADU POPESCU

We examine complements (inside products of a smooth projective complex curve of arbitrary genus) of unions of diagonals indexed by the edges of an arbitrary simple graph. We use Orlik–Solomon models associated to these quasiprojective manifolds to compute pairs of analytic germs at the origin, both for rank-1 and rank-2 representation varieties of their fundamental groups, and for degree-1 topological Green–Lazarsfeld loci. As a corollary, we describe all regular surjections with connected generic fiber, defined on the above complements onto smooth complex curves of negative Euler characteristic. We show that the nontrivial part at the origin, for both rank-2 representation varieties and their degree-1 jump loci, comes from curves of general type via the above regular maps. We compute explicit finite presentations for the Malcev Lie algebras of the fundamental groups, and we analyze their formality properties.

55N25, 55R80; 14F35, 20F38

1 Introduction and statement of results

Let Γ be a finite simple graph with cardinality n, vertex set V and edge set E. The *partial configuration space* of type Γ on a space Σ is

$$F(\Sigma,\Gamma)=\{z\in\Sigma^{\mathsf{V}}\mid z_i\neq z_j \text{ for all } ij\in\mathsf{E}\}. \tag{1}$$

When $\Gamma=K_n$, the complete graph with n vertices, $F(\Sigma,\Gamma)$ is the classical ordered configuration space of n distinct points in Σ. In this note, we analyze the interplay between geometry and topology when $\Sigma=\Sigma_g$ is a compact genus-g Riemann surface with partial configuration space denoted $F(g,\Gamma)$, with special emphasis on fundamental groups. The *partial pure braid groups* of type Γ in genus g, namely $P(g,\Gamma)=\pi_1(F(g,\Gamma))$, are natural generalizations of classical pure braid groups, which correspond to the case when $\Gamma=K_n$ and $\Sigma=\mathbb{C}$. When the graph is not complete,

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.1163

the classical approach to pure braid groups based on Fadell–Neuwirth fibrations does not work in full generality. Nevertheless, we are able in this note to compute rather delicate invariants of arbitrary partial pure braid groups, using techniques developed in Dimca and Papadima [11] and Măcinic, Papadima, Popescu and Suciu [18].

Viewing Σ_g as a smooth genus-g complex projective curve, $F(g,\Gamma)$ acquires the structure of an irreducible, smooth, quasiprojective complex variety (for short, a *quasiprojective manifold*). For such a quasiprojective manifold M, important geometric information is provided by maps onto manifolds of smaller dimension. Particularly interesting are the *admissible maps* in the sense of Arapura [2], ie the regular surjections onto quasiprojective curves, $f\colon M \to S$, having connected generic fiber. We say the admissible map f is of general type if $\chi(S) < 0$. We know from [2] that the set of admissible maps of general type on M, modulo reparametrization at the target, denoted $\mathcal{E}(M)$, is finite and is intimately related to the so-called *cohomology jump loci* of $\pi := \pi_1(M)$.

When $M = F(g,\Gamma)$, it is relatively easy to construct certain admissible maps of general type on M, associated to complete graphs $f\colon K_m \hookrightarrow \Gamma$ embedded in Γ; see Section 2. For $g \geq 2$, the relevant m equals 1, and $f_i\colon F(g,\Gamma) \to \Sigma_g$ is induced by the projection specified by the corresponding vertex $i \in \mathsf{V}$. For $g = 1$, the relevant m is 2, and $f_{ij}\colon F(1,\Gamma) \to \Sigma_1 \setminus \{0\}$ is given by the projection corresponding to $ij \in \mathsf{E}$, followed by the difference map on the elliptic curve Σ_1. For $g = 0$, the relevant m equals 4, and $f_{ijkl}\colon F(0,\Gamma) \to \mathbb{P}^1 \setminus \{0,1,\infty\}$ is the composition of the cross-ratio with the projection associated to the vertex set of the embedded K_4. Our first main result, proved in Section 2, establishes that there are no other admissible maps of general type on $M = F(g,\Gamma)$.

Theorem 1.1 *A complete set of representatives for $\mathcal{E}(F(g,\Gamma))$ is given by the admissible maps of general type described above.*

A basic topological invariant of a connected finite CW-complex M related to its cohomology jump loci is the *Malcev Lie algebra* of the fundamental group $\pi := \pi_1(M)$; cf [11]. The Malcev Lie algebra $\mathfrak{m}(\pi)$ of a group, over a characteristic zero field $\Bbbk$, defined by Quillen in [21], is a complete $\Bbbk$–Lie algebra whose filtration satisfies certain axioms, obtained by taking the primitives in the completion of the group ring $\Bbbk\pi$ with respect to the powers of the augmentation ideal.

Following Sullivan [23], we will say that a finitely generated group π is 1*–formal* if its Malcev Lie algebra is isomorphic to the completion with respect to the lower central series (lcs) filtration of a quadratic Lie algebra L (ie a Lie algebra presented by degree-1 generators and relations of degree 2): $\mathfrak{m}(\pi) \simeq \widehat{L}$. 1–formal groups enjoy many pleasant topological properties; see, for instance, Dimca, Papadima and Suciu [12]. The 1–formality of classical pure braid groups and pure welded braid groups also has strong consequences in the corresponding theories of finite-type invariants, as shown in Berceanu and Papadima [4].

In Section 3, we compute the Malcev Lie algebras of partial pure braid groups and determine precisely when they are 1–formal, as follows. Our next main result extends computations done by Bezrukavnikov [5] (for $g \geq 1$ and $\Gamma = K_n$) and Bibby and Hilburn [6] (for $g \geq 1$ and chordal graphs). Moreover, in our presentations below, redundant relations have been eliminated for $g \geq 1$.

Theorem 1.2 *The Malcev Lie algebra $\mathfrak{m}(P(g,\Gamma))$ is isomorphic to the* lcs *completion of a finitely presented Lie algebra, $L(g,\Gamma)$, with generators in degree 1 and relations in degrees 2 and 3, described in Proposition 3.2 for $g = 0$ and Proposition 3.4 for $g \geq 1$. The group $P(g,\Gamma)$ is not 1–formal if and only if $g = 1$ and the graph Γ contains a K_3 subgraph.*

Now, we move to our unifying theme: the interplay between the geometry of a quasiprojective manifold M, encoded by a smooth compactification $\overline{M}$, and the embedded topological jump loci of M. We start by recalling a couple of relevant definitions and facts related to the topological side of this story. Fix $q \in \mathbb{Z}_{>0} \cup \{\infty\}$. We will say that M is a q–finite space if (up to homotopy) M is a connected CW-complex with finite q–skeleton, whose (finitely generated) fundamental group will be denoted by π. Let $\iota\colon \mathbb{G} \to \mathrm{GL}(V)$ be a morphism of complex linear algebraic groups. The associated *characteristic varieties* (in degree $i \geq 0$ and depth $r \geq 0$),

$$\mathcal{V}_r^i(M,\iota) = \{\rho \in \mathrm{Hom}(\pi,\mathbb{G}) \mid \dim H^i(M, {}_{\iota\rho}V) \geq r\}, \tag{2}$$

are Zariski closed subvarieties (for $i \leq q$) of the affine *representation variety* $\mathrm{Hom}(\pi,\mathbb{G})$, for which the trivial representation provides a natural basepoint, $1 \in \mathrm{Hom}(\pi,\mathbb{G})$. These cohomology jump loci are called *topological Green–Lazarsfeld loci* for $r = 1$. They were introduced in the rank-one case (ie for $\iota = \mathrm{id}_{\mathbb{C}^\times}$) in Green and Lazarsfeld [14], for a smooth projective complex variety M. In the rank-one case, we simplify notation to $\mathcal{V}_r^i(M)$. Note that, in general, $\mathcal{V}_r^1(M,\iota) := \mathcal{V}_r^1(\pi,\iota)$ depends only on π for all r.

We go on by describing the infinitesimal analogs of the above notions, following [11]. Let $(A^\bullet, d)$ be a complex commutative differential graded algebra with positive grading (for short, a cdga). We will say that $A^\bullet$ is q–finite if $A^0 = \mathbb{C}\cdot 1$ and $\sum_{i=1}^{q} \dim A^i < \infty$. Let $\theta\colon \mathfrak{g} \to \mathfrak{gl}(V)$ be a finite-dimensional representation of a finite-dimensional complex Lie algebra. The affine variety of *flat connections*, $\mathscr{F}(A,\mathfrak{g})$, consists of the solutions in $A^1 \otimes \mathfrak{g}$ of the Maurer–Cartan equation, has the trivial flat connection 0 as a natural basepoint, and is natural in both A and $\mathfrak{g}$. For $\omega \in \mathscr{F}(A,\mathfrak{g})$, there is an associated covariant derivative, $d_\omega\colon A^\bullet \otimes V \to A^{\bullet+1} \otimes V$, with $d_\omega^2 = 0$, by flatness. The *resonance varieties*

$$\mathscr{R}^i_r(A,\theta) = \{\omega \in \mathscr{F}(A,\mathfrak{g}) \mid \dim H^i(A \otimes V, d_\omega) \geq r\} \tag{3}$$

are Zariski closed subvarieties (for $i \leq q$). We use the simplified notation $\mathscr{R}^i_r(A)$ in the rank-one case (ie when $\theta = \mathrm{id}_{\mathbb{C}}$).

We say that the cdga $A^\bullet$ is a q–model of M (and omit q from all terminology when $q = \infty$) if $A^\bullet$ has the same Sullivan q–minimal model as the de Rham cdga $\Omega^\bullet(M)$; cf [23]. In particular, $H^\bullet(A) \simeq H^\bullet(M)$ as graded algebras, when A is a model of M.

The link between topological and infinitesimal objects is provided by [11, Theorem B]. Assume that both A and M are q–finite and A is a q–model of M. Denote by θ the tangential representation of ι. Then for $i \leq q$ and $r \geq 0$, the embedded analytic germs $\mathscr{V}^i_r(M,\iota)_{(1)} \subseteq \mathrm{Hom}(\pi,\mathbb{G})_{(1)}$ at 1 are isomorphic to the corresponding embedded germs $\mathscr{R}^i_r(A,\theta)_{(0)} \subseteq \mathscr{F}(A,\mathfrak{g})_{(0)}$ at 0. Moreover, by [11, Theorem A], if π is a finitely generated group, then the germ $\mathrm{Hom}(\pi,\mathbb{G})_{(1)}$ depends only on the Malcev Lie algebra $\mathfrak{m}(\pi)$ and the Lie algebra of $\mathbb{G}$.

Finally, assume that M is a quasiprojective manifold, and $M = \overline{M} \setminus D$ is a smooth compactification obtained by adding at infinity a hypersurface arrangement D in $\overline{M}$ (in the sense of Dupont [13]). Then there is an associated (natural, finite) *Orlik–Solomon model* $A^\bullet(\overline{M}, D)$ of the finite space M, constructed in [13]. It follows from [11, Theorem C] that this model A determines $\mathscr{E}(M)$, which is in bijection with the positive-dimensional irreducible components through the origin, for both $\mathscr{R}^1_1(A)$ and $\mathscr{V}^1_1(M)$.

When $M = F(g,\Gamma)$, we may take $\overline{M} = \Sigma_g^{\mathsf{V}}$ and $D_\Gamma = \bigcup_{ij \in \mathsf{E}} \Delta_{ij}$ (the union of the diagonals associated to the edges of the graph). We prove Theorem 1.1 by computing the irreducible decomposition of $\mathscr{R}^1_1(A)$ for the Orlik–Solomon model $A = A(\overline{M}, D_\Gamma)$. When $g = 1$ and $\Gamma = K_n$, the result follows from a more precise description of all positive-dimensional components of $\mathscr{V}^1_1(M)$, obtained by Dimca in [10]. Given a 1–finite 1–model A of a connected CW-space M, we show in Theorem 3.1 that the Malcev Lie algebra $\mathfrak{m}(\pi_1(M))$ is isomorphic to the lcs completion of the *holonomy Lie algebra* of A, introduced in [18]. This general result is the basic tool for the proof of Theorem 1.2, where $M = F(g,\Gamma)$ and $A = A(\overline{M}, D_\Gamma)$.

$\mathrm{SL}_2(\mathbb{C})$–representation varieties received a lot of attention, both in topology and algebraic geometry. In order to describe their germs at 1 for partial pure braid groups, together with the embedded germs of associated nonabelian characteristic varieties (in degree 1 and depth 1), we use their infinitesimal analogs, described above. Let $\theta\colon \mathfrak{g} \to \mathfrak{gl}(V)$ be a finite-dimensional representation of $\mathfrak{g} = \mathfrak{sl}_2$ or $\mathfrak{sol}_2$, the Lie algebra of $\mathrm{SL}_2(\mathbb{C})$ or of its standard Borel subgroup. To state our next main result, we need two definitions from [18]. Denote by $\mathscr{F}^1(A, \mathfrak{g}) \subseteq \mathscr{F}(A, \mathfrak{g})$ the flat connections of the form $\omega = \eta \otimes g$, with $d\eta = 0$ and $g \in \mathfrak{g}$, and set $\Pi(A, \theta) = \{\omega \in \mathscr{F}^1(A, \mathfrak{g}) \mid \det \theta(g) = 0\}$. To have a uniform notation, denote by $f\colon F(g, \Gamma) \to S = \overline{S} \setminus F$ the admissible maps from Theorem 1.1, where $\overline{S} = \Sigma_g$ and $F \subseteq \overline{S}$ is a finite subset (in particular, a hypersurface arrangement in $\overline{S}$). To avoid trivialities, we will assume in genus 0 that $H^1(F(g, \Gamma)) \neq 0$. (The complete description of $H^1(F(g, \Gamma))$ may be found in Lemma 2.3; what happens in general with the embedded topological Green–Lazarsfeld loci in degree 1 of M at the origin, when $b_1(M) = 0$, is explained in Section 4.)

Theorem 1.3 *In the above setup, there is a regular extension* $\overline{f}\colon (\overline{M}, D) \to (\overline{S}, F)$ *of* f, *for all* $f \in \mathscr{E} := \mathscr{E}(F(g, \Gamma))$, *where* D *is a hypersurface arrangement in* $\overline{M}$ *with complement* $F(g, \Gamma)$, *which induces* cdga *maps between Orlik–Solomon models,* $f^*\colon A^\bullet(\overline{S}, F) \to A^\bullet(\overline{M}, D)$, *with the property that*

$$\mathscr{F}(A^\bullet(\overline{M}, D), \mathfrak{g}) = \mathscr{F}^1(A^\bullet(\overline{M}, D), \mathfrak{g}) \cup \bigcup_{f \in \mathscr{E}} f^* \mathscr{F}(A^\bullet(\overline{S}, F), \mathfrak{g}) \tag{4}$$

for $\mathfrak{g} = \mathfrak{sl}_2$ *or* $\mathfrak{sol}_2$, *and*

$$\mathscr{R}^1_1(A^\bullet(\overline{M}, D), \theta) = \Pi(A^\bullet(\overline{M}, D), \theta) \cup \bigcup_{f \in \mathscr{E}} f^* \mathscr{F}(A^\bullet(\overline{S}, F), \mathfrak{g}) \tag{5}$$

for any finite-dimensional representation $\theta\colon \mathfrak{g} \to \mathfrak{gl}(V)$.

This shows that for partial configuration spaces on smooth projective curves, the nontrivial part at the origin, for both $\mathrm{SL}_2(\mathbb{C})$–representation varieties and their degree-one topological Green–Lazarsfeld loci, "comes from curves of general type, via admissible maps". (The contribution of these curves, $f^* \mathscr{F}(A^\bullet(\overline{S}, F), \mathfrak{g})$, was computed in [18, Lemma 7.3].) A similar pattern is exhibited by quasiprojective manifolds with 1–formal fundamental group; cf [18, Corollary 7.2]. The geometric formulae from Theorem 1.3 seem to be quite satisfactory, since in genus 1, where non-1–formal examples appear (cf Theorem 1.2), the purely algebraic description from [18, Proposition 5.3] (obtained by assuming formality) may not hold, as we explain in Example 4.6.

2 Admissible maps and rank-one resonance

We devote this section to the proof of Theorem 1.1. Our strategy is to compute the irreducible decomposition of $\mathscr{R}^1_1(A(g, \Gamma))$, where $A^\bullet(g, \Gamma)$ is the Orlik–Solomon model of $M := F(g,\Gamma) = \overline{M} \setminus D_\Gamma$ from [13], $\overline{M} = \Sigma_g^{\mathsf{V}}$ and $D_\Gamma = \bigcup_{ij \in \mathsf{E}} \Delta_{ij}$. As a byproduct, we obtain a complete description of the irreducible components through 1, for the rank-one characteristic variety $\mathscr{V}_1^1(P(g,\Gamma))$, as explained in the introduction.

The Dupont models $A^\bullet(\overline{M}, D)$ are defined over $\mathbb{Q}$ and generalize Morgan's construction of *Gysin models* from [19], which corresponds to the case of a simple normal crossing divisor D. Among other things, the models of Dupont are natural with respect to regular morphisms $\overline{f}\colon (\overline{M}, D) \to (\overline{M}', D')$, in the following sense. When the regular map $\overline{f}\colon \overline{M} \to \overline{M}'$ has the property that $\overline{f}^{-1}(D') \subseteq D$, it induces a regular map $f\colon \overline{M} \setminus D \to \overline{M}' \setminus D'$, and a cdga map $f^*\colon A^\bullet(\overline{M}', D') \to A^\bullet(\overline{M}, D)$. Plainly, a graph inclusion $f\colon \Gamma' \hookrightarrow \Gamma$ (ie f embeds V' into V and E' into E) induces by projection a regular morphism $\overline{f}\colon (\Sigma_g^{\mathsf{V}}, D_\Gamma) \to (\Sigma_g^{\mathsf{V}'}, D_{\Gamma'})$, and a cdga map $f^*\colon A^\bullet(g,\Gamma') \to A^\bullet(g,\Gamma)$. Moreover, $A^\bullet(g,\Gamma) = A^\bullet_\bullet(g,\Gamma)$ is a bigraded cdga with *positive weights*, in the sense of Definition 5.1 from [11]. The lower degree, called *weight*, is preserved by cdga maps induced by graph inclusions. A simple example is $A^\bullet(g,\varnothing) = (H^\bullet(\Sigma_g^{\times n}), d = 0)$.

Now, we recall from [11; 18] a couple of facts about rank-1 resonance, needed in the sequel. Let $A^\bullet$ be a finite cdga. For $\xi \in A^1 \otimes \mathbb{C} = A^1$, the Maurer–Cartan equation reduces to $d\xi = 0$. Thus, $\mathscr{F}(A,\mathbb{C})$ is naturally identified with $H^1(A) \subseteq A^1$, since $A^0 = \mathbb{C} \cdot 1$. By definition, $\mathscr{R}_1^1(A) = \{\xi \in H^1(A) \mid H^1(A, d_\xi) \neq 0\}$, where $d_\xi \eta = d\eta + \xi\eta$ for $\eta \in A^1$. Clearly, $\mathscr{R}_1^1(A)$ depends only on the truncated cdga $A^{\leq 2} := A^\bullet / \bigoplus_{i>2} A^i$, and $\mathscr{R}_1^1(A) = \varnothing$ when $H^1(A) = 0$. We will use the following consequence of Theorem C from [11], applied to $M = F(g,\Gamma)$ and $A = A(g,\Gamma)$.

Theorem 2.1 *For a quasiprojective manifold M with finite model A having positive weights, $\mathscr{E}(M)$ is in bijection with the positive-dimensional (linear) irreducible components of $\mathscr{R}_1^1(A)$, via the correspondence $f \in \mathscr{E}(M) \mapsto \operatorname{im} H^1(f) \subseteq H^1(A)$.*

The maps from Theorem 1.1 are constructed in the following way. For a subset $\mathsf{V}' \subseteq \mathsf{V}$, we denote by $\mathrm{pr}_{\mathsf{V}'}\colon F(g,\Gamma) \to F(g,\Gamma')$ the regular map induced by the canonical projection, $\mathrm{pr}_{\mathsf{V}'}\colon \Sigma_g^{\mathsf{V}} \to \Sigma_g^{\mathsf{V}'}$, where Γ' is the full subgraph of Γ with vertex set V'. For an elliptic curve Σ_1, let $\overline{\delta}\colon (\Sigma_1^2, \Delta_{12}) \to (\Sigma_1, \{0\})$ be the regular morphism defined by $\overline{\delta}(z_1, z_2) = z_1 - z_2$. In genus 0, the regular map $\rho\colon F(0, K_4) \to \mathbb{P}^1 \setminus \{0,1,\infty\}$ is defined by $\rho(z_1,z_2,z_3,z_4) = \alpha(z_4)$, where $\alpha \in \mathrm{PSL}_2$ is the unique automorphism of $\mathbb{P}^1$ sending z_1, z_2 and z_3 to 0, 1 and ∞, respectively. For $g \geq 2$ and $f\colon K_1 \hookrightarrow \Gamma$, corresponding to $i \in \mathsf{V}$, set $f_i := \mathrm{pr}_i\colon F(g,\Gamma) \to \Sigma_g$. For $g = 1$ and $f\colon K_2 \hookrightarrow \Gamma$, corresponding to $ij \in \mathsf{E}$, set $f_{ij} := \delta \circ \mathrm{pr}_{ij}\colon F(1,\Gamma) \to \Sigma_1 \setminus \{0\}$. For $g = 0$ and $f\colon K_4 \hookrightarrow \Gamma$, with vertex subset $\{ijkl\} \subseteq \mathsf{V}$, set $f_{ijkl} := \rho \circ \mathrm{pr}_{ijkl}\colon F(0,\Gamma) \to \mathbb{P}^1 \setminus \{0,1,\infty\}$.

Lemma 2.2 *The above maps, f_i, f_{ij} and f_{ijkl}, are admissible, of general type.*

Proof In coordinates, $\rho(z_1, z_2, z_3, z_4) = (z_4 - z_1)/(z_2 - z_1) : (z_4 - z_3)/(z_2 - z_3)$ and $\rho(0, 1, \infty, z) = z$. Clearly, the maps $\rho\colon F(0, K_4) \to \mathbb{P}^1 \setminus \{0, 1, \infty\}$ and $\delta\colon F(1, K_2) \to \Sigma_1 \setminus \{0\}$, and the projections $\mathrm{pr}_*\colon F(g, \Gamma) \to F(g, K_{|*|})$ (where $*$ stands for i, ij or $ijkl$ and $|*|$ is 1, 2 or 4) are regular and surjective. The general-type condition is also clear: the spaces $\mathbb{P}^1 \setminus \{0, 1, \infty\} \simeq S^1 \vee S^1 \simeq \Sigma_1 \setminus \{0\}$ have Euler characteristic -1, and $\chi(\Sigma_g) \leq -2$ for $g \geq 2$.

In order to finish the proof, we show that all the fibers are connected. Let us denote by f_* any of the maps f_i, f_{ij} or f_{ijkl} and by φ_* the restriction of f_* to $F(g, K_n) \subseteq F(g, \Gamma)$. The fiber $\varphi_*^{-1}(z)$ is dense in $f_*^{-1}(z)$ (fix one or two or four points and move the other points outside the diagonals $z_p = z_q$), so it is enough to show that the fibers of φ_* are connected. The fibers of δ and ρ are path-connected:

$$\Sigma_1 \approx \delta^{-1}(z) \subseteq F(1, K_2), \quad F(0, K_3) \approx \rho^{-1}(z) \subseteq F(0, K_4).$$

The fibers of φ_* are path-connected as preimages of path-connected spaces through the locally trivial fibrations $\mathrm{pr}_*\colon F(g, K_n) \to F(g, K_{|*|})$ ($|*| = 1$, 2 or 4) with path-connected fibers $F(\Sigma_g \setminus \{z_*\}, K_{n-|*|})$. □

We recall from [13, Section 6] the complete description of the cdga $A^{\leq 2}$ for $A := A(g, \Gamma)$. We set $H^\bullet := H^\bullet(\Sigma_g)$, with $H^2 = \mathbb{C} \cdot \omega$ and with canonical symplectic basis $\{x^1, y^1, \dots, x^g, y^g\}$ of H^1 for $g \geq 1$, with $x^s y^s = \omega$ for all s. We know from [13] that $A^\bullet$ is generated as an algebra by $(H^\bullet)^{\otimes \mathsf{V}}$ (with weight equal to degree) and $G := \mathrm{span}\{G_{ij} \mid ij \in \mathsf{E}\}$ (with degree 1 and weight 2). The bigraded cdga map $f^*\colon A^\bullet(g, \Gamma') \to A^\bullet(g, \Gamma)$, associated to $f\colon \Gamma' \hookrightarrow \Gamma$, is determined by the canonical inclusions, $(H^\bullet)^{\otimes \mathsf{V}'} \hookrightarrow (H^\bullet)^{\otimes \mathsf{V}}$ and $G' \hookrightarrow G$. For $i \in \mathsf{V}$ and $g \geq 0$, we set $f_i^* \omega := \omega_i$, and for $g \geq 1$, we set $f_i^* x^s := x_i^s$ and $f_i^* y^s := y_i^s$ for all s. The structure of the truncated algebra $A^{\leq 2} = A^{\leq 2}(g, \Gamma)$ is described as follows:

- $A_1^1 = H^1(\Sigma_g^{\mathsf{V}}) = \bigoplus_{i \in \mathsf{V}} f_i^* H^1$ and $A_2^1 = G$;
- $A_2^2 = H^2(\Sigma_g^{\mathsf{V}})$;
- $A_3^2 = A_1^1 \otimes G$ modulo the relations (in genus $g \geq 1$) $(x_i^s - x_j^s) \otimes G_{ij}$ and $(y_i^s - y_j^s) \otimes G_{ij}$ for $s = 1, \dots, g$ and $ij \in \mathsf{E}$;
- $A_4^2 = \bigwedge^2 G$ modulo the relations $G_{jk} \wedge G_{ik} - G_{ij} \wedge G_{ik} + G_{ij} \wedge G_{jk}$ for $f\colon K_3 \hookrightarrow \Gamma$ (note that $A_4^2 = \mathrm{OS}^2(\mathcal{A}_\Gamma)$, the degree-2 piece of the Orlik–Solomon algebra [20] of the associated graphic arrangement of hyperplanes in $\mathbb{C}^{\mathsf{V}}$);
- $d(A_1^1) = 0$, $d(G_{ij}) = \omega_i + \omega_j + \sum_s (y_i^s \otimes x_j^s - x_i^s \otimes y_j^s) \in A_2^2$ when $g \geq 1$, and $d(G_{ij}) = \omega_i + \omega_j$ when $g = 0$;

- $\mu\colon \bigwedge^2 G \to A_4^2$ is the quotient map (exactly as in the graded algebra $\mathrm{OS}^\bullet(\mathcal{A}_\Gamma)$);
- $\mu\colon \bigwedge^2 A_1^1 \to A_2^2$ is the cup-product in the cohomology ring $H^\bullet(\Sigma_g^{\mathsf{V}})$;
- $\mu\colon A_1^1 \otimes G \to A_3^2$ is the quotient map.

(The lower indices of f, x, y, ω and G show the position in the cartesian or tensor product; the same convention will be used in Section 3 for a, b, z and C.)

Lemma 2.3 *In degree one, we have the following:*

(1) *If $g = 0$, then $H^1(F(0,\Gamma)) = 0$ if and only if every connected component of Γ is a tree or contains a unique cycle and this cycle has an odd length.*

(2) *If $g \geq 1$, then $H^1(F(g,\Gamma)) = H^1(\Sigma_g^{\mathsf{V}}) \neq 0$.*

Proof Due to the fact that A is a model of $F(g,\Gamma)$, we have

$$H^1(F(g,\Gamma)) = A_1^1 \oplus \ker(d\colon A_2^1 \to A_2^2) = H^1(\Sigma_g^{\mathsf{V}}) \oplus \ker(d\colon G \to H^2(\Sigma_g^{\mathsf{V}})).$$

We can split the differential according to the connected components of the graph $\Gamma = \amalg \Gamma(\alpha)$, $\mathsf{V} = \amalg \mathsf{V}(\alpha)$, $G = \amalg G(\alpha)$:

$$\ker(d\colon G \to H^2(\Sigma_g^{\mathsf{V}})) = \bigoplus_\alpha \ker(d\colon G(\alpha) \to H^2(\Sigma_g^{\mathsf{V}(\alpha)})),$$

so we give the proof for a connected graph Γ.

For $g \geq 1$, the coefficient of $y_i^s \otimes x_j^s$ in the differential of $\gamma = \sum_{ij \in \mathsf{E}} t_{ij} G_{ij}$ is t_{ij}; therefore, $d\colon G \to H^2(\Sigma_g^{\mathsf{V}})$ is injective.

For $g = 0$, we have that $\gamma = \sum_{ij \in \mathsf{E}} t_{ij} G_{ij}$ is a cocycle if and only if the coefficient of ω_i in $d(\gamma)$ is zero, ie

$$\sum_{j \in \mathsf{V},\, ij \in \mathsf{E}} t_{ij} = 0 \quad \text{for any } i \in \mathsf{V}. \tag{6}$$

This system of equations has n equations and $|\mathsf{E}|$ unknowns; if $\chi(\Gamma) = n - |\mathsf{E}| < 0$, one can find a nontrivial solution; hence $b_1(F(0,\Gamma)) \geq 1$. If $\chi(\Gamma) \geq 0$, we have to analyze only two cases (since Γ is connected):

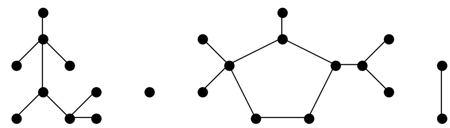
Case a ($\chi(\Gamma) = 1$) In this case, Γ is a (finite) tree; hence it has a vertex i of degree 1. One of the equations in the system (6) is $t_{ij} = 0$, and induction on $|\mathsf{V}|$ applied to the tree $\Gamma \setminus \{i\}$ shows that the system has only the trivial solution (the induction starts with $n = 1$, when $G = 0$).

Case b ($\chi(\Gamma) = 0$) In this case, $\Gamma \simeq S^1$ contains a unique cycle Γ_0 and, possibly, some branches; starting with a vertex of degree 1, we can eliminate these branches (if any), like in the previous case. The system is reduced to the equations corresponding to the vertices of Γ_0, say $1, 2, \ldots, l$:

$$t_{i-1,i} + t_{i,i+1} = 0, \quad i \equiv 1, \ldots, l \pmod{l}.$$

We get a nonzero solution $(a, -a, a, \ldots, -a)$ only for l even. $\square$

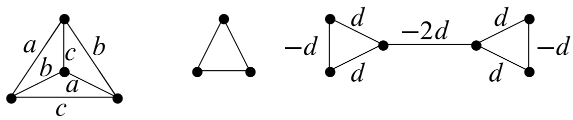
Example 2.4 Γ_1 :



$b_1(F(0, \Gamma_1)) = 0$

Example 2.5 Every edge is marked with its coefficient in an arbitrary cocycle; the unmarked edges have coefficient 0.

Γ_2 :



$(a + b + c = 0)$

$b_1(F(0, \Gamma_2)) = 3$

Remark 2.6 More generally, let Σ be an arbitrary complex projective manifold of dimension $m \geq 1$. The full configuration space $F(\Sigma, K_n)$ has a remarkable cdga model, $E^\bullet(\Sigma, n)$; when $m = 1$, we have $E^\bullet(\Sigma_g, n) = A^\bullet(g, K_n)$ (see eg [3] for details and references related to these models). As a graded algebra, $E^\bullet(\Sigma, n)$ is generated by $H^\bullet(\Sigma^n)$ and $G := \mathrm{span}\{G_{ij} \mid 1 \leq i < j \leq n\}$, taken in degree $2m - 1$. Denote by $EE^\bullet(\Sigma, n)$ the graded subalgebra of $E^\bullet(\Sigma, n)$ generated by G. It is shown in [3] that, when $\Sigma \neq \Sigma_0$, the restriction of d to $EE^+(\Sigma, n)$ is injective. This more general result gives an alternative proof of Lemma 2.3(2).

Proposition 2.7 *If $g \geq 2$, then $\mathscr{R}_1^1(A(g, \Gamma)) = \bigcup_{i \in \mathsf{V}} \mathrm{im}\, H^1(f_i)$ is the irreducible decomposition.*

Proof The inclusion $\bigcup_{i \in \mathsf{V}} \mathrm{im}\, H^1(f_i) \subseteq \mathscr{R}_1^1(A(g, \Gamma))$ is an obvious consequence of Theorem 2.1 and Lemma 2.2. For the proof of the opposite inclusion, we start with a nonzero cohomology class ξ in $H^1(A)$ and a d_ξ–cocycle $\eta \notin \mathbb{C} \cdot \xi$:

$$\xi = \sum_{i,s} (p_i^s x_i^s + q_i^s y_i^s), \quad \eta = \sum_{i,s} (u_i^s x_i^s + v_i^s y_i^s) + \sum_{ij \in \mathsf{E}} t_{ij} G_{ij}.$$

(From Lemma 2.3(2), ξ has no component in G.) For an arbitrary η, the differential $d_\xi \eta = d\eta + \xi \cdot \eta$ belongs to $A_2^2 \oplus A_3^2$; these two components are

$$A_2^2 \ni \sum_{ij \in \mathsf{E}} t_{ij} \Big(\omega_i + \omega_j + \sum_s (y_i^s \otimes x_j^s - x_i^s \otimes y_j^s) \Big) + \sum_{i,s} (p_i^s x_i^s + q_i^s y_i^s) \cdot \sum_{i,s} (u_i^s x_i^s + v_i^s y_i^s),$$

$$A_3^2 \ni \sum_{i,s} (p_i^s x_i^s + q_i^s y_i^s) \cdot \sum_{ij \in \mathsf{E}} t_{ij} G_{ij} = \xi \cdot \gamma.$$

We will show that the G–component of the d_ξ–cocycle η, namely $\gamma = \sum_{ij \in \mathsf{E}} t_{ij} G_{ij}$, is 0. Otherwise, there is an edge ij with $t_{ij} \neq 0$. Since the annihilator of G_{hk} is the span of $\{x_h^s - x_k^s, y_h^s - y_k^s\}_{1 \le s \le g}$, the vanishing of the A_3^2–component of $d_\xi \eta$ implies that ξ is reduced to

$$\xi = \sum_s p^s (x_i^s - x_j^s) + \sum_s q^s (y_i^s - y_j^s),$$

and also that γ has only one nonzero coefficient t_* (we can normalize it: $t_{ij} = 1$). In A_2^2, if $h \neq i, j$, the coefficients of $x_i^s \otimes x_h^r$, $x_i^s \otimes y_h^r$, $y_i^s \otimes x_h^r$ and $y_i^s \otimes y_h^r$ should be 0; hence $u_h^s = v_h^s = 0$ for any $h \neq i, j$ and any s. Hence, the A_2^2–component of $d_\xi \eta$ is reduced to

$$\omega_i + \omega_j + \sum_s (y_i^s \otimes x_j^s - x_i^s \otimes y_j^s) + \Big(\sum_s p^s (x_i^s - x_j^s) + \sum_s q^s (y_i^s - y_j^s) \Big) \cdot \sum_s (u_i^s x_i^s + u_j^s x_j^s + v_i^s y_i^s + v_j^s y_j^s);$$

the coefficients of the following elements in the canonical basis of A_2^2 are 0:

ω_i	$x_i^s \otimes y_j^s$	$y_i^s \otimes x_j^s$	$x_i^r \otimes x_j^s$
$1 + \sum_s p^s v_i^s - \sum_s q^s u_i^s$	$-1 + p^s v_j^s + q^s u_i^s$	$1 + q^s u_j^s + p^s v_i^s$	$p^r u_j^s + p^s u_i^r$

We show that this system has no solution. By the symmetry $(p, x) \leftrightarrow (q, y)$, we can suppose that there is an index s such that $p^s \neq 0$; if some $p^r = 0$, the second equation (for $s \to r$) implies that $u_i^r \neq 0$, and from the last equation we get $p^s = 0$, a contradiction. If all the coefficients p^s are nonzero, the last equation (for $s = r$) implies that $u_j^s = -u_i^s$ for any s, and the third equation shows that $1 - q^s u_i^s + p^s v_i^s = 0$ for any s. Adding these g equations, we find $g + \sum_s p^s v_i^s - \sum_s q^s u_i^s = 0$, and, comparing with the first equation, we obtain $g = 1$, again a contradiction.

Therefore, $\gamma = 0$; the nonvanishing of $H^1(A, d_\xi)$ is equivalent to

$$d_\xi \eta = \xi \cdot \eta = 0, \quad \eta \notin \mathbb{C} \cdot \xi.$$

This implies that $\xi \in \mathscr{R}_1^1(H^\bullet(\Sigma_g)^{\otimes \mathsf{V}}, d = 0)$. We infer from the Künneth formula for resonance [17, Proposition 5.6] that $\xi \in \operatorname{im} H^1(f_i)$ for some $i \in \mathsf{V}$.

In conclusion, $\mathscr{R}_1^1(A) = \bigcup_{i \in \mathsf{V}} \operatorname{im} H^1(f_i)$ is a finite union of linear subspaces. Since clearly there are no redundancies, this is the irreducible decomposition, as claimed. □

Proposition 2.8 *When $g = 1$, we have that $\mathscr{R}_1^1(A(1,\Gamma)) = \bigcup_{ij \in \mathsf{E}} \operatorname{im} H^1(f_{ij})$ is the irreducible decomposition if $\mathsf{E} \neq \varnothing$. Otherwise, $\mathscr{R}_1^1(A(1,\Gamma)) = \{0\}$.*

Proof Suppose that $\mathsf{E} = \varnothing$. As mentioned before, $A(1, \varnothing) = \big(\bigwedge(x_i, y_i), d = 0\big)$, and it is well known that the resonance variety $\mathscr{R}_1^1$ of an exterior algebra is reduced to 0.

Suppose that E is nonempty. Given $\xi = \sum_i p_i x_i + \sum_i q_i y_i$, a nonzero cohomology class in $\mathscr{R}_1^1(A)$ (see Lemma 2.3(2)), we may find

$$\eta = \sum_i u_i x_i + \sum_i v_i y_i + \sum_{ij \in \mathsf{E}} t_{ij} G_{ij}$$

such that $d_\xi \eta = 0$ and $\eta \notin \mathbb{C} \cdot \xi$. We may also suppose that there is one coefficient $t_{ij} \neq 0$ (otherwise we are in the previous case). Now we can apply the argument given in the proof of Proposition 2.7: there is only one nonzero coefficient t_* and $\xi \in \operatorname{Ann}(G_{ij})$; hence $\xi = p(x_i - x_j) + q(y_i - y_j)$. On the other hand, it is obvious that $H^1(f_{ij})(z) = z_i - z_j$ for $z \in H^1(\Sigma_1 \setminus \{0\}) = H^1(\Sigma_1)$.

We conclude, like in the proof of Proposition 2.7, that $\mathscr{R}_1^1(A) = \bigcup_{ij \in \mathsf{E}} \operatorname{im} H^1(f_{ij})$ is the irreducible decomposition, in this case. □

Proposition 2.9 *If $g = 0$ and $H^1(A(0,\Gamma)) = 0$, then $\mathscr{R}_1^1(A(0,\Gamma)) = \varnothing$.*

If $H^1(A(0,\Gamma)) \neq 0$, then $\mathscr{R}_1^1(A(0,\Gamma)) = \{0\} \cup \bigcup \operatorname{im} H^1(f_{ijkl})$ is the irreducible decomposition, where the union is taken over all K_4–subgraphs of Γ with vertex set $\{ijkl\}$, and $\{0\}$ is omitted when Γ contains such a subgraph.

Proof If $H^1(A(0,\Gamma)) = 0$ and $\xi \in \mathscr{R}_1^1(A)$, the definitions imply that $d_0 \eta = d\eta = 0$ for some $\eta \in A^1$. From this we get $\eta = 0$, which shows that $\mathscr{R}_1^1(A(0,\Gamma)) = \varnothing$.

From now on, we assume $H^1(A) \neq 0$. For any $K_4 \hookrightarrow \Gamma$ on vertices i, j, k and l, let us denote by $R_{ijkl} \subseteq H^1(A)$ the 2–dimensional subspace

$$\{a(G_{ij} + G_{kl}) + b(G_{ik} + G_{jl}) + c(G_{jk} + G_{il}) \mid a + b + c = 0\}.$$

When $\Gamma = K_4$, we find that $H^1(A(0, K_4)) = R_{1234}$, by solving the system (6). The map $H^1(f_{ijkl})$ is injective, since f_{ijkl} is admissible. Therefore, $\operatorname{im} H^1(f_{ijkl}) = R_{ijkl}$.

The inclusion $\mathscr{R}^1_1(A) \supseteq \{0\} \cup \bigcup R_{ijkl}$ follows from Theorem 2.1 and Lemma 2.2. Since plainly there are no redundancies in the above finite union of linear subspaces, we are left with proving that $\mathscr{R}^1_1(A) \setminus \{0\} \subseteq \bigcup R_{ijkl}$. To achieve this, we will also need to consider, for any $K_3 \hookrightarrow \Gamma$ on vertices i, j and k, the linear subspace $R_{ijk} \subseteq G = \mathrm{OS}^1(\mathcal{A}_\Gamma)$ defined by $R_{ijk} = \{aG_{ij} + bG_{jk} + cG_{ik} \mid a+b+c = 0\}$.

If $\xi \in \mathscr{R}^1_1(A) \setminus \{0\} \subseteq G \setminus \{0\}$, then $d\xi = 0$ and there is $\eta \in G \setminus \mathbb{C} \cdot \xi$ such that $d_\xi \eta = d\eta + \xi \cdot \eta = 0 \in A^2_2 \oplus A^2_4$, or, equivalently, $d\eta = 0$ and $\xi \cdot \eta = 0 \in \mathrm{OS}^2(\mathcal{A}_\Gamma)$. In particular, $\xi \in \mathscr{R}^1_1(\mathrm{OS}^\bullet(\mathcal{A}_\Gamma), d = 0) \setminus \{0\}$. It follows from [22, Section 3.5] that either $\xi \in R_{ijk}$ for some $K_3 \hookrightarrow \Gamma$, or $\xi \in R_{ijkl}$ for some $K_4 \hookrightarrow \Gamma$.

The first case cannot occur, since clearly $R_{ijk} \cap \ker(d) = 0$, by (6), and we are done. $\square$

Theorem 2.1 and Lemma 2.2, together with Propositions 2.7–2.9, prove Theorem 1.1 from the introduction. In the genus-0 case, the implication

$$H^1(A(0, \Gamma)) = 0 \implies \Gamma \text{ has no } K_4\text{–subgraphs}$$

is provided by Lemma 2.3(1).

3 Malcev completion and formality

We continue our analysis of partial pure braid groups with the proof of Theorem 1.2. Their Malcev Lie algebras are computed with the aid of the holonomy Lie algebras of their Orlik–Solomon models, $A^\bullet(g, \Gamma)$.

We will also consider a weaker notion of 1–formality: a finitely generated group π is *filtered formal* if its Malcev Lie algebra $\mathfrak{m}(\pi)$ is isomorphic to the lcs completion of a Lie algebra presentable with degree-1 generators and relations homogeneous with respect to bracket length. We recall that the free Lie algebra on a vector space, $\mathbb{L}^\bullet(W)$, is graded by bracket length. In low degrees, $\mathbb{L}^1(W) = W$, and the Lie bracket identifies $\mathbb{L}^2(W)$ with $\bigwedge^2 W$.

We are going to make extensive use of the following construction, introduced in [18, Definition 4.2]. The holonomy Lie algebra $\mathfrak{h}(A)$ of a 1–finite cdga A is the quotient of $\mathbb{L}(A^{1*})$ by the Lie ideal generated by $\operatorname{im}(d^* + \mu^*)$, where $d\colon A^1 \to A^2$ (respectively $\mu\colon \bigwedge^2 A^1 \to A^2$) is the differential (respectively the product) of the cdga $A^{\le 2}$, and $(\cdot)^*$ denotes vector space duals. This Lie algebra is functorial with respect to cdga maps, and has the following basic property. (A result similar to our theorem below was proved by Bezrukavnikov in [5], under the additional assumption that $A^\bullet$ is quadratic as a graded algebra; note that this condition is not satisfied in general by finite cdga models of spaces, in particular by the models $A^\bullet(0,\Gamma)$.)

Theorem 3.1 *If A is a 1–finite 1–model of a connected CW-space M, then $\mathfrak{m}(\pi_1(M))$ is isomorphic to the* lcs *completion of $\mathfrak{h}(A)$ as filtered Lie algebras.*

Proof Our approach is based on a key result obtained by Chen in [7] and refined by Hain in [15]. This result provides the following description for the Malcev completion of $\pi := \pi_1(M)$, over a characteristic zero field $\Bbbk$, in terms of iterated integrals and bar constructions.

Consider the complete Hopf algebra $\widehat{\Bbbk\pi}$, where the completion is taken with respect to the powers of the augmentation ideal of the group ring $\Bbbk\pi$. The complete Lie algebra $\mathfrak{m}(\pi)$ is the Lie algebra of primitives, $P\widehat{\Bbbk\pi}$, endowed with the induced filtration, defined by Quillen in [21, Appendix A]. On the other hand, let $B^\bullet(A)$ be the differential graded Hopf algebra obtained by applying the bar functor to the augmented cdga $A^\bullet$, where the augmentation sends A^+ to 0 and is the identity on $A^0=\Bbbk\cdot 1$; see eg [15, Section 1.1]. The dual Hopf algebra, $H^0B(A)^* = \mathrm{Hom}_{\Bbbk}(H^0B(A),\Bbbk)$, is a complete Hopf algebra, with filtration induced from the bar filtration of $H^0B(A)$; see [15, Section 2.4].

Next, let $f\colon A' \to A''$ be an augmented cdga map inducing an isomorphism in H^i for $i \le 1$ and a monomorphism in H^2 (for short, f is an augmented 1–equivalence). If $H^0(A')=\Bbbk\cdot 1$, we claim that the induced map, $H^0B(f)^*\colon H^0B(A'')^* \to H^0B(A')^*$, is a filtered isomorphism. Indeed, a standard argument based on the Eilenberg–Moore spectral sequence (like in Proposition 1.1.1 from [15]) shows that $H^0B(f)$ induces an isomorphism at the associated graded level, with respect to the bar filtrations, which clearly implies our assertion. The fact that $A^\bullet$ and $\Omega^\bullet(M)$ have the same Sullivan 1–minimal model, $\mathcal{M}^\bullet$, implies by rational homotopy theory [23] the existence of two augmented 1–equivalences, $\mathcal{M}^\bullet \to A^\bullet$ and $\mathcal{M}^\bullet \to \Omega^\bullet(M)$. Here, both $A^\bullet$ and $\mathcal{M}^\bullet$ are canonically augmented, as above, since $A^0=\mathcal{M}^0=\Bbbk\cdot 1$, and the augmentation of $\Omega^\bullet(M)$ is induced by the basepoint chosen for $\pi_1(M)$, as in [15].

It follows from [15, Corollary 2.4.5] that integration induces an isomorphism between $\widehat{\Bbbk\pi}$ and $H^0B(A)^*$, as complete Hopf algebras. This leads to the aforementioned description of the Malcev Lie algebra: $\mathfrak{m}(\pi) \simeq PH^0B(A)^*$, as complete Lie algebras.

Now, we claim that we may assume that $A^\bullet$ is of finite type, ie all graded pieces are finite dimensional. Indeed, the canonical cdga projection, $A^\bullet \twoheadrightarrow A^{\le 2}$, is clearly a 1–equivalence. Hence, $A^{\le 2}$ is also a 1–model of M, by [23]. It is equally easy to check that $\iota\colon \Bbbk\cdot 1 \oplus A^1 \oplus (\mathrm{im}(d)+\mathrm{im}(\mu)) \hookrightarrow A^{\le 2}$ is a cdga inclusion and a 1–equivalence.

Therefore, we may replace $A^{\le 2}$ by the above finite-type sub-cdga, without changing the holonomy Lie algebra, as claimed.

We may thus consider the dual cocommutative differential graded coalgebra, $A_\bullet := A^{\bullet *}$. By the standard duality between the bar construction for a cdga and the Adams cobar construction C for a cocommutative differential graded coalgebra [1], the complete Hopf algebras $H^0 B(A^\bullet)^*$ and $\widehat{H}_0 C(A_\bullet)$ are isomorphic. In concrete terms, the Hopf algebra $H_0 C(A_\bullet)$ is easily identified with the quotient of the primitively generated tensorial Hopf algebra on A_1, by the two-sided Hopf ideal generated by $\mathrm{im}(-d^* + \mu^*)$, and the completion is taken with respect to the descending filtration induced by tensor length.

Denote by $\mathfrak{q}(A)$ the quotient of the free Lie algebra $\mathbb{L}(A_1)$ by the Lie ideal generated by $\mathrm{im}(-d^* + \mu^*)$. The above discussion shows that the complete Hopf algebras $H^0 B(A)^*$ and $\widehat{U}\mathfrak{q}(A)$ are isomorphic, where $\widehat{U}$ is Quillen's completed universal enveloping algebra functor from [21, Appendix A].

Plainly, $-\mathrm{id}\colon A_1 \to A_1$ induces an isomorphism between the Lie algebras $\mathfrak{q}(A)$ and $\mathfrak{h}(A)$. We infer that $\mathfrak{m}(\pi) \simeq P\widehat{U}\mathfrak{h}(A)$, as complete Lie algebras.

Finally, let $\mathfrak{h}$ be a Lie algebra, and consider the canonical Lie homomorphism from [21, Appendix A], $\kappa\colon \mathfrak{h} \to P\widehat{U}\mathfrak{h}$. By [21, Corollary A3.9 and Remark A3.11], κ sends the lcs filtration of $\mathfrak{h}$ into the Malcev filtration of $P\widehat{U}\mathfrak{h}$, inducing an isomorphism at the associated graded level. Passing to completions, we infer that $\widehat{\kappa}\colon \widehat{\mathfrak{h}} \to P\widehat{U}\mathfrak{h}$ is a filtered Lie isomorphism. We conclude that $\mathfrak{m}(\pi) \simeq \widehat{\mathfrak{h}(A)}$, as filtered Lie algebras, thus finishing our proof. □

When $M = F(g,\Gamma)$ and $A = A(g,\Gamma)$, set $L(g,\Gamma) := \mathfrak{h}(A(g,\Gamma))$. We will denote, for $g \ge 0$, the basis dual to $\{G_{ij}\}_{ij\in\mathsf{E}}$ and $\{\omega_i\}_{i\in\mathsf{V}}$ by $\{C_{ij}\}_{ij\in\mathsf{E}}$ and $\{z_i\}_{i\in\mathsf{V}}$, respectively. For $g \ge 1$, the basis dual to $\{x_i^s, y_i^s \mid 1 \le i \le n, 1 \le s \le g\}$ will be denoted $\{a_i^s, b_i^s\}$.

Proposition 3.2 *The Malcev Lie algebra $\mathfrak{m}(P(0,\Gamma))$ is isomorphic to the lcs completion of $L(0,\Gamma)$, where the Lie algebra $L(0,\Gamma)$ is the quotient of the free Lie algebra on $\{C_{ij}\}_{ij\in\mathsf{E}}$ by the relations*

$$\sum_{j\in\mathsf{V},\, ij\in\mathsf{E}} C_{ij} \quad (i \in \mathsf{V}), \tag{7}$$

$$[C_{ij}, C_{kl}] \quad (ij, kl \in \mathsf{E}), \tag{8}$$

$$[C_{ij}, C_{jk}] \quad (ij, jk \in \mathsf{E} \text{ and } ik \notin \mathsf{E}), \tag{9}$$

$$[C_{ij} + C_{jk}, C_{ik}] \quad (ij, jk, ik \in \mathsf{E}). \tag{10}$$

In particular, the group $P(0,\Gamma)$ is always 1–formal.

	z_i $i \in \mathsf{V}$	$C_{ij} \wedge C_{kl}$ i,j,k,l distinct	$C_{ij} \wedge C_{jk}$ $ik \notin \mathsf{E}$	$C_{ij} \wedge C_{ik}$ $ij, ik, jk \in \mathsf{E}$	$C_{ij} \wedge C_{jk}$ $ij, ik, jk \in \mathsf{E}$
d^*	$\sum_{j \in \mathsf{V}, ij \in \mathsf{E}} C_{ij}$	0	0	0	0
μ^*	0	$[C_{ij}, C_{kl}]$	$[C_{ij}, C_{jk}]$	$[C_{ij}+C_{jk}, C_{ik}]$	$[C_{ij}+C_{ik}, C_{jk}]$
$\Downarrow$	(7)	(8)	(9)	(10)	(10)

Table 1: From the proof of Proposition 3.2. In the last two columns, $i < j < k$.

Proof We consider the following canonical basis in $(A^2)^*$:

$\{z_i\}_{i \in \mathsf{V}} \cup \{C_{ij} \wedge C_{kl}\}_{ij,kl \in \mathsf{E}} \cup \{C_{ij} \wedge C_{jk}\}_{ik \notin \mathsf{E}} \cup \{C_{ij} \wedge C_{ik}, C_{ij} \wedge C_{jk}\}_{ij,ik,jk \in \mathsf{E}}$

(in the product $C_{ij} \wedge C_{kl}$ we take $i < j$, $i < k < l$ and $j \neq k, l$, and in the last set we take $i < j < k$; see [5]). Dualizing d and μ, where

$$dG_{ij} = \omega_i + \omega_j, \quad \mu(G_{ik} \wedge G_{jk}) = G_{ij} \wedge G_{jk} - G_{ij} \wedge G_{ik},$$

we obtain the defining relations in the last row of Table 1. From the last two relations, we see $[C_{ik}+C_{jk}, C_{ij}] = 0$, hence the relation (10), where i, j, k are arbitrarily ordered. □

Remark 3.3 By [19, Corollary 10.3], if the quasiprojective manifold M has the vanishing property in degree 1, ie $W_1 H^1(M) = 0$, then $\pi_1(M)$ is 1–formal, where $W_\bullet$ denotes Deligne's weight filtration [8; 9]. According to [8; 9], $W_1 H^1(M) = 0$ whenever M admits a smooth compactification $\overline{M}$ with $b_1(\overline{M}) = 0$. Hence, $P(0, \Gamma)$ is actually 1–formal in this stronger sense.

Proposition 3.4 *For $g \geq 1$, the Malcev Lie algebra $\mathfrak{m}(P(g,\Gamma))$ is isomorphic to the* lcs *completion of $L(g,\Gamma)$, where the Lie algebra $L(g,\Gamma)$ is the quotient of the free Lie algebra on $\{a_i^s, b_i^s\}$ by the relations*

(11) $$C_{ij} := [a_i^s, b_j^s] = [a_j^t, b_i^t] \quad (\forall i \neq j, \forall s, t),$$

(12) $$C_{ij} = 0 \quad (ij \notin \mathsf{E}),$$

(13) $$[a_i^s, b_j^t] = [a_j^s, b_i^t] = 0 \quad (\forall i < j,\ \forall s \neq t),$$

(14) $$[a_i^s, a_j^t] = [b_i^s, b_j^t] = 0 \quad (\forall i \neq j,\ \forall s, t),$$

(15) $$\sum\nolimits_j C_{ij} = \sum\nolimits_s [b_i^s, a_i^s] \quad (i \in \mathsf{V}),$$

(16) $$[a_k^s, C_{ij}] = [b_k^s, C_{ij}] = 0 \quad (\forall k \neq i, j,\ \forall s).$$

In particular, $L(g,\Gamma)$ is generated in degree 1 with relations in degrees 2 and 3, and consequently, the group $P(g,\Gamma)$ is always filtered formal.

	1	2	3	4	5
	z_i $i \in \mathsf{V}$	$C_{ij} \wedge C_{kl}$ i,j,k,l distinct	$C_{ij} \wedge C_{jk}$ $ik \notin \mathsf{E}$	$C_{ij} \wedge C_{ik}$ $ij,ik,jk \in \mathsf{E}$	$C_{ij} \wedge C_{jk}$ $ij,ik,jk \in \mathsf{E}$
d^*	$\sum_{j \in \mathsf{V},\, ij \in \mathsf{E}} C_{ij}$	0	0	0	0
μ^*	$\sum_s [a_i^s, b_i^s]$	$[C_{ij}, C_{kl}]$	$[C_{ij}, C_{jk}]$	$[C_{ij}+C_{jk}, C_{ik}]$	$[C_{ij}+C_{ik}, C_{jk}]$
$\Downarrow$	(15)	(17)	(20)	(19)	(19)

6	7	8	9	10	11	12	13
$a_i^s \otimes b_j^t$ $i<j$	$b_i^s \otimes a_j^t$ $i<j$	$a_i^s \otimes a_j^t$ $i<j$	$b_i^s \otimes b_j^t$ $i<j$	$a_k^s \otimes C_{ij}$ $k \neq i,j$	$b_k^s \otimes C_{ij}$ $k \neq i,j$	$a_i^s \otimes C_{ij}$ $i<j$	$b_i^s \otimes C_{ij}$ $i<j$
$-\delta_{st} C_{ij}$	$\delta_{st} C_{ij}$	0	0	0	0	0	0
$[a_i^s, b_j^t]$	$[b_i^s, a_j^t]$	$[a_i^s, a_j^t]$	$[b_i^s, b_j^t]$	$[a_k^s, C_{ij}]$	$[b_k^s, C_{ij}]$	$[a_i^s + a_j^s, C_{ij}]$	$[b_i^s + b_j^s, C_{ij}]$
(11)–(13)	(11)–(13)	(14)	(14)	(16)	(16)	(18)	(18)

Table 2: From the proof of Proposition 3.4. The indices in columns 4 and 5 satisfy $i<j<k$. For any C_{pq} in the table, $pq \in \mathsf{E}$, and the entries in columns 6 and 7 are to be replaced by 0 in the second row when $ij \notin \mathsf{E}$.

Proof The canonical basis in $(A^2)^*$ contains the list in the proof of Proposition 3.2, and also (with indices $1 \leq i < j \leq n$, $1 \leq s,t \leq g$ and $k \neq i,j$)

$$\{a_i^s \otimes a_j^t, a_i^s \otimes b_j^t, b_i^s \otimes a_j^t, b_i^s \otimes b_j^t\} \cup \{a_k^s \otimes C_{ij}, b_k^s \otimes C_{ij}, a_i^s \otimes C_{ij}, b_i^s \otimes C_{ij}\}.$$

To dualize d and μ, the relevant relations are

$$dG_{ij} = \omega_i + \omega_j + \sum_s (y_i^s \otimes x_j^s - x_i^s \otimes y_j^s),$$

$$\mu(x_i^s \wedge y_i^s) = \omega_i, \quad \mu(x_i^s \wedge y_j^t) = x_i^s \otimes y_j^t, \quad \mu(y_i^s \wedge x_j^t) = y_i^s \otimes x_j^t \quad (i<j),$$

$$\mu(x_i^s \wedge x_j^t) = x_i^s \otimes x_j^t, \quad \mu(y_i^s \wedge y_j^t) = y_i^s \otimes y_j^t \quad (i<j),$$

$$\mu(x_i^s \wedge G_{jk}) = x_i^s \otimes G_{jk}, \quad \mu(y_i^s \wedge G_{jk}) = y_i^s \otimes G_{jk},$$

$$\mu(x_i^s \wedge G_{ij}) = x_i^s \otimes G_{ij} = \mu(x_j^s \wedge G_{ij}), \quad \mu(y_i^s \wedge G_{ij}) = y_i^s \otimes G_{ij} = \mu(y_j^s \wedge G_{ij}).$$

The defining relations are obtained in the last row of Table 2. Note that, when $ij \in \mathsf{E}$, in the relations (11) C_{ij} is the dual of G_{ij}. The relations (16) are obtained in columns 10 and 11 for $ij \in \mathsf{E}$ and, otherwise, are a trivial consequence of (12). It remains to prove that the relations (11)–(16) imply the following list:

(17) $$[C_{ij}, C_{kl}] = 0 \quad (\text{if } \operatorname{card}\{i,j,k,l\} = 4),$$

(18) $$[a_i^s + a_j^s, C_{ij}] = [b_i^s + b_j^s, C_{ij}] = 0 \quad (\forall i \neq j, \forall s),$$

(19) $$[C_{ij} + C_{jk}, C_{ik}] = 0 \quad (\text{if } ij, ik, jk \in \mathsf{E}),$$

(20) $$[C_{ij}, C_{jk}] = 0 \quad (\text{if } ij, jk \in \mathsf{E} \text{ and } ik \notin \mathsf{E}).$$

The first relation is obvious:

$$[C_{ij}, C_{kl}] = [C_{ij}, [a_k^s, b_l^s]] = 0 \quad (\text{by (11) and (16)}).$$

The second equation comes from the equalities

$$\begin{aligned}
[a_j^s, C_{ij}] &= \big[a_j^s, \textstyle\sum_k C_{ik}\big] && (\text{by (16)}) \\
&= \big[a_j^s, \textstyle\sum_t [b_i^t, a_i^t]\big] && (\text{by (15)}) \\
&= [a_j^s, [b_i^s, a_i^s]] && (\text{by (13) and (14)}) \\
&= [C_{ij}, a_i^s] && (\text{by (11) and (14)})
\end{aligned}$$

(by symmetry, we get $[b_i^s + b_j^s, C_{ij}] = 0$).

Using (18), we can finish the proof as follows:

$$\begin{aligned}
[C_{ij} + C_{jk}, C_{ik}] &= [[a_i^s, b_j^s] + [a_k^s, b_j^s], C_{ik}] && (\text{by (11)}) \\
&= [[a_i^s + a_k^s, b_j^s], C_{ik}] = 0 && (\text{by (16) and (18)}),
\end{aligned}$$

and finally (20) may be established as follows:

$$\begin{aligned}
[C_{ij}, C_{jk}] &= [C_{ij}, [a_j^s, b_k^s]] && (\text{by (11)}) \\
&= [[C_{ij}, a_j^s], b_k^s] && (\text{by (16)}) \\
&= -[[C_{ij}, a_i^s], b_k^s] && (\text{by (18)}) \\
&= -[C_{ij}, [a_i^s, b_k^s]] = 0 && (\text{by (16), (11) and (12)}).
\end{aligned}$$

□

Example 3.5 Note that filtered formality is strictly weaker than 1–formality, as shown by the Torelli group in genus 3, which has a cubic, non-1–formal Malcev Lie algebra; cf Hain's work from [16].

Proposition 3.6 *Suppose that either $g \geq 2$, or $g = 1$ and Γ contains no K_3. Then the group $P(g, \Gamma)$ is 1–formal.*

Proof The cubic relations (16) follow from the quadratic relations: if $g \geq 2$, take $t \neq s$; then

$$[a_k^s, C_{ij}] = [a_k^s, [a_i^t, b_j^t]] = 0 \quad \text{(by (11), (13) and (14))}.$$

If $g = 1$ and, say, $ik \notin \mathsf{E}$, we find

$$[a_k^1, C_{ij}] = [a_k^1, [a_j^1, b_i^1]] = 0 \quad \text{(by (11), (12) and (14))}. \qquad \square$$

Proposition 3.7 *If $g = 1$ and Γ contains a K_3 subgraph, then the group $P(1,\Gamma)$ is not 1–formal.*

Proof When $g \geq 1$ and $f\colon \Gamma' \hookrightarrow \Gamma$ is arbitrary, note that $f_*\colon H_1(\Sigma_g^{\mathsf{V}}) \twoheadrightarrow H_1(\Sigma_g^{\mathsf{V}'})$ extends to $f_*\colon \mathbb{L}^\bullet(H_1(\Sigma_g^{\mathsf{V}})) \twoheadrightarrow \mathbb{L}^\bullet(H_1(\Sigma_g^{\mathsf{V}'}))$, a graded Lie surjection which preserves the graded parts of the defining Lie ideals (11)–(16). Furthermore, the canonical injection $f_\dagger\colon H_1(\Sigma_g^{\mathsf{V}'}) \hookrightarrow H_1(\Sigma_g^{\mathsf{V}})$ extends to a graded Lie monomorphism, $f_\dagger\colon \mathbb{L}^\bullet(H_1(\Sigma_g^{\mathsf{V}'})) \hookrightarrow \mathbb{L}^\bullet(H_1(\Sigma_g^{\mathsf{V}}))$, which preserves the cubic relations (16). Therefore, the 1–formality of $P(1,\Gamma)$ would imply the 1–formality of $P(1, K_3)$, in contradiction with [12, Example 10.1]. $\square$

Remark 3.8 It follows from Proposition 2.7 and [17, Proposition 5.6] that when $g \geq 2$, we have $\mathscr{R}_1^1(A^\bullet(g,\Gamma)) = \mathscr{R}_1^1(H^\bullet(\Sigma_g^{\mathsf{V}}))$ for any graph Γ. Nevertheless, $\mathfrak{m}(P(g,\Gamma)) \not\simeq \mathfrak{m}(\pi_1(\Sigma_g^{\mathsf{V}}))$ if $\mathsf{E} \neq \varnothing$. Indeed, assuming the contrary, we infer from [23] that the spaces $F(g,\Gamma)$ and Σ_g^{V} have isomorphic decomposable subspaces in the cohomology ring in degree two: $DH^2(F(g,\Gamma)) \simeq DH^2(\Sigma_g^{\mathsf{V}})$. Plainly, $DH^2(\Sigma_g^{\mathsf{V}}) = H^2(\Sigma_g^{\mathsf{V}})$. The description of the Orlik–Solomon model $A^\bullet(g,\Gamma)$ from Section 2 readily implies that $DH^2(F(g,\Gamma)) = H^2(\Sigma_g^{\mathsf{V}})/dG$. By Lemma 2.3(2), the above two vector spaces DH^2 have different dimensions if $\mathsf{E} \neq \varnothing$, a contradiction.

4 Nonabelian representation varieties and jump loci

Finally, we analyze germs at 1 of rank-2 nonabelian representation varieties and their degree-one topological Green–Lazarsfeld loci for partial pure braid groups, via admissible maps and Orlik–Solomon models, and we prove Theorem 1.3. In this section, $\mathbb{G} = \mathrm{SL}_2(\mathbb{C})$ or its standard Borel subgroup, with Lie algebra $\mathfrak{g} = \mathfrak{sl}_2$ or $\mathfrak{sol}_2$. Key to our computations is the well-known fact that $[A,B] = 0$ in $\mathfrak{g}$ if and only if $\operatorname{rank}\{A,B\} \leq 1$.

If $S = \overline{S} \setminus F$ is a quasiprojective curve, where $\overline{S}$ is projective and $F \subseteq \overline{S}$ is a finite subset, then $(\overline{S}, F)$ is the unique smooth compactification of S. For a quasiprojective manifold M, it is known that there is a *convenient* smooth compactification,

$M = \overline{M} \setminus D$, where D is a hypersurface arrangement in $\overline{M}$, which has the property that every admissible map of general type, $f\colon M \to S$, is induced by a regular morphism, $\overline{f}\colon (\overline{M}, D) \to (\overline{S}, F)$. These in turn induce cdga maps between Orlik–Solomon models, denoted $f^*\colon A^\bullet(\overline{S}, F) \to A^\bullet(\overline{M}, D)$. By naturality, we obtain an inclusion

$$\mathscr{F}(A^\bullet(\overline{M}, D), \mathfrak{g}) \supseteq \mathscr{F}^1(A^\bullet(\overline{M}, D), \mathfrak{g}) \cup \bigcup_{f \in \mathscr{E}(M)} f^*\mathscr{F}(A^\bullet(\overline{S}, F), \mathfrak{g}). \tag{21}$$

For any finite-dimensional representation $\theta\colon \mathfrak{g} \to \mathfrak{gl}(V)$, we also know from [18, Corollary 3.8] that $\Pi(A, \theta) \subseteq \mathscr{R}_1^k(A, \theta)$ if $H^k(A) \neq 0$.

Let $\{f\colon B_f^\bullet \to A^\bullet\}$ be a finite family of cdga maps between finite objects.

Proposition 4.1 *Assume that $H^1(A) \neq 0$. For every f, suppose that $B_f^\bullet = B_f^{\leq 2}$, $\chi(H^\bullet(B_f)) < 0$ and f is a monomorphism. If $\mathscr{R}_1^1(A) = \bigcup_f \operatorname{im} H^1(f)$ and (21) holds as an equality for the family $\{f\colon B_f^\bullet \to A^\bullet\}$, then*

$$\mathscr{R}_1^1(A, \theta) = \Pi(A, \theta) \cup \bigcup_f f^*\mathscr{F}(B_f, \mathfrak{g}) \tag{22}$$

for any finite-dimensional representation $\theta\colon \mathfrak{g} \to \mathfrak{gl}(V)$.

Proof We first show the inclusion "$\supseteq$". The fact that $\Pi(A, \theta) \subseteq \mathscr{R}_1^1(A, \theta)$ is due to the assumption $H^1(A) \neq 0$. The equality $\mathscr{R}_1^1(B_f, \theta) = \mathscr{F}(B_f, \mathfrak{g})$ follows from [18, Proposition 2.4] since $B_f^\bullet = B_f^{\leq 2}$ and $\chi(H^\bullet(B_f)) < 0$. Lemma 2.6 from [18] implies that $f^*\mathscr{R}_1^1(B_f, \theta) \subseteq \mathscr{R}_1^1(A, \theta)$, since f is injective in degree 1. To verify the inclusion "$\subseteq$", pick $\omega \in \mathscr{R}_1^1(A, \theta) \setminus \bigcup_f f^*\mathscr{F}(B_f, \mathfrak{g})$. We infer from (21) that $\omega = \eta \otimes g$, with $d\eta = 0$ and $g \in \mathfrak{g}$. Theorem 1.2 from [18] says then that there is an eigenvalue λ of $\theta(g)$ such that $\lambda\eta \in \mathscr{R}_1^1(A)$. If $\det\theta(g) \neq 0$, then $\lambda \neq 0$. Since $\mathscr{R}_1^1(A) = \bigcup_f \operatorname{im} H^1(f)$, we deduce that $\eta = f^*\eta_f$ for some f and some $\eta_f \in H^1(B_f)$. Hence, $f^*(\eta_f \otimes g) \in \mathscr{F}(A, \mathfrak{g})$. The injectivity of f forces then $\eta_f \otimes g \in \mathscr{F}(B_f, \mathfrak{g})$. This implies that $\omega \in f^*\mathscr{F}(B_f, \mathfrak{g})$, a contradiction. Consequently, $\omega \in \Pi(A, \theta)$, and we are done. □

Let A be a finite model of the finite space M. If $b_1(M) = 0$, then it follows from [21] that $\mathfrak{m}(\pi_1(M)) = 0$. Theorems A and B in [11] together imply then that both germs $\operatorname{Hom}(\pi_1(M), \mathbb{G})_{(1)}$ and $\mathscr{F}(A, \mathfrak{g})_{(0)}$ contain only the origin. Furthermore, $b_1(M) = 0$ implies that $\mathscr{V}_1^1(M, \iota)_{(1)} = \mathscr{R}_1^1(A, \theta)_{(0)} = \varnothing$; cf [11, Theorem B] and [18, (15)]. For a quasiprojective manifold M with $b_1(M) > 0$, it follows from [11, Example 5.3] that we may always find a convenient compactification (by adding at infinity a normal crossing divisor) which satisfies all hypotheses from Proposition 4.1, for the family $\{f^*\colon A^\bullet(\overline{S}, F) \to A^\bullet(\overline{M}, D)\}_{f \in \mathscr{E}(M)}$, except possibly the last assumption.

In this way, we infer from Remark 3.3 and Proposition 4.1 that the genus-0 case of Theorem 1.3 becomes a consequence of the following general result.

Theorem 4.2 *If $b_1(M) > 0$ and $W_1 H^1(M) = 0$, then equality holds in (21) for a convenient compactification with normal crossings and for $\mathfrak{g} = \mathfrak{sl}_2$ or $\mathfrak{sol}_2$.*

Proof For every $f \in \mathscr{E}(M)$, note that $H^\bullet(\overline{f})\colon H^\bullet(\overline{S}) \to H^\bullet(\overline{M})$ is injective; see eg [11, Example 5.3]. Our vanishing assumption on $W_1 H^1(M)$ implies that $H^1(\overline{M}) = 0$; cf [8; 9]. Hence, $W_1 H^1(S) = 0$.

Let $A^\bullet_\bullet := A^\bullet(\overline{M}, D)$ be the Gysin model, and assume that $W_1 H^1(M) = 0$. Then $A^1 = A^1_2$, by [19]. Set $Z^1_2 := H^1(A) \subseteq A^1_2$, and denote by $A^\bullet_Z \subseteq A^{\leq 2}$ the sub-cdga with $d = 0$ defined by $A^0_Z = \mathbb{Q} \cdot 1$, $A^1_Z = Z^1_2$ and $A^2_Z = \mu(\bigwedge^2 Z^1_2) \subseteq A^2_4$. Note that $d(A^1_2) \subseteq A^2_2$. We infer that the cdga inclusion $\iota\colon A^\bullet_Z \hookrightarrow A^{\leq 2}$ is a 1–equivalence, ie it induces an isomorphism in H^1 and a monomorphism in H^2. On the other hand, it follows from the definitions that the variety $\mathscr{F}(A, \mathfrak{g})$ depends only on the corestrictions of $d\colon A^1 \to A^2$ and $\mu\colon \bigwedge^2 A^1 \to A^2$ to the subspace $\operatorname{im}(d) + \operatorname{im}(\mu) \subseteq A^2$ for any cdga A and any Lie algebra $\mathfrak{g}$. Therefore, we have an inclusion $\iota^*\colon \mathscr{F}(A_Z, \mathfrak{g}) \subseteq \mathscr{F}(A, \mathfrak{g})$.

Since ι is a 1–equivalence, it follows from Theorem 3.9 and Sections 7.3–7.5 in [11] that $\mathscr{F}(A_Z, \mathfrak{g})$ and $\mathscr{F}(A, \mathfrak{g})$ have the same analytic germs at 0. Now, we recall from [11] that each cdga, A and A_Z, has positive weights, and the associated $\mathbb{C}^\times$–actions preserve the varieties $\mathscr{F}(A_Z, \mathfrak{g})$ and $\mathscr{F}(A, \mathfrak{g})$, and the origin 0. This implies that all irreducible components of $\mathscr{F}(A, \mathfrak{g})$ pass through 0, and similarly for $\mathscr{F}(A_Z, \mathfrak{g})$. This in turn is enough to infer that actually $\mathscr{F}(A_Z, \mathfrak{g}) = \mathscr{F}(A, \mathfrak{g})$, since the germs at 0 are equal. Moreover, $\mathscr{F}(A_Z, \mathfrak{g}) = \mathscr{F}(H^\bullet(A), \mathfrak{g})$, by construction.

The equalities $\mathscr{F}(A^\bullet(\overline{M}, D), \mathfrak{g}) = \mathscr{F}(H^\bullet(M), \mathfrak{g})$ and $\mathscr{F}(A^\bullet(\overline{S}, F), \mathfrak{g}) = \mathscr{F}(H^\bullet(S), \mathfrak{g})$ are clearly compatible with the natural maps induced by $\overline{f}\colon (\overline{M}, D) \to (\overline{S}, F)$ for any $f \in \mathscr{E}(M)$. Plainly $\mathscr{F}^1(A^\bullet(\overline{M}, D), \mathfrak{g})$ depends only on $H^1(M)$ and $\mathfrak{g}$. Thus, we may replace in (21) $A^\bullet(\overline{M}, D)$ by $(H^\bullet(M), d = 0)$ and $A^\bullet(\overline{S}, F)$ by $(H^\bullet(S), d = 0)$. In this way, our claim reduces to the equality proved in [18, Corollary 7.2(55)]. □

In positive genus, we are going to describe explicitly the convenient compactifications from Theorem 1.3, and check that all hypotheses from Proposition 4.1 hold for the associated families of cdga maps, $\{f^*\colon A^\bullet(\overline{S}, F) \to A^\bullet(\overline{M}, D)\}_{f \in \mathscr{E}(M)}$, except the last assumption.

When $g \geq 2$, we have that $M := F(g, \Gamma) = \Sigma_g^{\mathsf{V}} \setminus D_\Gamma$ is a convenient compactification: for $i \in \mathsf{V}$, the regular morphism $\overline{f_i} := \mathrm{pr}_i\colon (\Sigma_g^{\mathsf{V}}, D_\Gamma) \to (\Sigma_g, \varnothing)$ extends the admissible map $f_i\colon F(g,\Gamma) \to \Sigma_g$ from Lemma 2.2. By Lemma 2.3(2), $H^1(A(g,\Gamma)) \neq 0$ for $g \geq 1$. Clearly, $B_f^\bullet = B_f^{\leq 2}$ and $\chi(H^\bullet(B_f)) < 0$ for any $f \in \mathscr{E}(M)$, since $B_f^\bullet = (H^\bullet(\Sigma_g), d = 0)$. It is easy to check that $f^*\colon A^\bullet(g,\Gamma') \to A^\bullet(g,\Gamma)$ is injective in degree $\bullet \leq 2$ for any $f\colon \Gamma' \hookrightarrow \Gamma$ and $g \geq 0$. Finally, the assumption on $\mathscr{R}_1^1(A(g,\Gamma))$ in Proposition 4.1 follows from Proposition 2.7.

In genus $g = 1$, we have that $M := F(1,\Gamma) = \Sigma_1^{\mathsf{V}} \setminus D_\Gamma$ is again a convenient compactification. For $ij \in \mathsf{E}$, denote by $\mathrm{pr}_{ij}\colon (\Sigma_1^{\mathsf{V}}, D_\Gamma) \to (\Sigma_1^2, D_{K_2})$ the regular morphism induced by projection. Let $\overline{\delta}\colon (\Sigma_1^2, D_{K_2}) \to (\Sigma_1, \{0\})$ be the regular morphism induced by the difference map of the elliptic curve Σ_1. Then clearly the regular morphism $\overline{f_{ij}} := \overline{\delta} \circ \mathrm{pr}_{ij}$ extends the admissible map $f_{ij}\colon F(1,\Gamma) \to \Sigma_1 \setminus \{0\}$ from Lemma 2.2. For any $f \in \mathscr{E}(M)$, we have that $B_f^\bullet = A^\bullet(\Sigma_1, \{0\}) = B_f^{\leq 2}$ is given by $B_f^0 = \mathbb{C} \cdot 1$, $B_f^1 = \mathrm{span}\{x, y, g\}$ and $B_f^2 = \mathbb{C} \cdot \mathcal{O}$. The differential is given by $dx = dy = 0$ and $dg = \mathcal{O}$, and the multiplication table is $xg = yg = 0$ and $xy = \mathcal{O}$. The hypotheses on $B_f^\bullet$ from Proposition 4.1 are clearly satisfied. It follows from naturality of Orlik–Solomon models [13] that $\delta^* x = x_1 - x_2$, $\delta^* y = y_1 - y_2$ and $\delta^* g = G_{12}$. In particular, $\delta^*\colon A^\bullet(\Sigma_1, \{0\}) \hookrightarrow A^\bullet(1, K_2)$ is injective, which proves the injectivity of $B_f^\bullet \to A^\bullet$ for any $f \in \mathscr{E}(M)$. Finally, the assumption on $\mathscr{R}_1^1(A(1,\Gamma))$ in Proposition 4.1 follows from Proposition 2.8, when $\mathsf{E} \neq \varnothing$. Otherwise, the claims in Theorem 1.3 follow from [18, Corollary 7.7].

By virtue of Proposition 4.1, we have thus reduced the proof of Theorem 1.3 in positive genus to checking that (21) holds as an equality for the families $\{f^*\colon A^\bullet(\overline{S}, F) \to A^\bullet(\overline{M}, D)\}_{f \in \mathscr{E}(M)}$ described above. To verify this equality, we will use another basic property of the holonomy Lie algebra of a cdga A, proved in Proposition 4.5 from [18]. This result allows us to naturally replace the variety of flat connections $\mathscr{F}(A, \mathfrak{g})$ by the variety of Lie homomorphisms, $\mathrm{Hom}_{\mathrm{Lie}}(\mathfrak{h}(A), \mathfrak{g})$, and $\mathscr{F}^1(A, \mathfrak{g})$ by $\mathrm{Hom}^1_{\mathrm{Lie}}(\mathfrak{h}(A), \mathfrak{g}) := \{\varphi \in \mathrm{Hom}_{\mathrm{Lie}}(\mathfrak{h}(A), \mathfrak{g}) \mid \dim \mathrm{im}(\varphi) \leq 1\}$.

Proposition 4.3 *If* $\varphi \in \mathrm{Hom}_{\mathrm{Lie}}\big(\mathfrak{h}(A(1,\Gamma)), \mathfrak{g}\big) \setminus \mathrm{Hom}^1_{\mathrm{Lie}}\big(\mathfrak{h}(A(1,\Gamma)), \mathfrak{g}\big)$, *there is* $ij \in \mathsf{E}$ *such that* $\varphi \in f_{ij}^* \,\mathrm{Hom}_{\mathrm{Lie}}\big(\mathfrak{h}(A(\Sigma_1, \{0\})), \mathfrak{g}\big)$.

Proof For $g \geq 1$, the holonomy Lie algebra $\mathfrak{h}(A(g,\Gamma))$ is isomorphic to the Lie algebra $L(g,\Gamma)$ from Proposition 3.4. By (14), a morphism $\varphi \in \mathrm{Hom}_{\mathrm{Lie}}\big(\mathfrak{h}(A(1,\Gamma)), \mathfrak{g}\big)$ satisfies

$$[\varphi(a_i), \varphi(a_j)] = [\varphi(b_i), \varphi(b_j)] = 0,$$

thus φ is defined by two elements $v, w \in \mathfrak{g}$ and two n–vectors $\alpha_* = (\alpha_i)$ and $\beta_* = (\beta_i)$:

$$\varphi(a_i) = \alpha_i v, \quad \varphi(b_i) = \beta_i w.$$

Equation (11) implies that $(\alpha_i\beta_j-\alpha_j\beta_i)[v,w]=0$. If $\varphi\notin \mathrm{Hom}^1_{\mathrm{Lie}}\big(\mathfrak{h}(A(1,\Gamma)),\mathfrak{g}\big)$, we have $\alpha_*\neq 0$, $\beta_*\neq 0$ and $[v,w]\neq 0$; hence $\mathrm{rank}\{\alpha_*,\beta_*\}=1$. Equation (15) is equivalent to

$$\sum_j [a_i,b_j]=\sum_j [a_j,b_i]=0 \quad (i\in \mathsf{V}).$$

Together with relation (14), these imply that $\sum_i a_i$ and $\sum_i b_i$ are central elements; therefore, their images $\sum_i \alpha_i v$ and $\sum_i \beta_i w$ are 0. In particular, at least two components of α_* (and the same components of β_*) are nonzero.

We will show that α_* and β_* have exactly two nonzero components. Relations (11) and (16) imply that, for any three distinct indices i, j and k,

$$\alpha_k\alpha_i\beta_j[v,[v,w]]=\beta_k\alpha_i\beta_j[w,[v,w]]=0.$$

The brackets $[v,[v,w]]$ and $[w,[v,w]]$ cannot be both 0 (otherwise $\mathrm{rank}\{v,w\}=1$); if $[v,[v,w]]\neq 0$, we have (for any three indices) $\alpha_k\alpha_i\beta_j=0$, which proves our claim (similarly if $[w,[v,w]]\neq 0$).

We infer that φ must be of the form

$$\begin{aligned} &\varphi(a_i)=\alpha v, \quad \varphi(a_j)=-\alpha v, \quad \varphi(a_k)=0,\\ &\varphi(b_i)=\beta w, \quad \varphi(b_j)=-\beta w, \quad \varphi(b_k)=0, \end{aligned} \tag{23}$$

with $\alpha,\beta\neq 0$ (where $k\neq i,j$). Therefore, $ij\in \mathsf{E}$, by (12).

The description of $A^\bullet(\Sigma_1,\{0\})$ implies, by a straightforward computation, that the Lie algebra $\mathfrak{h}(A(\Sigma_1,\{0\}))$ is the quotient of the free Lie algebra $\mathbb{L}(x^*,y^*,g^*)$ by the relation $g^*+[x^*,y^*]=0$, where $\{x^*,y^*,g^*\}$ is the basis dual to $\{x,y,g\}$. Therefore, $\mathfrak{h}(A(\Sigma_1,\{0\}))=\mathbb{L}(x^*,y^*)$. Moreover, the description of the action of δ^* and pr^*_{ij} on Orlik–Solomon models implies, by taking duals, that the Lie homomorphism $f_{ij*}\colon \mathfrak{h}(A(1,\Gamma))\to \mathfrak{h}(A(\Sigma_1,\{0\}))$ sends a_i to x^*, a_j to $-x^*$, b_i to y^*, b_j to $-y^*$, and a_k,b_k to 0 for $k\neq i,j$; see [18, Definition 4.2].

Define $\psi\in \mathrm{Hom}_{\mathrm{Lie}}\big(\mathfrak{h}(A(\Sigma_1,\{0\})),\mathfrak{g}\big)$ by $x^*\mapsto \alpha v$ and $y^*\mapsto \beta w$. By (23), we have $\varphi=f^*_{ij}(\psi)$. □

Proposition 4.4 *Assume that $g\geq 2$. If*

$$\varphi\in \mathrm{Hom}_{\mathrm{Lie}}\big(\mathfrak{h}(A(g,\Gamma)),\mathfrak{g}\big)\setminus \mathrm{Hom}^1_{\mathrm{Lie}}\big(\mathfrak{h}(A(g,\Gamma)),\mathfrak{g}\big),$$

there is $i\in \mathsf{V}$ such that $\varphi\in f_i^*\,\mathrm{Hom}_{\mathrm{Lie}}\big(\mathfrak{h}(A(\Sigma_g,\varnothing)),\mathfrak{g}\big)$.

Proof The holonomy Lie algebra of $A(\Sigma_g, \varnothing) = A(g, K_1)$ is generated by the elements $\{a^1, b^1, \dots, a^g, b^g\}$ modulo the relation $\sum_s [a^s, b^s] = 0$; hence a morphism $\psi \in \mathrm{Hom}_{\mathrm{Lie}}\big(\mathfrak{h}(A(\Sigma_g, \varnothing)), \mathfrak{g}\big)$ is defined by $2g$ elements $v^1, w^1, \dots, v^g, w^g \in \mathfrak{g}$ satisfying the relation $\sum_s [v^s, w^s] = 0$.

It is sufficient to show that for $\varphi \in \mathrm{Hom}_{\mathrm{Lie}}\big(\mathfrak{h}(A(g,\Gamma)), \mathfrak{g}\big) \setminus \mathrm{Hom}^1_{\mathrm{Lie}}\big(\mathfrak{h}(A(g,\Gamma)), \mathfrak{g}\big)$, there is an index i such that $\varphi(a_j^t) = \varphi(b_j^t) = 0$ for any $j \neq i$ and any t; this implies, via (11), that $\varphi(C_{jk}) = 0$ (for any $j \neq k$) and, using (15), that $\sum_s [\varphi(a_i^s), \varphi(b_i^s)] = 0$.

Denote by A and B the span of $\{\varphi(a_*^*)\}$ and $\{\varphi(b_*^*)\}$ respectively. As $\dim \operatorname{im}(\varphi) \geq 2$, we have to analyze only two cases:

Case 1 $(\dim(A) = \dim(B) = 1)$ In this case there are two linearly independent elements $v, w \in \mathfrak{g}$ and indices (i, s) and (k, t) such that

$$\varphi(a_j^r) = \alpha_j^r v \quad \text{and} \quad \varphi(b_j^r) = \beta_j^r w \quad \text{for any } j, r \text{ and } \alpha_i^s \neq 0 \neq \beta_k^t.$$

Relation (13) and $[v, w] \neq 0$ imply that $\beta_j^r = 0$ if $j \neq i$ and $r \neq s$; from the hypothesis $g \geq 2$ and relation (11), we obtain

$$\varphi(C_{ij}) = \alpha_i^s \beta_j^s [v, w] = \alpha_i^r \beta_j^r [v, w] = 0,$$

hence $\beta_j^r = 0$ for any $j \neq i$ and any r. This implies that $k = i$ and, by symmetry, that $\alpha_j^r = 0$ for any $j \neq i$ and any r.

Case 2 $(\dim(A) \geq 2)$ (By symmetry, the case $\dim(B) \geq 2$ can be treated in the same way.) In this case, there are indices $i = j$, $s \neq t$ and two linearly independent elements $v^s, v^t \in \mathfrak{g}$ such that

$$\varphi(a_i^s) = v^s, \quad \varphi(a_j^t) = v^t$$

($i \neq j$ contradicts relation (14), since $[v^s, v^t] \neq 0$). For any $k \neq i$ and any r, we obtain from (14) that

$$[\varphi(a_i^s), \varphi(a_k^r)] = [\varphi(a_i^t), \varphi(a_k^r)] = 0, \quad \text{hence} \quad \varphi(a_k^r) = 0.$$

Using relation (13), the same argument applied to b_k^r shows that $\varphi(b_k^r) = 0$ for any $k \neq i$ and any $r \neq s, t$. Again from (13), $[\varphi(a_i^t), \varphi(b_k^s)] = 0$. On the other hand, by (11), $[\varphi(a_i^s), \varphi(b_k^s)] = [\varphi(a_k^t), \varphi(b_i^t)] = 0$. Hence, $\varphi(b_k^r) = 0$ for any $k \neq i$ and $r = s, t$, and we are done. $\square$

Propositions 4.3 and 4.4 complete the proof of Theorem 1.3. Similar results were obtained in [18] for quasiprojective manifolds with 1–formal fundamental group. (Note that $(H^\bullet(S), d = 0)$ is a finite model of a quasiprojective curve S, and $\mathscr{F}((H^\bullet(S), d = 0), \mathfrak{g})$ is computed in Lemma 7.3 from [18] when $\chi(S) < 0$.) They were based on the following algebraic construction. Let $A^\bullet$ be a 1–finite cdga with linear resonance, ie $\mathscr{R}_1^1(A) = \bigcup_{C \in \mathcal{C}} C$ is a finite union of linear subspaces of $H^1(A)$. For each $C \in \mathcal{C}$, let $A_C^\bullet \hookrightarrow A^{\leq 2}$ be the sub-cdga defined by $A_C^0 = \mathbb{C} \cdot 1$, $A_C^1 = C$ and $A_C^2 = A^2$.

Proposition 4.5 [18, Proposition 5.3] *If in addition $d = 0$, then*

$$\mathscr{F}(A, \mathfrak{g}) = \mathscr{F}^1(A, \mathfrak{g}) \cup \bigcup_{C \in \mathcal{C}} \mathscr{F}(A_C, \mathfrak{g})$$

for $\mathfrak{g} = \mathfrak{sl}_2$ or $\mathfrak{sol}_2$.

Example 4.6 The geometric formulae from Theorem 1.3, based on Orlik–Solomon models, seem to be the right extension of the similar results in [18], beyond the 1–formal case. Indeed, let us consider for $A^\bullet = A^\bullet(1, \Gamma)$ the linear decomposition of $\mathscr{R}_1^1(A)$ from Proposition 2.8, case $\mathsf{E} \neq \varnothing$. For each $C = \operatorname{im} H^1(f_{ij})$, we claim that $\mathscr{F}(A_C, \mathfrak{g}) = \mathscr{F}^1(A_C, \mathfrak{g})$, when $\mathfrak{g} = \mathfrak{sl}_2$ or $\mathfrak{sol}_2$. This implies that the algebraic formula from Proposition 4.5 reduces in this case to the equality $\mathscr{F}(A, \mathfrak{g}) = \mathscr{F}^1(A, \mathfrak{g})$. On the other hand, we have seen that $\mathfrak{h}\big(A(\Sigma_1, \{0\})\big)$ is a free Lie algebra on two generators, and therefore $\mathscr{F}\big(A(\Sigma_1, \{0\}), \mathfrak{g}\big)$ contains an element not in $\mathscr{F}^1\big(A(\Sigma_1, \{0\}), \mathfrak{g}\big)$. Consequently, if $ij \in \mathsf{E}$ then it follows from Theorem 1.3 that $f_{ij}^* \mathscr{F}\big(A(\Sigma_1, \{0\}), \mathfrak{g}\big) \setminus \mathscr{F}^1\big(A(1, \Gamma), \mathfrak{g}\big) \neq \varnothing$. Thus, the algebraic formula does not hold.

To compute $\mathfrak{h}(A_C)$, we may replace A_C^2 by $\mu_C\big(\bigwedge^2 C\big)$. Note that $d_C = 0$, C is two-dimensional generated by $x_i - x_j$ and $y_i - y_j$, and $(x_i - x_j)(y_i - y_j) \neq 0$. It follows that the holonomy Lie algebra $\mathfrak{h}(A_C)$ is two-dimensional abelian. Therefore, $\mathrm{Hom}_{\mathrm{Lie}}(\mathfrak{h}(A_C), \mathfrak{g}) = \mathrm{Hom}^1_{\mathrm{Lie}}(\mathfrak{h}(A_C), \mathfrak{g})$ as claimed.

Acknowledgements Papadima was partially supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, PN-II-ID-PCE-2011-3-0288, grant 132/05.10.2011; Popescu was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PN-II-RU-TE-2012-3-0492.

References

[1] **J F Adams**, *On the cobar construction*, from "Colloque de topologie algébrique", Georges Thone, Liège (1957) 81–87 MR

[2] **D Arapura**, *Geometry of cohomology support loci for local systems, I*, J. Algebraic Geom. 6 (1997) 563–597 MR

[3] **S Ashraf**, **H Azam**, **B Berceanu**, *Representation theory for the Križ model*, Algebr. Geom. Topol. 14 (2014) 57–90 MR

[4] **B Berceanu**, **Ş Papadima**, *Universal representations of braid and braid-permutation groups*, J. Knot Theory Ramifications 18 (2009) 999–1019 MR

[5] **R Bezrukavnikov**, *Koszul DG-algebras arising from configuration spaces*, Geom. Funct. Anal. 4 (1994) 119–135 MR

[6] **C Bibby**, **J Hilburn**, *Quadratic-linear duality and rational homotopy theory of chordal arrangements*, Algebr. Geom. Topol. 16 (2016) 2637–2661 MR

[7] **K T Chen**, *Iterated path integrals*, Bull. Amer. Math. Soc. 83 (1977) 831–879 MR

[8] **P Deligne**, *Théorie de Hodge, II*, Inst. Hautes Études Sci. Publ. Math. 40 (1971) 5–57 MR

[9] **P Deligne**, *Théorie de Hodge, III*, Inst. Hautes Études Sci. Publ. Math. 44 (1974) 5–77 MR

[10] **A Dimca**, *Characteristic varieties and logarithmic differential* 1*–forms*, Compos. Math. 146 (2010) 129–144 MR

[11] **A Dimca**, **Ş Papadima**, *Non-abelian cohomology jump loci from an analytic viewpoint*, Commun. Contemp. Math. 16 (2014) 1350025, 47 MR

[12] **A Dimca**, **Ş Papadima**, **A I Suciu**, *Topology and geometry of cohomology jump loci*, Duke Math. J. 148 (2009) 405–457 MR

[13] **C Dupont**, *The Orlik–Solomon model for hypersurface arrangements*, Ann. Inst. Fourier (Grenoble) 65 (2015) 2507–2545 MR

[14] **M Green**, **R Lazarsfeld**, *Deformation theory, generic vanishing theorems, and some conjectures of Enriques, Catanese and Beauville*, Invent. Math. 90 (1987) 389–407 MR

[15] **R M Hain**, *The de Rham homotopy theory of complex algebraic varieties, I*, K–Theory 1 (1987) 271–324 MR

[16] **R Hain**, *Infinitesimal presentations of the Torelli groups*, J. Amer. Math. Soc. 10 (1997) 597–651 MR

[17] **A D Măcinic**, *Cohomology rings and formality properties of nilpotent groups*, J. Pure Appl. Algebra 214 (2010) 1818–1826 MR

[18] **D A Măcinic**, **Ş Papadima**, **C R Popescu**, **A I Suciu**, *Flat connections and resonance varieties: from rank one to higher ranks*, Trans. Amer. Math. Soc. 369 (2017) 1309–1343 MR

[19] **J W Morgan**, *The algebraic topology of smooth algebraic varieties*, Inst. Hautes Études Sci. Publ. Math. 48 (1978) 137–204 MR

[20] **P Orlik**, **H Terao**, *Arrangements of hyperplanes*, Grundl. Math. Wissen. 300, Springer, New York (1992) MR

[21] **D Quillen**, *Rational homotopy theory*, Ann. of Math. 90 (1969) 205–295 MR

[22] **H K Schenck**, **A I Suciu**, *Resonance, linear syzygies, Chen groups, and the Bernstein–Gelfand–Gelfand correspondence*, Trans. Amer. Math. Soc. 358 (2006) 2269–2289 MR

[23] **D Sullivan**, *Infinitesimal computations in topology*, Inst. Hautes Études Sci. Publ. Math. 47 (1977) 269–331 MR

Institute of Mathematics "Simion Stoilow", Romanian Academy
P.O. Box 1-764, 014700 Bucharest, Romania

barbu.berceanu@imar.ro, anca.macinic@imar.ro, stefan.papadima@imar.ro, radu.popescu@imar.ro

Received: 22 April 2016

Algebraic & Geometric Topology 17 (2017) 1189–1260

Symplectic embeddings of four-dimensional ellipsoids into integral polydiscs

DANIEL CRISTOFARO-GARDINER
DAVID FRENKEL
FELIX SCHLENK

In previous work, the second author and Müller determined the function $c(a)$ giving the smallest dilate of the polydisc $P(1,1)$ into which the ellipsoid $E(1,a)$ symplectically embeds. We determine the function of two variables $c_b(a)$ giving the smallest dilate of the polydisc $P(1,b)$ into which the ellipsoid $E(1,a)$ symplectically embeds for all integers $b \geqslant 2$.

It is known that, for fixed b, if a is sufficiently large then all obstructions to the embedding problem vanish except for the volume obstruction. We find that there is another kind of change of structure that appears as one instead increases b: the number-theoretic "infinite Pell stairs" from the $b=1$ case almost completely disappears (only two steps remain) but, in an appropriately rescaled limit, the function $c_b(a)$ converges as b tends to infinity to a completely regular infinite staircase with steps all of the same height and width.

53D05; 14B05, 32S05

1 Introduction and result

1.1 Introduction

Since Gromov's classic paper [15], it has been known that symplectic embedding problems are intimately related to many phenomena in symplectic geometry, Hamiltonian dynamics, and other fields. The smallest interesting dimension is four, and all our results are in this dimension. So consider the standard four-dimensional symplectic vector space $(\mathbb{R}^4,\omega)$, where $\omega = dx_1 \wedge dy_1 + dx_2 \wedge dy_2$. Open subsets in $\mathbb{R}^4$ are endowed with the same symplectic form. Given two such sets U and V, a symplectic embedding of U into V is a smooth embedding $\varphi\colon U \to V$ that preserves the symplectic form: $\varphi^*\omega = \omega$. We write $U \overset{s}{\hookrightarrow} V$ if there exists a symplectic embedding $U \to V$. Deciding whether $U \overset{s}{\hookrightarrow} V$ is very hard in general.

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.1189

One thus looks at simple sets, such as the open ball $B^4(a)$ of radius $\sqrt{a}$, or polydiscs $P(a,b) = B^2(a) \times B^2(b) \subset \mathbb{R}^2(x_1,y_1) \times \mathbb{R}^2(x_2,y_2)$, or ellipsoids

$$E(a,b) := \left\{ \frac{x_1^2+y_1^2}{a} + \frac{x_2^2+y_2^2}{b} < 1 \right\}.$$

In four dimensions, Gromov's nonsqueezing theorem states that

$$B^4(a) \overset{s}{\hookrightarrow} B^2(b) \times \mathbb{R}^2(x_2,y_2)$$

only if $a \leqslant b$. In other words, one cannot do better than the identity mapping. After this rough rigidity result, the "fine structure of symplectic rigidity" was investigated by looking at other embedding problems. The first important results were on the "packing problem", where U is a disjoint union of balls; see Biran [2; 3], Gromov [15] and McDuff and Polterovich [25]. Further understanding on the fine structure came with the study of embeddings of ellipsoids; see Choi, Cristofaro-Gardiner, Frenkel, Hutchings and Ramos [9], Frenkel and Müller [14], Hutchings [18], McDuff [23; 24], McDuff and Schlenk [26] and Schlenk [27; 28]. Note that $E(a,b) \overset{s}{\hookrightarrow} V$ if and only if $E\left(1, \frac{b}{a}\right) \overset{s}{\hookrightarrow} (1/\sqrt{a})V$. We can thus take $E(1,a)$ with $a \geqslant 1$ as U. Encode the embedding problems $E(1,a) \overset{s}{\hookrightarrow} B^4(b)$ and $E(1,a) \overset{s}{\hookrightarrow} P(b,b) =: C^4(b)$ in the functions

$$c_B(a) := \inf\{\lambda > 0 \mid E(1,a) \overset{s}{\hookrightarrow} B^4(\lambda)\},$$
$$c_C(a) := \inf\{\lambda > 0 \mid E(1,a) \overset{s}{\hookrightarrow} C^4(\lambda)\}.$$

Since symplectic embeddings are volume-preserving, $c_B(a) \geqslant \sqrt{a}$ and $c_C(a) \geqslant \sqrt{\frac{a}{2}}$. The functions $c_B(a)$ and $c_C(a)$ were computed in [26; 14]:

The function $c_B(a)$ has three parts: On $[1,\tau^4]$, with $\tau = \frac{1}{2}(1+\sqrt{5})$ the golden ratio, c_B is given by the "Fibonacci stairs", namely an infinite stairs each of whose steps is made of a segment on a line going through the origin and a horizontal segment, with feet (endpoints) on the volume constraint $\sqrt{a}$, and both the feet and the edge determined by Fibonacci numbers. Then there is one step over $\left[\tau^4, 7\frac{1}{9}\right]$, whose left part over $[\tau^4, 7]$ is affine but nonlinear: $c_B(a) = \frac{a+1}{3}$. Finally, for $a \geqslant 7\frac{1}{9}$ the graph of $c_B(a)$ is given by eight strictly disjoint steps made of two affine segments, and $c_B(a) = \sqrt{a}$ for $a \geqslant 8\frac{1}{36}$. See Figures 1–6 later in the introduction for similar pictures in our setting.

The function $c_C(a)$ has a similar structure: On $[1,\sigma^2]$, with $\sigma = 1+\sqrt{2}$ the silver ratio, c_C is given by the "Pell stairs", namely an infinite stairs each of whose steps is made of a segment on a line going through the origin and a horizontal segment, with feet on the volume constraint $\sqrt{\frac{a}{2}}$, and both the feet and the edge determined by Pell numbers. Then there is one step over $\left[\sigma^2, 6\frac{1}{8}\right]$, whose left part over $[\sigma^2, 6]$ is affine but nonlinear: $c_C(a) = \frac{a+1}{4}$. Finally, for $a \geqslant 6\frac{1}{8}$ the graph of $c_C(a)$ is given by six strictly disjoint steps made of two affine segments, and $c_C(a) = \sqrt{\frac{a}{2}}$ for $a \geqslant 7\frac{1}{32}$.

1.2 Result

We are interested in understanding what happens with the rich structure of the functions c_B and c_C if we take as targets "longer" sets. To this end, we look at the embedding problems $E(1,a) \overset{s}{\hookrightarrow} P(b,c)$ for $c = kb$ with $k \geqslant 2$ an integer, which we encode in the functions

$$c_b(a) := \inf\{\lambda > 0 \mid E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)\}, \quad b \in \mathbb{N}_{\geqslant 2}. \tag{1-1}$$

Note that $c_1 = c_C$. The volume constraint is now $c_b(a) \geqslant \sqrt{\frac{a}{2b}}$. To formulate our result, we define for $b \in \mathbb{N}_{\geqslant 2}$ and for $k \in \{0, 1, 2, \dots, \lfloor \sqrt{2b} \rfloor\}$ the numbers

$$u_b(k) := \frac{(2b+k)^2}{2b} = 2b + 2k + \frac{k^2}{2b}, \quad v_b(k) := 2b\left(\frac{2b+2k+1}{2b+k}\right)^2$$

and

$$\alpha_b := \frac{1}{b}\left(b^2 + 2b + \sqrt{(b^2+2b)^2 - 1}\right), \quad \beta_b := 2b + 4 + \frac{1}{2b(b+1)^2}.$$

Note that $u_b(k) \leqslant 2b + 2k + 1 \leqslant v_b(k)$ with strict inequalities for $k^2 < 2b$ and equalities for $k^2 = 2b$, and that

$$2b + 2k < u_b(k) \leqslant v_b(k) < 2b + 2k + 2 \quad \text{for } k \geqslant 1.$$

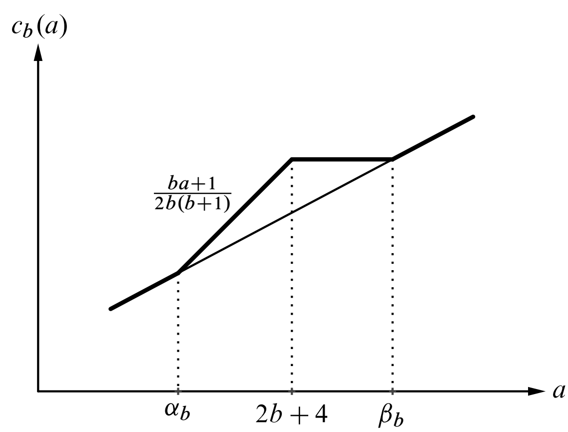


Figure 1: The affine step in our result

Further, $v_b(1) < \alpha_b < 2b+4 < \beta_b < u_b(2)$. The intervals $I_b(k) := [u_b(k), v_b(k)]$ thus have positive length except for $k^2 = 2b$, and the intervals

$$I_b(0),\ \ I_b(1),\ \ [\alpha_b, \beta_b],\ \ I_b(2),\ \ \dots,\ \ I_b(\lfloor\sqrt{2b}\rfloor)$$

are in the right order and are disjoint except that $I_b(0)$ touches $I_b(1)$.

Theorem 1.1 *For every integer $b \geqslant 2$ the function $c_b(a)$ describing the symplectic embedding problem $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ is given by the volume constraint $\sqrt{\frac{a}{2b}}$ except for the following $\lceil\sqrt{2b}\,\rceil + 2$ intervals:*

(i) $c_b(a) = 1$ *if* $a \in [1, 2b]$.

(ii) *For* $k \in \{0, 1, 2, \dots, \lfloor\sqrt{2b}\rfloor\}$ *and on the interval* $I_b(k)$,

$$c_b(a) = \begin{cases} \dfrac{a}{2b+k} & \text{if } a \in [u_b(k), 2b+2k+1], \\[2mm] \dfrac{2b+2k+1}{2b+k} & \text{if } a \in [2b+2k+1, v_b(k)]. \end{cases}$$

(iii) *On the interval* $[\alpha_b, \beta_b]$,

$$c_b(a) = \begin{cases} \dfrac{ba+1}{2b(b+1)} & \text{if } a \in [\alpha_b, 2b+4], \\[2mm] 1 + \dfrac{2b+1}{2b(b+1)} & \text{if } a \in [2b+4, \beta_b]. \end{cases}$$

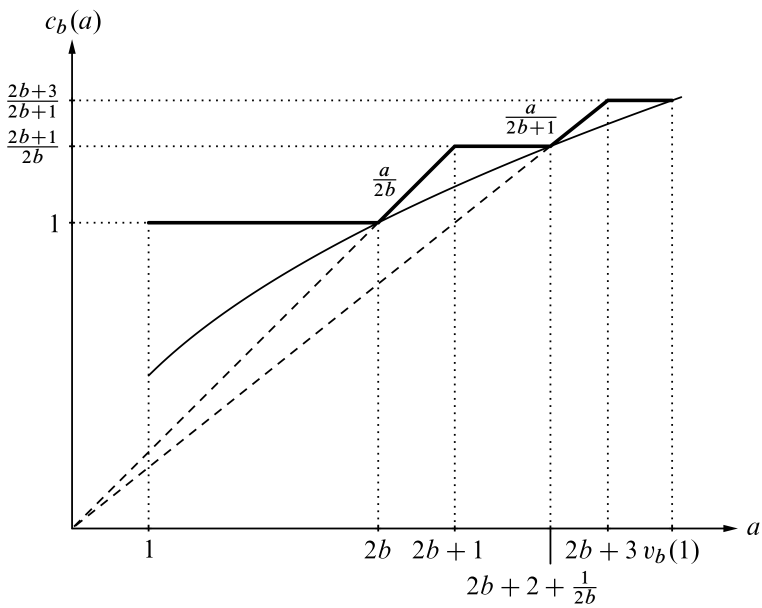


Figure 2: The graph of $c_b(a)$ on $[1, v_b(1)]$

Remarks 1.2 (1) Theorem 1.1 also solves the problem $E(1,a) \overset{s}{\hookrightarrow} E(\lambda, \lambda 2b)$ for integers $b \geqslant 2$, since, for every integer b,

$$E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b) \iff E(1,a) \overset{s}{\hookrightarrow} E(\lambda, \lambda 2b). \tag{1-2}$$

This has been shown by Frenkel and Müller [14, Corollary 1.6] for $b = 1$ by using that ECH capacities provide a complete set of invariants for the embedding problem $E(1,a) \overset{s}{\hookrightarrow} P(b,c)$, and this proof generalizes to all $b \in \mathbb{N}$. In Section 4.1 we shall prove (1-2) by using the "reduction method" (Method 2 of Section 2.2).

(2) One can replace the infimum in definition (1-1) by the minimum. This follows from the previous remark and from the fact that $E(1,a) \overset{s}{\hookrightarrow} E(\lambda, \lambda 2b)$ also for $\lambda = c_b(a)$; see McDuff [23, Corollary 1.6] and also Cristofaro-Gardiner [11, Corollary 1.6] for a generalization. Altogether, we see that

$$E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b) \iff E(1,a) \overset{s}{\hookrightarrow} E(\lambda, \lambda 2b) \iff \lambda \geqslant c_b(a).$$

Geometric description of the result We proceed with describing the functions $c_b(a)$ given in Theorem 1.1 more geometrically. The left part of the steps described in part (ii) of the theorem lie on a line passing through the origin, while the left part of the step described in part (iii) lies on a line crossing the y–axis at $\frac{1}{2b(b+1)}$. We call the steps in (ii) the "linear steps" and the step in (iii) the "affine step".

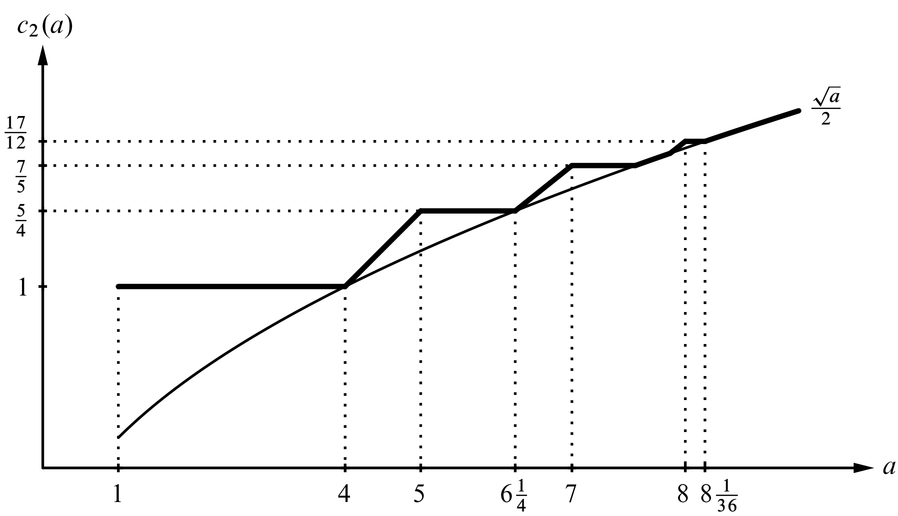


Figure 3: The graph of $c_2(a)$

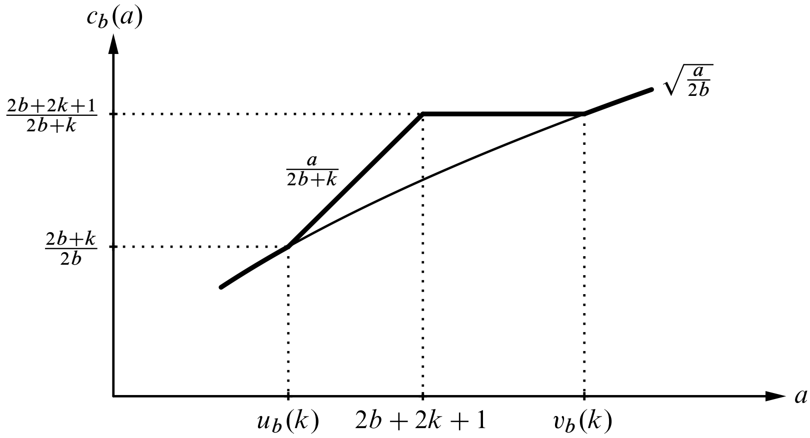


Figure 4: One of the $\lceil \sqrt{2b}\, \rceil$ linear steps

The graph of $c_b(a)$ on $[1, v_b(1)]$ is given by

$$c_b(a) = \begin{cases} 1 & \text{if } a \in [1, 2b], \\ \dfrac{a}{2b} & \text{if } a \in [2b, 2b+1], \\ \dfrac{2b+1}{2b} & \text{if } a \in \left[2b+1, 2b+2+\dfrac{1}{2b}\right], \\ \dfrac{a}{2b+1} & \text{if } a \in \left[2b+2+\dfrac{1}{2b}, 2b+3\right], \\ \dfrac{2b+3}{2b+1} & \text{if } a \in \left[2b+3, 2b+4-\dfrac{4}{(2b+1)^2}\right]; \end{cases}$$

see Figure 2. This part of the graph touches the volume constraint only in three points. Then follows a "volume interval", and then the affine step described in part (iii) and Figure 1. For $b = 2$ there are no further obstructions (Figure 3), but for $b \geqslant 3$ there are $\lceil \sqrt{2b}\, \rceil - 2$ more linear steps, that are strictly disjoint and made of a linear and a horizontal segment (Figures 4 and 5).

The length of the affine step is

$$\beta_b - \alpha_b < \beta_b - v_b(1) = \frac{1}{2b(b+1)^2} + \frac{4}{(2b+1)^2},$$

and hence this step becomes very small for b large. The length of the k^{th} linear step is

$$\ell_b(k) := v_b(k) - u_b(k) = (2b - k^2)\frac{8b^2 + k^2 + (2+8k)b}{2b(2b+k)^2}.$$

For fixed b, the function $\ell_b(k)$ is strictly decreasing, with $\ell_b(\sqrt{2b}) = 0$. For fixed k, however, $\lim_{b\to\infty} \ell_b(k) = 2$. More precisely, $\ell_b(0)$ is strictly decreasing to 2, and

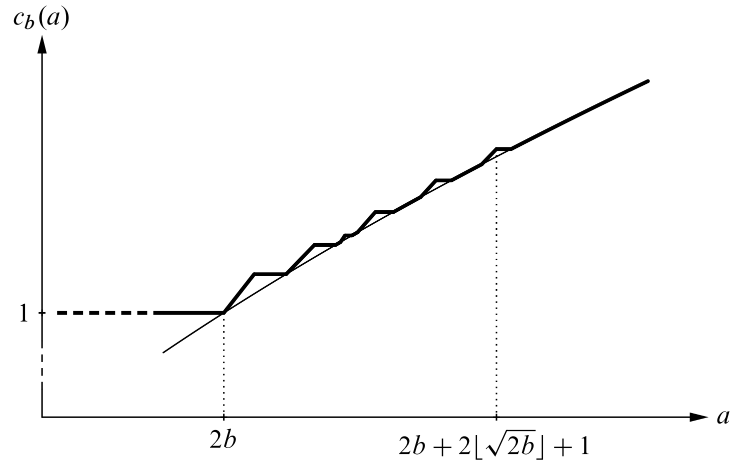


Figure 5: The graph of $c_9(a)$

$\ell_b(k)$ is strictly increasing to 2 for every $k \geqslant 1$. Since the edge of the k^{th} step is at $2b+2k+1$, we see that, for $b \to \infty$, an arbitrarily large (but fixed) part of the graph of $c_b(a)$ consists of linear steps of length almost 2, which almost form a connected staircase (Figure 6).

We reformulate this behaviour of $c_b(a)$ for large b in terms of a rescaled limit function: Consider the rescaled functions

$$\hat{c}_b(a) = 2bc_b(a+2b) - 2b, \quad a \geqslant 0,$$

which are obtained from $c_b(a)$ by first forgetting about the horizontal line $c_b(a) = 1$ over $[1, 2b]$ that comes from the nonsqueezing theorem, then vertically rescaling by $2b$, and finally translating the graph by the vector $(-2b, -2b)$. Further, consider the

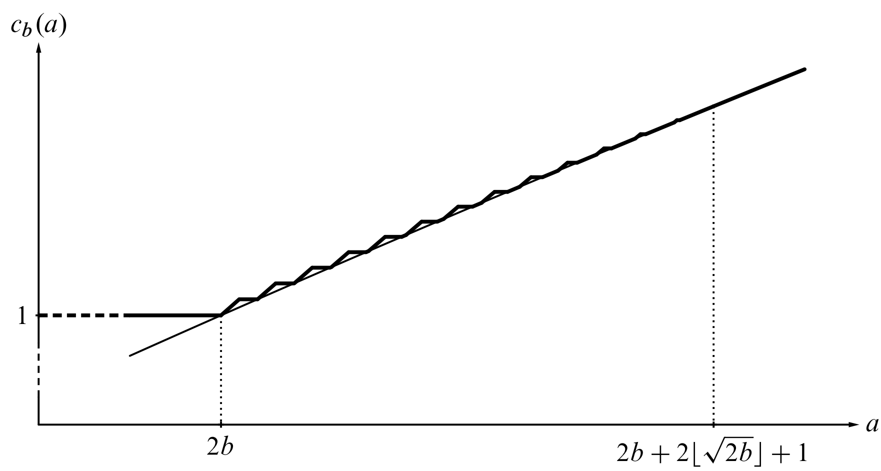


Figure 6: The graph of $c_{85}(a)$

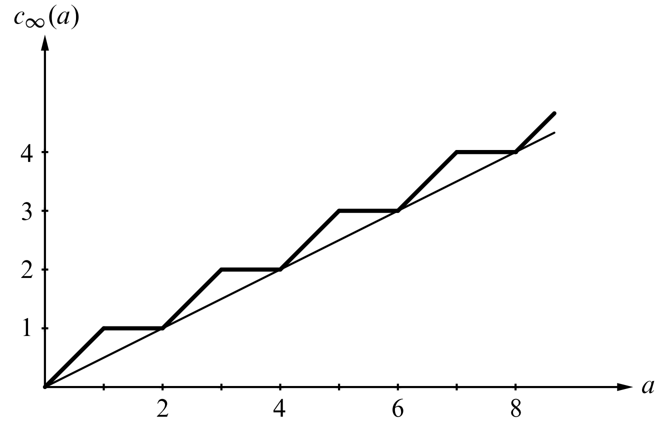


Figure 7: The graph of the rescaled limit function $c_\infty(a)$

function $c_\infty\colon [0,\infty) \to \mathbb{R}$ drawn in Figure 7; its graph consists of infinitely many steps of width 2 and slope 1 that are based at the line $\frac{a}{2}$. Then

$$\lim_{b\to\infty} \hat{c}_b(a) = c_\infty(a), \quad a \in [0,\infty), \tag{1-3}$$

uniformly on bounded sets. Indeed, applying the same rescaling to $\sqrt{\frac{a}{2b}}$ yields $2b\sqrt{(a+2b)/2b} - 2b$, which is $\frac{a}{2} + O\big(\frac{a^2}{2b}\big)$ for $b \geqslant a$. One can also check that $\hat{c}_b(a)$ is increasing to $c_\infty(a)$ for all a.

1.3 Interpretation

Recall from the introduction that the graph of $c_C(a)$ has three parts: First the infinite Pell stairs, then one affine step, and then six more steps.

If we take $b=1$ in the above description of $c_b(a)$ on $[1, v_b(1)]$, we exactly obtain $c_C(a)$ on $[1, v_1(1)] = \big[1, \frac{50}{9}\big]$. Further, if we take $b=1$ in the description (iii) of the affine step of $c_b(a)$, we exactly obtain the affine step of $c_C(a)$ over $\big[\sigma^2, 6\frac{1}{8}\big]$. Hence $c_b(a)$ generalizes $c_C(a)$ on the first two steps and on the affine step. This is not a coincidence. Indeed, the two exceptional classes giving rise to the first two steps of the Pell stairs are the first two in the sequence (1-4) of exceptional classes E_n giving rise to all the linear steps of $c_b(a)$, and the exceptional class giving rise to the affine step of $c_C(a)$ is the first in a sequence of exceptional classes F_b giving rise to the affine step in $c_b(a)$; see Section 3.

On the other hand, the remaining infinitely many steps of the Pell stairs have no counterpart for $b \geqslant 2$. Similarly, the linear steps described in Theorem 1.1(ii) are more regular than the affine steps on the right part of $c_C(a)$, none of which consists of a linear and a horizontal segment. We thus see that the first two steps and the affine step of $c_C(a) = c_1(a)$ are stable under the deformations of b we consider, while the other steps are not.

By Theorem 1.1, $c_b(a)$ equals the volume constraint $\sqrt{\frac{a}{2b}}$ for $a \geqslant v_b(\lfloor\sqrt{2b}\rfloor) = 2b + O(\sqrt{b})$, that is, there are no packing obstructions for the embedding problem $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ for a sufficiently large. This is not a surprise. Indeed, this phenomenon was already observed for the embedding problems $E(1,a) \overset{s}{\hookrightarrow} B^4(b)$ and $E(1,a) \overset{s}{\hookrightarrow} C^4(b)$, and it fits well with previous results: It is known for many closed connected symplectic manifolds (M,ω) that there is a number $N(M,\omega)$ such that (M,ω) admits a full symplectic packing by k equal balls for every $k \geqslant N(M,\omega)$ ("packing stability"; see Biran [2; 3], Buse and Hind [5; 6], Buse, Hind and Opshtein [7] and Buse and Pinsonnault [8]). Similarly, an explicit construction implies that for any connected symplectic manifold (M,ω) of finite volume, the proportion of the volume that can be filled by a dilate of the ellipsoid $E(1,\dots,1,a)$ tends to 1 as $a \to \infty$; see Schlenk [28, Section 6]: The packing obstruction tends to zero as the *domain* is more and more elongated.

Theorem 1.1 exhibits a different phenomenon: If in the problem $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ the *target* is elongated ($b \to \infty$), then the regular Pell stairs in the graph of $c_1(a)$ first almost disappears (only two linear steps and the affine step remain), but then for large b the graph of $c_b(a)$ reorganizes to a staircase that asymptotically is infinite and completely regular.

1.4 Stabilization and connection with symplectic folding

Let a, $b \geqslant 1$ be real numbers. Following Cristofaro-Gardiner and Hind [12] we consider for each $N \geqslant 3$ the stabilized problem

$$c_b^N(a) := \inf\{\lambda > 0 \mid E(1,a) \times \mathbb{C}^{N-2} \overset{s}{\hookrightarrow} P(\lambda, \lambda b) \times \mathbb{C}^{N-2}\}.$$

Then $c_b^N(a) \leqslant c_b(a)$.

Lemma 1.3 *For every $N \geqslant 3$ and all real numbers a, $b \geqslant 1$,*

$$c_b^N(a) \leqslant f_b(a) := \frac{2a}{a+2b-1}.$$

Proof Set $\mu = \frac{a(2b-1)}{a+2b-1}$ and $\lambda = 2\left(1 - \frac{\mu}{a}\right)$. Then $\mu + \frac{\lambda}{2} = b\lambda$. Since $b \geqslant 1$ we have $\mu \geqslant \frac{\lambda}{2}$. Note that $\frac{\lambda}{2} = 1 - \frac{\mu}{a}$ is the area of a z_2–disc in $E(a,1)$ over a point z_1 on the boundary of the disc $D(\mu)$ of area μ. Applying Hind's folding construction in [16, Section 2] with μ — instead of $\frac{S}{S+1}$ — we obtain for every $\varepsilon > 0$ a symplectic embedding

$$E(1,a) \times \mathbb{C} \overset{s}{\hookrightarrow} P\Big(\mu + \frac{\lambda}{2} + \varepsilon, 2\frac{\lambda}{2} + \varepsilon\Big) \times \mathbb{C}.$$

Now recall that $\mu + \frac{\lambda}{2} = b\lambda$ and note that $\lambda = f_b(a)$. □

In view of the above proof, we call the graph of $f_b(a)$ the folding curve. Now note that

$$f_b(2b + 2k + 1) = \frac{2b+2k+1}{2b+k}, \quad k \geqslant 0.$$

For $b \in \mathbb{N}$ this is also the value of c_b at the edge points of the k^{th} linear step. In other words, the linear steps oscillate between the volume constraint $\sqrt{\frac{a}{2b}}$ and the folding curve; see Figures 4 and 8.

Conjecture 1.4 *The edge points of the linear steps are stable, in the sense that at these points we have $c_b^N = c_b$ for all $N \geqslant 3$.*

This conjecture is based on the main result of [12], where it is shown that the edge points of the Fibonacci stairs for the problem $E(a,1) \overset{s}{\hookrightarrow} B^4(\lambda)$ are stable. It is likely that one can prove it by a similar method as in [12]; see also the discussion at the end of the next section. A proof of Conjecture 1.4 is not the concern of the present work, but a positive answer would imply that the folding construction in the proof of Lemma 1.3 is sharp at the edge points of the linear steps.

Recall that $c_b(a) = 1$ for $a \in [1, 2b]$. As we shall see in Proposition 3.5(ii), $c_b(a) = \sqrt{\frac{a}{2b}}$ for all $a \geqslant (\sqrt{2b}+1)^2$ and all real $b \geqslant 2$. Now notice that $f_b(a) \geqslant \sqrt{\frac{a}{2b}}$ if and only if $a \in [(\sqrt{2b}-1)^2, (\sqrt{2b}+1)^2]$. It follows that

$$c_b^N(a) < c_b(a) \quad \text{if } a \notin [2b-1, (\sqrt{2b}+1)^2]$$

for all $b \geqslant 2$ and $N \geqslant 3$.

We finally notice that under the rescaling yielding the limit function $c_\infty(a)$, we have $\widehat{f}_b(a) = 2bf_b(a+2b) - 2b = \frac{2b(a+1)}{a+4b-1}$, and so

$$f_\infty(a) := \lim_{b \to \infty} \widehat{f}_b(a) = \frac{a+1}{2}.$$

This means that also the limit function c_∞ oscillates, between the limit function $\frac{a}{2}$ of the volume constraint $\sqrt{\frac{a}{2b}}$ and the limit function $\frac{a+1}{2}$ of the folding curve.

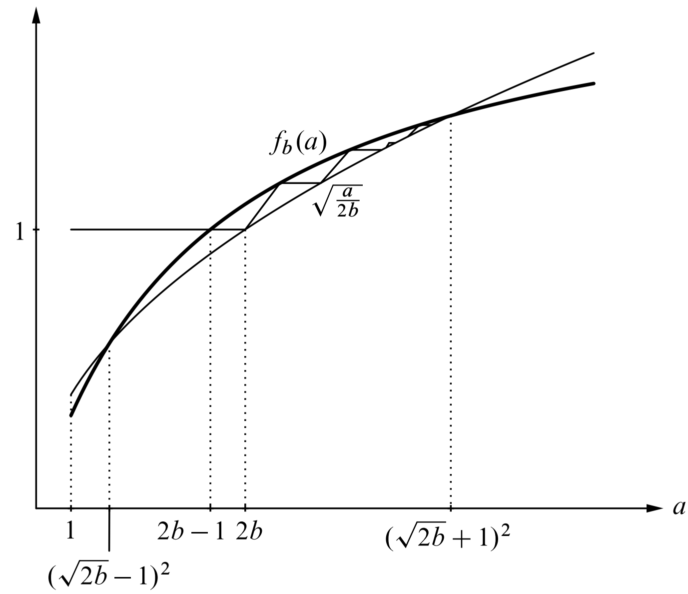


Figure 8: The volume constraint, $c_b(a)$, and the folding curve for $b=5$

1.5 Method

In principle, there are two methods to prove Theorem 1.1: The first method (Method 1 in Section 2.2, which was used by Frenkel and Müller [14] and McDuff and Schlenk [26]) is to find the strongest obstruction for the embedding problem $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ coming from exceptional classes (ie homology classes in a certain multiple blow-up of $\mathbb{C}\mathrm{P}^2$ represented by embedded J–holomorphic -1 spheres). The second method (Method 2 in Section 2.2, that was first used by Buse and Pinsonnault [8]) is a cohomological version of the first method: One associates to a hypothetical embedding $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ a cohomology class, and checks whether this class transforms to a "reduced vector" under Cremona transforms. While the first method is sufficient for solving the problems $E(1,a) \overset{s}{\hookrightarrow} B^4(\lambda)$ and $E(1,a) \overset{s}{\hookrightarrow} C^4(\lambda)$ — see [14; 26] — it does not lead to a proof of the entire Theorem 1.1, because the known upper bound for the number of obstructive exceptional classes tends to infinity with b. On the other hand, Method 2 does yield a proof of Theorem 1.1, as will become clear from our proof. We shall not follow such a puristic approach, however, but an opportunistic one, which uses both methods: Given b, we first write down a finite set of exceptional classes that yield embedding obstructions, namely $E_0=(1,0;1)$ and

$$\begin{aligned} E_n &:= (n,1;1^{\times(2n+1)}), \quad n=b,\dots,b+\lfloor\sqrt{2b}\rfloor, \\ F_b &:= (b(b+1), b+1; b+1, b^{\times(2b+3)}) \end{aligned} \tag{1-4}$$

(see Section 2.2 for the notation), and then use Method 2 to show that the obstruction $h_b(a)$ given by these classes is complete. In other words, we use Method 1 to show that $c_b(a) \geqslant h_b(a)$ and Method 2 to show that $c_b(a) \leqslant h_b(a)$ — with the exception that for a large and for $b = 2$ and $a \in \left[8\frac{1}{36}, 9\right]$ we use Method 1 to show that $c_b(a)$ equals the volume constraint $\sqrt{\frac{a}{2b}}$.

This hybrid approach yields the shortest proof of Theorem 1.1 we know. Further, knowing a set of exceptional classes that provide all embedding obstructions is interesting for at least two reasons: First, the holomorphic spheres underlying these classes provide a geometric explanation of the graphs of the functions $c_b(a)$. Second, one should be able to use these holomorphic spheres to prove Conjecture 1.4; it is probably the case that one can find the needed obstructions by stretching these spheres and then "stabilizing" as in Cristofaro-Gardiner and Hind [12] and Hind and Kerman [17].

1.6 Outlook

Our ultimate goal is to see the *continuous* film of graphs $c_b(a)$ for $b \geqslant 1$ real. It would be particularly interesting to understand this film for $b \in [1, 2]$, or just for $b \in [1, 1+\varepsilon]$ for some $\varepsilon > 0$, namely to understand how the Pell stairs disappear. In Burkhart, Panescu and Timmons [4], ECH capacities are used to compute $c_b(a)$ for $b = \frac{13}{2}$ and to get an idea of this film. In accordance with Theorem 1.1, Conjecture 6.3 in [4] and further investigations we make:

Conjecture 1.5 *For any real $b \geqslant 2$ the function $c_b(a)$ is given by the maximum of the volume constraint $\sqrt{\frac{a}{2b}}$ and the obstructions coming from the exceptional classes E_n and F_n in* (1-4).

The obstructions given by the exceptional classes E_n and F_n are readily computed; see Section 3.3: While the classes E_n again give rise to a finite staircase with linear steps, the classes F_n give an obstruction only for $b \in \left(n - \frac{n}{(n+1)^2}, n + \frac{1}{n+2}\right)$. While our proof of Theorem 1.1 should extend to a proof of Conjecture 1.5, the analysis is more involved, since fractional parts arise, that are harder to estimate.

Our only definite result for b real is that for every real $b \geqslant 2$ we have $c_b(a) = \sqrt{\frac{a}{2b}}$ for all $a \geqslant (\sqrt{2b} + 1)^2$; see Proposition 3.5(ii).

Acknowledgements We cordially thank Dusa McDuff, who already in 2010 suggested to us to use the reduction method to analyze the embedding problem $E(1, a) \overset{s}{\hookrightarrow} C^4(\lambda)$. Cristofaro-Gardiner was partially supported by NSF grant DMS-1402200. Schlenk was partially supported by SNF grant 200021-163419.

2 Methods of proof

In this section we describe the methods we will use in the proof of Theorem 1.1. For more details we refer to the surveys [10; 19; 29].

2.1 Translation to a ball packing problem

Fix $b \geqslant 1$. Since the function $c_b(a)$ is continuous in a, it suffices to compute $c_b(a)$ for $a \geqslant 1$ rational. The *weight expansion* $\boldsymbol{w}(a)$ of such an a is the finite decreasing sequence

$$\boldsymbol{w}(a) := (\underbrace{1,\dots,1}_{\ell_0}, \underbrace{w_1,\dots,w_1}_{\ell_1}, \dots, \underbrace{w_N,\dots,w_N}_{\ell_N}) \equiv (1^{\times \ell_0}, w_1^{\times \ell_1}, \dots, w_N^{\times \ell_N}) \tag{2-1}$$

such that $w_1 = a - \ell_0 < 1$, $w_2 = 1 - \ell_1 w_1 < w_1$, and so on. For example, $a = \frac{25}{9}$ has weight expansion $\boldsymbol{w}(a) = \left(1, 1, \frac{7}{9}, \frac{2}{9}, \frac{2}{9}, \frac{2}{9}, \frac{1}{9}, \frac{1}{9}\right) \equiv \left(1^{\times 2}, \frac{7}{9}, \frac{2}{9}^{\times 3}, \frac{1}{9}^{\times 2}\right)$.

Write $B(\boldsymbol{w}(a))$ for the disjoint union of balls $B(1) \amalg \cdots \amalg B(w_N)$ whose weights are those appearing in $\boldsymbol{w}(a)$, with multiplicities. Based on [23] it was shown in [14, Proposition 1.4] that $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ if and only if

$$B(\boldsymbol{w}(a)) \amalg B(\lambda) \amalg B(\lambda b) \overset{s}{\hookrightarrow} B(\lambda(b+1)); \tag{2-2}$$

compare the moment map picture on the left of Figure 9.

2.2 Three translations to a combinatorial problem

In order to reformulate problem (2-2), we look at the general ball packing problem

$$\coprod_{i=1}^{n} B(a_i) \overset{s}{\hookrightarrow} B(\mu). \tag{2-3}$$

We shall describe three combinatorial solutions of (2-3).

Denote by X_n the n–fold complex blow-up of $\mathbb{C}\mathrm{P}^2$, endowed by the orientation induced by the complex structure. Its homology group $H_2(X_n;\mathbb{Z})$ has the canonical basis $\{L, E_1, \dots, E_n\}$, where $L = [\mathbb{C}\mathrm{P}^1]$ and the E_i are the classes of the exceptional divisors. The Poincaré duals of these classes are denoted by $\ell, e_1, \dots, e_n$. Let $K := -3L + \sum_{i=1}^n E_i$ be the Poincaré dual of $-c_1(X_n)$, and consider the K–symplectic cone $\mathcal{C}_K(X_n) \subset H^2(X_n;\mathbb{R})$, namely the set of cohomology classes that can be represented by symplectic forms ω on X_n that are compatible with the orientation of X_n and have first Chern class $c_1(\omega) = c_1(X_n) = \mathrm{PD}(-K)$. Denote by $\bar{\mathcal{C}}_K(X_n)$ its closure in $H^2(X_n;\mathbb{R})$.

McDuff and Polterovich [25] proved that an embedding (2-3) exists if and only if

$$\mu\ell - \sum_{i=1}^{n} a_i e_i \in \bar{\mathcal{C}}_K(X_n).$$

We thus need to describe $\bar{\mathcal{C}}_K(X_n)$. For this consider the set $\mathcal{E}_K(X_n) \subset H_2(X_n;\mathbb{Z})$ of classes E with $-K\cdot E = c_1(E) = 1$ and $E \cdot E = -1$ that can be represented by smoothly embedded spheres. Li and Liu [22] characterized $\bar{\mathcal{C}}_K(X_n)$ as

(2-4) $$\bar{\mathcal{C}}_K(X_n) = \{\alpha \in H^2(X_n;\mathbb{R}) \mid \alpha^2 \geqslant 0 \text{ and } \alpha(E) \geqslant 0 \text{ for all } E \in \mathcal{E}_K(X_n)\}.$$

We thus need to describe $\mathcal{E}_K(X_n)$. For this define, for $n \geqslant 3$, the *Cremona transform* Cr: $\mathbb{R}^{1+n} \to \mathbb{R}^{1+n}$ as the linear map taking $(x_0; x_1, \dots, x_n)$ to

(2-5) $$(2x_0 - x_1 - x_2 - x_3;\ x_0 - x_2 - x_3,\ x_0 - x_1 - x_3,\ x_0 - x_1 - x_2,\ x_4,\ \dots,\ x_n).$$

A vector $(x_0; x_1, \dots, x_n)$ is *ordered* if $x_1 \geqslant \cdots \geqslant x_n$. The *standard Cremona move* takes an ordered vector $(x_0; \boldsymbol{x})$ to the vector obtained by ordering $\mathrm{Cr}(x_0; \boldsymbol{x})$. More generally, a *Cremona move* is a Cremona transform followed by any permutation of the components of $\boldsymbol{x}$.

For later use we recall the geometric origin of Cr and of Cremona moves. For any nonzero vector u in an inner-product space, the map $r_u(x) = x - 2\frac{\langle u,x\rangle}{\langle u,u\rangle}u$ is the reflection about u, and hence an involution. Similarly, for a class $A \in H_2(X_n;\mathbb{R})$ with $A\cdot A \neq 0$ the map $r_A(B) = B - 2\frac{A\cdot B}{A\cdot A}A$ is an involution of $H_2(X_n;\mathbb{R})$. For $|A\cdot A| \in \{1,2\}$, this map is also an automorphism of $H_2(X_n;\mathbb{Z})$. Now take the classes $A_0 = L - E_1 - E_2 - E_3$ and $A_{ij} = E_i - E_j$ for $1 \leqslant i < j \leqslant n$. Their self-intersection number is -2 and so, for these classes,

(2-6) $$r_A(B) = B + (A\cdot B)A.$$

With respect to the basis $\{L, E_1, \dots, E_n\}$ we have that r_{A_0} is given by (2-5), that is, $r_{A_0} = \mathrm{Cr}\colon \mathbb{Z}^{1+n} \to \mathbb{Z}^{1+n}$ takes the integral vector $(d; \boldsymbol{m}) = (d; m_1, \dots, m_n)$ to

(2-7) $$(2d - m_1 - m_2 - m_3;\ d - m_2 - m_3,\ d - m_1 - m_3,\ d - m_1 - m_2,\ m_4,\ \dots,\ m_n),$$

and $r_{A_{ij}}$ is the transposition τ_{ij} interchanging the i^{th} and j^{th} coordinate. These involutions of $H_2(X_n;\mathbb{Z})$ are induced by orientation-preserving diffeomorphisms of X_n. This is clear for τ_{ij} (lift to X_n an isotopy of $\mathbb{C}\mathrm{P}^2$ interchanging holomorphically small discs around the i^{th} and j^{th} blow-up points), and it holds for all classes A_0 and A_{ij} because each of them can be represented by a smoothly embedded sphere S, and the smooth version of the Dehn–Seidel twist along S [30] is a diffeomorphism inducing (2-6), in view of the Picard–Lefschetz formula [1, page 26]. Since the maps

Cr and τ_{ij} preserve both the intersection product on $H_2(X_n;\mathbb{Z})$ and the class K, they preserve the set $\mathcal{E}_K(X_n)$.

Based on [21; 22] it was shown in [26, Proposition 1.2.12] that a homology class $E = dL - \sum_{i=1}^n m_i E_i$ belongs to $\mathcal{E}_K(X_n)$ if and only if the vector $(d;\boldsymbol{m}) = (d; m_1,\dots,m_n)$ is equal to $(0;-1,0,\dots,0)$ up to a permutation of the m_i, or if $(d;\boldsymbol{m}) \in \mathbb{N} \cup (\mathbb{N}\cup\{0\})^n$ satisfies the Diophantine system

$$\sum_{i=1}^n m_i = 3d-1, \quad \sum_{i=1}^n m_i^2 = d^2+1, \tag{2-8}$$

and reduces to $(0;-1,0,\dots,0)$ under repeated standard Cremona moves. Summarizing, we find:

Method 1 (obstructive classes) *An embedding* (2-3) *exists if and only if* $\sum_{i=1}^n a_i^2 \leqslant \mu^2$ *and* $\sum_{i=1}^n a_i m_i \leqslant \mu d$ *for all vectors* $(d;\boldsymbol{m})$ *of nonnegative integers satisfying* (2-8) *and reducing to* $(0;-1,0,\dots,0)$ *under repeated standard Cremona moves.*

Remark 2.1 It is shown in [23] (see also [19]) that (2-3) is also equivalent to $\sum_{i=1}^n a_i m_i \leqslant \mu d$ *for all vectors* $(d;\boldsymbol{m})$ *of nonnegative integers satisfying the Diophantine system* (2-8). It follows that if we use exceptional classes only to give *lower* bounds for $c_b(a)$ (as we do in this paper), then we do not need to show that these classes reduce to $(0;-1,0,\dots,0)$ under repeated standard Cremona moves. We shall nevertheless perform these reductions, since they are readily done (see Section 3.2) and since we wish to know explicit exceptional classes responsible for the embedding obstructions beyond the volume constraint.

In view of (2-2) we find that $E(1,a) \overset{s}{\hookrightarrow} P(\lambda,\lambda b)$ if and only if $\lambda \geqslant \sqrt{\frac{a}{2b}}$ and

$$\lambda(b+1) \geqslant \lambda(bm_1+m_2) + m_3 w_1 + \cdots + m_{k+2} w_k \tag{2-9}$$

for all vectors $(d;\boldsymbol{m})$ of nonnegative integers satisfying (2-8) with $n=k+2$ and reducing to $(0;-1,0,\dots,0)$ under repeated standard Cremona moves.

Condition (2-9) is not handy, since λ appears on both sides. We thus better work directly in $P(\lambda,\lambda b)$ or in its compactification $S^2\times S^2$ endowed with the product symplectic form of the same volume. Let Y_{k+1} be the complex blow-up of $S^2\times S^2$ in $k+1$ points. Then the classes $S_1 = [S^2\times \mathrm{pt}]$ and $S_2 = [\mathrm{pt}\times S^2]$ and the classes $F_1,\dots,F_{k+1}$ of the exceptional divisors form a basis of $H_2(Y_{k+1})$. As one can guess from the picture on the right of Figure 9, there exists a diffeomorphism $\psi\colon Y_{k+1}\to X_{k+2}$ such that the induced map $\psi_*\colon H_2(Y_{k+1}) \to H_2(X_{k+2})$ is given by

$$\begin{aligned}
S_1 &\mapsto L - E_1, \\
S_2 &\mapsto L \qquad - E_2, \\
F_1 &\mapsto L - E_1 - E_2, \\
F_i &\mapsto \qquad\qquad - E_{i+1}, \quad i \geqslant 2.
\end{aligned}$$

If we write $(d, e; m_1, \dots, m_{k+1})$ for $dS_1 + eS_2 - m_1 F_1 - \dots - m_{k+1} F_{k+1}$, we thus have

$$\psi_*(d, e; \boldsymbol{m}) = (d + e - m_1;\ d - m_1,\ e - m_1,\ m_2,\ \dots, m_{k+1}). \tag{2-10}$$

Given vectors $\boldsymbol{u} \in \mathbb{R}^{n_1}$ and $\boldsymbol{v} \in \mathbb{R}^{n_2}$ we write $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \sum_{i=1}^{\max(n_1,n_2)} u_i v_i$. In the basis S_1, S_2, $F_1, \dots, F_{k+1}$, we can reformulate Method 1 as:

Method 1$'$ (obstructive classes) *An embedding $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ exists if and only if $\lambda \geqslant \sqrt{\frac{a}{2b}}$ and*

$$\lambda \geqslant \frac{\langle \boldsymbol{m}, \boldsymbol{w}(a) \rangle}{d + be} =: \mu_b(d, e; \boldsymbol{m})(a) \tag{2-11}$$

for all vectors $(d, e; \boldsymbol{m})$ of nonnegative integers that satisfy the Diophantine system

$$\sum m_i = 2(d + e) - 1, \quad \sum m_i^2 = 2de + 1 \tag{2-12}$$

and for which $\psi_*(d, e; \boldsymbol{m})$ reduces to $(0; -1, 0, \dots, 0)$ under repeated standard Cremona moves.

For the detailed translation of Method 1 to Method 1$'$ we refer to the proof of [14, Proposition 3.9]. As we shall see in Section 3, the obstructions to embeddings $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ beyond the volume (that is, the steps in the graphs $c_b(a)$) are all given by the two series of exceptional classes $(d, e, \boldsymbol{m})$

$$\begin{aligned}
E_n &:= (n, 1; 1^{\times(2n+1)}), \\
F_n &:= (n(n+1), n+1; n+1, n^{\times(2n+3)}).
\end{aligned} \tag{2-13}$$

In Method 1, the Cremona moves acted on integral homology classes $(d; \boldsymbol{m})$. The second method applies Cremona moves to real cohomology classes α, and verifies by a finite algorithm whether $\alpha \in \overline{\mathcal{C}}_K(X_n)$.

For convenience, we write $(\mu; a_1, \dots, a_n)$ instead of $\mu\ell - \sum_{i=1}^n a_i e_i$. Recall that the Cremona transform Cr on $H_2(X_n; \mathbb{Z})$ is induced by an orientation-preserving diffeomorphism φ of X_n. Since $\mathrm{Cr} = \varphi_*$ is an involution, the map φ^* induced on cohomology $H^2(X_n;\mathbb{R})$ is also given by formula (2-5) with respect to the Poincaré dual basis $\{\ell, e_1, \dots, e_n\}$, that is, $\varphi^* = \mathrm{Cr}\colon \mathbb{R}^{1+n} \to \mathbb{R}^{1+n}$ takes the vector $(\mu; a_1, \dots, a_n)$ to

$$(2\mu - a_1 - a_2 - a_3;\ \mu - a_2 - a_3,\ \mu - a_1 - a_3,\ \mu - a_1 - a_2,\ a_4,\ \dots,\ a_n). \tag{2-14}$$

Call an ordered vector $(\mu; a_1, \dots, a_n)$ *reduced* if $\mu \geqslant a_1 + a_2 + a_3$. Using the characterisation (2-4) and building on [21; 22], Buse and Pinsonnault [8, Section 2.3] and Karshon and Kessler [20, Section 6.3] designed the following algorithm to decide whether an embedding (2-3) exists:

Method 2 (reduction at a point) *Let $\alpha = (\mu; a_1, \dots, a_n)$ be an ordered vector with $\mu \geqslant 0$ and $\alpha^2 \geqslant 0$. The sequence obtained from applying to α standard Cremona moves contains a reduced vector. Let $(\widehat{\mu}; \widehat{a}_1, \dots, \widehat{a}_n)$ be the first reduced vector in this sequence. Then $\alpha \in \overline{\mathcal{C}}_K(X_n)$ if and only if $\widehat{a}_1, \dots, \widehat{a}_n \geqslant 0$.*

We shall only need the "if" part of this equivalence. In fact, we shall use a version thereof that will permit us to avoid finding the reordering after each Cremona transform:

Proposition 2.2 *Let $\alpha = (\mu; a_1, \dots, a_n)$ be a vector with $\mu \geqslant 0$ and $\alpha^2 \geqslant 0$, and assume that there is a sequence $\alpha = \alpha_0, \alpha_1, \dots, \alpha_m$ of vectors such that α_{j+1} is obtained from α_j by a Cremona move. If $\alpha_m = (\widehat{\mu}; \widehat{a}_1, \dots, \widehat{a}_n)$ is reduced and $\widehat{a}_1, \dots, \widehat{a}_n \geqslant 0$, then $\alpha \in \overline{\mathcal{C}}_K(X_n)$.*

Proof According to [22, Proposition 4.9(3)], a reduced vector with nonnegative coefficients belongs to $\overline{\mathcal{C}}_K(X_n)$. Hence $\alpha_m \in \overline{\mathcal{C}}_K(X_n)$. By assumption we have $\alpha_m = (\pi \circ \mathrm{Cr})(\alpha_{m-1})$, where π is a coordinate permutation of $\mathbb{R}^n$. Write π as a product $\tau_s \circ \dots \circ \tau_1$ of transpositions. Since Cr and τ_i are involutions,

$$\alpha_{m-1} = (\mathrm{Cr} \circ \tau_1 \circ \dots \circ \tau_s)(\alpha_m).$$

Recall that Cr and τ_i preserve the set $\mathcal{E}_K(X_n)$. By (2-4), these maps also preserve $\overline{\mathcal{C}}_K(X_n)$. Thus $\alpha_{m-1} \in \overline{\mathcal{C}}_K(X_n)$. Iterating this argument yields $\alpha = \alpha_0 \in \overline{\mathcal{C}}_K(X_n)$. □

It turns out that for transforming a (reducible) vector to a reduced vector by Cremona moves, it is best to reorder every vector in the process. In our reduction schemes in Sections 5–8 we will usually do this, but not always, to avoid distinguishing even more cases. The point of Proposition 2.2 is that even when we do restore the order of a vector, we do not need to prove this, except for the head of the last vector: All we need to make sure is that we eventually arrive at a vector $(\widehat{\mu}; \widehat{a}_1, \widehat{a}_2, \widehat{a}_3, \widehat{a}_4, \dots)$ that is reduced and has $\widehat{a}_j \geqslant 0$ for all j, ie is such that

$$\min\{\widehat{a}_1, \widehat{a}_2, \widehat{a}_3\} \geqslant \max\{\widehat{a}_4, \dots, \widehat{a}_n\}, \qquad \widehat{\mu} \geqslant \widehat{a}_1 + \widehat{a}_2 + \widehat{a}_3, \qquad \widehat{a}_j \geqslant 0 \quad \text{for all } j.$$

On the other hand, we will always immediately check in each step that the new coefficients are nonnegative, since otherwise we may easily forget checking a coefficient at the end.

Recall that an embedding $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ exists if and only if an embedding (2-2) exists. Together with Proposition 2.2 we find the following recipe:

Proposition 2.3 *An embedding $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ exists if there exists a finite sequence of Cremona moves that transforms the vector* (4-1) *to an ordered vector with nonnegative entries and defect $\delta \geqslant 0$.*

In our applications of this proposition we will have $\lambda \in (1,2)$. The first Cremona transform thus maps

$$((b+1)\lambda;\, b\lambda,\, \lambda,\, 1^{\times \lfloor a \rfloor},\, w_1^{\times \ell_1}, \dots)$$

with $\delta = -1$ to the vector

$$((b+1)\lambda - 1;\, b\lambda - 1,\, \lambda - 1,\, 0,\, 1^{\times(\lfloor a \rfloor - 1)},\, w_1^{\times \ell_1}, \dots),$$

which reorders to

$$((b+1)\lambda - 1;\, b\lambda - 1,\, 1^{\times(\lfloor a \rfloor - 1)} \,\|\, \lambda - 1,\, w_1^{\times \ell_1}, \dots).$$

The action of this Cremona move on the balls

$$B(\boldsymbol{w}(a)) \amalg B(\lambda) \amalg B(b\lambda) \overset{s}{\hookrightarrow} B((b+1)\lambda)$$

with $B(\boldsymbol{w}(a)) \overset{s}{\hookrightarrow} P(\lambda, b\lambda)$ is illustrated in Figure 9.

Notation 2.4 The symbol $\|$ indicates that the terms before $\|$ are ordered, while the terms after $\|$ are possibly not ordered, and that all terms before $\|$ are not less than the terms after $\|$.

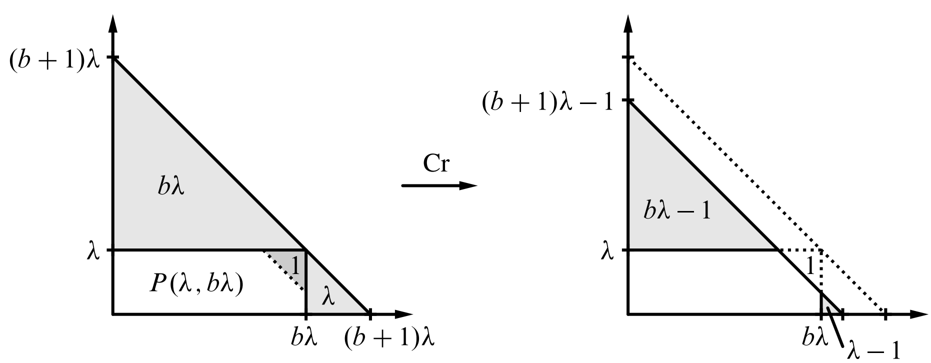


Figure 9: The effect of the first Cremona move

Method 3 (ECH capacities) *Hutchings* [18] *used his embedded contact homology to associate with every bounded starlike domain $U \subset \mathbb{R}^4$ a sequence of symplectic capacities $c_1(U) \leqslant c_2(U) \leqslant \cdots$. For an ellipsoid $E(a,b)$, this sequence is given by arranging the numbers of the form $ma+nb$ with m, $n \geqslant 0$ in nondecreasing order, with multiplicities. For instance,*

$$(c_k(E(1,1))) = (1,1,2,2,2,3,3,3,3,4,\dots).$$

McDuff [24] showed that ECH capacities provide a complete set of invariants for the embedding problem $E(a,b) \overset{s}{\hookrightarrow} E(c,d)$:

$$E(a,b) \overset{s}{\hookrightarrow} E(c,d) \quad \iff \quad c_k(E(a,b)) \leqslant c_k(E(c,d)) \quad \text{for all } k \geqslant 1.$$

Since the embedding problems $E(1,a) \overset{s}{\hookrightarrow} E(\lambda, \lambda 2b)$ and $E(1,a) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ are equivalent, it follows that

$$c_b(a) = \sup_{k \geqslant 1} \left\{ \frac{c_k(E(1,a))}{c_k(E(1,2b))} \right\}. \tag{2-15}$$

It is not clear, though, how to derive from this description of $c_b(a)$ the graphs given in Theorem 1.1.

We say that an exceptional class $E = (d,e;\boldsymbol{m}) \in \mathcal{E}_K(X_n)$ is *b–obstructive* if there is some $a \geqslant 1$ such that the obstruction function (2-11) is larger than the volume constraint:

$$\mu_b(d,e;\boldsymbol{m})(a) > \sqrt{\frac{a}{2b}}.$$

According to Method 1, it suffices to find all b–obstructive classes: the graph of $c_b(a)$ is given as the supremum of the constraints of the b–obstructive classes and of the volume constraint. Since exceptional classes are represented by holomorphic spheres, this method gives insight into the nature of the obstruction to a full embedding. It is also useful for guessing the graph of $c_b(a)$, by first guessing a relevant set of b–obstructive classes (see Section 3). On the other hand, it is sometimes hard to find all b–obstructive classes for a point a. Method 2 is very efficient at a given point a, at least if one has an idea what $c_b(a)$ should be. However, the reduction scheme often depends rather subtly on the point a; see Sections 5–8. The reduction method is thus quite "local in a". While it is usually impossible to compute $c_b(a)$ by Method 3 (see however [4; 13]), this method is very useful for guessing the graph of $c_b(a)$, since, using (2-15) and a computer, one gets good lower bounds for $c_b(a)$.

Accordingly, we have found Theorem 1.1 as follows: We first found the exceptional classes E_n and F_n in (2-13), then used ECH capacities to convince ourselves that there are no further constraints besides the volume, and then proved this by the reduction method. This seems to be a convenient procedure for solving symplectic embedding problems for which ECH capacities are known to form a complete set of invariants, such as those studied in [11].

3 Applications of Method 1

Fix a real number $b \geqslant 1$. As in (2-11) we associate with every solution $(d,e;\boldsymbol{m})$ of the Diophantine system (2-12) the obstruction function

$$\mu_b(d,e;\boldsymbol{m})(a) = \frac{\langle \boldsymbol{m}, \boldsymbol{w}(a)\rangle}{d+be}, \tag{3-1}$$

where as before $\boldsymbol{w}(a)$ is the weight expansion of $a \geqslant 1$. Further, define the error vector $\varepsilon := \varepsilon(a)$ by

$$\boldsymbol{m} = \frac{d+be}{\sqrt{2ba}}\boldsymbol{w}(a) + \varepsilon.$$

(Here, we add zeros to $\boldsymbol{m}$ or $\boldsymbol{w}(a)$ if they do not have the same length.)

3.1 Recollections

The following proposition generalizes [14, Lemma 4.8]:

Proposition 3.1 *Fix a real number $b \geqslant 1$. Given a nonnegative solution $(d,e;\boldsymbol{m})$ of (2-12) and $a \geqslant 1$, we have:*

(i) $\mu_b(d,e;\boldsymbol{m})(a) \leqslant \sqrt{2de+1}\sqrt{a}/(d+be)$.

(ii) $\mu_b(d,e;\boldsymbol{m})(a) > \sqrt{\frac{a}{2b}}$ *if and only if* $\langle \varepsilon, \boldsymbol{w}(a)\rangle > 0$.

(iii) *If* $\mu_b(d,e;\boldsymbol{m})(a) > \sqrt{\frac{a}{2b}}$, *then* $d = be+h$ *with* $|h| < \sqrt{2b}$ *and* $\langle \varepsilon,\varepsilon\rangle < 1 - \frac{h^2}{2b}$.

Proof By the Cauchy–Schwarz inequality and since $\sum w_i^2 = a$,

$$(d+be)\mu_b(d,e;\boldsymbol{m})(a) = \langle \boldsymbol{m}, \boldsymbol{w}(a)\rangle \leqslant \|\boldsymbol{m}\|\,\|\boldsymbol{w}(a)\| = \sqrt{2de+1}\sqrt{a},$$

proving (i). Assertion (ii) is immediate. To prove (iii), we compute

$$\begin{aligned} 2(be+h)e+1 = 2de+1 = \langle \boldsymbol{m},\boldsymbol{m}\rangle &= \left\langle \frac{2be+h}{\sqrt{2ba}}\boldsymbol{w}(a)+\varepsilon, \frac{2be+h}{\sqrt{2ba}}\boldsymbol{w}(a)+\varepsilon \right\rangle \\ &= \frac{(2be+h)^2}{2ba}a + 2\frac{2be+h}{\sqrt{2ba}}\langle \boldsymbol{w}(a),\varepsilon\rangle + \langle \varepsilon,\varepsilon\rangle. \end{aligned}$$

The first of the three summands is $2be^2 + 2eh + \frac{h^2}{2b}$, and so

$$1 = \frac{h^2}{2b} + 2\frac{2be+h}{\sqrt{2ba}}\langle \boldsymbol{w}(a), \varepsilon\rangle + \langle \varepsilon, \varepsilon\rangle.$$

Hence, if $\mu_b(d,e;\boldsymbol{m})(a) > \sqrt{\frac{a}{2b}}$, then, by (ii), $\langle \boldsymbol{w}(a), \varepsilon\rangle > 0$, whence $0 \leqslant \langle \varepsilon, \varepsilon\rangle < 1 - \frac{h^2}{2b}$. This also shows $|h| < \sqrt{2b}$. □

3.2 Two sequences of exceptional classes and their constraints

In our analysis of the functions $c_b(a)$, two sequences of exceptional homology classes will play a role. For each $n \in \mathbb{N}$ we define the classes

$$E_n := (n, 1; 1^{\times(2n+1)}),$$
$$F_n := (n(n+1),\ n+1;\ n+1,\ n^{\times(2n+3)}).$$

Notice that E_n is a perfect class at $a = 2n+1$, in the sense that $\boldsymbol{m}$ is a multiple of $\boldsymbol{w}(a)$. Similarly, F_n is nearly perfect at $a = 2n+4$. While the constraints of the classes $E_b, E_{b+1}, \dots, E_{b+\lfloor\sqrt{2b}\rfloor}$ will give the $\lceil\sqrt{2b}\,\rceil$ linear steps in the graph of $c_b(a)$ centred at $2b+2k+1$, the constraint of F_b will give the affine step of $c_b(a)$ centred at $2b+4$.

Lemma 3.2 *The classes E_n and F_n satisfy the Diophantine system* (2-12) *and their image under ψ_* reduces to* $(0; -1, 0, \dots, 0)$ *under repeated standard Cremona moves.*

Proof One readily checks that the classes E_n and F_n satisfy the Diophantine system (2-12).

For the sequel it is useful to rewrite the Cremona transform Cr as follows: Define the *defect* of a vector $(d;\boldsymbol{m}) = (d; m_1, \dots, m_k)$ by $\delta := d - m_1 - m_2 - m_3$. Then (2-7) can be written as

$$\mathrm{Cr}(d;\boldsymbol{m}) = (d+\delta;\ m_1+\delta,\ m_2+\delta,\ m_3+\delta,\ m_4,\ \dots,\ m_k).$$

The isomorphism ψ_* from (2-10) maps $E_n = (n, 1; 1^{\times(2n+1)})$ to $(n;\ n-1,\ 1^{\times 2n})$, which under one standard Cremona move is mapped to $(n-1;\ n-2,\ 1^{\times 2(n-1)})$, and thence, under n such moves, to $(0; -1)$. Next, ψ_* maps F_1 to the class $(2; 1^{\times 5})$, which reduces to $(0; -1)$ under two standard Cremona moves, Further, for $n \geqslant 2$,

$$\psi_*(F_n) = (n^2+n;\ n^2-1,\ n^{\times(2n+3)}).$$

Under n standard Cremona moves with $\delta = -n+1$ this vector reduces to

$$(2n;\ n^{\times 3},\ n-1,\ 1^{\times 2n}).$$

Applying one more standard Cremona move with $\delta = -n$ yields $(n;\, n-1,\, 1^{\times 2n})$, which reduces in n steps to $(0;-1)$, as we have seen above. $\square$

We next compute the constraints given by the classes E_n and F_n. In view of definition (3-1) and the definition of these classes,

$$\mu_b(E_{b+k})(a) = \frac{\langle 1^{\times(2b+2k+1)}, \boldsymbol{w}(a)\rangle}{2b+k}, \quad \mu_b(F_b)(a) = \frac{\langle (b+1, b^{\times(2b+3)}), \boldsymbol{w}(a)\rangle}{2b(b+1)}.$$

From this we readily find:

Lemma 3.3 *Fix an integer $b \geqslant 2$.*

(i) *For $k \in \{0, 1, 2, \dots, \lfloor\sqrt{2b}\rfloor\}$,*

$$\mu_b(E_{b+k})(a) = \begin{cases} \dfrac{a}{2b+k} & \text{if } a \in [2b+2k, 2b+2k+1], \\ \dfrac{2b+2k+1}{2b+k} & \text{if } a \geqslant 2b+2k+1. \end{cases}$$

(ii)
$$\mu_b(F_b)(a) = \begin{cases} \dfrac{ba+1}{2b(b+1)} & \text{if } a \in [2b+3, 2b+4], \\ 1 + \dfrac{2b+1}{2b(b+1)} & \text{if } a \geqslant 2b+4. \end{cases}$$

We in particular see that the class E_{b+k} gives rise to the linear step over $I_b(k)$ and F_b gives rise to the affine step over $[\alpha_b, \beta_b]$.

3.3 The constraints of E_n and F_n for real $b \geqslant 2$

In this subsection we compute the obstructions to the problem $E(1,a) \to P(\lambda, \lambda b)$ given by the exceptional classes E_n and F_n for all real $b \geqslant 2$. This is not used in the proof of Theorem 1.1, but supports Conjecture 1.5.

Let $b \geqslant 2$ be a real number. Recall that for $a \geqslant 1$ every exceptional class $E = (d, e; \boldsymbol{m})$ yields the constraint

$$\mu_b(E)(a) = \frac{\langle \boldsymbol{m}, \boldsymbol{w}(a)\rangle}{d+be}.$$

For $E_0 = (1, 0; 1)$ we have

$$\mu_b(E_0)(a) = 1 \tag{3-2}$$

and for $E_n = (n, 1; 1^{\times(2n+1)})$ with $n \geqslant 1$ we have

$$\mu_b(E_n)(a) = \begin{cases} \dfrac{a}{n+b} & \text{if } a \in [2n, 2n+1], \\ \dfrac{2n+1}{n+b} & \text{if } a \geqslant 2n+1. \end{cases}$$

The class E_n is b–obstructive on $[2n,\infty)$ only if $\frac{2n+1}{n+b} > \sqrt{\frac{2n+1}{2b}}$, and in view of (3-2) we can also assume that $\frac{2n+1}{n+b} > 1$, or $n > b-1$. The relevant values of n are thus

$$n \in \{\lfloor b \rfloor, \dots, \lfloor b+\sqrt{2b} \rfloor\},$$

where $\lfloor b \rfloor$ is the largest integer not greater than b. The constraint 1 of E_0 meets the first linear step, given by $E_{\lfloor b \rfloor}$, at $a = b + \lfloor b \rfloor$, and is thus strictly above $\sqrt{\frac{a}{2b}}$ if $b \notin \mathbb{N}$. For $n \geqslant \lfloor b \rfloor$ the step of E_n meets the step of E_{n+1} at $a = (2n+1)(n+b+1)/(n+b)$, which is above $\sqrt{\frac{a}{2b}}$ if and only if $b-n \geqslant (b-n)^2$. The step of $E_{\lfloor b \rfloor}$ thus meets the one of $E_{\lfloor b \rfloor +1}$ above the volume constraint, with equality if and only if $b \in \mathbb{N}$, and all other linear steps are strictly disjoint.

Next, let $\boldsymbol{b}$ be the "integer closest to b", namely $b = \boldsymbol{b} + \varepsilon$ with $\varepsilon \in \left(-\frac{1}{2}, \frac{1}{2}\right]$. Then

$$\mu_b(F_{\boldsymbol{b}})(a) = \begin{cases} \dfrac{\boldsymbol{b}a+1}{(\boldsymbol{b}+b)(\boldsymbol{b}+1)} & \text{if } a \in [2\boldsymbol{b}+3, 2\boldsymbol{b}+4], \\ \dfrac{2\boldsymbol{b}^2+4\boldsymbol{b}+1}{(\boldsymbol{b}+b)(\boldsymbol{b}+1)} & \text{if } a \geqslant 2\boldsymbol{b}+4. \end{cases}$$

But notice that this constraint is stronger than $\sqrt{\frac{a}{2b}}$ only if

$$\mu_b(F_{\boldsymbol{b}})(2\boldsymbol{b}+4) = \frac{2\boldsymbol{b}^2+4\boldsymbol{b}+1}{(2\boldsymbol{b}+\varepsilon)(\boldsymbol{b}+1)} > \sqrt{\frac{\boldsymbol{b}+2}{\boldsymbol{b}+\varepsilon}}$$

or, equivalently, $\varepsilon \in \left(-\frac{\boldsymbol{b}}{(\boldsymbol{b}+1)^2}, \frac{1}{\boldsymbol{b}+2}\right)$. One readily checks that the affine step defined by $\mu_b(F_{\boldsymbol{b}})$ is strictly disjoint from the two neighbouring linear steps given by $E_{\boldsymbol{b}+1}$ and $E_{\boldsymbol{b}+2}$.

For $a \geqslant 1$ and $b \geqslant 2$ let $d_b(a)$ be the maximum of the volume constraint $\sqrt{\frac{a}{2b}}$ and the obstructions $\mu_b(E_n)(a)$ and $\mu_b(F_{\boldsymbol{b}})$ discussed above. Then $d_b(a) \geqslant c_b(a)$ of course, and Conjecture 1.5 claims that $d_b(a) = c_b(a)$ for all real $b \geqslant 2$.

3.4 $c_b(a)$ at $a = 2b+2+\frac{1}{2b}$

Set $a_b := 2b+2+\frac{1}{2b}$. We will show in Section 4.2 by the reduction method that $c_b(a_b) = \frac{2b+1}{2b}$. (Notice that this value equals the volume constraint $\sqrt{a_b/2b}$.) Here we show this by using positivity of intersection with the class

$$G_b := (b(2b+1), 2b+1; (2b)^{\times(2b+2)}, 1^{\times(2b+1)}), \quad b \in \mathbb{N}.$$

The $\boldsymbol{m}$ of G_b is obtained from $2b\boldsymbol{w}(a_b)$ by adding one 1, whence G_b is nearly perfect at a_b. One readily checks that G_b satisfies the Diophantine system (2-12) and that its image under ψ_* reduces to $(0;-1,0,\dots,0)$ under repeated standard Cremona moves. Hence G_b is an exceptional class. Its obstruction at a_b is

$$\mu_b(G_b)\Big(2b+2+\frac{1}{2b}\Big)=\frac{2b(2b+2)+1}{2b(2b+1)}=\frac{2b+1}{2b}.$$

Write $G_b=(b(2b+1),2b+1;\boldsymbol{m}_b,1)$ with $\boldsymbol{m}_b:=((2b)^{\times(2b+2)},1^{\times 2b})=2b\boldsymbol{w}(a_b)$. Recall that exceptional classes are represented by embedded J–holomorphic spheres, whence, by positivity of intersection, $E\cdot E'\geqslant 0$ for any two different exceptional classes $E\neq E'$. Applying this to G_b and any different exceptional class $(d,e;\boldsymbol{m})$, we obtain

$$(be+d)(2b+1)=b(2b+1)e+(2b+1)d\geqslant\langle\boldsymbol{m},(\boldsymbol{m}_b,1)\rangle\geqslant\langle\boldsymbol{m},\boldsymbol{m}_b\rangle=2b\langle\boldsymbol{m},\boldsymbol{w}(a_b)\rangle.$$

Hence

$$\mu_b(d,e;\boldsymbol{m})(a_b)=\frac{\langle\boldsymbol{m},\boldsymbol{w}(a_b)\rangle}{be+d}\leqslant\frac{2b+1}{2b},$$

as we wished to show.

Remarks 3.4 (i) The classes E_1 and E_2 also give rise to the first two steps of $c_C(a)=c_1(a)$, and the class F_1 gives rise to the affine step of $c_C(a)$; see [14]. This is the "holomorphic reason" why the first two steps of the Pell stairs and the affine step of $c_C(a)$ survive to all functions $c_b(a)$ for $b\geqslant 2$. On the other hand, none of the classes E_n with $n\geqslant 3$ and F_n with $n\geqslant 2$ is obstructive for the problem $E(1,a)\xrightarrow{s}C^4(\lambda)$, and none of the classes giving rise to the other steps of the Pell stairs, nor any of the classes giving rise to the six exceptional steps of $c_C(a)$, gives an obstruction for the problems $E(1,a)\xrightarrow{s}P(\lambda,\lambda b)$ with $b\geqslant 2$.

Similarly, G_1 is the first of the sequence of exceptional classes $E(\alpha_n)$ in [14] that imply, via positivity of intersection, that at the feet of the Pell stairs there is no embedding obstruction beyond the volume constraint.

(ii) We do not know all b–obstructive classes. However, using positivity of intersection and the analogues of Lemmas 3.8 and 3.11 we checked that $\mu_b(E)(2b+2k+1)<\frac{2b+2k+1}{2b+k}$ for any exceptional class $E\neq E_{b+k}$, and that $\mu_b(E)(2b+4)\leqslant\sqrt{\frac{2b+4}{2b}}$ for any exceptional class $E\neq F_b$, that is, F_b is the only b–obstructive class at $2b+4$. For F_2 this is carried out in Lemma 3.10.

3.5 $c_b(a)$ for a large

For $b\in\mathbb{N}_{\geqslant 2}$ we abbreviate

$$v_b^+:=v_b(\lfloor\sqrt{2b}\rfloor)=2b\bigg(\frac{2b+2\lfloor\sqrt{2b}\rfloor+1}{2b+\lfloor\sqrt{2b}\rfloor}\bigg)^2.$$

Assertion (ii) of the following proposition improves [4, Theorem 1.1].

Proposition 3.5 (i) *For every* $b \in \mathbb{N}_{\geqslant 2}$,

$$c_b(a) = \begin{cases} \dfrac{2b+2\lfloor\sqrt{2b}\rfloor+1}{2b+\lfloor\sqrt{2b}\rfloor} & \text{if } a \in [2b+2\lfloor\sqrt{2b}\rfloor+1, v_b^+], \\ \sqrt{\dfrac{a}{2b}} & \text{if } a \geqslant v_b^+. \end{cases}$$

(ii) *For every real* $b \geqslant 2$ *we have* $c_b(a) = \sqrt{\frac{a}{2b}}$ *for all* $a \geqslant (\sqrt{2b}+1)^2$.

Notice that the length of the interval $[2b+2\lfloor\sqrt{2b}\rfloor+1, v_b^+]$ in (i) is

$$\frac{(2b+2\lfloor\sqrt{2b}\rfloor+1)(2b-\lfloor\sqrt{2b}\rfloor^2)}{(2b+\lfloor\sqrt{2b}\rfloor)^2}$$

and hence positive if and only if $\lfloor\sqrt{2b}\rfloor < \sqrt{2b}$, ie $2b$ is not a perfect square.

Proof Assume that $(d,e;\boldsymbol{m})$ is a nonnegative solution of (2-12). If $e=0$, then $(d,e;\boldsymbol{m}) = (1,0;1)$, and so $\mu_b(d,e;\boldsymbol{m})(a) = 1$ is smaller than the values of $c_b(a)$ claimed in (i) and (ii). We can thus assume that $e \geqslant 1$.

Suppose that $\mu_b(d,e;\boldsymbol{m})(a) > \sqrt{\frac{a}{2b}}$ for some $a \geqslant 1$. Then, by Proposition 3.1(iii), $d < be + \sqrt{2b}$. We estimate

$$\mu_b(d,e;\boldsymbol{m})(a) = \frac{\langle \boldsymbol{m}, \boldsymbol{w}(a)\rangle}{be+d} \leqslant \frac{\sum m_i}{be+d} = \frac{2(d+e)-1}{be+d} =: f_{b,e}(d). \tag{3-3}$$

The function $d \mapsto f_{b,e}(d)$ is increasing. We can thus further estimate

$$\mu_b(d,e;\boldsymbol{m})(a) \leqslant f_{b,e}(be+\sqrt{2b}) = \frac{2(be+\sqrt{2b}+e)-1}{2be+\sqrt{2b}} =: L(b,e). \tag{3-4}$$

Claim 1 $\frac{\partial}{\partial e} L(b,e) \leqslant 0$.

Proof We compute

$$\frac{\partial}{\partial e} L(b,e) = \frac{2(b+1)(2be+\sqrt{2b}) - 2b(2(be+\sqrt{2b}+e)-1)}{(2be+\sqrt{2b})^2},$$

which is nonpositive if and only if the numerator is nonpositive. Expanding the numerator, we see that this holds if and only if $b+\sqrt{2b} \leqslant b\sqrt{2b}$, which holds true because $b \geqslant 2$. □

Proof of Proposition 3.5(ii) Assume that $(d,e;\boldsymbol{m})$ is an exceptional class with $e \geqslant 1$ and such that $\mu_b(d,e;\boldsymbol{m})(a) > \sqrt{\frac{a}{2b}}$ for some $a \geqslant (\sqrt{2b}+1)^2$. By (3-4) and Claim 1,

$$\mu_b(d,e;\boldsymbol{m})(a) \leqslant L(b,e) \leqslant L(b,1) = \frac{\sqrt{2b}+1}{\sqrt{2b}} \leqslant \sqrt{\frac{a}{2b}},$$

a contradiction.

Proof of Proposition 3.5(i) Assume from now on that $b \in \mathbb{N}_{\geqslant 2}$. If $e=1$, then (2-12) becomes

$$\sum m_i = \sum m_i^2 = 2d+1$$

and so $(d,e;\boldsymbol{m})$ is the exceptional class $E_d = (d,1;1^{\times(2d+1)})$. Recall that on the interval $[2d,2d+2]$ the obstruction function

$$\mu_b(E_d)(a) = \frac{\langle \boldsymbol{w}(a), 1^{\times(2d+1)}\rangle}{b+d}$$

gives a linear step with edge at $2d+1$. If $\lfloor\sqrt{2b}\rfloor < \sqrt{2b}$, then the largest k for which E_{b+k} yields a constraint strictly stronger than the volume is $k = \lfloor\sqrt{2b}\rfloor$, because

$$\frac{2b+2k+1}{2b+k} > \sqrt{\frac{2b+2k+1}{2b}} \iff 2b > k^2.$$

We are left with showing that for $e \geqslant 2$ we have $\mu_b(d,e;\boldsymbol{m})(a) \leqslant \sqrt{\frac{a}{2b}}$ for all solutions $(d,e;\boldsymbol{m})$ of (2-12) and all $a \geqslant v_b^+$. Assume first that $e \geqslant 3$. Then (3-4) and Claim 1 yield

$$\mu_b(d,e;\boldsymbol{m})(a) \leqslant L(b,e) \leqslant L(b,3).$$

Claim 2 $L(b,3) \leqslant \sqrt{\frac{a}{2b}}$ *for all* $b \in \mathbb{N}_{\geqslant 2}$ *and* $a \geqslant v_b^+$.

Proof It suffices to prove the claim for $a = v_b^+$. We have

$$L(b,3) - 1 = \frac{\sqrt{2b}+5}{6b+\sqrt{2b}} \quad \text{and} \quad \sqrt{\frac{v_b^+}{2b}} - 1 = \frac{\lfloor\sqrt{2b}\rfloor+1}{2b+\lfloor\sqrt{2b}\rfloor}.$$

For $b \in \{2,3,4\}$ the inequality

$$\frac{\lfloor\sqrt{2b}\rfloor+1}{2b+\lfloor\sqrt{2b}\rfloor} \geqslant \frac{\sqrt{2b}+5}{6b+\sqrt{2b}}$$

is readily verified. For $b \geqslant 5$ we use that $x \mapsto \frac{x+1}{2b+x}$ is increasing, and estimate

$$\sqrt{\frac{v_b^+}{2b}} - L(b,3) \geqslant \frac{(\sqrt{2b}-1)+1}{2b+(\sqrt{2b}-1)} - \frac{\sqrt{2b}+5}{6b+\sqrt{2b}}.$$

The right-hand side multiplied with the product of the denominators equals $f(b) := 4b\sqrt{2b} - 10b - 4\sqrt{2b} + 5$. Since $bf'(b) = 6b\sqrt{2b} - 2\sqrt{2b} - 10b \geqslant 0$ for $b \geqslant 2$ and $f(5) > 0$, the claim follows. □

Assume now that $e = 2$. We first treat the case $b \geqslant 5$. In view of (3-4) it suffices to show that $L(b,2) \leqslant \sqrt{v_b^+/2b}$, or

$$\frac{\sqrt{2b}+3}{4b+\sqrt{2b}} \leqslant \frac{\lfloor\sqrt{2b}\rfloor + 1}{2b + \lfloor\sqrt{2b}\rfloor}.$$

This inequality is readily verified for $b = 5$. For $b \geqslant 6$ the stronger inequality

$$\frac{\sqrt{2b}+3}{4b+\sqrt{2b}} \leqslant \frac{(\sqrt{2b}-1)+1}{2b+(\sqrt{2b}-1)}$$

holds true. Indeed, this inequality is equivalent to $g(b) := 2b\sqrt{2b} - 6b - 2\sqrt{2b} + 3 \geqslant 0$, which holds true since $bg'(b) = 3b\sqrt{2b} - \sqrt{2b} - 6b \geqslant 0$ for $b \geqslant 6$ and $g(6) \geqslant 0$.

Assume now that $b \in \{2,3,4\}$. Then $\sqrt{v_b^+/2b} = 1 + \frac{3}{2b+2}$. Using (3-3) this time with $d \leqslant \lfloor be + \sqrt{2b}\rfloor$, we find

$$\mu_b(d,2;\boldsymbol{m})(a) \leqslant f_{b,2}(\lfloor 2b + \sqrt{2b}\rfloor) = \frac{2\lfloor 2b + \sqrt{2b}\rfloor + 3}{2b + \lfloor 2b + \sqrt{2b}\rfloor}.$$

For $b \in \{2,3,4\}$ the right-hand side is at most $1 + \frac{3}{2b+2}$. Proposition 3.5 is proven. □

3.6 The interval $\left[8\frac{1}{36}, 9\right]$ for $b = 2$

Proposition 3.6 $c_2(a) = \frac{1}{2}\sqrt{a}$ *for* $a \in \left[8\frac{1}{36}, 9\right]$.

Proof The arguments in this section are close to those in [26, Section 5.3] and [14, Section 7.3]. In fact, the last step of $c_{\boldsymbol{B}}(a)$ and of $c_2(a)$ both end at $8\frac{1}{36}$ and are given by the class F_2. There are some differences, however, and so we give a complete exposition for the convenience of the reader.

Fix a rational number $a = \frac{p}{q} \in (8,9)$, with $\frac{p}{q}$ in reduced form, with weight expansion

$$(1^{\times \ell_0}, w_1^{\times \ell_1}, \ldots, w_N^{\times \ell_N}). \tag{3-5}$$

Then $w_N = \frac{1}{q}$ and $\sum_{j=0}^N \ell_j w_j = a + 1 - \frac{1}{q}$ by [26, Lemma 1.2.6]. Set $M := \ell(a) := \sum_{j=0}^N \ell_j$ and $L = \sum_{j=1}^N \ell_j = \ell(a) - 8$. Then $q \geqslant L$ by [26, Sublemma 5.1.1].

For $b = 2$ the error vector ε of an exceptional class $(d,e;\boldsymbol{m})$ at a is

$$\boldsymbol{m} = \frac{d+2e}{2\sqrt{a}}\boldsymbol{w}(a) + \varepsilon. \tag{3-6}$$

Define the partial error sums

$$\sigma := \sum_{i=\ell_0+1}^{M} \varepsilon_i^2 \quad \text{and} \quad \sigma' := \sum_{i=\ell_0+1}^{M-\ell_N} \varepsilon_i^2 \leqslant \sigma.$$

Recall from Proposition 3.1(iii) that for an obstructive class $(d, e; \boldsymbol{m})$ we have $d = 2e + h$ with $h \in \{-1, 0, 1\}$, and $\sigma < 1$ if $h = 0$ and $\sigma < \frac{3}{4}$ if $|h| = 1$. For the function

$$y(a) := a - 3\sqrt{a} + 1$$

we have $y\left(\frac{p}{q}\right) > \frac{1}{q}$ for all $\frac{p}{q} \in (8, 9)$. Write $\ell(\boldsymbol{m})$ for the number of positive entries in $\boldsymbol{m}$.

Lemma 3.7 *Let $(d, e; \boldsymbol{m})$ be an exceptional class such that there exists $a = \frac{p}{q} \in (8, 9)$ with $\ell(a) = \ell(\boldsymbol{m})$ and $\mu_2(d, e; \boldsymbol{m})(a) > \frac{1}{2}\sqrt{a}$. Set $v_M := (d + 2e)/(2q\sqrt{a})$. Then:*

(i) $\left|\sum \varepsilon_i\right| \leqslant \sqrt{\sigma L}$.

(ii) *If $v_M < 1$, then $\left|\sum \varepsilon_i\right| \leqslant \sqrt{\sigma' L}$.*

(iii) *If $v_M \leqslant \frac{1}{2}$, then $v_M > \frac{1}{3}$ and $\sigma' \leqslant \frac{1}{2}$. If $v_M \leqslant \frac{2}{3}$, then $\sigma' \leqslant \frac{7}{9}$.*

(iv) *With $\delta := y(a) - \frac{1}{q}$ we have*

$$4e + h \leqslant \frac{2\sqrt{a}}{\delta}\left(\sqrt{\sigma q} - \left(1 - \frac{h}{2}\right)\right) \leqslant \frac{2\sqrt{a}}{\delta}\left(\frac{\sigma}{\delta v_M} - \left(1 - \frac{h}{2}\right)\right).$$

If $v_M < 1$, then σ can be replaced by σ'.

Proof The proofs of (i), (ii) and (iii) are as for [26, Lemma 5.1.2]. To prove (iv) we compute

$$\begin{aligned} -\sum_{i=1}^{M} \varepsilon_i = \frac{d+2e}{2\sqrt{a}} \sum_{j=0}^{N} \ell_j w_j - \sum_{i=1}^{M} m_i &= \frac{d+2e}{2\sqrt{a}}\left(a + 1 - \frac{1}{q}\right) - (2d + 2e - 1) \\ &= \frac{4e+h}{2\sqrt{a}}\left(a + 1 - \frac{1}{q}\right) - (6e + 2h - 1) \\ &= \frac{4e+h}{2\sqrt{a}}\left(y(a) - \frac{1}{q}\right) + \left(1 - \frac{h}{2}\right), \end{aligned}$$

where we have used (3-6) and (2-12). Then, using $q \geqslant L$ and (i), we find

$$\sqrt{\sigma q} \geqslant \sqrt{\sigma L} \geqslant \frac{4e+h}{2\sqrt{a}}\left(y(a) - \frac{1}{q}\right) + \left(1 - \frac{h}{2}\right) = \frac{4e+h}{2\sqrt{a}}\delta + \left(1 - \frac{h}{2}\right) > \delta v_M q.$$

Thus $\sqrt{q} < \sqrt{\sigma}/\delta v_M$, and so

$$4e + h \leqslant \frac{2\sqrt{a}}{\delta}\Big(\sqrt{\sigma q} - \Big(1 - \frac{h}{2}\Big)\Big) < \frac{2\sqrt{a}}{\delta}\Big(\frac{\sigma}{\delta v_M} - \Big(1 - \frac{h}{2}\Big)\Big).$$

If $v_M < 1$, the same arguments go through when replacing σ by σ'. □

The following lemma is proven as in [26, Lemma 2.1.7]:

Lemma 3.8 *Let $(d, e; \boldsymbol{m})$ be an exceptional class such that $\mu_2(d, e; \boldsymbol{m})(a) > \frac{1}{2}\sqrt{a}$ for some $a \in [8, 9)$. Then:*

(i) *The vector $(m_1, \dots, m_8)$ is of the form*

$$(m, \dots, m) \quad \text{or} \quad (m, \dots, m, m-1) \quad \text{or} \quad (m+1, m, \dots, m).$$

(ii) *If $m_1 \neq m_8$, then $\sum_{i=1}^{8} \varepsilon_i^2 \geqslant \frac{7}{8}$.*

Lemma 3.9 *There is no exceptional class $(d, e; \boldsymbol{m})$ such that $\mu_2(d, e; \boldsymbol{m})(a) > \frac{1}{2}\sqrt{a}$ for some $a \in (8, 9)$ with $\ell(a) = \ell(\boldsymbol{m})$.*

Proof Assume that $(d, e; \boldsymbol{m})$ is an exceptional class such that $\mu_2(d, e; \boldsymbol{m})(a) > \frac{1}{2}\sqrt{a}$ for some $a \in (8, 9)$ with $\ell(a) = \ell(\boldsymbol{m})$.

We first show that $m_1 = \cdots = m_8$. Assume the contrary. By Lemma 3.8, $\langle \varepsilon, \varepsilon \rangle \geqslant \frac{7}{8}$ and $\sigma \leqslant \frac{1}{8}$. The inequality $\langle \varepsilon, \varepsilon \rangle \geqslant \frac{7}{8}$ and Proposition 3.1(iii) show that $h = 0$. Since $M > 8$ and $\sigma \leqslant \frac{1}{8}$, we find $v_M \geqslant 1 - \frac{1}{\sqrt{8}} > \frac{1}{2}$. Further, since $a \geqslant 8\frac{1}{q}$,

$$\delta = y(a) - \tfrac{1}{q} \geqslant y\big(8\tfrac{1}{q}\big) - \tfrac{1}{q} = 9 - 3\sqrt{8\tfrac{1}{q}} \geqslant 9 - 3\sqrt{8\tfrac{1}{2}} \geqslant \tfrac{1}{4}.$$

Altogether, $\frac{\sigma}{\delta v_M} < 1$, in contradiction with Lemma 3.7(iv).

We are now going to show that e must be small. For this we first notice that, by Lemma 3.7(iii),

$$\begin{aligned} v_M \in \Big[\frac{1}{3}, \frac{1}{2}\Big] \quad &\Longrightarrow \quad \frac{\sigma'}{v_M} \leqslant \frac{1/2}{1/3} = \frac{3}{2}, \\ v_M \in \Big[\frac{1}{2}, \frac{2}{3}\Big] \quad &\Longrightarrow \quad \frac{\sigma'}{v_M} \leqslant \frac{7/9}{1/2} = \frac{14}{9}, \\ v_M \geqslant \frac{2}{3} \quad &\Longrightarrow \quad \frac{\sigma}{v_M} \leqslant \frac{3}{2}. \end{aligned}$$

For fixed q and h, the functions

$$\begin{aligned} F(a, q, h) &:= \frac{2\sqrt{a}}{\delta}\Big(\sqrt{q} - \Big(1 - \frac{h}{2}\Big)\Big), \\ G(a, q, h) &:= \frac{2\sqrt{a}}{\delta}\Big(\frac{14}{9}\frac{1}{\delta} - \Big(1 - \frac{h}{2}\Big)\Big) \end{aligned}$$

are strictly decreasing for $a \in (8,9)$. Since $a \geqslant 8\frac{1}{q}$, we see from Lemma 3.7(iv) that

$$4e + h \leqslant f(q,h),\ g(q,h),$$

where $f(q,h) := F\big(8\frac{1}{q}, q, h\big)$ and $g(q,h) := G\big(8\frac{1}{q}, q, h\big)$. Explicitly,

$$\begin{aligned} f(q,h) &:= \frac{2\sqrt{8\frac{1}{q}}}{\delta(q)}\left(\sqrt{q} - \left(1 - \frac{h}{2}\right)\right), \\ g(q,h) &:= \frac{2\sqrt{8\frac{1}{q}}}{\delta(q)}\left(\frac{14}{9}\frac{1}{\delta(q)} - \left(1 - \frac{h}{2}\right)\right), \end{aligned}$$

where $\delta(q) := y(8\frac{1}{q}) - \frac{1}{q} = 9 - 3\sqrt{8\frac{1}{q}}$. We have that $\frac{\partial f}{\partial q}(q,h) > 0$ for $q \geqslant 3$ and $\frac{\partial g}{\partial q}(q,h) < 0$ for $q \geqslant 2$, and $f(q,h) < g(q,h)$ for $q \in \{2,3\}$. In fact, $f(q,h) = g(q,h)$ if and only if $\sqrt{q} = \frac{14}{9}\frac{1}{\delta(q)}$, which happens at $q \approx 11.1$. One readily checks that

$$f(11,-1),\ g(12,-1) < 23, \quad f(11,0),\ g(12,0) < 29, \quad f(11,1),\ g(12,1) < 35.$$

It follows that

$$4e + h \leqslant 22, 28, 34 \quad \text{for } h = -1, 0, 1, \text{ respectively},$$

and so

$$e \leqslant 5 \quad \text{if } h = -1, \qquad e \leqslant 7 \quad \text{if } h = 0, \qquad e \leqslant 8 \quad \text{if } h = 1. \tag{3-7}$$

However, one readily checks that there are no solutions $(2e + h, e; \boldsymbol{m})$ of (2-12) satisfying (3-7) and $m_1 = \cdots = m_8$. To illustrate the computation, we take $e = 8$ and $h = 1$. The Diophantine system then becomes

$$\sum_{i \geqslant 1} m_i = 49, \quad \sum_{i \geqslant 1} m_i^2 = 273.$$

Since $m := m_1 = \cdots = m_8$, we must have $m \leqslant 5$. For $m = 5$ we get

$$\sum_{i \geqslant 9} m_i = 9, \quad \sum_{i \geqslant 9} m_i^2 = 73,$$

which has no solution for $m_i \leqslant 5$. Similarly there are no solutions for $m \in \{1,2,3,4\}$. □

Lemma 3.10 *The only exceptional class* $(d, e; \boldsymbol{m})$ *with* $\mu_2(d, e; \boldsymbol{m})(8) > \frac{1}{2}\sqrt{8}$ *is* $F_2 = (6, 3; 3, 2^{\times 7})$.

Proof Consider an exceptional class with $\mu_2(d, e; \boldsymbol{m})(8) > \frac{1}{2}\sqrt{a}$. By Lemma 3.11 below, $\ell(\boldsymbol{m}) \leqslant 8$. If $\ell(\boldsymbol{m}) \leqslant 7$, Lemma 3.8(i) shows that $\boldsymbol{m} = (1^{\times 7})$; but the only solution of (2-12) with this $\boldsymbol{m}$ is $(3,1;1^{\times 7})$, and $\mu_2(3,1;1^{\times 7})(8) = \frac{7}{5} < \frac{1}{2}\sqrt{8}$. We can thus assume that $\ell(\boldsymbol{m}) = 8$. By Lemma 3.8, the vector $\boldsymbol{m}$ has the form

$$\boldsymbol{m} = (m^{\times 8}) \quad \text{or} \quad \boldsymbol{m} = (m^{\times 7}, m-1) \quad \text{or} \quad \boldsymbol{m} = (m+1, m^{\times 7})$$

for some $m \in \mathbb{N}$.

If $\boldsymbol{m} = (m^{\times 8})$, then the linear of the Diophantine equations yields $8m = 2(d+e)-1$, which is impossible since $8m$ is even and $2(d+e)-1$ is odd.

In the two other cases, Proposition 3.1(iii) and Lemma 3.8(ii) show that $d = 2e$.

If $\boldsymbol{m} = (m^{\times 7}, m-1)$, the Diophantine system becomes

$$8m = 6e, \quad 8m^2 - 2m = 4e^2.$$

Inserting $e = \frac{4}{3}m$ into the second equation leads to $4m^2 = 9m$, which has no solution in $\mathbb{N}$.

If $\boldsymbol{m} = (m+1, m^{\times 7})$, the Diophantine system becomes

$$8m + 2 = 6e, \quad 8m^2 + 2m = 4e^2.$$

Inserting $e = \frac{1}{3}(4m+1)$ into the second equation leads to $4m^2 - 7m - 2 = 0$, whose only integral solution is $m = 2$. Hence $(d,e;\boldsymbol{m}) = (6,3;3,2^{\times 7}) = F_2$. □

The following lemma is a version of [26, Lemma 2.1.3]:

Lemma 3.11 *Let $(d,e;\boldsymbol{m})$ be an exceptional class, and suppose that I is a maximal nonempty open interval such that $\frac{1}{2}\sqrt{a} < \mu_2(d,e;\boldsymbol{m})(a)$ for all $a \in I$. Then there is a unique $a_0 \in I$ such that $\ell(a_0) = \ell(\boldsymbol{m})$. Moreover, $\ell(a) \geqslant \ell(\boldsymbol{m})$ for all $a \in I$.*

Here, the last assertion is proven as follows: If $\ell(a) < \ell(\boldsymbol{m})$, then $\sum_{i \leqslant \ell(a)} m_i^2 < 2de+1$, so that $\langle \boldsymbol{w}(a), \boldsymbol{m} \rangle \leqslant \|\boldsymbol{w}(a)\| \sqrt{2de} = \sqrt{a}\sqrt{2de}$. Hence

$$\mu_2(d,e;\boldsymbol{m})(a) \leqslant \frac{\sqrt{2de}\sqrt{a}}{d+2e} \leqslant \frac{\sqrt{a}}{2}.$$

End of the proof of Proposition 3.6 Suppose to the contrary that $\mu_2(d,e;\boldsymbol{m})(a) > \frac{1}{2}\sqrt{a}$ for some $a \in \left[8\frac{1}{36}, 9\right)$. By Lemma 3.11 we may choose a_0 with $\ell(a_0) = \ell(\boldsymbol{m})$ in the interval I containing a on which this inequality holds.

Assume that $a_0 \leqslant 8$. Then $a_0 \leqslant 8 < a$, and so $8 \in I$. Then Lemma 3.10 shows that $(d,e;\boldsymbol{m}) = F_2$. But F_2 is not obstructive for $a \geqslant 8\frac{1}{36}$.

Hence $a_0 > 8$. We already know from Proposition 3.5 that $c_2(a) = \frac{1}{2}\sqrt{a}$ for $a \geqslant 9$. Hence $a_0 \in (8,9)$. Hence Lemma 3.9 applies, and yields the desired contradiction. □

4 First applications of the reduction method

In this section we first use the reduction method to prove the equivalence (1-2). We then use this method to prove that the obstructions given by the exceptional classes E_n are sharp at their edges, and then to compute $c_b(a)$ at the end points of the first linear step.

As in Section 3.2 we define the *defect* of $(\mu;\boldsymbol{a})=(\mu;a_1,\dots,a_k)$ by $\delta:=\mu-a_1-a_2-a_3$. Then the Cremona transform (2-14) can be written as

$$\mathrm{Cr}(\mu;\boldsymbol{a})=(\mu+\delta;\,a_1+\delta,\,a_2+\delta,\,a_3+\delta,\,a_4,\,\dots,\,a_k).$$

4.1 Proof of the equivalence (1-2)

By continuity we can assume that a is rational. Recall that $E(1,a)\overset{s}{\hookrightarrow}P(\lambda,\lambda b)$ if and only if there exists an embedding (2-2). By the nonsqueezing theorem we must have $\lambda\geqslant 1$. Hence Method 2 formulated in Section 2.2 shows that an embedding (2-2) exists if and only if $\lambda\geqslant\sqrt{\frac{a}{2b}}$ and if the first reduced vector in the orbit of

$$(\lambda(b+1);\lambda b,\lambda,\boldsymbol{w}(a)) \tag{4-1}$$

under standard Cremona moves has no negative entries.

The weight decomposition of the ellipsoid $E((2b-1)\lambda,2b\lambda)$ is $((2b-1)\lambda,\lambda^{\times(2b-1)})$. The main result of [23] thus shows that $E(1,a)\overset{s}{\hookrightarrow}E(\lambda,2b\lambda)$ if and only if

$$B(\boldsymbol{w}(a))\amalg B((2b-1)\lambda)\amalg\coprod_{2b-1}B(\lambda)\overset{s}{\hookrightarrow}B(2b\lambda).$$

Method 2 shows that such an embedding exists if and only if $\lambda\geqslant\sqrt{\frac{a}{2b}}$ and if the first reduced vector in the orbit of

$$(2b\lambda;(2b-1)\lambda,\lambda^{\times(2b-1)},\boldsymbol{w}(a)) \tag{4-2}$$

under standard Cremona moves has no negative entries. Applying $b-1$ standard Cremona moves with defect $\delta=-\lambda$ to the vector (4-2) we reach the vector (4-1). $\square$

In the rest of this paper we will show that besides for the volume constraint $\sqrt{\frac{a}{2b}}$ there are no other obstructions to the embedding problem $E(1,a)\overset{s}{\hookrightarrow}P(\lambda,\lambda b)$ than those given by the exceptional classes E_n and F_n. For this it suffices to show that if we take for λ the value claimed for $c_b(a)$ in Theorem 1.1, then there exists an embedding $E(1,a)\overset{s}{\hookrightarrow}P(\lambda,\lambda b)$. This problem, in turn, we solve by the recipe formulated in Proposition 2.3.

4.2 $c_b(a)$ at $a = 2b + 2k + 1$, and at $a = 2b$ and $a = 2b + 2 + \frac{1}{2b}$

Lemma 4.1 $c_b(2b + 2k + 1) \leqslant \frac{2b+2k+1}{2b+k}$ *for* $k \in \{0, 1, 2, \ldots, \lfloor\sqrt{2b}\rfloor\}$.

Proof Set $\lambda = \frac{2b+2k+1}{2b+k} = 1 + \frac{k+1}{2b+k} \in (1, 2)$. Then one standard Cremona move with $\delta = -1$ takes the vector $(\lambda(b + 1); \lambda b, \lambda, 1^{\times(2b+2k+1)})$ to

$$(\lambda(b + 1) - 1; \lambda b - 1, 1^{\times(2b+2k)}, \lambda - 1).$$

Since $\lambda b - 1 + (b + k)(\lambda - 2) = 0$, applying $b + k$ Cremona moves with $\delta = \lambda - 2$ to this vector yields the vector $(\lambda; (\lambda - 1)^{\times(2b+2k+1)})$, which is reduced, since $\delta = 3 - 2\lambda = \frac{2b-k-2}{2b+k} \geqslant 0$ for $k \leqslant \sqrt{2b}$ and $b \geqslant 2$. □

Lemma 4.2 $c_b(2b) = 1$ *and* $c_b\big(2b + 2 + \frac{1}{2b}\big) = \frac{2b+1}{2b}$.

Proof In view of the volume constraint $c_b(a) \geqslant \sqrt{\frac{a}{2b}}$, it suffices to show the inequalities $c_b(2b) \leqslant 1$ and $c_b\big(2b + 2 + \frac{1}{2b}\big) \leqslant \frac{2b+1}{2b}$.

Set $\lambda = 1$. Then b Cremona moves with $\delta = -1$ take the vector $(b + 1; b, 1^{\times(2b+1)})$ to $(1; 1)$, which is reduced.

Set $\lambda = \frac{2b+1}{2b} = 1 + \frac{1}{2b}$. Then one standard Cremona move with $\delta = -1$ takes the vector $\big(\lambda(b + 1); \lambda b, \lambda, 1^{\times(2b+2)}, \big(\frac{1}{2b}\big)^{\times 2b}\big)$ to

$$\Big(\lambda(b + 1) - 1; \lambda b - 1, 1^{\times(2b+1)}, \Big(\frac{1}{2b}\Big)^{\times(2b+1)}\Big).$$

Since $\lambda b - 1 + b(\lambda - 2) = 0$, applying b Cremona moves with $\delta = \lambda - 2$ yields the vector $\big(\lambda; 1, \big(\frac{1}{2b}\big)^{\times(4b+1)}\big)$. Applying $2b$ Cremona moves with $\delta = \frac{1}{2b}$ yields $\big(\frac{1}{2b}; \frac{1}{2b}\big)$, which is reduced. □

Corollary 4.3 *Theorem 1.1 holds for* $a \in [1, 2b + 3]$.

Proof By Gromov's nonsqueezing theorem, $E(1, 1) \overset{s}{\hookrightarrow} P(\lambda, \lambda b)$ implies $\lambda \geqslant 1$. (In our language this reads $\mu_b(E_0)(1) = 1$ for $E_0 := (1, 0; 1)$.) Since the function c_b is monotone increasing, this and $c_b(2b) = 1$ show that $c_b(a) = 1$ for $a \in [1, 2b]$.

The functions c_b have the scaling property

$$\frac{c_b(\lambda a)}{\lambda a} \leqslant \frac{c_b(a)}{a} \quad \text{for all } \lambda \geqslant 1;$$

see [26, Lemma 1.1.1] for the easy proof. Therefore:

Lemma 4.4 *If for two values $a_0 < a_1$ the points $(a_0, c_b(a_0))$ and $(a_1, c_b(a_1))$ lie on a line through the origin, then the whole segment between these two points belongs to the graph of c_b, that is, c_b is linear on $[a_0, a_1]$.*

Lemmas 3.3(i), 4.1 and 4.2 thus show that the graph of c_b on $[1, 2b+3]$ is as in Figure 2. □

4.3 Organization of the proof of Theorem 1.1

We order the rest of the proof by increasing difficulty.

For $b \in \mathbb{N}_{\geqslant 5}$ and $k = 2, \dots, \lfloor \sqrt{2b} \rfloor - 1$, the intervals $I_b(k)$ and $I_b(k+1)$ enclose the interval $[v_b(k), u_b(k+1)]$, that contains the point $2b+2k+2$. We first show that $c_b(a) = \sqrt{\frac{a}{2b}}$ on this interval. More precisely, we subdivide this interval into its left and right part,

$$L_b(k) := [v_b(k), 2b+2k+2] \quad \text{and} \quad R_b(k) := [2b+2k+2, u_b(k+1)],$$

and show in Section 5 and Section 6 that $c_b(a) = \sqrt{\frac{a}{2b}}$ on $L_b(k)$ and $R_b(k)$, respectively. Theorem 1.1 then follows for all $a \geqslant 2b+5$. Indeed, together with Lemmas 3.3(i) and 4.1, we now know that for $k \geqslant 2$ the edge point and the two end points of the linear steps lie on the graph of $c_b(a)$, and hence by Lemma 4.4 these linear steps belong to $c_b(a)$ entirely. Further, by Proposition 3.5(i), Theorem 1.1 holds for $a \geqslant v_b(\lfloor \sqrt{2b} \rfloor)$.

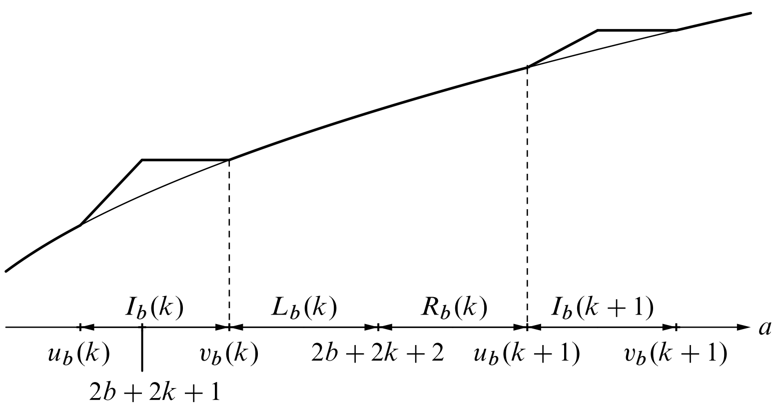


Figure 10

We already know from Corollary 4.3 that Theorem 1.1 holds for $a \leqslant 2b+3$. We are thus left with the interval $[2b+3, 2b+5]$. It suffices to treat the subinterval $[v_b(1), u_b(2)]$. Indeed, we then know that $c_b(2b+3) = c_b(v_b(1))$, whence the second linear step is established, and we already know that the third linear step, which begins at $u_b(2)$, belongs to $c_b(a)$. (Note that for $b=2$ there is no third linear step, but then $u_b(2)=2b+5=9$.) Recall that

$$v_b(1)<\alpha_b<2b+4<\beta_b<u_b(2).$$

We shall treat the interval $[v_b(1),2b+4]$ in Section 7. The case $b=2$ is then complete, since $c_2(8)=\frac{17}{12}=c_2\bigl(8\frac{1}{36}\bigr)$ and in view of Proposition 3.6. The interval $[2b+4,u_b(2)]$ for $b\geqslant 3$ is treated in Section 8. Showing $c_b(a)=\sqrt{\frac{a}{2b}}$ on the intervals $[v_b(1),\alpha_b]$ and $[\beta_b,u_b(2)]$ is the hardest part of the paper, since on these intervals the reduction algorithm is rather intricate. On the other hand, establishing the affine segment over $[\alpha_b,2b+4]$ will be easier, and it turns out that the reduction method establishes the affine steps of $c_B(a)$ and $c_C(a)$ much faster than the positivity of intersection argument used in [26; 14].

Since the embedding functions $c_b(a)$ are continuous, it suffices to compute them on a dense set. In the rest of the paper we shall assume that $a\geqslant 1$ is rational. Hence a has a finite weight expansion $\boldsymbol{w}(a)=(1^{\times\lfloor a\rfloor},w_1^{\times\ell_1},w_2^{\times\ell_2},\dots)$. Sometimes it will be convenient to assume also that $\ell_1\geqslant 1$ or $\ell_2\geqslant 1$ or $\ell_3\geqslant 2$, which holds for a dense set of rational a.

5 The intervals $L_b(k)=[v_b(k),2b+2k+2]$

Recall that

$$v_b(k)=2b\Bigl(\frac{2b+2k+1}{2b+k}\Bigr)^2.$$

Theorem 5.1 *Assume that $b\in\mathbb{N}_{\geqslant 5}$ and $k\in\{2,\dots,\lfloor\sqrt{2b}\rfloor-1\}$. Then $c_b(a)=\sqrt{\frac{a}{2b}}$ for $a\in L_b(k)$.*

Proof The weight expansion at $a\in L_b(k)$ is

$$\boldsymbol{w}(a)=(1^{\times 2(b+k)+1},w_1^{\times\ell_1},w_2^{\times\ell_2},\dots).$$

Define the numbers λ, z_1 and z_2 by

$$\lambda=\sqrt{\frac{a}{2b}}=\sqrt{\frac{2(b+k)+1+w_1}{2b}}=:1+z_1,$$

$$z_2:=(2b+k)\lambda-(2b+2k+1)=\sqrt{\frac{2(b+k)+1+w_1}{2b}}(2b+k)-(2b+2k+1).$$

Lemma 5.2 (i) $2z_1\leqslant 1+z_2$.

(ii) *$z_2\geqslant 0$ and $z_2\leqslant w_1$.*

(iii) *For $k\geqslant 3$ and $\ell_1=1$ we have $w_2+z_2-z_1\geqslant 0$.*

Proof (i) We wish to show that

$$2b+2k \leqslant 2+(2b+k-2)\lambda.$$

We show that this inequality even holds if $w_1 \leqslant 0$ in λ is set to zero, ie that

$$2b+2k \leqslant 2+\sqrt{\frac{2(b+k)+1}{2b}}(2b+k-2).$$

After solving for the root, squaring and multiplying with $2b(2b+k-2)^2$, we find that this inequality is equivalent to

$$4b^2+(2k+1)(k-2)^2+2b(k^2-2k-4) \geqslant 0,$$

which holds true since $k \geqslant 2$ and $b \geqslant 2$.

(ii) Note that $z_2 = 0$ at the left boundary $v_b(k)$ of $L_b(k)$. Since z_2 is increasing on $L_b(k)$, we see that $z_2 \geqslant 0$.

At $v_b(k)$ we have $w_1 \geqslant 0 = z_2$. In order to show that $z_2 \leqslant w_1$ on $L_b(k)$, it thus suffices to check that the derivative of the function $f_{b,k}(w_1) = w_1 - z_2(b,k,w_1)$ is nonnegative, ie

$$f'_{b,k}(w_1) = 1-\frac{1}{2}\sqrt{\frac{2b}{2(b+k)+1+w_1}}\frac{2b+k}{2b} \geqslant 0.$$

This holds if it holds for $w_1 = 0$, ie if

$$\frac{4b}{2b+k} \geqslant \sqrt{\frac{2b}{2b+2k+1}}.$$

This is equivalent to

$$8b(2b+2k+1) \geqslant 4b^2+4bk+k^2,$$

which holds true since $k^2 \leqslant 2b$.

(iii) Fix $k \geqslant 3$ and $b \geqslant 2$. Define the function $f_{b,k}$ on $[v_b(k)-\lfloor v_b(k)\rfloor, 1]$ by

(5-1) $$f_{b,k}(w_1) := w_2+z_2-z_1 = -w_1+(2b+k-1)\lambda-(2b+2k)+1.$$

Then $f'_{b,k} \leqslant 0$. Indeed, this is equivalent to

$$2b+k-1 \leqslant 2\sqrt{2b(2b+2k+1+w_1)},$$

which follows from

$$2b+k \leqslant 2\sqrt{2b(2b+2k+1)}.$$

It therefore suffices to show that $f_{b,k}(1) \geqslant 0$, ie

$$\sqrt{\frac{b+k+1}{b}} \geqslant \frac{2b+2k}{2b+k-1}.$$

Squaring and multiplying by $b(2b+k-1)^2$ this becomes

$$(1+k)((k-3)b+(k-1)^2) \geqslant 0,$$

which holds true since $k \geqslant 3$. $\square$

In view of Proposition 2.3 we wish to transform the vector

$$((b+1)\lambda; b\lambda, \lambda, \boldsymbol{w}(a))$$

to a reduced vector by a finite sequence of Cremona moves. One Cremona move yields

$$\big((b+1)\lambda - 1; b\lambda - 1, 1^{\times 2(b+k)} \parallel z_1, w_1^{\times \ell_1}, w_2^{\times \ell_2}, \dots\big).$$

Here and in the sequel we use the notation explained in Notation 2.4. Next, $b+k$ Cremona moves with $\delta = \lambda - 2 = z_1 - 1$ yield

$$(\lambda + z_2; z_2, w_1^{\times \ell_1}, z_1^{\times 2(b+k)+1}, w_2^{\times \ell_2}, \dots). \tag{5-2}$$

Assume that $z_1 \geqslant w_1$. Since $z_2 \leqslant w_1$, the vector (5-2) reorders to

$$(\lambda + z_2; z_1^{\times 2(b+k)+1}, w_1^{\times \ell_1} \parallel z_2, w_2^{\times \ell_2}, \dots). \tag{5-3}$$

Then $\delta = \lambda + z_2 - 3z_1 = 1 + z_2 - 2z_1 \geqslant 0$ by Lemma 5.2(i). Since all entries of (5-3) are nonnegative, this vector is reduced.

From now on we thus assume that $w_1 \geqslant z_1$. Then the vector (5-2) becomes

$$(\lambda + z_2; w_1^{\times \ell_1} \parallel z_1^{\times 2(b+k)+1}, z_2, w_2^{\times \ell_2}, \dots). \tag{5-4}$$

If $\ell_1 \geqslant 3$, then $\delta = 1 + z_1 + z_2 - 3w_1 \geqslant z_1 + z_2 \geqslant 0$. If $\ell_1 = 2$, then

$$\delta = 1 + z_1 + z_2 - 2w_1 - (z_1 \text{ or } z_2 \text{ or } w_2) \geqslant 1 - (2w_1 + w_2) \geqslant 0.$$

So assume that $\ell_1 = 1$, that is, the vector (5-4) is

$$(\lambda + z_2; w_1 \parallel z_1^{\times 2(b+k)+1}, z_2, w_2^{\times \ell_2}, \dots).$$

Case 1 $(z_1 \geqslant z_2, w_2)$ Then the vector at hand is

$$(\lambda + z_2; w_1, z_1^{\times 2(b+k)+1} \parallel z_2, w_2^{\times \ell_2}, \dots).$$

Hence $\delta = 1+z_1+z_2-w_1-2z_1 = w_2+z_2-z_1$. For $k \geqslant 3$ this number is nonnegative by Lemma 5.2(iii). Assume now that $k=2$ and that $\delta = w_2+z_2-z_1 < 0$. We reduce the above vector $b+2$ times by δ and get

$$(w_2+z_1+z_2+*; * = w_1+(b+2)(w_2+z_2-z_1), z_1, (w_2+z_2)^{\times(2b+4)} \parallel z_2, w_2^{\times \ell_2}, \dots).$$

The order is right by the assumption $w_2+z_2 \leqslant z_1$ and the following lemma. For this vector, $\delta = 0$.

Lemma 5.3 $$w_1+(b+2)(w_2+z_2-z_1) \geqslant z_1.$$

Proof Define the function f_b on $[v_b(2) - \lfloor v_b(2) \rfloor, 1]$ by

$$f_b(w_1) := w_1 + (b+2)(w_2+z_2-z_1) - z_1. \tag{5-5}$$

We compute

$$f_b(w_1) = -(b+1)w_1 + (2b^2+5b+1)\lambda - (2b^2+7b+5), \quad \text{where } \lambda = \sqrt{\frac{2b+5+w_1}{2b}}.$$

We wish to show that $f_b(w_1) \geqslant 0$. We estimate

$$f_b'(w_1) = -(b+1) + \frac{2b^2+5b+1}{2\sqrt{2b(2b+5+w_1)}} \leqslant -(b+1) + \frac{2b^2+5b}{4b} \leqslant 0.$$

Hence $f_b(w_1) \geqslant f_b(1) = -(b+1) + (2b^2+5b+1)\sqrt{\frac{b+3}{b}} - (2b^2+7b+5)$. The right-hand side is nonnegative if and only if

$$\sqrt{\frac{b+3}{b}} \geqslant \frac{2(b^2+4b+3)}{2b^2+5b+1}.$$

Squaring and multiplying by $b(2b^2+5b+1)^2$ we find that this is equivalent to the inequality $(b+3)(b-1)^2 \geqslant 0$, which holds true. □

Case 2 $(z_2 \geqslant z_1, w_2)$ Then $\delta = 1 + z_1 - w_1 - (z_1 \text{ or } w_2) \geqslant 0$.

Case 3 $(w_2 \geqslant z_1, z_2)$ The vector at hand is

$$(\lambda + z_2; w_1, w_2^{\times \ell_2} \parallel z_1^{\times 2(b+k)+1}, z_2, w_3^{\times \ell_3}, w_4^{\times \ell_4}, \dots).$$

Subcase 3(a) $(\ell_2 \geqslant 2)$ Then $\delta = z_1+z_2-w_2$. Assume that $\delta < 0$, ie $w_2 > z_1+z_2$.

If $\ell_2 = 2m_2 \geqslant 2$ is even, we reduce m_2 times by δ and get

$$\big(z_1+z_2+w_2+*; * = w_1 + m_2(z_1+z_2-w_2) \parallel \\ (z_1+z_2)^{\times m_2}, z_1^{\times 2(b+k)+1}, z_2, w_3^{\times \ell_3}, w_4^{\times \ell_4}, \dots\big).$$

Here, $* \geqslant z_1 + z_2$ and $* \geqslant w_3 = w_1 - \ell_2 w_2$ because $m_2 w_2 \leqslant \ell_2 w_2 \leqslant w_1$.

If $z_1 + z_2 \geqslant w_3$, then

$$\delta = w_2 - (z_1 + z_2 \text{ or } z_1 \text{ or } z_2 \text{ or } w_3) \geqslant w_2 - (z_1 + z_2) > 0.$$

If $w_3 \geqslant z_1 + z_2$, then

$$\delta = z_1 + z_2 + w_2 - w_3 - (w_3 \text{ if } \ell_3 \geqslant 2 \text{ and } z_1 + z_2 \text{ or } w_4 \text{ if } \ell_3 = 1).$$

In the first case, $\delta \geqslant 0$ since $w_2 = \ell_3 w_3 + w_4 \geqslant 2w_3$, and in the second case, $\delta = w_2 - w_3 \geqslant 0$ or $\delta = z_1 + z_2 \geqslant 0$.

If $\ell_2 = 2m_2 + 1 \geqslant 3$ is odd, we again reduce m_2 times by δ and get

$$\big(z_1 + z_2 + w_2 + *;\ * = w_1 + m_2(z_1 + z_2 - w_2),\ w_2 \parallel \\ (z_1 + z_2)^{\times m_2}, z_1^{\times 2(b+k)+1}, z_2, w_3^{\times \ell_3}, w_4^{\times \ell_4}, \dots\big).$$

If $z_1 + z_2 \geqslant w_3$, then $\delta = 0$. If $w_3 > z_1 + z_2$, then $\delta = z_1 + z_2 - w_3 < 0$. The vector at hand is

$$(z_1 + z_2 + w_2 + *;\ * = w_1 + m_2(z_1 + z_2 - w_2),\ w_2,\ w_3^{\times \ell_3} \parallel (z_1 + z_2)^{\times m_2},\ w_4^{\times \ell_4},\ \dots),$$

and applying one more Cremona transform yields the vector

$$(z_1 + z_2 + w_2 + *;\ *,\ w_2 + z_1 + z_2 - w_3,\ w_3^{\times \ell_3 - 1} \parallel (z_1 + z_2)^{\times m_2 + 1},\ w_4^{\times \ell_4},\ \dots),$$

where now $* = w_1 + m_2(z_1 + z_2 - w_2) + (z_1 + z_2 - w_3)$. The ordering holds since if $\ell_3 \geqslant 2$ then $w_2 + z_1 + z_2 - w_3 \geqslant w_2 - w_3 \geqslant w_3$, and if $\ell_3 = 1$ then $w_2 + z_1 + z_2 - w_3 = z_1 + z_2 + w_4$. Now

$$\delta = w_3 - (w_3 \text{ or } z_1 + z_2 \text{ or } w_4) \geqslant 0.$$

Subcase 3(b) ($\ell_2 = 1$) Then $\delta = 1 + z_1 + z_2 - w_1 - w_2 - x = z_1 + z_2 - x$ with $x \in \{z_1, z_2, w_3\}$. If $x \in \{z_1, z_2\}$ then $\delta \in \{z_2, z_1\} \geqslant 0$. If $x = w_3$, then the vector at hand is

$$(\lambda + z_2; w_1, w_2, w_3^{\times \ell_3} \parallel z_1^{\times 2(b+k)+1}, z_2, w_4^{\times \ell_4}, \dots).$$

Notice that $w_2 = 1 - w_1$ and $w_3 = w_1 - w_2$. We have $\delta = z_1 + z_2 - w_3$. If $w_3 > z_1 + z_2$, we apply one more Cremona transform and obtain

$$(z_1 + z_2 + w_2 + *;\ * = z_1 + z_2 + w_2,\ z_1 + z_2 + w_2 - w_3,\ w_3^{\times(\ell_3 - 1)} \parallel \\ z_1 + z_2,\ z_1^{\times 2(b+k)+1},\ z_2,\ w_4^{\times \ell_4},\ \dots).$$

The ordering is right since if $\ell_3 \geqslant 2$ then $w_2 \geqslant 2w_3$, and if $\ell_3 = 1$ then $w_2 - w_3 = w_4$.

If $\ell_3 \geqslant 2$ then $\delta = 0$.

If $\ell_3 = 1$ then $\delta = w_3 - (z_1 + z_2) > 0$ or $\delta = w_3 - w_4 \geqslant 0$.

The proof of Theorem 5.1 is complete. □

6 The intervals $R_b(k) = [2b + 2k + 2\,,u_b(k+1)]$

Theorem 6.1 *Assume that $b \in \mathbb{N}_{\geqslant 5}$ and that $k \in \{2, \dots, \lfloor\sqrt{2b}\rfloor - 1\}$. Then $c_b(a) = \sqrt{\frac{a}{2b}}$ for $a \in R_b(k)$.*

Proof For notational convenience we shift the index k by one, and prove that $c_b(a) = \sqrt{\frac{a}{2b}}$ for $a \in R_b(k-1)$ and $k \in \{3, \dots, \lfloor\sqrt{2b}\rfloor\}$.

We start with three inequalities that will be useful later on.

Lemma 6.2 *We have:*

(i) $$\frac{2b+2k}{2b} \geqslant \left(\frac{2b+2k-2}{2b+k-2}\right)^2 \quad \text{for } k \geqslant 4.$$

(ii) $$\sqrt{\frac{2b+2k}{2b}} \geqslant \frac{2b+2k}{2b+k}.$$

(iii) $$\frac{2b+k}{2b} \leqslant \frac{2b+2k}{2b+k-1} \qquad \text{if } k^2 \leqslant 2b.$$

Proof (i) This is equivalent to

$$\frac{(2b+2k)(2b+k-2)^2 - 2b(2b+2k-2)^2}{2b(2b+k-2)^2} \geqslant 0,$$

which holds true for $k \geqslant 4$ because the numerator of the left-hand side can be written as $2k(b(k-4) + (k-2)^2)$.

(ii) This follows from $(2b+k)^2 - 2b(2b+2k) = k^2$.

(iii) This follows from

$$(2b+2k)(2b) - (2b+k-1)(2b+k) = 2b + k - k^2,$$

since $2b + k - k^2 \geqslant k > 0$ by assumption. □

Except possibly for the right endpoint, which we can neglect, the weight expansion at $a \in R_b(k-1) = \left[2b+2k, 2b+2k+\frac{k^2}{2b}\right]$ is

$$\boldsymbol{w}(a) = (1^{\times 2b+2k}, w_1^{\times \ell_1}, w_2^{\times \ell_2}, \dots).$$

Set $\lambda = \sqrt{\frac{a}{2b}}$. We wish to transform the vector

$$((b+1)\lambda; b\lambda, \lambda, \boldsymbol{w}(a)) \tag{6-1}$$

to a reduced vector by a sequence of Cremona moves. Define the numbers

$$\begin{aligned} z_1 &:= \lambda - 1, \\ y_1 &:= (2b+k)\lambda - (2b+2k-1), \\ z_2 &:= y_1 - \lambda. \end{aligned}$$

Then $z_1, y_1 \geqslant 0$ and $z_2 \in [0,1]$. Indeed, as we have seen in Lemma 5.2(ii), $z_2 = 0$ at the left endpoint of $L_b(k-1)$, and $z_2 \leqslant 1$ since $\lambda \leqslant \frac{2b+k}{2b}$, using also Lemma 6.2(iii). Applying one Cremona move to (6-1) we obtain

$$((b+1)\lambda - 1;\ b\lambda - 1,\ 1^{\times 2b+2k-1} \parallel z_1,\ w_1^{\times \ell_1},\ \dots).$$

Applying $b+k-1$ Cremona transforms with $\delta = \lambda - 2$ and reordering we obtain

$$(y_1; 1 \parallel z_1^{\times 2b+2k-1}, z_2, w_1^{\times \ell_1}, \dots). \tag{6-2}$$

6.1 The case $z_1 \geqslant w_1$

For $z_1 \geqslant w_1$, assume first that $k \geqslant 4$, or that $k = 3$ and $z_2 \geqslant z_1$. If $z_2 \geqslant z_1$, then the vector (6-2) reorders to

$$(y_1; 1, z_2, z_1^{\times 2b+2k-1}, w_1^{\times \ell_1}, \dots).$$

This vector has defect $\delta = y_1 - 1 - z_2 - z_1 = 0$ and hence is reduced. If $z_1 \geqslant z_2$, then the vector (6-2) reorders to the vector

$$(y_1; 1, z_1^{\times 2b+2k-1} \parallel z_2, w_1^{\times \ell_1}, \dots), \tag{6-3}$$

which for $k \geqslant 4$ is reduced, since then $\delta = y_1 - 1 - 2z_1 = y_1 - (2\lambda - 1) \geqslant 0$ by Lemma 6.2(i) and the fact that $\lambda \geqslant \sqrt{\frac{2b+2k}{2b}}$.

Assume now that $k = 3$ and $z_1 \geqslant z_2$. If $\widehat{\delta} := y_1 - 1 - 2z_1 \geqslant 0$, the vector (6-3) is reduced. Otherwise, we apply $b+2$ Cremona moves to obtain

$$(y_1 + (b+2)\widehat{\delta};\ 1 + (b+2)\widehat{\delta},\ z_1,\ z_2^{\times 2b+5} \parallel w_1^{\times \ell_1},\ \dots). \tag{6-4}$$

The ordering is right by the following claim, and the defect is $y_1 - 1 - z_1 - z_2 = 0$, whence this vector is reduced.

Claim *Assume that $k = 3$. Then:*

(i) $1 + (b+2)\widehat{\delta} \geqslant z_1$.

(ii) *If $z_1 \geqslant w_1$, then $z_2 \geqslant w_1$.*

Proof Inequality (i) is equivalent to

$$(2b^2+5b+1)\lambda \geqslant (2b^2+8b+6).$$

It suffices to check this inequality for $\lambda = \sqrt{\frac{2b+6}{2b}}$, where it is equivalent to the inequality $3b^2+10b+3 \geqslant 0$, which holds true for all $b \geqslant 1$.

For (ii), we know that $\lambda - 1 \geqslant w_1$, ie

$$a \leqslant \lambda + (2b+5). \tag{6-5}$$

Since $a = 2b\lambda^2$, this is equivalent to

$$\lambda \geqslant \frac{1+\sqrt{1+8b(2b+5)}}{4b}. \tag{6-6}$$

We wish to show that $z_2 - 2 \geqslant w_1$, ie $a \leqslant 1+(2b+2)\lambda$. In view of (6-5), this will hold if $\lambda + (2b+5) \leqslant 1+(2b+2)\lambda$, ie

$$\frac{2b+4}{2b+1} \leqslant \lambda. \tag{6-7}$$

By (6-6), this would follow from

$$\frac{2b+4}{2b+1} \leqslant \frac{1+\sqrt{1+8b(2b+5)}}{4b}.$$

Isolating the root and squaring, this becomes the true inequality $72b/(2b+1)^2 \geqslant 0$. □

6.2 The case $w_1 \geqslant z_1$

Assume now that

$$w_1 \geqslant z_1. \tag{6-8}$$

The vector (6-2) in question is

$$(y_1; 1 \parallel w_1^{\times \ell_1}, z_1^{\times 2b+2k-1}, z_2, \dots). \tag{6-9}$$

Define

$$z_3 := y_1 - 1 - w_1.$$

Note that $z_3 \geqslant 0$ on our interval, and $z_3 = 0$ at the right endpoint $a = \frac{(2b+k)^2}{2b}$. The significance of z_3 and of the following lemma will become clear later.

Lemma 6.3 *If $z_3 \geqslant w_1$, then the vector (6-9) is reduced.*

Proof For $\delta := y_1 - 1 - z_2 - w_1$ we have $z_2 + \delta = z_3$ and $w_1 + \delta = z_1$. Applying one Cremona move to

$$(y_1; 1 \parallel z_2, w_1^{\times \ell_1}, z_1^{\times 2b+2k-1}, \dots)$$

we thus obtain

$$(y_1 + \delta;\ 1 + \delta,\ z_3 \parallel w_1^{\times \ell_1 - 1},\ z_1^{\times 2b+2k},\ \dots). \tag{6-10}$$

The ordering is right because $z_3 \geqslant w_1 \geqslant z_1$ by assumption and by (6-8). The defect of (6-10) is thus

$$y_1 - 1 - z_3 - (w_1 \text{ or } z_1 \text{ or } w_2) \geqslant y_1 - 1 - z_3 - w_1 = 0. \qquad \square$$

From now on we thus assume that

$$w_1 \geqslant z_3. \tag{6-11}$$

Lemma 6.4 $$z_2 \geqslant z_1, z_3.$$

Proof The inequality $z_2 \geqslant z_1$ translates to

$$(2b + k - 1)\lambda - (2b + 2k - 1) \geqslant \lambda - 1,$$

or, equivalently,

$$\lambda \geqslant \frac{2b + 2k - 2}{2b + k - 2}. \tag{6-12}$$

But we know that $\lambda \geqslant \sqrt{\frac{2b+2k}{2b}}$, whence in the case $k \geqslant 4$ the inequality (6-12) follows from Lemma 6.2(i). In the case $k = 3$, (6-12) is (6-7).

The inequality $z_2 \geqslant z_3$ is $-\lambda \geqslant -1 - w_1$. This is equivalent to $\lambda - 1 \leqslant w_1$, which follows from (6-8). $\square$

The rest of the proof of Theorem 6.1 is divided into the cases $\ell_1 = 2m$ even and $\ell_1 = 2m + 1$ odd.

Case I ($\ell_1 = 2m$ even) We can assume by continuity that $\ell_1 > 0$, so that $m \geqslant 1$. By applying m Cremona transforms to the vector (6-9) with $\delta = y_1 - 1 - 2w_1$ we obtain

$$(y_2 + y_1 - 1; y_2 \parallel z_1^{\times 2b+2k-1}, z_2, z_3^{\times \ell_1}, w_2^{\times \ell_2}, \dots), \tag{6-13}$$

where $y_2 := 1 + m(y_1 - 1 - 2w_1)$. The ordering is right by the previous and the next lemma.

Lemma 6.5 $$y_2 \geqslant z_2, w_2.$$

Proof The inequality $y_2 \geqslant z_2$ is equivalent to

$$1 + m(y_1 - 1 - 2w_1) \geqslant y_1 - \lambda.$$

Since $\ell_1 w_1 \leqslant 1$ and $\lambda \geqslant 1$, it suffices to show that $(m-1)(y_1-1) \geqslant 0$. This follows since $y_1 \geqslant 1$, by Lemma 6.2(ii).

The inequality $y_2 \geqslant w_2$ is equivalent to

$$1 + m(y_1 - 1 - 2w_1) \geqslant w_2.$$

Since $\ell_1 w_1 \leqslant 1$, it suffices to show that $m(y_1 - 1) \geqslant w_2$. For this, it suffices to show that $y_1 - 1 \geqslant w_1$, ie $a \leqslant \lambda(2b+k)$. This follows from the fact that $a \leqslant \frac{(2b+k)^2}{2b}$. $\square$

Lemma 6.6 *If $z_3 \geqslant w_2$, then the vector* (6-13) *is reduced.*

Proof Assume that $z_3 \geqslant w_2$. If $z_1 \geqslant z_3$, then (6-13) is

$$(y_2 + y_1 - 1; y_2, z_2, z_1^{\times 2b+2k-1}, z_3^{\times \ell_1}, w_2^{\times \ell_2} \parallel \dots),$$

which is reduced. Hence we can assume that $z_3 \geqslant z_1$. In this case, we apply one Cremona transform to

$$(y_2 + y_1 - 1;\ y_2,\ z_2,\ z_3^{\times \ell_1} \parallel z_1^{\times 2b+2k-1},\ w_2^{\times \ell_2}, \dots)$$

with $\delta = z_1 - z_3$ and obtain

$$(y_2 + y_1 - 1 + \delta;\ y_2 + \delta,\ z_2 + \delta,\ z_3^{\times \ell_1 - 1} \parallel z_3 + \delta,\ z_1^{\times 2b+2k-1},\ w_2^{\times \ell_2}, \dots)$$

since $\ell_1 \geqslant 2$. First note that $z_3 + \delta = z_1 \geqslant 0$. To see that the ordering is right, we need to check that $z_2 + \delta \geqslant z_3$. This is equivalent to $y_1 - 1 \geqslant 2z_3$, which is equivalent to $y_1 - 1 - 2w_1 \leqslant 0$, which holds by (6-11). Since the defect vanishes, this vector is reduced. $\square$

From now on we thus assume that

$$w_2 \geqslant z_3. \tag{6-14}$$

Lemma 6.7 *If $z_1 \geqslant w_2$, then the vector* (6-13) *is reduced.*

Proof Assume that $z_1 \geqslant w_2$. Then the vector (6-13) is

$$(y_2 + y_1 - 1;\ y_2,\ z_2,\ z_1^{\times 2b+2k-1},\ w_2^{\times \ell_2},\ z_3^{\times \ell_1} \parallel \dots)$$

with defect $y_1 - 1 - z_2 - z_1 = 0$. $\square$

From now on we thus assume that

$$w_2 \geqslant z_1. \tag{6-15}$$

By now, our vector is

$$(y_2+y_1-1;\ y_2,\ w_2^{\times \ell_2} \,\|\, z_2,\ z_1^{\times 2b+2k-1},\ z_3^{\times \ell_1},\ w_3^{\times \ell_3},\ \dots) \quad \text{if } w_2 \geqslant z_2, \tag{6-16}$$

$$(y_2+y_1-1;\ y_2,\ z_2,\ w_2^{\times \ell_2} \,\|\, z_1^{\times 2b+2k-1},\ z_3^{\times \ell_1},\ w_3^{\times \ell_3},\ \dots) \quad \text{if } z_2 \geqslant w_2. \tag{6-17}$$

Subcase 1 $(\ell_2 \geqslant 2)$ In case (6-16) we have $\delta \geqslant y_1-1-w_1$, since $2w_2 \leqslant w_1$. Since $y_1-1-w_1 = z_3 \geqslant 0$, the vector is reduced.

In case (6-17) we have $\delta = z_1 - w_2 < 0$. Applying one Cremona transform yields

$$(y_2+y_1-1+\delta;\ y_2+\delta,\ z_2+z_1-w_2,\ w_2^{\times \ell_2-1} \,\|\, z_1^{\times 2b+2k},\ z_3^{\times \ell_1},\ w_3^{\times \ell_3},\ \dots). \tag{6-18}$$

The ordering is right since $z_2+z_1 \geqslant 2w_2$. Indeed, this is equivalent to $y-1 \geqslant 2w_2$. Since $w_1 \geqslant 2w_2$, this follows from $y_1-1 \geqslant w_1$, which holds because $y_1-1-w_1 = z_3 \geqslant 0$. The defect of (6-18) vanishes.

Subcase 2 $(\ell_2 = 1)$ We again distinguish two cases.

Assume first that $w_3 \geqslant z_2$. We are then in case (6-16), and, since $z_2 \geqslant z_1$ and $z_2 \geqslant z_3$, the vector at hand is

$$(y_2+y_1-1;\ y_2,\ w_2,\ w_3^{\times \ell_3},\ z_2 \,\|\, \dots).$$

This vector is reduced, since $w_1 = w_2 + w_3$ and hence $\delta = y_1 - 1 - w_2 - w_3 = z_3$.

Assume now that either $w_2 \geqslant z_2 \geqslant w_3$ or $z_2 \geqslant w_2$. Since also $z_2 \geqslant z_1$ and $z_2 \geqslant z_3$, in both (6-16) and (6-17) we have $\delta = z_1 - w_2$. Further, $w_2 = w_1 - w_3$ since $\ell_2 = 1$, and so $z_2 + \delta = z_2 + z_1 - w_2 = w_3 + z_3$. Hence both vectors transform to

$$(y_2+y_1-1+\delta;\ y_2+\delta \,\|\, w_3+z_3,\ z_1^{\times 2b+2k},\ z_3^{\times \ell_1},\ w_3^{\times \ell_3},\ \dots).$$

This vector is reduced after reordering: if $w_3 + z_3 \geqslant z_1$, then

$$\delta = z_1 + z_2 - w_3 - z_3 - (z_1 \text{ or } z_3 \text{ or } w_3) = w_2 - (z_1 \text{ or } z_3 \text{ or } w_3) \geqslant 0$$

by (6-14) and (6-15), and if $z_1 \geqslant w_3 + z_3$, then $\delta = z_1 + z_2 - 2z_1 = z_2 - z_1 \geqslant 0$.

Case II $(\ell_1 = 2m+1 \text{ odd})$ We start from the vector (6-9). By applying $m \geqslant 0$ Cremona transforms with $\delta = y_1 - 1 - 2w_1$ we obtain

$$(\hat{y}_2+y_1-1;\ \hat{y}_2,\ z_3^{\times(\ell_1-1)},\ w_1,\ z_1^{\times 2b+2k-1},\ z_2,\ w_2^{\times \ell_2},\ \dots),$$

where $\hat{y}_2 := 1 + m(y_1 - 1 - 2w_1)$.

Now apply another Cremona transform to the partially reordered vector

$$(\hat{y}_2 + y_1 - 1;\ \hat{y}_2,\ w_1,\ z_2,\ z_1^{\times 2b+2k-1},\ z_3^{\times(\ell_1-1)},\ w_2^{\times \ell_2},\ \dots).$$

With $\delta = y_1 - 1 - w_1 - z_2 = z_1 - w_1$ we obtain

$$(\hat{y}_2 + y_1 - 1 + \delta;\ \hat{y}_2 + \delta \parallel z_1^{\times 2b+2k},\ z_3^{\times \ell_1},\ w_2^{\times \ell_2},\ \dots) \tag{6-19}$$

since $w_1 + \delta = z_1$ and $z_2 + \delta = z_3$. We are again assuming, by continuity, that $\ell_2 \geqslant 1$. The ordering is right in view of the following lemma:

Lemma 6.8 (i) $\hat{y}_2 + \delta \geqslant z_1$.

(ii) $\hat{y}_2 + \delta \geqslant z_3$.

(iii) $\hat{y}_2 + \delta \geqslant w_2$.

Proof Using $1 = \ell_1 w_1 + w_2$ and $y_1 - 1 = z_1 + z_2$, we compute

$$\hat{y}_2 + \delta = (m+1)(z_1 + z_2) - z_2 + w_2.$$

Assertions (i) and (iii) follow at once. Assertion (ii) follows at once for $m \geqslant 1$, and for $m = 0$ it also holds, since then $w_1 + w_2 = 1 \geqslant z_2$. □

We now show that the vector (6-19) is reduced, or can be transformed in one step to a reduced vector. (We will need to transform the vector in only one case). In view of Lemma 6.8, we just have to consider the various possibilities for the orderings of z_1, z_3 and w_2. Denote by δ_* the defect of the reordering of (6-19).

Case 1 $(z_1 \geqslant z_3, w_2)$ Then $\delta_* = y_1 - 1 - 2z_1 = z_2 - z_1 \geqslant 0$ by Lemma 6.4.

Case 2 $(z_3 \geqslant z_1, w_2)$ Then $\delta_* \geqslant y_1 - 1 - 2z_3 = w_1 - z_3 \geqslant 0$ by (6-11).

Case 3 $(w_2 \geqslant z_1, z_3)$ Then the vector (6-19) is

$$(\hat{y}_2 + y_1 - 1 + \delta;\ \hat{y}_2 + \delta,\ w_2^{\times \ell_2} \parallel z_1^{\times 2b+2k},\ z_3^{\times \ell_1},\ w_3^{\times \ell_3},\ \dots). \tag{6-20}$$

Subcase $(\ell_2 \geqslant 2)$ Then (6-20) is reduced if $y_1 - 1 \geqslant 2w_2$. We know that $2w_2 \leqslant w_1$. Hence it suffices to show that $y_1 - 1 \geqslant w_1$, which follows from the fact that $z_3 \geqslant 0$.

Subcase $(\ell_2 = 1)$ We distinguish three cases.

Assume first that $w_3 \geqslant z_1, z_3$. Then (6-20) is reduced, since

$$\delta_* = y_1 - 1 - (w_2 + w_3) = y_1 - 1 - w_1 = z_3.$$

Assume next that $z_3 \geqslant z_1, w_3$. Then (6-20) is reduced, since

$$\delta_* = y_1 - 1 - w_2 - z_3 = w_1 - w_2.$$

Assume finally that $z_1 \geqslant z_3, w_3$. Then the vector in question is

$$(\hat{y}_2 + y_1 - 1 + \delta;\ \hat{y}_2 + \delta,\ w_2,\ z_1^{\times 2b+2k} \,\|\, z_3^{\times \ell_1},\ w_3^{\times \ell_3},\ \dots).$$

If $\widehat{\delta} := y_1 - 1 - w_2 - z_1 = z_2 - w_2 \geqslant 0$, this vector is reduced. Otherwise, we apply one Cremona transform and obtain

$$(\hat{y}_2 + y_1 - 1 + \delta + \widehat{\delta};\ \hat{y}_2 + \delta + \widehat{\delta},\ w_2 + \widehat{\delta},\ z_1 + \widehat{\delta},\ z_1^{\times 2b+2k-1},\ \dots). \tag{6-21}$$

Note that $z_1 + \widehat{\delta} = y_1 - 1 - w_2 \geqslant y_1 - 1 - w_1 = z_3 \geqslant 0$ and that $w_2 + \widehat{\delta} = z_2 \geqslant z_1$ by Lemma 6.4. Hence (6-21) reorders to the vector

$$(\hat{y}_2 + y_1 - 1 + \delta + \widehat{\delta};\ \hat{y}_2 + \delta + \widehat{\delta},\ z_2,\ z_1^{\times 2b+2k-1},\ \dots)$$

which is reduced, since its defect is $y_1 - 1 - z_2 - z_1 = 0$.

The proof of Theorem 6.1 is finally complete. □

7 The interval $[v_b(1), 2b+4]$

Recall that for $b \in \mathbb{N}_{\geqslant 2}$ we defined $v_b(1) := 2b\big(\frac{2b+3}{2b+1}\big)^2$ and

$$\alpha_b := \frac{1}{b}\big(b^2 + 2b + \sqrt{(b^2+2b)^2 - 1}\big) \in \,]v_b(1), 2b+4[.$$

Theorem 7.1 *For every $b \in \mathbb{N}_{\geqslant 2}$ we have*

$$c_b(a) = \begin{cases} \sqrt{\dfrac{a}{2b}} & \text{if } a \in [v_b(1), \alpha_b], \\[2ex] \dfrac{ba+1}{2b(b+1)} & \text{if } a \in [\alpha_b, 2b+4]. \end{cases}$$

In particular, $c_b(\alpha_b) = \sqrt{\frac{\alpha_b}{2b}}$ and $c_b(2b+4) = 1 + \frac{2b+1}{2b(b+1)}$.

Proof Let $a \in [v_b(1), 2b+4]$ be a rational number. For $w_1(b) = v_b(1) - (2b+3)$ we compute $w_1'(b) = 16/(2b+1)^3$. Hence $w_1(b) \geqslant w_1(2) = \frac{21}{25} > \frac{5}{6}$ for $b \geqslant 2$, and so $\ell_1 = 1$ and $\ell_2 \geqslant 5$. The weight expansion of a thus has the form

$$\boldsymbol{w}(a) = (1^{\times(2b+3)}, w_1, w_2^{\times \ell_2}, \dots, w_N^{\times \ell_N}).$$

We wish to show that for $\lambda = c_b(a)$ as in the theorem, the vector $((b+1)\lambda; b\lambda, \lambda, \boldsymbol{w}(a))$ can be reduced to a reduced vector.

7.1 The interval $[v_b(1), \alpha_b]$

Assume that $a \in [v_b(1), \alpha_b]$. Then $\lambda = \sqrt{\frac{a}{2b}}$. Define the numbers

$$\begin{aligned} z_1 &:= \lambda - 1, \\ z_2 &:= (2b+1)\lambda - (2b+3), \\ z_3 &:= (2b+1)\lambda - (a-1), \\ z_4 &:= b(z_3 - z_1) + w_1, \\ z_5 &:= 2b(b+1)\lambda - (ba+1), \\ z_6 &:= b(2z_5 + z_1 - z_4 - 2z_3) + z_4. \end{aligned}$$

In the following, the symbol $\overset{e}{=}$ means that an identity is readily checked by expanding the relevant z_i as polynomials of degree two in λ with coefficients polynomials in b. For instance,

$$z_3 = 1 + z_2 - w_1 \overset{e}{=} z_1 + z_5 - z_4, \tag{7-1}$$

$$z_6 \overset{e}{=} b(2b(b+1)-1)\lambda - (b^2 a - w_1). \tag{7-2}$$

In this section, all newly created numbers will be one of $z_1, \dots, z_6$ or 0, and we shall write down each z_i of every vector. In other words, the dots $\dots$ in any vector are either w_j or 0.

7.1.1 Inequalities

Lemma 7.2 *On the interval $[v_b(1), \alpha_b]$ the following inequalities hold true:*

(i) $b\lambda - 1 \geqslant 1$ *and* $w_1 \geqslant z_1 \geqslant w_2$.

(ii) $w_1 \geqslant 1 - z_1 + z_2 \geqslant z_1 \geqslant z_2$.

(iii) $z_1 \geqslant z_3 \geqslant z_2, w_2$.

(iv) $z_1 \geqslant z_5$. *Moreover,* $z_5 \geqslant z_3$ *is equivalent to* $z_4 \geqslant z_1$.

(v) $z_4 \geqslant z_3$.

(vi) $z_6 \geqslant z_2$, z_5, w_2.

(vii) *If* $b \geqslant 3$, *then* $z_1 - z_4 + 2z_5 - 2w_2 \geqslant 0$.

(viii) $z_i \geqslant 0$ *for all* $i \in \{1, \dots, 6\}$.

Proof (i) We have $b\lambda - 1 \geqslant b - 1 \geqslant 1$. In order to prove $w_1 \geqslant z_1$, we show that the function

$$f_b(a) := w_1 - z_1 = a - (2b+2) - \sqrt{\frac{a}{2b}}$$

is nonnegative. Since $f_b'(a) = 1 - \frac{1}{4b}\sqrt{\frac{2b}{a}} > 0$, it suffices to see that $f_b(v_b(1)) = (4b^2-5)/(2b+1)^2 \geqslant 0$, which holds true for $b \geqslant 2$.

To prove $z_1 \geqslant w_2$, define the function $f_b(a) := z_1 - w_2 = \sqrt{\frac{a}{2b}} + a - (2b+5)$. Since $f_b'(a) = \frac{1}{4b}\sqrt{\frac{2b}{a}} + 1 > 0$, it suffices to see that $f_b(v_b(1)) = (4b-2)/(2b+1)^2 \geqslant 0$, which holds true for $b \geqslant 2$.

(ii) We compute

$$1 - z_1 + z_2 = 2b(\lambda - 1) - 1 \geqslant \lambda - 1 = z_1.$$

This proves the second inequality, and that the first inequality $w_1 \geqslant 1 - z_1 + z_2$ is equivalent to $2b\lambda^2 - 2b\lambda - 2 \geqslant 0$. Since the left-hand side is increasing for $\lambda \geqslant 1$, it suffices to check this inequality at $\lambda(v_b(1)) = \frac{2b+3}{2b+1}$, where it becomes $(4b-2)/(2b+1)^2 \geqslant 0$.

The third inequality $z_1 \geqslant z_2$ is equivalent to $\sqrt{2ab} \leqslant 2b+2$. Squaring this leads to $a \leqslant 2b + 4 + \frac{2}{b}$, which is verified for $a \leqslant \alpha_b < 2b+4$.

(iii) The inequality $z_1 \geqslant z_3$ is equivalent to $w_1 \geqslant 1 - z_1 + z_2$, hence true. The other two inequalities follow from $z_3 = z_2 + w_2$.

(iv) The inequality $z_1 \geqslant z_5$ is equivalent to $a \geqslant (2b^2+2b-1)^2/2b^3$. This inequality is satisfied since $(2b^2+2b-1)^2/2b^3 \leqslant v_b(1)$ is equivalent to $8b^3 + 12b^2 - 1 \geqslant 0$, which is true for $b \geqslant 2$.

The inequality $z_5 \geqslant z_3$ is equivalent to $z_4 \geqslant z_1$ since $z_3 = z_1 + z_5 - z_4$.

(v) Define the function $f_b(\lambda) := z_4 - z_3 \stackrel{\text{e}}{=} \lambda(2b^2 - 2b - 1) - (b-2)2b\lambda^2 - 4$. For $b = 2$ we compute $f_2(\lambda) = 3\lambda - 4 \geqslant f_2(\lambda(v_2(1))) = \frac{1}{5} > 0$. For $b \geqslant 3$ we have

$$f_b'(\lambda) = 2b^2 - 2b - 1 - 4b(b-2)\lambda \leqslant -2b^2 + 6b - 1 \leqslant -1$$

since $\lambda \geqslant 1$. It thus suffices to show that $f_b(\lambda) > 0$ at $\lambda = \sqrt{\frac{2b+4}{2b}}$, that is,

$$\sqrt{\frac{2b+4}{2b}}(2b^2 - 2b - 1) \geqslant 2b^2 - 4.$$

Squaring both sides leads to $4b^2 - 7b + 2 \geqslant 0$ which is verified for $b \geqslant 3$.

(vi) The first inequality means that the function

$$f_b(a) = z_6 - z_2 \stackrel{\text{e}}{=} (2b^3 + 2b^2 - 3b - 1)\lambda + (1 - b^2)a$$

is nonnegative for $a \in [v_b(1), \alpha_b]$. Equivalently,

$$\frac{1}{\sqrt{2b}}(2b^3 + 2b^2 - 3b - 1) \geqslant \sqrt{a}(b^2 - 1).$$

It suffices to show this inequality for $a = 2b+4$, ie

$$\frac{1}{2b}(2b^3+2b^2-3b-1)^2 \geqslant (2b+4)(b^2-1)^2.$$

This is equivalent to $(b-1)^2 \geqslant 0$, which holds true.

We next show that the function

$$f_b(a) = z_6 - z_5 \overset{\mathrm{e}}{=} -2-2b+(2b^3-3b)\lambda+(1+b-b^2)a$$

is nonnegative for $a \in [v_b(1), \alpha_b]$.

If $b=2$, then $f_b(a) = -a+5\sqrt{a}-6 > 0$ on $[2b+3, 2b+4] = [7,8]$.

For $b \geqslant 3$ we compute that

$$f_b'(a) = (2b^3-3b)\lambda_b'(a)+(1+b-b^2)$$

is negative on $[v_b(1), \alpha_b]$, since $\lambda_b'(a) = 1/(2\sqrt{2ab})$ is decreasing and since $f_b'(2b) = \frac{1}{4}+b-\frac{b^2}{2} < 0$ for $b \geqslant 3$. It thus suffices to show that

$$f_b(2b+4) = 2(1+2b-b^2-b^3)+(2b^3-3b)\sqrt{\frac{b+2}{b}}$$

is positive. This is equivalent to $b^2+2b-4 \geqslant 0$, which holds true.

We finally show that the function

$$f_b(a) = z_6 - w_2 \overset{\mathrm{e}}{=} -7-4b+b(2b^2+2b-1)\lambda+(2-b^2)a$$

is nonnegative for $a \in [v_b(1), \alpha_b]$.

If $b=2$, then $f_b(a) = -2a+11\sqrt{a}-15 > 0$ on $[2b+3, 2b+4] = [7,8]$.

For $b \geqslant 3$ we compute that

$$f_b'(a) = b(2b^2+2b-1)\lambda_b'(a)+2-b^2$$

is negative on $[v_b(1), \alpha_b]$, since $f_b'(2b) = \frac{1}{4}(2b^2+2b-1)+2-b^2 < 0$ for $b \geqslant 3$. It thus suffices to show that

$$f_b(2b+4) = 1-2b^2(b+2)+b(2b^2+2b-1)\sqrt{\frac{b+2}{b}}$$

is positive. This is equivalent to $b^2+2b-1 \geqslant 0$, which holds true.

(vii) We compute

$$\delta_b(a) := z_1 - z_4 + 2z_5 - 2w_2 \overset{\mathrm{e}}{=} -8 - 4b + (1+4b+2b^2)\lambda + (1-b)a$$

and

$$\delta_b'(a) = 1 - b + \frac{2b^2+4b+1}{2\sqrt{2}\sqrt{ab}}.$$

Assume first that $b = 3$. Then $\delta_3(a) = -20 + \frac{31}{\sqrt{6}}\sqrt{a} - 2a$. Since

$$\delta_3'(a) = -2 + \frac{31}{2\sqrt{6}\sqrt{a}}$$

is positive for $a \in [2b+3, 2b+4] = [9, 10]$, and since $\delta_3(v_3(1)) = \frac{1}{49} > 0$, the function $\delta_3(a)$ is positive on $[v_3(1), \alpha_3]$.

Assume now that $b = 4$. Then

$$\delta_4(a) = -24 + \frac{49\sqrt{a}}{2\sqrt{2}} - 3a.$$

Hence $\delta_4(2b) = \delta_4(8) = 1$ and $\delta_4(2b+4) = \delta_4(12) = -60 + 49\sqrt{\frac{3}{2}} > 0$, and so $\delta_4(a) > 0$ for all $a \in [2b, 2b+4]$.

Assume finally that $b \geqslant 5$. Then $\delta_b'(a) < 0$ for $a \in [2b, 2b+4]$. Indeed, $\delta_b'(a)$ is decreasing and $\delta_b'(2b) = 1 - b + (2b^2+4b+1)/4b < 0$. We are left with showing that

$$\delta_b(2b+4) = -(4+6b+2b^2) + (1+4b+2b^2)\sqrt{\frac{b+2}{b}}$$

is positive, which is true since equivalent to $\frac{b+2}{b} > 0$.

(viii) We show that z_2, $z_5 \geqslant 0$. The other inequalities then follow from the previous items. The inequality $z_2 \geqslant 0$ is equivalent to $\lambda \geqslant \frac{2b+3}{2b+1}$, which holds true. Moreover, $z_5 \geqslant 0$ is equivalent to

$$\lambda \geqslant \frac{ba+1}{2b(b+1)}, \tag{7-3}$$

which means that the line $a \mapsto \frac{ba+1}{2b(b+1)}$ of the affine step is below the volume constraint $\sqrt{\frac{a}{2b}}$. This holds true on $[2b, \alpha_b]$, since $\sqrt{\frac{a}{2b}}$ is convex and since (7-3) is an equality at α_b and a strict inequality at $2b$. □

7.1.2 Reductions Reducing the vector $((b+1)\lambda; b\lambda, \lambda, 1^{\times(2b+3)}, w_1, w_2^{\times \ell_2} \,\|\, \dots)$ with $\delta = -1$ yields

$$((b+1)\lambda - 1; b\lambda - 1, \underbrace{\lambda - 1}_{=z_1}, 0, 1^{\times(2b+2)}, w_1, w_2^{\times \ell_2}, \dots).$$

By Lemma 7.2(i) this vector reorders to

$$((b+1)\lambda-1; b\lambda-1, 1^{\times(2b+2)}, w_1, z_1, w_2^{\times \ell_2} \parallel \dots, 0).$$

Applying b Cremona transforms with $\delta=\lambda-2$ and regrouping the produced z_1's, we get

$$(\underbrace{(2b+1)\lambda-(2b+1)}_{=z_2+2}; \underbrace{2b\lambda-(2b+1)}_{=1-z_1+z_2}, 1^{\times 2}, w_1, z_1^{\times(2b+1)}, w_2^{\times \ell_2}, \dots).$$

By Lemma 7.2(ii), this vector reorders to

$$(z_2+2; 1^{\times 2}, w_1, 1-z_1+z_2, z_1^{\times(2b+1)}, w_2^{\times \ell_2} \parallel \dots).$$

Applying one Cremona transform with $\delta=z_2-w_1$ yields the vector

$$(2z_2+2-w_1; \underbrace{(1+z_2-w_1)}_{=z_3 \text{ by (7-1)}}{}^{\times 2}, z_2, 1-z_1+z_2, z_1^{\times(2b+1)}, w_2^{\times \ell_2}, \dots),$$

which by Lemma 7.2(iii) reorders to

$$(2z_2+2-w_1;\ 1-z_1+z_2,\ z_1^{\times(2b+1)},\ z_3^{\times 2} \parallel z_2,\ w_2^{\times \ell_2}, \dots).$$

Applying $b-1$ Cremona transforms with $\delta=z_3-z_1$ and regrouping the produced z_3's, we get

$$(\underbrace{(b-1)(z_3-z_1)+2z_2+2-w_1}_{\stackrel{e}{=}\,2z_1+z_5};\ \underbrace{b(z_3-z_1)+w_1}_{=z_4},\ z_1^{\times 3},\ z_3^{\times 2b},\ z_2,\ w_2^{\times \ell_2}, \dots). \tag{7-4}$$

We now distinguish the cases $z_4 \geqslant z_1$ and $z_1 \geqslant z_4$.

Case 1 $(z_4 \geqslant z_1)$ The ordered vector is then

$$(2z_1+z_5; z_4, z_1^{\times 3}, z_3^{\times 2b} \parallel z_2, w_2^{\times \ell_2}, \dots).$$

One more Cremona transform with $\delta=z_5-z_4$ yields

$$(2(z_1+z_5)-z_4; z_5, (\underbrace{z_1+z_5-z_4}_{=z_3 \text{ by (7-1)}})^{\times 2}, z_1, z_3^{\times 2b}, z_2, w_2^{\times \ell_2}, \dots),$$

which by Lemma 7.2(iv) reorders to

$$(2(z_1+z_5)-z_4; z_1, z_5, z_3^{\times(2b+2)} \parallel z_2, w_2^{\times \ell_2}, \dots).$$

We already know that all entries of this vector are nonnegative, and its defect is $\delta=z_1+z_5-z_4-z_3=0$. Hence this vector is reduced.

Case 2 $(z_1 \geqslant z_4)$ Reorder the vector (7-4) as

$$(2z_1 + z_5; z_1^{\times 3}, z_4, z_3^{\times 2b} \parallel z_2, w_2^{\times \ell_2}, \dots).$$

Recall from Lemma 7.2(iv) that $z_1 \geqslant z_5$. Apply one Cremona transform with $\delta = z_5 - z_1$ to obtain

$$(2z_5 + z_1; z_5^{\times 3}, z_4, z_3^{\times 2b}, z_2, w_2^{\times \ell_2}, \dots).$$

Since $z_3 \geqslant z_5$ by Lemma 7.2(iv), this vector reorders to

$$(2z_5 + z_1; z_4, z_3^{\times 2b} \parallel z_2, z_5^{\times 3}, w_2^{\times \ell_2}, \dots).$$

Applying b Cremona transforms with $\delta = 2z_5 + z_1 - z_4 - 2z_3$ and regrouping the produced z_5's, we obtain the vector

$$(\underbrace{(b+1)(2z_5+z_1)-b(z_4+2z_3)}_{=:\mu}; \underbrace{b(2z_5+z_1-z_4-2z_3)+z_4}_{=z_6}, z_2, z_5^{\times(2b+3)}, w_2^{\times \ell_2}, \dots),$$

which by Lemma 7.2(vi) reorders to

$$(\mu; z_6 \parallel z_2, z_5^{\times(2b+3)}, w_2^{\times \ell_2}, \dots). \tag{7-5}$$

Notice that this vector does not contain z_1, z_3 or z_4.

Proposition 7.3 *Assume that $a \leqslant \alpha_b$ and $z_1 \geqslant z_4$. If $b = 2$, also assume that $w_2 \leqslant \max\{z_2, z_5\}$. Then the vector* (7-5) *is reduced.*

Proof We already know that all entries of (7-5) are nonnegative. Using (7-1) we compute

$$\mu - z_6 = z_1 - z_4 + 2z_5 = z_3 + z_5. \tag{7-6}$$

Subcase 1 $(z_5 \geqslant w_2)$ Then

$$\delta = \mu - z_6 - z_5 - (z_2 \text{ or } z_5) = z_3 - (z_2 \text{ or } z_5) \geqslant 0,$$

where in the last step we have used Lemma 7.2(iii)–(iv).

Subcase 2 $(z_2 \geqslant w_2 \geqslant z_5)$ Then

$$\delta = \mu - z_6 - (z_2 + w_2) = z_3 + z_5 - z_3 = z_5 \geqslant 0.$$

Subcase 3 $(w_2 \geqslant z_2, z_5)$ This is the case where we assume that $b \geqslant 3$. Recall that $\ell_2 \geqslant 2$. Hence

$$\delta_b(a) = \mu - z_6 - 2w_2 = z_1 - z_4 + 2z_5 - 2w_2$$

is nonnegative by Lemma 7.2(vii). ◻

In view of Proposition 7.3 we can assume that $b=2$ and that $w_2 \geqslant \max\{z_2, z_5\}$. The vector at hand then is

$$(\mu; z_6, w_2^{\times \ell_2} \parallel z_2, z_5^{\times(2b+3)}, \dots). \tag{7-7}$$

We set $z_7 := z_2 + z_5$ and compute

$$\delta = \mu - z_6 - 2w_2 \overset{(7\text{-}6)}{=} z_3 + z_5 - 2w_2 \overset{(7\text{-}1)}{=} 1 + z_2 - w_1 + z_5 - 2w_2 = z_7 - w_2.$$

If $\delta \geqslant 0$ we are done. So assume that $\delta = z_7 - w_2 < 0$, and set $m := \lfloor \frac{1}{2}\ell_2 \rfloor$ and $\widehat{\mu} := \mu + m\delta$, $\widehat{z}_6 := z_6 + m\delta$. Applying m Cremona transforms and swapping the position of w_2 and $z_7^{\times 2m}$ if ℓ_2 is odd, we obtain

$$(\widehat{\mu}; \widehat{z}_6, z_7^{\times 2m}, z_2, z_5^{\times(2b+3)}, w_3^{\times \ell_3}, \dots) \quad \text{if } \ell_2 = 2m, \tag{7-8}$$

$$(\widehat{\mu}; \widehat{z}_6, w_2, z_7^{\times 2m}, z_2, z_5^{\times(2b+3)}, w_3^{\times \ell_3}, \dots) \quad \text{if } \ell_2 = 2m+1. \tag{7-9}$$

Proposition 7.4 *After reordering, the vector* (7-8) *is reduced. After reordering, the vector* (7-9) *is reduced if* $z_7 \geqslant w_3$*, and transforms to a reduced vector by one Cremona move if* $w_3 > z_7$*.*

Proof We first show the inequalities

$$\widehat{z}_6 \geqslant w_2 \geqslant z_7 \geqslant z_2, z_5. \tag{7-10}$$

Then also $\widehat{z}_6$, $z_7 \geqslant 0$. We have $w_2 - z_7 = -\delta > 0$ and $z_7 = z_2 + z_5 \geqslant z_2, z_5$. We are thus left with proving $\widehat{z}_6 \geqslant w_2$. For $m \in \mathbb{N}$ we compute

$$f_m(a) := \widehat{z}_6 - w_2 = z_6 + mz_7 - (m+1)w_2 = -(m+2)a + \left(\tfrac{17}{2}m + 11\right)\sqrt{a} - (16m+15).$$

Then $f_m'(a) = -(m+2) + \left(\frac{17}{2}m+11\right)/(2\sqrt{a}) > 0$ for all $m \in \mathbb{N}$ and $a \in [2b+3, 2b+4] = [7,8]$, since this holds true for $a = 8$. Recall that $\ell_2 \geqslant 5$. Since $\ell_2 = \lfloor w_1/w_2 \rfloor = \lfloor \frac{-7+a}{8-a} \rfloor$ and $\ell_2(\alpha_2) = 30$, we can assume that $2 \leqslant m \leqslant 15$. If the multiplicity of w_2 is ℓ_2, then $w_1 \in [\ell_2/(\ell_2+1), (\ell_2+1)/(\ell_2+2)[$. Thus $\widehat{z}_6 - w_2$ is given by f_m for $a \in \left[7 + \frac{2m}{2m+1}, 7 + \frac{2m+2}{2m+3}\right[\cap [v_2(1), \alpha_2]$. Since each f_m is increasing on $[7,8]$, it now suffices to note that $f_2(v_2(1)) = f_2\left(\left(\frac{14}{5}\right)^2\right) = \frac{1}{25} > 0$ and to check that $f_m\left(7 + \frac{2m}{2m+1}\right) \geqslant 0$ for $m \in \{3, \dots, 15\}$, which is readily done, for instance by noticing that $m \mapsto f_m\left(7 + \frac{2m}{2m+1}\right)$ is increasing.

Case 1 ($z_7 \geqslant w_3$) The part $(\mu; a_1, a_2, a_3)$ of the ordered vectors is then as in (7-8) and (7-9). Therefore, $\widehat{\delta} = \mu - z_6 - 2z_7 = \delta - 2(z_7 - w_2) = -\delta > 0$ if ℓ_2 is even, and

$\widehat{\delta} = \mu - z_6 - w_2 - z_7 = \delta - (z_7 - w_2) = 0$ if ℓ_2 is odd. Hence the vectors (7-8) and (7-9) are reduced.

Case 2 $(w_3 > z_7)$ In this case, the vectors at hand are

(7-11) $$(\widehat{\mu}; \widehat{z}_6, w_3^{\times \ell_3} \parallel z_7^{\times 2m}, w_4^{\times \ell_4}, z_2, z_5^{\times(2b+3)}, \dots) \quad \text{if } \ell_2 = 2m,$$

(7-12) $$(\widehat{\mu}; \widehat{z}_6, w_2, w_3^{\times \ell_3} \parallel z_7^{\times 2m}, w_4^{\times \ell_4}, z_2, z_5^{\times(2b+3)}, \dots) \quad \text{if } \ell_2 = 2m+1.$$

Assume first that ℓ_2 is even. If $\ell_3 = 1$, then (7-10) shows that

$$\widehat{\delta} = \mu - z_6 - w_3 - (z_7 \text{ or } w_4) = w_2 + z_7 - w_3 - (z_7 \text{ or } w_4) = (w_2 - w_3 \text{ or } z_7) \geqslant 0.$$

If $\ell_3 \geqslant 2$, then $\widehat{\delta} = \mu - z_6 - 2w_3 = w_2 + z_7 - 2w_3 \geqslant z_7 \geqslant 0$.

Assume now that ℓ_2 is odd. Then $\widehat{\delta} = \mu - z_6 - w_2 - w_3 = z_7 - w_3 < 0$. Applying one more Cremona move to the vector (7-12) yields

$$(\widehat{\mu} + \widehat{\delta};\ \widehat{z}_6 + \widehat{\delta},\ w_2 + \widehat{\delta},\ w_3^{\times \ell_3 - 1} \parallel z_7^{\times 2m+1},\ w_4^{\times \ell_4},\ z_2,\ z_5^{\times(2b+3)}, \dots).$$

The ordering is right because if $\ell_3 = 1$, then $w_2 + \widehat{\delta} = w_2 + z_7 - w_3 = z_7 + w_4$, and if $\ell_3 \geqslant 2$, then $w_2 + \widehat{\delta} = w_2 + z_7 - w_3 \geqslant w_3$.

If $\ell_3 = 1$, then the defect is now $\widetilde{\delta} = \mu - z_6 - w_2 - \widehat{\delta} - (z_7 \text{ or } w_4) = w_3 - (z_7 \text{ or } w_4) > 0$, and if $\ell_3 \geqslant 2$, then $\widetilde{\delta} = w_3 - w_3 = 0$. □

This completes the proof of Theorem 7.1 for $a \leqslant \alpha_b$.

7.2 The interval $[\alpha_b, 2b+4]$

It turns out that the reduction process for $a \in [\alpha_b, 2b+4]$ is the same as for $a \in [v_b(1), \alpha_b]$ in Case 2. Set $\lambda = \frac{ba+1}{2b(b+1)}$ and define $z_1, \dots, z_6$ as in Section 7.1. Applying the same Cremona moves (ie the same sequence of Cremona transforms and reorderings) as in Case 2, we obtain the vector (7-5), namely

(7-13) $$(\mu; z_6 \parallel z_2, z_5^{\times(2b+3)}, w_2^{\times \ell_2}, \dots).$$

It suffices to prove the following statement:

Proposition 7.5 *If $a \geqslant \alpha_b$, then the vector* (7-13) *is reduced.*

Proof The identity $\lambda = \frac{ba+1}{2b(b+1)}$ is equivalent to $z_5 = 0$. We now show $z_6, z_2 \geqslant w_2$, implying $z_6, z_2 \geqslant 0$. Using (7-2) we find that the inequality $z_6 \geqslant w_2$ is equivalent to the inequality

$$w_1 \geqslant \frac{3b+3}{3b+4},$$

which is satisfied since $\frac{3b+3}{3b+4} \leqslant \alpha_b - (2b+3)$ for all $b \geqslant \frac{2}{3}(-1+\sqrt{7})$. The inequality $z_2 \geqslant w_2$ is equivalent to the inequality

$$w_1 \geqslant \frac{4b^2+3b-1}{4b^2+3b},$$

which is satisfied since $\frac{4b^2+3b-1}{4b^2+3b} \leqslant \alpha_b - (2b+3)$ for all $b \geqslant \frac{5}{4}$.

The ordered vector is thus

$$(\mu; z_6, z_2, w_2^{\times \ell_2}, \dots, 0^{\times(2b+3)}).$$

(The inequality $z_6 \geqslant z_2$ holds true, but there is no need to prove it). Using again $\mu - z_6 = z_1 - z_4 + 2z_5$ and $z_1 + z_5 - z_4 = 1 + z_2 - w_1$ from (7-1) we find, since $z_5 = 0$,

$$\delta = (\mu - z_6) - (z_2 + w_2) = (z_1 - z_4) - (z_2 + 1 - w_1) = 0.$$

Hence the vector (7-13) is reduced. □

8 The interval $[2b+4, u_b(2)]$ for $b \geqslant 3$

Recall that $\gamma_b := u_b(2) = \frac{(2b+2)^2}{2b} = 2b+4+\frac{2}{b}$ and that

$$\beta_b := \frac{(2b^2+4b+1)^2}{2b(b+1)^2} = 2b+4+\frac{1}{2b(b+1)^2} \in \,]2b+4, \gamma_b[.$$

Throughout this section we assume that $b \geqslant 3$.

Theorem 8.1 *For $b \geqslant 3$ we have*

$$c_b(a) = \begin{cases} 1 + \dfrac{2b+1}{2b(b+1)} & \text{if } a \in [2b+4, \beta_b], \\ \sqrt{\dfrac{a}{2b}} & \text{if } a \in [\beta_b, \gamma_b]. \end{cases}$$

Proof In view of Theorem 7.1 it suffices to prove that $c_b(a) = \sqrt{\frac{a}{2b}}$ on $[\beta_b, \gamma_b]$. Let $a \in [\beta_b, \gamma_b]$ be a rational number with weight expansion

$$\boldsymbol{w}(a) = (1^{\times(2b+4)}, w_1^{\times \ell_1}, w_2^{\times \ell_2}, \dots, w_n^{\times \ell_n}).$$

8.1 Inequalities

Set $\lambda = \sqrt{\frac{a}{2b}}$. We wish to show that the vector $((b+1)\lambda; b\lambda, \lambda, \boldsymbol{w}(a))$ can be reduced to a reduced vector. Notice that

$$\lambda(\beta_b) = 1 + \frac{2b+1}{2b(b+1)}, \quad \lambda(\gamma_b) = 1 + \frac{1}{b}.$$

Define the numbers

$$\begin{aligned} z_1 &:= \lambda - 1, \\ z_2 &:= (2b+1)\lambda - (2b+3), \\ z_3 &:= 1 + b(z_2 - z_1), \\ z_4 &:= 1 + (b+1)(z_2 - z_1), \\ z_5 &:= z_1 + z_2 - w_1 \end{aligned}$$

and $m = \left\lfloor \frac{1}{2}\ell_1 \right\rfloor$, where $\ell_1 = \left\lfloor \frac{1}{w_1} \right\rfloor$.

Lemma 8.2 *On the interval $[\beta_b, \gamma_b]$ the following inequalities hold true:*

(i) $1 - z_1 + z_2 \geqslant z_1 \geqslant z_2 \geqslant 0$.

(ii) $1 - z_1 + z_2 \geqslant w_1$.

(iii) $z_3 \geqslant z_2, z_4, w_1$ *and* $z_4 \geqslant 0$.

(iv) $z_3 + b(z_4 - z_2) \geqslant z_2$.

(v) $z_2 + z_4 - w_1 \geqslant w_1$.

(vi) $2z_2 \geqslant w_1$ *and* $z_2 \geqslant w_3$.

(vii) $1 - z_1 + z_2 + m(z_1 + z_2 - 2w_1) \geqslant w_1$.

In particular, $z_i \geqslant 0$ for all i.

Proof (i) The inequality $z_1 \geqslant z_2$ was already shown in the proof of Lemma 7.2(ii). The inequality $z_2 \geqslant 0$ is equivalent to $(2b+1)\lambda \geqslant 2b+3$. Since λ is increasing, it suffices to verify this in $a = \beta_b$, that is, that

$$(2b+1)\left(1 + \frac{2b+1}{2b(b+1)}\right) \geqslant 2b+3,$$

or, equivalently, $(2b+1)^2 \geqslant 4b(b+1)$, which holds true.

The inequality $1 - z_1 + z_2 \geqslant z_1$ is equivalent to $(2b-1)\lambda \geqslant 2b$. It suffices to verify this in $a = \beta_b$, that is, that

$$(2b-1)\left(1 + \frac{2b+1}{2b(b+1)}\right) \geqslant 2b,$$

or, equivalently, $2b^2 \geqslant 2b+1$, which holds true.

(ii) This is equivalent to $a-3 \leqslant 2b\lambda$. Since the slope of $2b\lambda = \sqrt{2ba}$ is $\sqrt{\frac{b}{2a}} < 1$, it suffices to check this inequality at $a = \gamma_b$, ie that $2b+2 \geqslant a-3$, which holds true.

(iii) The inequality $z_3 \geqslant z_2$ is equivalent to $(2b^2-2b-1)\lambda \geqslant 2b^2-4$. It suffices to verify this at $a = \beta_b$, that is, that

$$(2b^2-2b-1)\left(1+\frac{2b+1}{2b(b+1)}\right) \geqslant 2b^2-4,$$

or, equivalently, $2b \geqslant 1$, which holds true.

The inequality $z_3 \geqslant z_4$ follows from $z_1 \geqslant z_2$.

The inequality $z_3 \geqslant w_1$ is equivalent to $2b^2\lambda \geqslant 2b^2+a-5$ or, using $a = 2b\lambda^2$, to

$$f_b(\lambda) := -2b\lambda^2 + 2b^2\lambda - 2b^2 + 5 \geqslant 0.$$

Since $b \geqslant 3$ we have $f_b'(\lambda) = 2b(b-2\lambda) > 0$, and $f_b(\lambda(\beta_b)) = \frac{2b^2+2b-1}{2b(b+1)^2} > 0$.

The inequality $z_4 \geqslant 0$ is equivalent to $2b(b+1)\lambda \geqslant 2b^2+4b+1$, which holds true, since this is an equality at $a = \beta_b$.

(iv) This is equivalent to $(2b^2+4b+1)\lambda \geqslant 2(b^2+3b+2)$. At $a = \beta_b$, this inequality is equivalent to

$$(2b^2+4b+1)(2b+1) \geqslant 2b(b+1)(2b+3),$$

which in turn simplifies to $1 \geqslant 0$.

(v) This is equivalent to $(2b^2+4b+1)\lambda \geqslant 2(a+b^2+b-2)$, or, using $a = 2b\lambda^2$, to

$$f_b(\lambda) := 4b\lambda^2 - (2b^2+4b+1)\lambda + 2(b^2+b-2) \leqslant 0 \tag{8-1}$$

on $[\beta_b, \gamma_b]$. Its derivative is $f_b'(\lambda) = 8b\lambda - (2b^2+4b+1)$.

Assume first that $b = 3$. Then $f_b'(\lambda) = 24\lambda - 31 \geqslant 0$ since this holds true in $\lambda(\beta_b) = \frac{31}{24}$. Hence (8-1) follows from $f_b(\lambda(\gamma_b)) = f_3\left(\frac{4}{3}\right) = 0$.

Assume now that $b \geqslant 4$. Then $f_b'(\lambda(\gamma_b)) = 8(b+1) - (2b^2+4b+1) \leqslant 0$, whence $f_b'(\lambda) \leqslant 0$. Thus (8-1) follows from $f_b(\lambda(\beta_b)) = -\frac{b-1}{2b(b+1)^2} \leqslant 0$.

(vi) This is equivalent to

$$f_b(\lambda) := b\lambda^2 - (2b+1)\lambda + (b+1) \leqslant 0 \tag{8-2}$$

on $[\beta_b, \gamma_b]$. Since $f_b'(\lambda) = 2b\lambda - (2b+1) \geqslant 2b\lambda(\beta_b) - (2b+1) = \frac{b}{b+1} \geqslant 0$ on $[\beta_b, \gamma_b]$, the inequality (8-2) follows from $f_b(\lambda(\gamma_b)) = 0$.

Further, $z_2 \geqslant \frac{1}{2}w_1 \geqslant w_3$ since $w_1 = \ell_2 w_2 + w_3 \geqslant w_2 + w_3 \geqslant 2w_3$.

(vii) Recall that $1 = \ell_1 w_1 + w_2$. If $\ell_1 = 2m+1$, then (vii) becomes

$$w_2 - z_1 + z_2 + m(z_1 + z_2) \geqslant 0,$$

which holds true. If $\ell_1 = 2m$, then (vii) becomes $w_2 - z_1 + z_2 + m(z_1 + z_2) \geqslant w_1$. This holds true since it holds true for $m = 1$ by assertion (vi). □

The following lemma will be very useful:

Lemma 8.3 *If $w_2 \geqslant z_2$, then $\ell_2 = 1$.*

Proof Recall that we can assume $\ell_3 \geqslant 1$, that is, $w_3 > 0$. If $\ell_2 \geqslant 2$, then $w_1 = \ell_2 w_2 + w_3 > 2w_2 \geqslant 2z_2 \geqslant w_1$, by Lemma 8.2(vi). □

8.2 Reductions

Applying one Cremona transform to

$$((b+1)\lambda; b\lambda, \lambda, 1^{\times(2b+4)}, w_1^{\times \ell_1}, \dots)$$

with $\delta = -1$ yields

$$((b+1)\lambda - 1; b\lambda - 1, \underbrace{\lambda - 1}_{=z_1}, 0, 1^{\times(2b+3)}, w_1^{\times \ell_1}, \dots),$$

which we reorder to

$$((b+1)\lambda - 1; b\lambda - 1, 1^{\times(2b+3)} \parallel z_1, w_1^{\times \ell_1}, \dots, 0).$$

Applying b Cremona transforms with $\delta = \lambda - 2$ we obtain

$$(\underbrace{(2b+1)\lambda - (2b+1)}_{=z_2+2}; \underbrace{2b\lambda - (2b+1)}_{=1-z_1+z_2}, 1^{\times 3}, z_1^{\times(2b+1)}, w_1^{\times \ell_1}, \dots, 0),$$

which by Lemma 8.2 reorders to

$$(z_2 + 2; 1^{\times 3}, 1 - z_1 + z_2 \parallel z_1^{\times(2b+1)}, w_1^{\times \ell_1}, \dots, 0).$$

Applying one Cremona transform with $\delta = z_2 - 1$ yields

$$(2z_2 + 1; z_2^{\times 3}, 1 - z_1 + z_2, z_1^{\times(2b+1)}, w_1^{\times \ell_1}, \dots, 0),$$

which we reorder to

$$(2z_2 + 1; 1 - z_1 + z_2 \parallel z_1^{\times(2b+1)}, z_2^{\times 3}, w_1^{\times \ell_1}, \dots, 0). \tag{8-3}$$

We now distinguish several cases, according to the order of $z_1 \geqslant z_2$ and w_1.

Case 1 $(z_1 \geqslant z_2, w_1)$ Applying $b-1$ Cremona moves to the vector (8-3) with $\delta = z_2 - z_1$ we get the vector

$$(z_1 + z_2 + z_3; z_3, z_1^{\times 3}, z_2^{\times(2b+1)}, w_1^{\times \ell_1}, \dots). \tag{8-4}$$

Case 1(a) $(z_1 \geqslant z_2 \geqslant w_1)$ If $z_3 \geqslant z_1$, we apply one more Cremona move with $\delta = z_2 - z_1$ and obtain

$$(2z_2 + z_3; \underbrace{z_3 + z_2 - z_1}_{=z_4}, z_1, z_2^{\times(2b+3)}, w_1^{\times \ell_1}, \dots).$$

The assumption $z_3 \geqslant z_1$ is equivalent to $z_4 \geqslant z_2$. Hence this vector is ordered up to possibly swapping z_4 and z_1, and in either case $\delta = 0$, whence this vector is reduced. We can thus assume for the rest of Case 1(a) that

$$z_1 \geqslant z_3 \quad \text{and} \quad z_2 \geqslant z_4. \tag{8-5}$$

By Lemma 8.2(iii) the vector (8-4) reorders to

$$(z_1 + z_2 + z_3; z_1^{\times 3}, z_3 \parallel z_2^{\times(2b+1)}, w_1^{\times \ell_1}, \dots). \tag{8-6}$$

One Cremona transform with $\delta = z_4 - z_1$ yields the vector

$$(2z_4 + z_1; z_4^{\times 3}, z_3, z_2^{\times(2b+1)}, w_1^{\times \ell_1}, \dots),$$

which by (8-5) reorders to

$$(2z_4 + z_1; z_3, z_2^{\times(2b+1)} \parallel z_4^{\times 3}, w_1^{\times \ell_1}, \dots).$$

Under b Cremona transforms with $\delta = z_4 - z_2$ this vector becomes

$$(2z_4 + z_1 + b(z_4 - z_2); z_3 + b(z_4 - z_2), z_2 \parallel z_4^{\times(2b+3)}, w_1^{\times \ell_1}, \dots),$$

where the ordering follows from Lemma 8.2(iv). Then

$$\delta = z_4 - (z_4 \text{ or } w_1).$$

If $z_4 \geqslant w_1$ we are done. If $w_1 \geqslant z_4$, one more Cremona transform with $\delta = z_4 - w_1$ yields the vector

$$(2z_4 + z_1 + b(z_4 - z_2) + \delta;\ z_3 + b(z_4 - z_2) + \delta,\ z_2 + z_4 - w_1,\ w_1^{\times(\ell_1 - 1)} \parallel z_4^{\times(2b+4)}, \dots),$$

which is ordered by Lemma 8.2(v) and has defect 0.

Case 1(b) $(z_1 \geqslant w_1 \geqslant z_2)$ Assume first that $z_1 \geqslant z_3$. The vector (8-4) then reorders to

$$(z_1+z_2+z_3; z_1^{\times 3}, z_3, w_1^{\times \ell_1} \parallel z_2^{\times(2b+1)}, w_2^{\times \ell_2}, \dots). \tag{8-7}$$

Since $z_4 \leqslant z_3 \leqslant z_1$, we also have $z_4 \leqslant z_1$, and so $\delta = z_4 - z_1 \leqslant 0$. One Cremona transform yields

$$(2z_4+z_1; z_4^{\times 3}, z_3, w_1^{\times \ell_1}, z_2^{\times(2b+1)}, w_2^{\times \ell_2}, \dots).$$

Since $z_1+z_4 = z_2+z_3$ and $z_1 \geqslant z_3$, we have $z_4 \leqslant z_2$, whence this vector reorders to

$$(2z_4+z_1; z_3, w_1^{\times \ell_1} \parallel z_2^{\times(2b+1)}, z_4^{\times 3}, w_2^{\times \ell_2}, \dots).$$

By Lemma 8.2(v) we can estimate

$$\delta = (z_4+z_2-w_1) - (w_1 \text{ or } z_2 \text{ or } z_4 \text{ or } w_2) \geqslant w_1 - w_1 = 0.$$

For the rest of Case 1(b) we can thus assume that

$$z_3 \geqslant z_1 \quad \text{and} \quad z_4 \geqslant z_2.$$

The vector (8-4) then reorders to

$$(z_1+z_2+z_3; z_3, z_1^{\times 3}, w_1^{\times \ell_1} \parallel z_2^{\times(2b+1)}, w_2^{\times \ell_2}, \dots).$$

Applying one Cremona transform with $\delta = -z_1 + z_2$ yields

$$(2z_2+z_3; z_4 \leftrightarrow z_1, w_1^{\times \ell_1} \parallel z_2^{\times(2b+3)}, w_2^{\times \ell_2}, \dots).$$

The ordering is right up to possibly swapping $z_4 \leftrightarrow z_1$, since $z_4 \geqslant w_1$ by Lemma 8.2(v). Abbreviate

$$* := z_2+z_4-w_1 \quad \text{and} \quad z_5 := z_1+z_2-w_1.$$

Then $z_5 \geqslant z_2$. Applying one Cremona transform with $\delta = z_2 - w_1$ we obtain

$$(*+z_1+z_2; *, z_5, w_1^{\times(\ell_1-1)}, z_2^{\times(2b+4)}, w_2^{\times \ell_2}, \dots). \tag{8-8}$$

By Lemma 8.2(v) we have $* \geqslant w_1$. If also $z_5 \geqslant w_1$, then

$$\delta = w_1 - (w_1 \text{ or } z_2 \text{ or } w_2) \geqslant 0.$$

So assume that $z_5 \leqslant w_1$. Then the vector (8-8) reorders to

$$(z_1+z_2+*; *, w_1^{\times(\ell_1-1)} \parallel z_5, z_2^{\times(2b+4)}, w_2^{\times \ell_2}, \dots). \tag{8-9}$$

Subcase 1 ($\ell_1 = 2m+1$ with $m \geqslant 0$) Applying m Cremona transforms with $\delta_* := z_5 - w_1$ we get

$$(z_1 + z_2 + * + m\delta_*;\ * + m\delta_*,\ z_5^{\times \ell_1},\ z_2^{\times(2b+4)} \leftrightarrow w_2^{\times \ell_2}, \dots). \tag{8-10}$$

We claim that this vector is reduced after reordering.

Assume that $z_5 \geqslant w_2$. Then the ordering in (8-10) is right by Lemma 8.4(i) below, and

$$\delta = w_1 - (z_5 \text{ or } z_2 \text{ or } w_2) \geqslant 0.$$

Assume that $w_2 \geqslant z_5$. Recall that $z_5 = z_1 + z_2 - w_1 \geqslant z_2 \geqslant w_3$. By Lemma 8.3 we have $\ell_2 = 1$, and so by Lemma 8.4(i) the vector (8-10) reorders to

$$(z_1 + z_2 + * + m\delta_*;\ * + m\delta_* \leftrightarrow w_2,\ z_5^{\times \ell_1},\ z_2^{\times(2b+4)} \parallel \dots).$$

Now $\delta = z_1 + z_2 - w_2 - z_5 = w_1 - w_2 \geqslant 0$.

Subcase 2 ($\ell_1 = 2m$ with $m \geqslant 1$) Applying $m-1$ Cremona transforms to (8-9) with $\delta_* = z_5 - w_1$ we get

$$(z_1+z_2+*+(m-1)\delta_*;\ *+(m-1)\delta_*,\ w_1,\ z_5^{\times(\ell_1-1)},\ z_2^{\times(2b+4)},\ w_2^{\times \ell_2}, \dots). \tag{8-11}$$

Assume that $z_5 \geqslant w_2$. Then Lemma 8.4(ii) shows that (8-11) reorders to

$$(z_1 + z_2 + * + (m-1)\delta_*;\ * + (m-1)\delta_* \leftrightarrow w_1,\ z_5^{\times(\ell_1-1)} \parallel z_2^{\times(2b+4)},\ w_2^{\times \ell_2}, \dots),$$

and $\delta = 0$.

Assume that $w_2 \geqslant z_5$. Then $\ell_2 = 1$ by Lemma 8.3, and we reorder (8-11) to

$$(z_1 + z_2 + * + (m-1)\delta_*;\ * + (m-1)\delta_*,\ w_1,\ w_2,\ z_5^{\times(\ell_1-1)},\ z_2^{\times(2b+4)}, \dots).$$

One Cremona transform with $\widehat{\delta} = z_5 - w_2$ yields the vector

$$(z_1 + z_2 + * + (m-1)\delta_* + \widehat{\delta};\ * + (m-1)\delta_* + \widehat{\delta},\ z_1 + z_2 - w_2,\ z_5^{\times \ell_1},\ z_2^{\times(2b+4)}, \dots).$$

Recall that $z_1 + z_2 - w_2 \geqslant z_5 \geqslant z_2 \geqslant w_3$ (by Lemma 8.2(vi)) and note that

$$* + (m-1)\delta_* + \widehat{\delta} \geqslant z_5 + \widehat{\delta} = 2z_1 + 2z_2 - 2w_1 - w_2 \geqslant 0$$

by Lemma 8.4(ii), the assumption $z_1 \geqslant w_1$ and Lemma 8.2(vi).

If $* + (m-1)\delta_* + \widehat{\delta} \geqslant z_5$, then $\delta = w_2 - z_5 \geqslant 0$.

If $* + (m-1)\delta_* + \widehat{\delta} \leqslant z_5$, then $\delta = * + (m-1)\delta_* - z_5 \geqslant 0$.

Lemma 8.4 *Assume that* $z_1 \geqslant w_1 \geqslant z_5$.

(i) *If* $\ell_1 = 2m+1$, *then* $* + m\delta_* \geqslant z_5$.

(ii) *If* $\ell_1 = 2m$, *then* $* + (m-1)\delta_* \geqslant z_5$.

The proof is given in Section 8.3.

Case 2 $(w_1 \geqslant z_1 \geqslant z_2)$ Then $z_1 \geqslant z_2 \geqslant z_5$. Recall from Lemma 8.2(vi) that $z_2 \geqslant w_3$. We shall therefore not display $w_3^{\times \ell_3}$ in the vectors below. The vector (8-3) reorders to

$$(2z_2+1;\ 1-z_1+z_2,\ w_1^{\times \ell_1} \parallel z_1^{\times(2b+1)},\ z_2^{\times 3},\ w_2^{\times \ell_2},\dots). \tag{8-12}$$

Case 2(a) $(\ell_1 = 2m+1$ is odd$)$ Applying m Cremona transforms with $\delta_* = z_5 - w_1 \leqslant 0$ we obtain the vector

$$(2z_2+1+m\delta_*;\ 1-z_1+z_2+m\delta_*,\ w_1,\ z_5^{\times(\ell_1-1)},\ z_1^{\times(2b+1)},\ z_2^{\times 3},\ w_2^{\times \ell_2},\dots).$$

By assumption, $z_1 \geqslant z_2 \geqslant z_5$. By Lemma 8.2(vii) this vector reorders to

$$(2z_2+1+m\delta_*;\ 1-z_1+z_2+m\delta_*,\ w_1 \parallel z_1^{\times(2b+1)},\ z_2^{\times 3},\ z_5^{\times(\ell_1-1)},\ w_2^{\times \ell_2},\dots). \tag{8-13}$$

Subcase 1 $(z_1 \geqslant w_2)$ Applying one Cremona move with $\delta = z_2 - w_1$ we obtain

$$(3z_2+1-w_1+m\delta_*;\ 1-z_1+2z_2-w_1+m\delta_*,\ z_1^{\times 2b},\ z_2^{\times 4},\ z_5^{\times \ell_1},\ w_2^{\times \ell_2},\dots).$$

Applying b Cremona transforms with $\delta = z_2 - z_1$ and setting

$$*_1 := 1 + m\delta_* + (b+1)(z_2 - z_1) + z_2 - w_1,$$

we obtain

$$(*_1 + z_1 + z_2; *_1, z_2^{\times(2b+4)}, z_5^{\times \ell_1}, w_2^{\times \ell_2},\dots). \tag{8-14}$$

We claim that this vector is reduced after reordering. To see this, assume first that $z_2 \geqslant w_2$. If $*_1 \geqslant z_2$ then $\delta = z_1 - z_2 \geqslant 0$, and if $z_2 \geqslant *_1$ then $\delta = *_1 + z_1 - 2z_2 \geqslant 0$ by Lemma 8.5. Assume now that $w_2 \geqslant z_2$. Then $\ell_2 = 1$ by Lemma 8.3. If $*_1 \geqslant z_2$ then $\delta = z_1 - w_2 \geqslant 0$, and if $z_2 \geqslant *_1$ then $\delta = *_1 + z_1 - z_2 - w_2 \geqslant 0$, by Lemma 8.5.

Subcase 2 $(w_2 \geqslant z_1)$ Then $\ell_2 = 1$ by Lemma 8.3, and

$$w_1 \geqslant w_2 \geqslant z_1 \geqslant z_2 \geqslant z_1 + z_2 - w_2 \geqslant z_1 + z_2 - w_1 = z_5. \tag{8-15}$$

The vector (8-13) becomes

$$(2z_2+1+m\delta_*;\ 1-z_1+z_2+m\delta_*,\ w_1,\ w_2,\ z_1^{\times(2b+1)},\ z_2^{\times 3},\ z_5^{\times(\ell_1-1)},\dots).$$

Applying one Cremona move with $\delta = z_1 + z_2 - w_1 - w_2$ we obtain

$$(* + z_1 + z_2; *, z_1^{\times(2b+1)}, z_2^{\times 3}, z_1 + z_2 - w_2, z_5^{\times \ell_1},\dots),$$

where $* := 1 + 2z_2 + m\delta_* - w_1 - w_2$. Applying b Cremona transforms with $\delta = z_2 - z_1$ we obtain the vector

$$(*_2 + z_1 + z_2; *_2, z_1, z_2^{\times(2b+3)}, z_1 + z_2 - w_2, z_5^{\times \ell_1},\dots),$$

where

$$*_2 := 1 + m\delta_* + b(z_2 - z_1) + 2z_2 - w_1 - w_2 = *_1 + z_1 - w_2.$$

This vector is reduced after reordering. Indeed, if $*_2 \geqslant z_2$ then $\delta = 0$, and if $z_2 \geqslant *_2$ then $\delta = *_2 - z_2 = *_1 + z_1 - z_2 - w_2 \geqslant 0$, by Lemma 8.5.

Lemma 8.5 *Assume that $w_1 \geqslant z_1 \geqslant z_2 \geqslant z_5$ and that $\ell_1 = 2m+1$. Then*

$$*_1 \geqslant 2z_2 - z_1,\ w_2 + z_2 - z_1.$$

The proof is given in Section 8.3.

Case 2(b) ($\ell_1 = 2m$ is even) Applying to the vector (8-12) m Cremona transforms with $\delta_* = z_5 - w_1 \leqslant 0$ we obtain the vector

$$(2z_2 + 1 + m\delta_*;\ 1 - z_1 + z_2 + m\delta_*,\ z_5^{\times \ell_1},\ z_1^{\times(2b+1)},\ z_2^{\times 3},\ w_2^{\times \ell_2}, \dots).$$

By Lemma 8.2(vii) this vector reorders to

$$(2z_2 + 1 + m\delta_*;\ 1 - z_1 + z_2 + m\delta_* \parallel z_1^{\times(2b+1)},\ z_2^{\times 3},\ z_5^{\times \ell_1},\ w_2^{\times \ell_2}, \dots). \tag{8-16}$$

Subcase 1 ($z_1 \geqslant w_2$) Applying b Cremona transforms with $\delta = z_2 - z_1$ and setting

$$*_3 := 1 + m\delta_* + (b+1)(z_2 - z_1)$$

we obtain

$$(*_3 + z_1 + z_2; *_3, z_1, z_2^{\times(2b+3)}, z_5^{\times \ell_1}, w_2^{\times \ell_2}, \dots). \tag{8-17}$$

If $z_2 \geqslant w_2$, then Lemma 8.6 shows that the ordering is

$$(*_3 + z_1 + z_2;\ *_3 \leftrightarrow z_1,\ z_2^{\times(2b+3)} \parallel z_5^{\times \ell_1},\ w_2^{\times \ell_2}, \dots),$$

and this vector is reduced since $\delta = 0$. So assume that $z_1 \geqslant w_2 \geqslant z_2$. Then $\ell_2 = 1$ by Lemma 8.3, and we reorder the vector (8-17) to

$$(*_3 + z_1 + z_2; *_3, z_1, w_2, z_2^{\times(2b+3)}, z_5^{\times \ell_1}, \dots).$$

Applying one Cremona transform with $\delta = z_2 - w_2$ we obtain

$$(*_3 + z_2 - w_2 + z_1 + z_2;\ *_3 + z_2 - w_2 \leftrightarrow z_1 + z_2 - w_2,\ z_2^{\times(2b+4)},\ z_5^{\times \ell_1}, \dots).$$

Note that $z_1 + z_2 - w_2 \geqslant z_2$ by assumption. If the ordering is right, then $\delta = w_2 - z_2 \geqslant 0$. Otherwise, $z_2 > *_3 + z_2 - w_2$, and then $\delta = *_3 - z_2 \geqslant 0$ by Lemma 8.6.

Subcase 2 ($w_2 \geqslant z_1$) By Lemma 8.3 we have $\ell_2 = 1$, and the vector (8-16) becomes

$$(2z_2 + 1 + m\delta_*;\ 1 - z_1 + z_2 + m\delta_*,\ w_2,\ z_1^{\times(2b+1)},\ z_2^{\times 3} \parallel z_5^{\times \ell_1}, \dots).$$

Applying one more Cremona move with $\delta = z_2 - w_2$ we obtain

$$(*_4 + z_1 + z_2; *_4, z_1^{\times 2b}, z_2^{\times 4}, z_1 + z_2 - w_2, z_5^{\times \ell_1}, \dots),$$

where $*_4 := 1 + m\delta_* - z_1 + 2z_2 - w_2$. Applying b Cremona transforms with $\delta = z_2 - z_1$ we obtain the vector

$$(*_4 + b(z_2 - z_1) + z_1 + z_2;\ *_4 + b(z_2 - z_1),\ z_2^{\times(2b+4)},\ z_1 + z_2 - w_2,\ z_5^{\times \ell_1}, \dots).$$

We claim that this vector is reduced after reordering. Indeed, if the ordering is right, then $\delta = z_1 - z_2 \geqslant 0$. Otherwise, $z_2 > *_4 + b(z_2 - z_1)$, and then

$$\delta = *_4 + b(z_2 - z_1) + z_1 - 2z_2 = *_3 + z_1 - z_2 - w_2 \geqslant 0$$

in view of Lemma 8.6.

Lemma 8.6 *Assume that $w_1 \geqslant z_1 \geqslant z_2 \geqslant z_5$ and that $\ell_1 = 2m$. Then*

$$*_3 \geqslant z_2,\ w_2 + z_2 - z_1.$$

8.3 Proof of Lemmas 8.4, 8.5 and 8.6

In this section we prove Lemmas 8.4, 8.5 and 8.6. Recall that $\delta_* = z_1 + z_2 - 2w_1$ and

$$* = z_2 + z_4 - w_1 = 1 + (b+1)(z_2 - z_1) + z_2 - w_1.$$

Hence

$$\begin{aligned} * + m\delta_* = *_1 &= 1 + m\delta_* + (b+1)(z_2 - z_1) + z_2 - w_1, \\ *_3 &= 1 + m\delta_* + (b+1)(z_2 - z_1). \end{aligned}$$

Note that $\delta_* \leqslant 0$ in all three lemmas. The proofs are along the following lines. All inequalities are, roughly, of the form

$$1 + m\delta_* + b(z_2 - z_1) \geqslant 0 \tag{8-18}$$

or, using $1 = (2m(+1))w_1 + w_2$,

$$m(z_1 + z_2) + b(z_2 - z_1) \geqslant 0. \tag{8-19}$$

In Lemma 8.4, the assumption $z_1 \geqslant w_1$ translates, roughly, to $m \succcurlyeq \frac{b}{2}$. Further, $w_1 \geqslant z_5$ translates to $3z_2 \geqslant z_1$, which together with (8-19) implies Lemma 8.4 for $m \succcurlyeq \frac{b}{2} + 1$. For the remaining one or two $m \approx \frac{b+1}{2}$ we prove the lemma using (8-18) and $\delta_* \leqslant 0$.

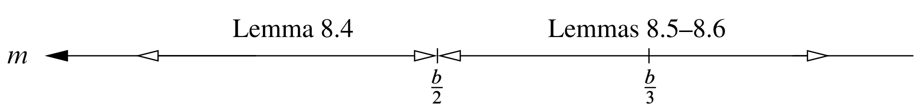


Figure 11

Lemmas 8.5 and 8.6 are proven similarly: the case $m \succcurlyeq \frac{b}{3}$ is settled using $2z_2 \geqslant z_1$ and (8-19), and the case $m \preccurlyeq \frac{b}{3} - 1$ is settled using (8-18) and $\delta_* \leqslant 0$.

Proof of Lemma 8.4 The inequality $z_1 \geqslant w_1$ implies that

$$\ell_1 \geqslant b. \tag{8-20}$$

Indeed, $z_1 \geqslant w_1$ is equivalent to $\sqrt{\frac{a}{2b}} \geqslant a - (2b+3)$, or

$$a \leqslant 2b + 3 + \frac{1 + \sqrt{16b^2 + 24b + 1}}{4b},$$

which in turn translates to

$$\frac{1}{w_1} \geqslant \frac{4b}{1 + \sqrt{16b^2 + 24b + 1} - 4b}.$$

Since the right-hand side is larger than b, inequality (8-20) follows.

We next observe that $w_1 \geqslant z_5$ implies that

$$3z_2 \geqslant z_1. \tag{8-21}$$

Indeed, $(3z_2 - z_1) - (w_1 - z_5) = 2(2z_2 - w_1) \geqslant 0$ by Lemma 8.2(vi). This is the main ingredient for proving the next two claims.

Claim 1 *Lemma 8.4(i) holds for $m \geqslant \frac{b}{2} + 1$.*

Proof This follows from $* + m\delta_* \geqslant z_1$, and, since $1 = (2m+1)w_1 + w_2$, this inequality follows from

$$(b+2)(z_2 - z_1) + m(z_1 + z_2) \geqslant 0.$$

Using (8-21) we estimate

$$(b+2)(z_2-z_1)+m(z_1+z_2) = (-b+m-2)z_1 + (b+m+2)z_2 \geqslant (-b+2m-2)\tfrac{2}{3}z_1,$$

which is nonnegative if $m \geqslant \frac{b}{2} + 1$. □

Claim 2 *Lemma 8.4(ii) holds for $m \geqslant \frac{b}{2} + \frac{3}{2}$.*

Proof This follows from $* + (m-1)\delta_* \geqslant w_1$, and since $1 = 2mw_1 + w_2$, this inequality follows from

$$(b+1)(z_2 - z_1) + (m-1)(z_1 + z_2) + z_2 \geqslant 0.$$

Using (8-21) we estimate

$$\begin{aligned}(b+1)(z_2-z_1)+(m-1)(z_1+z_2)+z_2 &= (-b+m-2)z_1+(b+m+1)z_2 \\ &\geqslant (-2b+4m-5)\tfrac{1}{3}z_1,\end{aligned}$$

which is nonnegative if $m \geqslant \frac{b}{2}+\frac{5}{4}$. □

Proof of Lemma 8.4(i) In view of (8-20) and Claim 1 we can assume that m is in $\left[\frac{b-1}{2}, \frac{b+1}{2}\right]$. We wish to show that for these m (of which there are one or two) we have $*+m\delta_* \geqslant z_5$. Since $\delta_* \leqslant 0$, this follows if $*+\frac{b+1}{2}\delta_* \geqslant z_5$, that is,

$$f_b(\lambda) := -2b(b+1)\lambda^2+b(3b+4)\lambda-(b^2+b-2) \geqslant 0$$

for $a \in \left[2b+4+\frac{1}{2m+2}, 2b+4+\frac{1}{2m+1}\right]$ and $m \in \left[\frac{b-1}{2}, \frac{b+1}{2}\right]$. Since $f_b'(\lambda) \leqslant -b^2 < 0$ and $m \geqslant \frac{b-1}{2}$, it suffices to show that $f_b(\lambda) \geqslant 0$ at

$$\lambda = \sqrt{\frac{2b+4+1/b}{2b}},$$

that is,

$$1+\frac{2}{b}+\frac{1}{2b^2} \geqslant \left(\frac{3b^2+7b+3+1/b}{3b^2+4b}\right)^2.$$

Subtracting 1 and multiplying by $2b^2(3b^2+4b)^2$ this inequality becomes

$$3b^4-8b^3-30b^2-12b-2 \geqslant 0,$$

which holds true for $b \geqslant 5$.

To deal with the cases $b \in \{3,4\}$ we return to $*+m\delta_* \geqslant z_5$, ie

$$1+(b+1)(z_2-z_1)+m(z_1+z_2-2w_1)-z_1 \geqslant 0. \tag{8-22}$$

Assume that $b=4$. Then $m=2$, and (8-22) becomes

$$7z_2+1 \geqslant 4z_1+4w_1 \quad \text{on } I := \left[12+\tfrac{1}{6}, 12+\tfrac{1}{5}\right],$$

ie $f(a) := -a+\frac{59}{8}\sqrt{\frac{a}{2}}-6 \geqslant 0$ on I. This holds true since $f'(a) < 0$ on I and $f\left(12+\frac{1}{5}\right) > 0$. Finally, if $b=3$, then $m \in \{1,2\}$. For $m=2$, (8-22) becomes

$$-2a+\frac{13}{2}\sqrt{\frac{3a}{2}}-5 \geqslant 0$$

on $\left[10+\frac{1}{6}, 10+\frac{1}{5}\right]$, which holds true; and for $m=1$, (8-22) becomes $-a+\frac{31}{2}\sqrt{\frac{a}{6}}-10 \geqslant 0$ on $\left[10+\frac{1}{4}, 10+\frac{1}{3}\right]$, which holds true too. □

Proof of Lemma 8.4(ii) In this case, (8-20) and Claim 2 show that we can assume that $m \in \left[\frac{b}{2}, \frac{b}{2}+1\right]$. We wish to show that for these m we have $* + (m-1)\delta_* \geqslant z_5$. Since $\delta_* \leqslant 0$, this follows if $* + \frac{b}{2}\delta_* \geqslant z_5$, that is,

$$f_b(\lambda) := -2b^2\lambda^2 + (3b^2+3b-1)\lambda - b(b+2) \geqslant 0$$

for $a \in \left[2b+4+\frac{1}{2m+1}, 2b+4+\frac{1}{2m}\right]$ and $m \in \left[\frac{b}{2}, \frac{b}{2}+1\right]$. Since $f_b'(\lambda) \leqslant -b^2+3b-1 < 0$ and since $m \geqslant \frac{b}{2}$, it suffices to show that $f_b(\lambda) \geqslant 0$ at

$$\lambda = \sqrt{\frac{2b+4+1/b}{2b}},$$

that is,

$$1 + \frac{2}{b} + \frac{1}{2b^2} \geqslant \left(\frac{3b^2+6b+1}{3b^2+3b-1}\right)^2.$$

Subtracting 1 and multiplying by $2b^2(3b^2+3b-1)^2$ this becomes

$$3b^4 - 6b^3 - 21b^2 - 2b + 1 \geqslant 0,$$

which holds true for $b \geqslant 4$.

Assume that $b = 3$. Then $m = 2$, and $* + (m-1)\delta_* \geqslant z_5$ becomes $-a + \frac{31}{2}\sqrt{\frac{a}{6}} - 10 \geqslant 0$ on $\left[10+\frac{1}{5}, 10+\frac{1}{4}\right]$, which holds true. □

Proof of Lemma 8.5 This is equivalent to

$$1 + m\delta_* + b(z_2 - z_1) + z_2 - w_1 \geqslant z_2, w_2. \tag{8-23}$$

Since $1 = (2m+1)w_1 + w_2$, this is equivalent to

$$m(z_1+z_2) + b(z_2-z_1) + z_2 + w_2 \geqslant z_2, w_2,$$

which follows if

$$m(z_1+z_2) + b(z_2-z_1) \geqslant 0. \tag{8-24}$$

Claim 1 (8-24) *holds for* $m \geqslant \frac{b}{3}$.

Indeed, since $2z_2 \geqslant w_1 \geqslant z_1$ by Lemma 8.2 and by assumption,

$$m(z_1+z_2) + b(z_2-z_1) = (m-b)z_1 + (m+b)z_2 \geqslant (3m-b)\frac{z_1}{2}.$$

Claim 2 (8-23) *holds for* $m \leqslant \frac{b}{3} - 1$.

Proof Since $\delta_* \leqslant 0$ and $w_1 \geqslant z_2,\ w_2$, it suffices to show that

$$1+\left(\frac{b}{3}-1\right)\delta_* + b(z_2-z_1)+z_2-w_1 \geqslant w_1, \tag{8-25}$$

or, equivalently, that

$$f_b(\lambda) := -4b^2\lambda^2 + (8b^2+2b-3)\lambda - 2(2b^2+b-3) \geqslant 0. \tag{8-26}$$

Note that $f_b'(\lambda) = -8b^2\lambda + (8b^2+2b-3) < 0$ for $\lambda \geqslant \lambda(\beta_b)$ since $(b+1)\,f_b'(\lambda(\beta_b)) = -(6b^2+5b+3) < 0$. Hence (8-26) follows from $bf_b(\lambda(\gamma_b)) = b-3 \geqslant 0$. □

Claim 3 (8-23) *holds for* $m \leqslant \frac{b-1}{3}$ *if* $b \geqslant 7$.

Proof It suffices to show that

$$1+\frac{b-1}{3}\delta_* + b(z_2-z_1)+z_2-w_1 \geqslant w_1,$$

or, equivalently, that

$$g_b(\lambda) := -(4b^2+8b)\lambda^2 + (8b^2+6b+1)\lambda - 4b^2+2b+14 \geqslant 0. \tag{8-27}$$

Since $g_b'(\lambda) < 0$ for $\lambda \geqslant 1$, (8-27) follows from $bg_b(\lambda(\gamma_b)) = b-7$. □

In view of the three claims above we are left with showing the lemma for $b \in \{4,5\}$ and $m=1$.

Assume that $b=5$. It suffices to show that $1+\delta_*+5(z_2-z_1)+z_2 \geqslant 2w_1$ for $a \in [\beta_b, \gamma_b]$, that is,

$$f(\lambda) := -40\lambda^2 + 73\lambda - 30 \geqslant 0 \quad \text{for } a \in [\beta_b, \gamma_b].$$

This holds true since $f'(\lambda) < 0$ for $\lambda \geqslant 1$ and $f(\lambda(\gamma_b)) = 0$.

Assume that $b=4$. Then $*_1 = 1+\delta_*+5(z_2-z_1)+z_2-w_1$. The inequality $*_1 \geqslant 2z_2 - z_1$ becomes $1+5z_2 \geqslant 3w_1+3z_1$, or

$$f(\lambda) := -8\lambda^2 + 14\lambda - 5 \geqslant 0,$$

which holds true since $f'(\lambda) < 0$ for $\lambda \geqslant 1$ and $f(\lambda(\gamma_b)) = 0$. The inequality $*_1 \geqslant w_2+z_2-z_1 = 1-3w_1+z_2-z_1$ becomes $6z_2 \geqslant 3z_1$, which holds true. □

Proof of Lemma 8.6 This is equivalent to

$$\zeta := 1+m\delta_* + (b+1)(z_2-z_1) \geqslant z_2,\ w_2+z_2-z_1. \tag{8-28}$$

Since $1 = 2mw_1 + w_2$, this is equivalent to $m(z_1+z_2)+b(z_2-z_1) \geqslant z_1 - w_2, 0$, which follows if

$$m(z_1+z_2)+b(z_2-z_1) \geqslant z_1,\ 0. \tag{8-29}$$

Claim 1 (8-29) *holds for* $m \geqslant \frac{b+2}{3},\ \frac{b}{3}$.

Claim 2 (8-28) *holds for* $m \leqslant \frac{b}{3},\ \frac{b-2}{3}$.

Proof For $m \leqslant \frac{b}{3}$, the inequality $\zeta \geqslant z_2$ in (8-28) follows from the inequality $1 + \frac{b}{3}\delta_* + (b+1)(z_2 - z_1) \geqslant z_2$, which is equivalent to (8-25). For $m \leqslant \frac{b-2}{3}$, the inequality $\zeta \geqslant w_2 + z_2 - z_1$ in (8-28) follows from $1 + \frac{b-2}{3}\delta_* + b(z_2-z_1) \geqslant w_1$, or

$$f_b(\lambda) := (-4b^2+2b)\lambda^2 + (8b^2-2b-4)\lambda - 4b^2 + 7 \geqslant 0. \tag{8-30}$$

Note that $f_b'(\lambda) < 0$ for $\lambda \geqslant \lambda(\beta_b)$ since $(b+1)f_b'(\lambda(\beta_b)) = -2(3b^2+b+1) < 0$. Hence (8-30) follows from $bf_b(\lambda(\gamma_b)) = b-2$. □

Claim 3 (8-28) *holds for* $m = \frac{b+1}{3}$ *if* $b \geqslant 5$, *and for* $m = \frac{b-1}{3}$ *if* $b \geqslant 4$.

Proof The first assertion is that $1 + \frac{b+1}{3}\delta_* + (b+1)(z_2-z_1) \geqslant z_2$ for $b \geqslant 5$, or

$$g_b(\lambda) := (-4b^2-4b)\lambda^2 + (8b^2+4b-1)\lambda - 4b^2 + 10 \geqslant 0. \tag{8-31}$$

Since $g_b'(\lambda) < 0$ for $\lambda \geqslant 1$, (8-31) follows from $bg_b(\lambda(\gamma_b)) = b-5$.

The second assertion follows if $1 + \frac{b-1}{3}\delta_* + b(z_2-z_1) \geqslant w_1$ for $b \geqslant 4$, that is,

$$h_b(\lambda) := (-4b^2-2b)\lambda^2 + (8b^2-2)\lambda - 4b^2 + 2b + 11 \geqslant 0. \tag{8-32}$$

Since $h_b'(\lambda) < 0$ for $\lambda \geqslant 1$, (8-32) follows from $bh_b(\lambda(\gamma_b)) = b-4$. □

The three claims above imply Lemma 8.6. □

Remark 8.7 One can use the reduction method also for showing that $c_2(a) = \frac{1}{2}\sqrt{a}$ on $[\beta_2, u_2(2)] = \left[8\frac{1}{36}, 9\right]$, of course. Contrary to all other assertions in Lemma 8.2, assertion (v) does not hold for $b=2$ if $a \geqslant 8.0831$, however. The reduction scheme for $b=2$ on $[\beta_b, u_b(2)]$ is therefore quite different from the one for $b \geqslant 3$, in particular in Case 1(b).

References

[1] **V I Arnol'd**, **S M Guseĭn-Zade**, **A N Varchenko**, *Singularities of differentiable maps, II*, Monographs in Mathematics 83, Birkhäuser, Boston (1988) MR

[2] **P Biran**, *Symplectic packing in dimension* 4, Geom. Funct. Anal. 7 (1997) 420–437 MR

[3] **P Biran**, *A stability property of symplectic packing*, Invent. Math. 136 (1999) 123–155 MR

[4] **M Burkhart**, **P Panescu**, **M Timmons**, *Symplectic embeddings of* 4*–dimensional ellipsoids into polydiscs*, Involve 10 (2017) 219–242

[5] **O Buse**, **R Hind**, *Symplectic embeddings of ellipsoids in dimension greater than four*, Geom. Topol. 15 (2011) 2091–2110 MR

[6] **O Buse**, **R Hind**, *Ellipsoid embeddings and symplectic packing stability*, Compos. Math. 149 (2013) 889–902 MR

[7] **O Buse**, **R Hind**, **E Opshtein**, *Packing stability for symplectic* 4*–manifolds*, Trans. Amer. Math. Soc. 368 (2016) 8209–8222 MR

[8] **O Buse**, **M Pinsonnault**, *Packing numbers of rational ruled four-manifolds*, J. Symplectic Geom. 11 (2013) 269–316 MR

[9] **K Choi**, **D Cristofaro-Gardiner**, **D Frenkel**, **M Hutchings**, **V G B Ramos**, *Symplectic embeddings into four-dimensional concave toric domains*, J. Topol. 7 (2014) 1054–1076 MR

[10] **D Cristofaro-Gardiner**, *ECH capacities and dynamics*, in preparation

[11] **D Cristofaro-Gardiner**, *Symplectic embeddings from concave toric domains into convex ones*, preprint (2014) arXiv

[12] **D Cristofaro-Gardiner**, **R Hind**, *Symplectic embeddings of products*, preprint (2015) arXiv To appear in Comment. Math. Helv.

[13] **D Cristofaro-Gardiner**, **A Kleinman**, *Ehrhart polynomials and symplectic embeddings of ellipsoids*, preprint (2013) arXiv

[14] **D Frenkel**, **D Müller**, *Symplectic embeddings of* 4*–dim ellipsoids into cubes*, J. Symplectic Geom. 13 (2015) 765–847 MR

[15] **M Gromov**, *Pseudo holomorphic curves in symplectic manifolds*, Invent. Math. 82 (1985) 307–347 MR

[16] **R Hind**, *Some optimal embeddings of symplectic ellipsoids*, J. Topol. 8 (2015) 871–883 MR

[17] **R Hind**, **E Kerman**, *New obstructions to symplectic embeddings*, Invent. Math. 196 (2014) 383–452 MR

[18] **M Hutchings**, *Quantitative embedded contact homology*, J. Differential Geom. 88 (2011) 231–266 MR

[19] **M Hutchings**, *Recent progress on symplectic embedding problems in four dimensions*, Proc. Natl. Acad. Sci. USA 108 (2011) 8093–8099 MR

[20] **Y Karshon**, **L Kessler**, *Distinguishing symplectic blowups of the complex projective plane*, preprint (2014) arXiv To appear in J. Symplectic Geom.

[21] **B-H Li**, **T-J Li**, *Symplectic genus, minimal genus and diffeomorphisms*, Asian J. Math. 6 (2002) 123–144 MR

[22] **T-J Li**, **A-K Liu**, *Uniqueness of symplectic canonical class, surface cone and symplectic cone of* 4*–manifolds with* $B^+ = 1$, J. Differential Geom. 58 (2001) 331–370 MR

[23] **D McDuff**, *Symplectic embeddings of* 4*–dimensional ellipsoids*, J. Topol. 2 (2009) 1–22 MR

[24] **D McDuff**, *The Hofer conjecture on embedding symplectic ellipsoids*, J. Differential Geom. 88 (2011) 519–532 MR

[25] **D McDuff**, **L Polterovich**, *Symplectic packings and algebraic geometry*, Invent. Math. 115 (1994) 405–434 MR

[26] **D McDuff**, **F Schlenk**, *The embedding capacity of* 4*–dimensional symplectic ellipsoids*, Ann. of Math. 175 (2012) 1191–1282 MR

[27] **F Schlenk**, *Symplectic embeddings of ellipsoids*, Israel J. Math. 138 (2003) 215–252 MR

[28] **F Schlenk**, *Embedding problems in symplectic geometry*, De Gruyter Expositions in Mathematics 40, de Gruyter, Berlin (2005) MR

[29] **F Schlenk**, *Symplectic embedding problem, old and new*, book in preparation

[30] **P Seidel**, *Lagrangian two-spheres can be symplectically knotted*, J. Differential Geom. 52 (1999) 145–171 MR

DCG: *Mathematics Department, Harvard University*
1 Oxford Street, Cambridge, MA 02138, United States

DF and FS: *Institut de Mathématiques, Université de Neuchâtel*
Rue Émile-Argand 11, CH-2000 Neuchâtel, Switzerland

gardiner@math.harvard.edu, david@frenkel.ch, schlenk@unine.ch

Received: 26 April 2016

Algebraic & Geometric Topology 17 (2017) 1261–1281

A Khovanov stable homotopy type for colored links

ANDREW LOBB
PATRICK ORSON
DIRK SCHÜTZ

We extend Lipshitz and Sarkar's definition of a stable homotopy type associated to a link L whose cohomology recovers the Khovanov cohomology of L. Given an assignment c (called a *coloring*) of a positive integer to each component of a link L, we define a stable homotopy type $\mathcal{X}_{\mathrm{col}}(L_c)$ whose cohomology recovers the c–colored Khovanov cohomology of L. This goes via Rozansky's definition of a categorified Jones–Wenzl projector P_n as an infinite torus braid on n strands.

We then observe that Cooper and Krushkal's explicit definition of P_2 also gives rise to stable homotopy types of colored links (using the restricted palette $\{1, 2\}$), and we show that these coincide with $\mathcal{X}_{\mathrm{col}}$. We use this equivalence to compute the stable homotopy type of the $(2, 1)$–colored Hopf link and the 2–colored trefoil. Finally, we discuss the Cooper–Krushkal projector P_3 and make a conjecture of $\mathcal{X}_{\mathrm{col}}(U_3)$ for U the unknot.

57M27

1 Introduction

1.1 Categorification

Given a semisimple Lie algebra $\mathfrak{g}$ and a link $L \subset S^3$ in which each component of L is decorated by an irreducible representation of $\mathfrak{g}$, the Reshetikhin–Turaev construction returns an invariant of that link that can, in principle, be computed combinatorially from any diagram of L. The standard example is the Jones polynomial, which arises from decorating all components with the fundamental representation $V = V^1$ of $\mathfrak{sl}_2$ (here the superscript 1 on the representation refers to the highest weight of V being 1). There are then two obvious first directions in which one can generalize.

On the one hand, one might vary the Lie algebra and consider instead $\mathfrak{sl}_n$, but still with the fundamental representation of $\mathfrak{sl}_n$. Each invariant obtained this way is a 1–variable specialization of the 2–variable HOMFLYPT polynomial, and satisfies an oriented skein relation, which yields the benefit of easy computability.

Published: 14 March 2017 DOI: 10.2140/agt.2017.17.1261

On the other hand, one might stick with $\mathfrak{sl}_2$, but vary the irreducible representation. There is one irreducible $(n+1)$–dimensional representation V^n (of highest weight n) for each $n \geq 1$. Decorating with V^n gives rise to the so-called n–colored Jones polynomial. The colored Jones polynomials no longer satisfy such pleasant skein relations, but they are powerful — for example giving rise to 3–manifold invariants (also called Reshetikhin–Turaev invariants or, in another form, Turaev–Viro invariants).

Both the $\mathfrak{sl}_n$ polynomials and the colored Jones polynomials admit categorifications — that is, they can be exhibited as the graded Euler characteristic of bigraded cohomology theories. In the case of $\mathfrak{sl}_n$, this is Khovanov–Rozansky cohomology [6]. In the case of the colored Jones polynomial there are constructions due to many authors, some inequivalent, although the two we shall be considering in fact give isomorphic cohomologies. The first is due to Rozansky [10] and the second due to Cooper and Krushkal [3]. In both cases, the fundamental representation of $\mathfrak{sl}_2$ gives Khovanov cohomology [5].

1.2 Spacification

Recently it has been shown that Khovanov cohomology admits a *spacification*, that is, for any link there is a stable homotopy type $\mathcal{X}(L)$ whose cohomology gives Khovanov cohomology (the *bi*grading of Khovanov cohomology is recovered from a splitting of $\mathcal{X}(L)$ into wedge of spaces indexed by the integers). This is work due to Lipshitz and Sarkar [8]. We note that the term "spacification" is not yet well-defined, since it is unclear exactly what properties one should require of it. (For example: should just taking a wedge of the Moore spaces determined by the cohomology count as a spacification?) Nevertheless, we find it a convenient shorthand for now.

It is a natural question if other Reshetikhin–Turaev invariants admitting categorifications can further be spacified. In the $\mathfrak{sl}_n$ case, work by two of the authors with Dan Jones [4] has constructed an $\mathfrak{sl}_n$ stable homotopy type given the input of a *matched* knot diagram. There is good evidence that this stable homotopy type should be diagram-independent. For $n = 2$ it agrees with the stable homotopy type due to Lipshitz and Sarkar.

The case of the colored Jones invariants is, in a sense, a little easier. In particular, Rozansky's categorification admits spacification. In the case of the c–colored unknot whose categorification is, in Rozansky's construction, the stable limit of the Khovanov cohomology of c–stranded torus links as the number of twists goes to infinity, this has been observed by Willis [12], whose paper appeared on the arXiv while this one was being written. The case of a c–colored link in general is no harder, and in fact Rozansky has already taken care of the difficult work.

Since the Cooper–Krushkal and the Rozansky categorifications are equivalent, the natural expectation is that one can lift the Cooper–Krushkal categorification to a spacification equivalent to the Rozansky spacification. This turns out to be straightforward in the 2–colored case, but at least the more obvious attempt fails in the 3–colored case, as we discuss later.

1.3 Computational results

We shall define a stable homotopy type $\mathcal{X}_{\mathrm{col}}(L_c)$, where L_c is a framed link with a coloring c of its components by positive integers. Picking the coloring 1 for each component returns the stable homotopy type $\mathcal{X}_{\mathrm{col}}(L_1)$, a grading-shifted version of Lipshitz and Sarkar's stable homotopy type $\mathcal{X}(L)$.

We make some computations for certain links and colorings in Section 4. Already in the simplest case these show interesting behavior: the link with the lowest positive crossing number is the Hopf link and the first coloring which has not yet been considered by Lipshitz and Sarkar is where one component is colored with 2 and the other with 1. The tail of the colored Khovanov cohomology of the $(2,1)$–colored Hopf link agrees with the tail of the colored Khovanov cohomology of the $(2,1)$–colored 2–component unlink. Nevertheless, we observe that even these tails can be distinguished by the stable homotopy type.

Although we are not yet able to compute fully the stable homotopy type of the 3–colored unknot, we make a conjecture based on some partial computations. This conjecture is interesting because its truth would imply that the periodicity of the tail of the stable homotopy type of a colored link (even in the case of the 3–colored unknot) can be longer than the periodicity of the tail of its cohomology.

1.4 Plan of the paper

In Section 2 we first observe that we can combine Rozansky's insight with the work of Lipshitz and Sarkar. This combination is straightforward and yields a stable homotopy type of a framed colored link whose cohomology recovers colored Khovanov cohomology. Secondly, we give ourselves a framework in which to make computations. For this it makes more sense to use the Cooper–Krushkal categorification, which, at least in the case of colors 2 and 3, is entirely explicit. We define what we mean by a lift of the Cooper–Krushkal categorification to a spacification and show that any such lift gives the same stable homotopy type as that arising from Rozansky's construction.

In Section 3, we construct such a lift of the Cooper–Krushkal categorification for colorings taken from the restricted palette $\{1,2\}$. The case of 3–colored cannot be made to work in the way that one might expect (there is an explicit obstruction to this). Finally, in Section 4 we make computations as already discussed in Section 1.3. At the end of this section we give a discussion of the Cooper–Krushkal 3–colored case.

Acknowledgements We thank the anonymous referee, whose comments much improved our exposition. The authors were partially supported by the EPSRC Grant EP/M000389/1.

2 Two approaches to a colored stable homotopy type

The colored Jones polynomial is an invariant of framed links L in which each component of L has been assigned a *color*, or in other words a positive integer weight. We write the color of a component k of L as $c(k)$, and often keep track of the coloring as a subscript L_c.

To compute the polynomial one takes a diagram of L_c in which the self-writhe of each component is equal to its framing. Then one replaces each component k by $c(k)$ parallel copies following the blackboard framing. Finally, one places on each component a *Jones–Wenzl projector*. This projector is an element of the relevant Temperley–Lieb algebra, with coefficients in rational functions of q. Finally, one applies the Kauffman bracket, and obtains an element of $\mathbb{Z}[\![q, q^{-1}]$ by expanding in powers of q (or an element of $\mathbb{Z}[q, q^{-1}]\!]$ by expanding in powers of q^{-1}).

The Jones–Wenzl projector is idempotent and satisfies turnback-triviality. It turns out that these two universal properties are enough to determine it completely. The Jones–Wenzl projector should in principle lift, in a categorification of the colored Jones polynomial, to a complex in Bar-Natan's tangles-and-cobordisms category [1], satisfying idempotence and turnback-triviality up to chain homotopy equivalence. Cooper and Krushkal [3] and Rozansky [10] give ways of achieving such a lift. Cooper and Krushkal proceed explicitly and give a categorified projector that they define inductively, while Rozansky realizes the categorified projector as a limit of the complexes associated to torus braids. It is surprising that the latter approach had apparently not been considered even at the decategorified level until Rozansky's insight! As observed by Cooper and Krushkal, categorified universal properties imply that the two competing categorifications give identical cohomological groups.

2.1 Grading and other conventions

We note that there is a discrepancy in the grading conventions between the original paper of Khovanov's [5], Rozansky's torus braids paper [10] and Cooper and Krushkal's

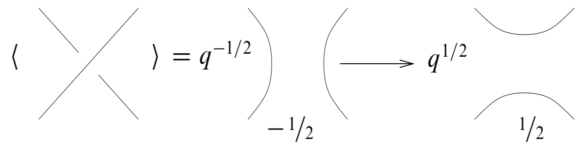


Figure 1: We follow the grading conventions as depicted in the complex that we associate to a single crossing. The complex is supported in cohomological degrees $\pm\frac{1}{2}$, and a quantum grading shift is applied. The differential increases the cohomological degree by 1 and preserves the quantum grading. A crossingless circle has complex supported in cohomological degree 0 and quantum degrees $+1$ and -1.

paper [3]. We apologize for possibly adding to the confusion. We shall essentially work with the bigrading conventions used by Bar-Natan [1] up to an overall shift. The overall shift makes it easier to treat the colored Khovanov cohomology as an invariant of a colored framed link, with no choice of orientation. The convention is depicted in Figure 1.

With these conventions, the Khovanov complex $\langle D\rangle$ of a diagram D is invariant up to bigraded homotopy equivalence under the second and third Reidemeister moves, but it is only invariant up to an overall shift under the first Reidemeister move. Hence it becomes a chain homotopy invariant of framed links (where the framing is given by the blackboard-framing of a diagram). If, on the other hand, D and D' differ by the first Reidemeister move with the writhes satisfying $w(D) = w(D') + 1$, then there is a bigraded homotopy equivalence between

$$\langle D\rangle \quad \text{and} \quad h^{-1/2}q^{-3/2}\langle D'\rangle,$$

where the powers of h and q represent cohomological and quantum degree shifts in the usual way.

2.2 Rozansky spacification

Rozansky [10] has given an approach to colored Khovanov cohomology that expresses the c–colored cohomology of a link L as the limit of the Khovanov cohomologies of a c–strand cable of L in which one puts an increasing number of twists. The stabilization of the cohomology was observed earlier by Stošić in the case of L being the unknot, which amounts to the stabilization of the cohomology of the (p,c)–torus link as $p \to \infty$.

We now summarize the construction. We shall be sticking with the convention of right-handed full twists, although there is an analogous story for left-handed twists. In

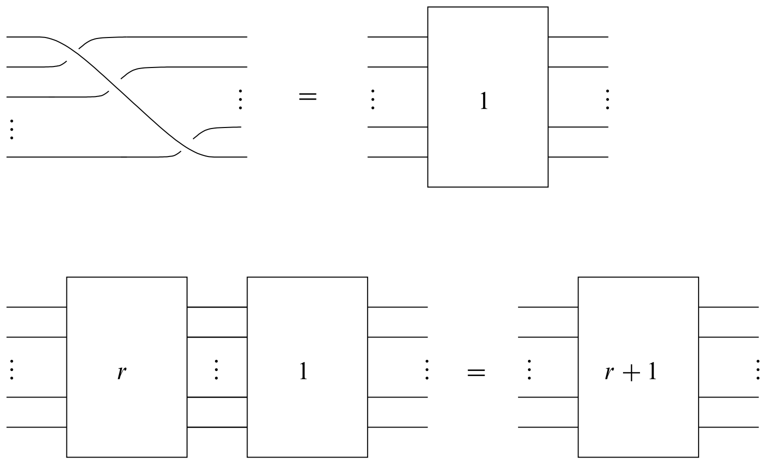


Figure 2: This shows inductively what is meant by twisting r times positively on an n–stranded braid.

Figure 2, we describe what is meant by twisting r times on an n–stranded braid. We write this braid as $B_{r,n}$. To each such braid, Bar-Natan's construction [1] associates a complex, which we shall denote by $\langle B_{r,n}\rangle$. In this complex, each cochain group is a vector of tangle smoothings, each such smoothing coming with a quantum degree. We shall apply a bigrading shift to this complex so that the resolution which is the identity braid group element is in cohomological degree 0 and comes with quantum degree shift 0. We write the shifted complex as $h^{r(n-1)/2}q^{r(n-1)/2}\langle B_{r,n}\rangle$, where the exponents of h and q denote cohomological and quantum degree shifts, respectively. Note that all other resolutions of the braid now lie in positive cohomological degrees.

For each $r \geq 1$ there is a map of complexes

$$F_r\colon (hq)^{rn(n-1)/2}\langle B_{rn,n}\rangle \to (hq)^{(r\text{-}1)n(n-1)/2}\langle B_{(r-1)n,n}\rangle,$$

given by taking F_1 to be the identity in cohomological degree 0, and then defining F_r to be the tensor product of F_1 with the identity on $(hq)^{(r-1)n(n-1)/2}\langle B_{(r-1)n,n}\rangle$.

Rozansky shows that for large r the cone complex $\mathrm{Cone}(F_r)$ is homotopy equivalent to a complex in which each smoothing that appears has high cohomological and quantum degrees. For our purposes, we are mainly interested in the quantum degree; we have:

Proposition 2.1 [10, Theorem 4.4] *The cone* $\mathrm{Cone}(F_r)$ *is homotopy equivalent to a complex made up of circleless smoothings, where each such smoothing is shifted in quantum degree by at least* $2n(r-1)+1$.

The precise form of the quantum degree shift is unimportant for us; rather we note that it increases at least linearly with r.

Definition 2.2 Let D_c be an unoriented link diagram in which each component is colored by a positive integer weight (we write the coloring by weights as c), and each component k carries a basepoint. Let the diagram D^r_c be given by the blackboard-framed c–stranded cable of D_c in which each component k receives $c(k)r$ positive twists at the basepoint. In other words, the diagram is cut open at each basepoint and $B_{rc(k),c(k)}$ is inserted.

Definition 2.3 Let

$$G_r\colon (hq)^{\sum_k rc(k)(c(k)-1)/2}\langle D^r_c\rangle \to (hq)^{\sum_k (r-1)c(k)(c(k)-1)/2}\langle D^{r-1}_c\rangle$$

be induced by the tensor product of the maps F_r at each basepoint.

Lemma 2.4 *It follows from Proposition 2.1 that, for fixed j and for all large enough r, the map of cohomologies*

$$H^{i,j}\big((hq)^{\sum_k rc(k)(c(k)-1)/2}\langle D^r_c\rangle\big) \to H^{i,j}\big((hq)^{\sum_k (r-1)c(k)(c(k)-1)/2}\langle D^{r-1}_c\rangle\big)$$

induced by G_r is an isomorphism.

Proof There is more than one way to see this. For example, label the components $k_1,\dots,k_s$ and write

$$e_\alpha = \sum_{\beta=1}^{\beta=\alpha-1} \tfrac{1}{2}(r-1)c(k_\beta)(c(k_\beta)-1) + \sum_{\beta=\alpha}^{\beta=s} \tfrac{1}{2}rc(k_\beta)(c(k_\beta)-1),$$

and denote by $D^{k_\alpha}_c$ the result of taking the c–cable of D and adding rc twists at the basepoints of $k_\alpha,\dots,k_s$ and $(r-1)c$ twists at the basepoints of $k_1,\dots,k_{\alpha-1}$. Then we can write

$$G_r = \overline{F^{k_s}_r}\circ\cdots\circ\overline{F^{k_1}_r},$$

where

$$\overline{F^{k_\alpha}_r}\colon (hq)^{e_\alpha}\langle D^{k_\alpha}_c\rangle \to (hq)^{e_{\alpha+1}}\langle D^{k_{\alpha+1}}_c\rangle$$

is induced by F_r at a chosen basepoint. The cone $\mathrm{Cone}(\overline{F^{k_\alpha}_r})$ is homotopy equivalent to a complex made up of the tensor product of three Bar-Natan complexes of tangles. Namely:

- A complex of circleless smoothings at the chosen basepoint whose quantum degree increases linearly with r.

- At the other basepoints, the complexes $(hq)^{rn(n-1)/2}\langle B_{rn,n}\rangle$. After circle removal, these consist of circleless smoothings each in a nonnegative quantum degree. This can be seen by observing that the identity braid is in cohomological and quantum degree 0. Smoothings in cohomological degree d differ from the identity braid by exactly d surgeries and so contain at most $d-1$ circles.
- A complex independent of r arising from the Bar-Natan complex of the diagram away from the basepoints.

Finally we recall the homological algebra fact that

$$\mathrm{Cone}(k \circ l) = \mathrm{Cone}(\Sigma^{-1}\,\mathrm{Cone}(k) \to \mathrm{Cone}(l))$$

for maps of complexes $k\colon C \to C'$ and $l\colon C'' \to C$. This implies that $\mathrm{Cone}(G_r)$ can be represented by circleless smoothings such that the minimal quantum degree among them increases at least linearly with r. Since j was fixed, we can choose r large enough that cohomology of $\mathrm{Cone}(G_r)$ is 0 in quantum degree j, which means G_r induces an isomorphism in quantum degree j. □

Hence we can make the following definition:

Definition 2.5 For fixed j, the c–colored Khovanov cohomology of the diagram D framed by the componentwise writhe is defined to be the group

$$\mathrm{Kh}^{i,j}_{\mathrm{col}}(D_c) = H^{i,j}\big((hq)^{\sum_k rc(k)(c(k)-1)/2}\langle D^r_c\rangle\big)$$

for sufficiently large r.

Independence of the cohomology under Reidemeister moves II and III and under choice of basepoints follows immediately from the independence under Reidemeister moves II and III of standard Khovanov cohomology. The fact that a suitable Euler characteristic of the cohomology agrees with the c–colored Jones polynomial of D is due to Rozansky.

Since $H^{i,j}\big((hq)^{\sum_k rc(k)(c(k)-1)/2}\langle D^r_c\rangle\big)$ is simply a grading-shifted version of the usual Khovanov cohomology of D^r_c, the construction of Lipshitz and Sarkar gives rise to a stable homotopy type $\bar{\mathcal{X}}^j(D^r_c)$ which recovers the Khovanov cohomology as its (suitably shifted) singular cohomology groups.

Furthermore, observe that the map G_r is induced by quotienting out a subcomplex generated by standard generators of the Khovanov complex. This subcomplex corresponds to an upward-closed subcategory of the framed flow category associated by Lipshitz and Sarkar to D^r_c. It follows that G_r is induced by a map

$$g_r\colon \bar{\mathcal{X}}^j(D^{r-1}_c) \to \bar{\mathcal{X}}^j(D^r_c).$$

Since g_r gives an isomorphism on cohomology for all sufficiently large r, Whitehead's theorem implies that g_r is a stable homotopy equivalence for sufficiently large r.

Definition 2.6 We can now define the colored stable homotopy type for fixed j to be

$$\mathcal{X}^j_{\mathrm{col}}(D_c) = \overline{\mathcal{X}}^j(D^r_c)$$

for sufficiently large r. In other words, this is the homotopy colimit of the directed system of maps g_r.

The invariance of this stable homotopy type under choice of basepoints and under Reidemeister moves II and III follows from the invariance of the Lipshitz–Sarkar homotopy type under Reidemeister moves II and III.

Remark 2.7 Willis [12] gave Definition 2.6 in the case that D is the unknot and gave an independent argument that the limit of the system g_r exists. Using his own estimates of quantum degree rather than Rozansky's, Willis has independently defined the Rozansky spacification, in a paper appearing on the arXiv shortly after ours [13]. He further showed a stabilization of the spacification of the c–colored unknot as $c \to \infty$.

Remark 2.8 Definition 2.6 implies that the framing of the link components only affects the colored stable homotopy type up to an overall shift in bigrading, as is the case for the colored Khovanov cohomology. This is because the blackboard-framed c–cable of a 1–crossing Reidemeister 1–tangle is equivalent to a full twist in a c–stranded braid by a sequence of Reidemeister moves involving c Reidemeister I moves. Reidemeister moves preserve the stable homotopy type according to Lipshitz and Sarkar, but Reidemeister I moves introduce a shift (with our grading conventions).

2.3 Cooper–Krushkal spacification

In this subsection we give the properties that one might expect of a spacification based on the Cooper–Krushkal categorification. These properties are enough to imply that any such spacification is stably homotopy equivalent to the Rozansky spacification, as is verified in Section 2.4. The construction of such spacifications is, however, not straightforward, and we leave discussion of these to Section 3.

Suppose for each $n \geq 1$ that P_n is a complex of (n,n)–tangle smoothings in the sense of Bar-Natan [1], so that each P_n is a *universal projector* by [3, Definition 3.1]. Cooper and Krushkal have given a way of constructing such universal projectors. We note that a part of their definition of P_n is that the identity n–braid smoothing appears only once and in degree $(0,0)$, and that the quantum and cohomological degrees of every smoothing in the complex are nonnegative.

Suppose that T is a tangle diagram in the plane punctured by k discs with $2n_i$ ordered boundary points on the i^{th} disc. Then we may define the Khovanov cochain complex (of free abelian groups) $\langle T_{\boldsymbol{P}} \rangle$ by taking the tensor product of the Bar-Natan complex $\langle T \rangle$ and P_{n_i} for $i = 1, \dots, k$ in the obvious way.

Definition 2.9 A *Cooper–Krushkal framed flow category* (CKffc) is a choice of finite-object framed flow category (see [8] for the definition and references) $\mathcal{C}(T_{\boldsymbol{P}})$ refining the Khovanov cochain complex $\langle T_{\boldsymbol{P}} \rangle$ for each such T. Choosing a particular crossing of the tangle T we write T^0 and T^1 for the 0– and 1–resolutions of that crossing. We require that the standard generators corresponding to the subcomplex $\langle T^1_{\boldsymbol{P}} \rangle$ (resp. the quotient complex $\langle T^0_{\boldsymbol{P}} \rangle$) correspond to upward-closed (resp. downward-closed) framed flow subcategories of $\mathcal{C}(T_{\boldsymbol{P}})$ such that the associated CW–complex is stably homotopy equivalent to $|\mathcal{C}(T^1_{\boldsymbol{P}})|$ (resp. $|\mathcal{C}(T^0_{\boldsymbol{P}})|$).

Furthermore, if we denote by T^{id} the tangle diagram produced by filling the k^{th} boundary disc of T with the identity n_k–braid, then $\langle T^{\text{id}}_{\boldsymbol{P}} \rangle$ is naturally a quotient complex of $\langle T_{\boldsymbol{P}} \rangle$ generated by standard generators of $\langle T_{\boldsymbol{P}} \rangle$. We require this quotient complex to correspond to a downward-closed subcategory of $\mathcal{C}(T_{\boldsymbol{P}})$ with associated CW–complex stably homotopy equivalent to $|\mathcal{C}(T^{\text{id}}_{\boldsymbol{P}})|$.

Remark 2.10 We can restrict this definition, if we like, to certain values of n. In particular in this paper we give a genuine CKffc only for the color $n = 2$. For the color $n = 3$ we may slightly alter the definition of a CKffc, to arrive at a framed flow category spacifying a cohomology theory that has its graded Euler characteristic a nonstandard normalization of the 3–colored Jones polynomial. If we insist on the standard normalization we run into difficulties. We discuss this in Section 3.

Remark 2.11 The condition that a CKffc assigns a *finite-object* framed flow category is equivalent to the condition that the minimal quantum degree of the circleless smoothings in the i^{th} cochain group of P_n tends to infinity as $i \to \infty$. Although this is true for the explicit examples of universal projectors constructed by Cooper and Krushkal, it is not required by them axiomatically.

2.4 The equivalence

CKffcs are nice since, *if they exist*, they give an honest framed flow category whose associated stable homotopy type recovers colored Khovanov cohomology as its singular

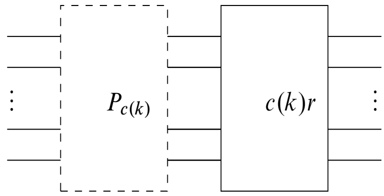


Figure 3: We describe how to form the complex $C^{i,j}_{P,r}(D)$ from a based c–colored diagram D. We take the blackboard-framed c–cable of D and at the basepoint of each component k of D we tensor in $P_{c(k)}$ and add $c(k)r$ twists as shown in the diagram. We then take the corresponding cochain complex and shift by $hq^{\sum_k rc(k)(c(k)-1)/2}$.

cohomology. Going via Rozansky's construction we are producing instead a stable homotopy type as a homotopy colimit of spaces arising from framed flow categories.

Nevertheless, we shall next see that CKffcs, if they exist, would give rise to the same stable homotopy types as does $\mathcal{X}^j_{\mathrm{col}}$. More precisely, let D be a link diagram framed by the componentwise writhe with each component k having a basepoint, and each being colored by a positive integer weight $c(k)$. We write D^{cab} for the tangle formed by cutting D open at each basepoint and then taking the blackboard-framed c–cable. Then we can consider the Bar-Natan cochain complex of free abelian groups formed by tensoring in each $P_{c(k)}$ corresponding to component k in the obvious way. This cochain complex is the Cooper–Krushkal complex that categorifies the colored Jones polynomial of D, and it is refined by the framed flow category $\mathcal{C}(D^{\mathrm{cab}}_P)$ provided by the CKffc. Writing $\mathcal{C}^j(D^{\mathrm{cab}}_P)$ for the part of this framed flow category in quantum degree j, we have the following result:

Proposition 2.12 *With the diagram D as above we have*

$$\mathcal{X}^j_{\mathrm{col}}(D_c) \simeq |\mathcal{C}^j(D^{\mathrm{cab}}_P)|.$$

Proof We fix j. We write $C^{i,j}_{P,r}(D)$ to be the cochain complex of free abelian groups formed by following the procedure as outlined in Figure 3. By the definition of a CKffc, there is a framed flow category $\mathcal{A}^j_{P,r}(D)$ that refines $C^{i,j}_{P,r}(D)$.

Consider the quotient complex C_1 of $C^{i,j}_{P,r}(D)$ consisting of all generators corresponding to taking the 0–resolution at each of the crossings of the twist regions at the basepoints. This corresponds to a downward-closed subcategory $\mathcal{A}_1$ of $\mathcal{A}^j_{P,r}(D)$. We observe firstly that $|\mathcal{A}_1|$ is stably homotopy equivalent to $|\mathcal{A}^j_{P,0}(D)|$, which is exactly $|\mathcal{C}^j(D^{\mathrm{cab}}_P)|$, and secondly that the corresponding upward-closed subcategory has trivial cohomology by the turnback-triviality condition on the projectors $P_{c(k)}$. Hence we have that

$$|\mathcal{A}^j_{P,r}(D)| \simeq |\mathcal{A}_1| \simeq |\mathcal{C}^j(D^{\mathrm{cab}}_P)|.$$

On the other hand, for any value of r, the complex $C^{i,j}_{P,r}(D)$ can be written as the total complex

$$(hq)^{\sum_k rc(k)(c(k)-1)/2}\langle D^r_c\rangle \to \Gamma^r_1 \to \cdots \to \Gamma^r_s \to \cdots,$$

where each Γ^r_s carries an internal differential arising from all crossings of D^r_c, while the part of the differential from Γ^r_s to Γ^r_{s+1} is induced by the differentials of the $P_{c(k)}$. This is because the identity braid smoothing is the only smoothing appearing in cohomological degree zero of each complex $P_{c(k)}$.

Now the minimal quantum degree of a generator in $\bigoplus_r \Gamma^r_s$ tends to $+\infty$ as s tends to $+\infty$ (see Remark 2.11). On the other hand, each Γ^r_s is chain-homotopy equivalent by Gauss-elimination to a complex in which the minimal quantum degree is bounded below by $b(r)$, a function independent of s and tending to $+\infty$ as r tends to $+\infty$. This follows from [10, Formula (4.9)] (taking into account our different grading conventions) and the observation that the cohomological Reidemeister I and II relations can be proved by Gauss-elimination.

Hence the lowest quantum degree of the support of the cohomology of the subcomplex

$$\Gamma^r_1 \to \cdots \to \Gamma^r_s \to \cdots$$

tends to $+\infty$ as r tends to $+\infty$. The quotient complex $(hq)^{\sum_k rc(k)(c(k)-1)/2}\langle D^r_c\rangle$ corresponds to a downward-closed subcategory of $\mathcal{A}^j_{P,r}(D)$ with associated stable homotopy type $\bar{\mathcal{X}}^j(D^r_c)$. So for large enough r we have

$$\mathcal{X}^j_{\mathrm{col}}(D_c) \simeq \bar{\mathcal{X}}^j(D^r_c) \simeq |\mathcal{A}^j_{P,r}(D)| \simeq |\mathcal{A}_1| \simeq |\mathcal{C}^j(D^{\mathrm{cab}}_P)|. \qquad \square$$

Remark 2.13 We have worked here with colored links, but all of what we have done applies, *mutatis mutandis*, to more general (in other words, not just diagrams obtained by cabling) closed diagrams containing Jones–Wenzl projectors.

3 Lifting the Cooper–Krushkal projectors

In this section we give a CKffc associated to link diagrams colored with colors drawn from the palette $\{1,2\}$. It would seem a priori very likely that the methods used in this construction should extend to the color 3, since for this color we have (due to Cooper and Krushkal [3]) an explicit and fairly simple cohomological projector. However, it turns out that there is an unexpected nontrivial obstruction to this extension. The

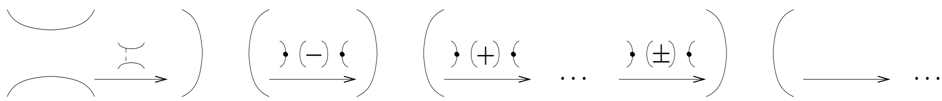


Figure 4: We show here the Cooper–Krushkal projector. We suppress the degree shifts for ease of visualization. The degree shifts can be determined by noting that the identity-braid or horizontal smoothing on the far left is in cohomological degree 0 and quantum degree 0, and all differentials raise the cohomological degree by 1 and preserve the quantum degree.

obstruction can be obviated by renormalizing the 3–colored Jones invariant of the 0–framed unknot to be

$$(q^{-2}+1+q^2)(1-q^2+q^4-q^6+\cdots) \quad \text{rather than} \quad q^{-2}+1+q^2.$$

We briefly discuss the obstruction and renormalization at the end of Section 4, but we do not give in this paper the full construction of the renormalized spacification.

3.1 A 2–colored Cooper–Krushkal projector

Two of the authors and Dan Jones [4] considered the 2–stranded braid of k crossings, each of the same sign. The Bar-Natan complex of this tangle has a particularly simple form: it is homotopy equivalent to a complex which has one circleless smoothing in each cohomological degree from $-\frac{1}{2}k$ to $\frac{1}{2}k$ (with the grading conventions used in this paper). Indeed, in Figure 4, we give the Cooper–Krushkal projector for the color 2; the Bar-Natan complex for the positively twisted k–crossing 2–braid is, up to an overall shift, the quotient complex of this projector consisting of all tangles of cohomological degree less than $k+1$.

Decomposing a closed link diagram D into a tensor product of such tangles, one can consider the tensor product of their simplified chain homotopy class representatives. This gives a cochain complex $\langle D\rangle^{\mathrm{simp}}$ (depending on the decomposition of D) of free abelian groups, and $\langle D\rangle^{\mathrm{simp}}$ is refined by a framed flow category given in [4]. The associated stable homotopy type was shown to be independent of the choice of decomposition, and it was observed that the decomposition in which each tangle has a single crossing returns the Lipshitz–Sarkar framed flow category.

Taking a suitably normalized version of this construction for $k=\infty$ gives a construction of a CKffc. In particular, this construction enables us to make nontrivial calculations of the colored stable homotopy types of the $(2,1)$–colored Hopf link as well as of the 2–colored trefoil.

Suppose that T is a tangle diagram in the plane punctured by k discs each with 4 ordered boundary points. Let the closed diagram T^r be given by filling in each disc with r positive full twists.

We consider a particular decomposition of T^r into a tensor product of tangles — specifically, we take one tangle (of $2r$ crossings) at each filled disc, one tangle for every other crossing of T^r, and finally the rest of the diagram which is crossingless.

Such a decomposition into tangles is exactly the input into the construction of [4]. So, incorporating now an overall shift and fixing a quantum degree j, there is a framed flow category $\mathcal{A}^j(T^r)$ refining the quantum degree j part of the simplified cochain complex $(hq)^{kr}\langle T^r\rangle^{\mathrm{simp}}$.

Finally we note that for fixed j and large enough r, the quantum degree j part of $(hq)^{kr}\langle T^r\rangle^{\mathrm{simp}}$ agrees with the quantum degree j part of the Cooper–Krushkal complex $\langle T_P\rangle$. This is because, in the $\mathfrak{sl}_2$ case, the construction of [4] gives a framed flow category refining the simplified Bar-Natan complex of link diagrams decomposed into tangles, each of which is a 2–braid. So, taking r to be large, the framed flow category $\mathcal{A}^j(T^r)$ provides our candidate for a CKffc. The remaining properties required of a CKffc are now straightforward to verify.

4 Examples

4.1 The 2–colored unknot

Consider a diagram of the blackboard framed 2–cable of the 0–crossing unknot containing a Cooper–Krushkal projector P_2. The generators in the resulting cochain complex come from smoothings with two circles in homological degree 0, and one circle in homological degree bigger than 0; compare Figure 4. The minimal quantum degree in which we get a generator is therefore $q = -2$ with one generator in homological degree 0. For $q = 0$ we get two generators in homological degree 0 and one in homological degree 1. For $q = 2$ there is one generator in homological degrees 0, 1 and 2 each.

For $q = 2j$ with $j \geq 2$ we get two generators, one in homological degree $j-1$ and one in degree j. The coboundary map alternates between multiplication by 0 and 2. The cohomology is therefore easily calculated, and determines the stable homotopy types because of thinness. We thus get

$$\begin{gathered}\mathcal{X}^{-2}_{\mathrm{col}}(U_2) = S^0, \quad \mathcal{X}^{0}_{\mathrm{col}}(U_2) = S^0, \quad \mathcal{X}^{2}_{\mathrm{col}}(U_2) = S^2,\\ \mathcal{X}^{4j}_{\mathrm{col}}(U_2) = M(\mathbb{Z}/2, 2j) \qquad \text{for } j \geq 1,\\ \mathcal{X}^{4j+2}_{\mathrm{col}}(U_2) = S^{2j+1} \vee S^{2j+2} \quad \text{for } j \geq 1.\end{gathered}$$

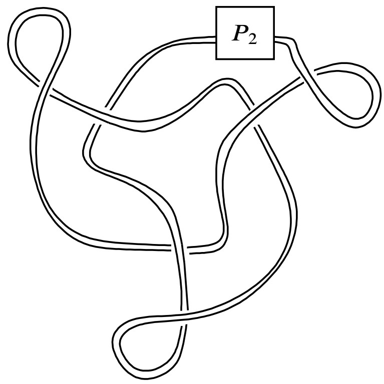


Figure 5: The 0–framed 2–cable of the right-handed trefoil with a Cooper–Krushkal projector placed on it

Note that the notation $M(G,n)$ stands for a Moore space, a space whose only nontrivial integral homology group is G in degree n.

4.2 The 2–colored trefoil

In Figure 5 we give a diagram of a 2–cable of the right-handed trefoil T containing a Cooper–Krushkal projector P_2. The extra loops ensure that we get the 0–framed 2–cable, and we denote it by T_2^0. For each quantum degree j this diagram gives rise to a framed flow category $\mathcal{A}$ as described in Section 3.1.

For calculational purposes, we want to remove the three double loops. Performing two Reidemeister I moves and one Reidemeister III move turns each double loop into $B_{-2,2}$, which can be absorbed by the projector P_2. However, because of the Reidemeister I moves, we get a shift in homological and quantum degrees. More precisely, we get $\langle D_2^r \rangle = h^3 q^9 \langle D_2'^{\,r-3} \rangle$, where D' is the standard 3–crossing diagram of the right-handed trefoil. Denoting the 2–colored right-hand trefoil with framing 3 by T_2^3, we get $\mathrm{Kh}^{i,j}_{\mathrm{col}}(T_2^0) = \mathrm{Kh}^{i-6,j-12}_{\mathrm{col}}(T_2^3)$.

Taking these shifts into account and working with the diagram for T_2^3, we see that the least quantum degree in $\mathcal{A} = \mathcal{A}_0$ which admits an object is given by $q = 2$ with homological degree $h = 0$, coming from a smoothened diagram with 4 circles. This is indeed the only object in this quantum degree.

The projector P_2 gives rise to upward-closed subcategories $\mathcal{A}_k$ for $k \geq 0$ generated by objects that arise from a tangle in P_2 of cohomological degree at least k. The highest quantum degree of an object in $\mathcal{A}_0 - \mathcal{A}_1$ is $q = 24$ coming from 6 circles in the smoothened diagram. It follows that for quantum degree $q \geq 26$ the relevant flow category $\mathcal{A}^q$ is a full subcategory of $\mathcal{A}_1$.

The quotient category $\mathcal{A}_k/\mathcal{A}_{k+1}$ for $k \geq 1$ is, up to degree shifts, the Lipshitz–Sarkar flow category of a diagram of the unknot with 12 crossings. Furthermore, this diagram can be transformed into the standard unknot diagram by performing six Reidemeister II moves. The category $\mathcal{A}_k/\mathcal{A}_{k+1}$ for $k \geq 1$ is therefore stably equivalent to a flow category containing two objects of homological degree $k+6$, one of quantum degree $2k+12$, the other of quantum degree $2k+10$.

Also notice that the associated cochain complexes to the flow categories $\mathcal{A}^q$ and $\mathcal{A}^{q+4}$ for $q \geq 26$ only differ in a cohomological shift by 2. If the tail turns out to be cohomologically thin (as it does), it follows that the stable homotopy types for q up to 28 determine all the stable homotopy types. The stable homotopy types for q up to 28 may be determined using the diagram $D_2'^{\,r}$ for large r. It turns out that $r = 8$ is sufficient, and the following calculations have been done using the program `KnotJob` [11].

We can identify all stable homotopy types from cohomology and Steenrod square calculations using the classification result of Baues and Hennes [2] with the exception of $q = 10$, where $\mathcal{X}^{10}_{\mathrm{col}(2)}(T)$ is either $S^3 \vee S^4 \vee S^6$ or $X(\varepsilon, 3) \vee S^4$. Recall that $X(\varepsilon, n)$ is the space obtained by attaching an $(n{+}3)$–cell to S^n using the nontrivial element of $\pi_2^{\mathrm{st}} \cong \mathbb{Z}/2$. Excluding this, we get

$$\mathcal{X}^2_{\mathrm{col}}(T_2^0) = S^0, \qquad \mathcal{X}^4_{\mathrm{col}}(T_2^0) = S^0,$$
$$\mathcal{X}^6_{\mathrm{col}}(T_2^0) = S^2, \qquad \mathcal{X}^8_{\mathrm{col}}(T_2^0) = X({}_2\eta, 2),$$
$$\mathcal{X}^{12}_{\mathrm{col}}(T_2^0) = X(\eta 2, 5) \vee S^6, \qquad \mathcal{X}^{14}_{\mathrm{col}}(T_2^0) = X(\eta 2, 5) \vee S^7 \vee S^8 \vee S^8,$$
$$\mathcal{X}^{16}_{\mathrm{col}}(T_2^0) = S^7 \vee M(\mathbb{Z}/4, 8) \vee M(\mathbb{Z}/2, 8), \qquad \mathcal{X}^{18}_{\mathrm{col}}(T_2^0) = S^9 \vee S^9 \vee M(\mathbb{Z}/2, 9) \vee S^{10},$$
$$\mathcal{X}^{20}_{\mathrm{col}}(T_2^0) = M(\mathbb{Z}/2, 9) \vee M(\mathbb{Z}/2, 10) \vee S^{11}, \qquad \mathcal{X}^{22}_{\mathrm{col}}(T_2^0) = S^{11} \vee M(\mathbb{Z}/2, 11) \vee S^{12},$$
$$\mathcal{X}^{24}_{\mathrm{col}}(T_2^0) = S^{12} \vee M(\mathbb{Z}/2, 12)$$

The tail is given by

$$\mathcal{X}^{4j+2}_{\mathrm{col}}(T_2^0) = S^{2j+1} \vee S^{2j+2} \quad \text{for } j \geq 6,$$
$$\mathcal{X}^{4j}_{\mathrm{col}}(T_2^0) = M(\mathbb{Z}/2, 2j) \qquad \text{for } j \geq 7.$$

Notice that for $j \geq 26$ we have $\mathcal{X}^j_{\mathrm{col}}(T_2^0) = \mathcal{X}^j_{\mathrm{col}}(U_2)$.

The notation $X(\eta 2, n)$ is taken from [2], and stands for an elementary Chang complex. It is an appropriate suspension of $\mathbb{RP}^4/\mathbb{RP}^1$ such that the first nontrivial homology

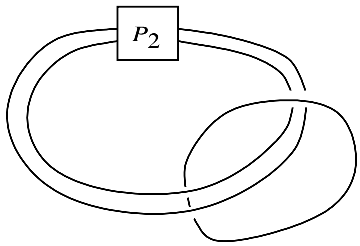


Figure 6: This is the 0–framed $(2,1)$–cable of the Hopf link, in which the 2–cabled component receives a Cooper–Krushkal projector.

group is in degree n. Similarly, $X({}_2\eta, m)$ is a suspension of $\mathbb{RP}^5/\mathbb{RP}^2$ such that the first nontrivial homology group is in degree m. Both spaces have nontrivial Sq^2 and are therefore not wedges of Moore spaces.

4.3 The $(2,1)$–colored Hopf link

We denote the $(2,1)$–colored 0–framed Hopf link by $H_{2,1}$. In Figure 6 we give a diagram of the Hopf link, in which one of the components has been replaced by a 0–framed 2–cable containing a Cooper–Krushkal projector P_2. For each quantum degree j this diagram gives rise to a framed flow category, as described in Section 3.1. The associated stable homotopy type is $\mathcal{X}_{\mathrm{col}}(H_{2,1})$.

Note that the diagram consists of the tensor product of three parts: the projector P_2 and then two tangles, each of which is a 2–crossing 2–braid. As before, we can filter the flow category via the projector, leading to categories $\mathcal{A}_j$ for $j \geq 0$.

For actual calculations, we replace the projector with a $(2r)$–tangle, so the resulting diagram is that of the $P(-2,2,2r)$ pretzel link. For a given quantum degree we can then use the method of [4] to get a flow category built from three tangles. The lowest quantum degree for which we can get an object is $q = -5$, for which there is exactly one object of homological degree -2.

For $q \geq 7$, all objects are contained in $\mathcal{A}_1$, and the categories $\mathcal{A}^{2j-1}$ and $\mathcal{A}^{2j+3}$ for $j \geq 4$ have the following similarity. If α is an object of $\mathcal{A}^{2j-1}$ which also sits in $\mathcal{A}_k$ for $k \geq 1$, there is a corresponding object $\overline{\alpha}$ in $\mathcal{A}^{2j+3}$ also in $\mathcal{A}_{k+2}$ with $|\overline{\alpha}| = |\alpha| + 2$. It is clear from the framing formulas in [4] that $\mathcal{M}(\alpha,\beta) \cong \mathcal{M}(\overline{\alpha},\overline{\beta})$ as framed manifolds, provided these are at most 1–dimensional.

Therefore the colored Khovanov cohomology of the tail is periodic, and since we only get nontrivial cohomology groups in three adjacent degrees, we also get periodicity of the stable homotopy type in the tail. This uses that the 1–dimensional moduli spaces agree with framing for $\mathcal{A}^{2j-1}$ and $\mathcal{A}^{2j+3}$. Calculation of Khovanov cohomology and the second Steenrod Square shows that

$$\begin{aligned} \mathcal{X}^{-5}_{\mathrm{col}}(H_{2,1}) &= S^{-2}, & \mathcal{X}^{-3}_{\mathrm{col}}(H_{2,1}) &= S^{-2}, \\ \mathcal{X}^{-1}_{\mathrm{col}}(H_{2,1}) &= S^{0}, & \mathcal{X}^{1}_{\mathrm{col}}(H_{2,1}) &= X({}_2\eta, 0), \\ \mathcal{X}^{3}_{\mathrm{col}}(H_{2,1}) &= S^1 \vee S^2 \vee S^2, & \mathcal{X}^{5}_{\mathrm{col}}(H_{2,1}) &= X({}_2\eta, 2). \end{aligned}$$

The tail is given by

$$\begin{aligned} \mathcal{X}^{4j-1}_{\mathrm{col}}(H_{2,1}) &= X(\eta 2, 2j-1) \vee S^{2j} && \text{for } j \geq 2, \\ \mathcal{X}^{4j+1}_{\mathrm{col}}(H_{2,1}) &= X({}_2\eta, 2j) \vee S^{2j+1} && \text{for } j \geq 2. \end{aligned}$$

The $(2,1)$–colored unlink $U_{2,1}$ is the disjoint union of the 1–colored 0–framed unknot U_1 and the 2–colored 0–framed unknot U_2. The stable homotopy type can therefore be derived using [7, Theorem 1]. More precisely, we get

$$\mathcal{X}^{j}_{\mathrm{col}}(U_{2,1}) = (\mathcal{X}^1(U) \wedge \mathcal{X}^{j-1}_{\mathrm{col}}(U_2)) \vee (\mathcal{X}^{-1}(U) \wedge \mathcal{X}^{j+1}_{\mathrm{col}}(U_2)).$$

Since both $\mathcal{X}^1(U) = S^0 = \mathcal{X}^{-1}(U)$, we have that $\mathcal{X}^{j}_{\mathrm{col}}(U_{2,1})$ is a wedge of Moore spaces for all j. In the tail we have

$$\begin{aligned} \mathcal{X}^{4j-1}_{\mathrm{col}}(U_{2,1}) &= S^{2j-1} \vee S^{2j} \vee M(\mathbb{Z}/2, 2j) && \text{for } j \geq 2, \\ \mathcal{X}^{4j+1}_{\mathrm{col}}(U_{2,1}) &= M(\mathbb{Z}/2, 2j) \vee S^{2j+1} \vee S^{2j+2} && \text{for } j \geq 2. \end{aligned}$$

In particular, we have

$$\mathrm{Kh}^{i,j}_{\mathrm{col}}(U_{2,1}) = \mathrm{Kh}^{i,j}_{\mathrm{col}}(H_{2,1})$$

for all $j \geq 7$ (a result that for high enough j is not unexpected, and that can be derived in ways other than brute calculation), but

$$\mathcal{X}^{j}_{\mathrm{col}}(U_{2,1}) \not\simeq \mathcal{X}^{j}_{\mathrm{col}}(H_{2,1}).$$

4.4 A conjecture on the 3–colored unknot

The stable homotopy type of the 0–framed 3–colored unknot $\mathcal{X}^{j}_{\mathrm{col}}(U_3)$ was partially computed by Willis [12], who showed that it was not a wedge of Moore spaces and so, in some sense, more interesting than just the colored Khovanov cohomology.

The 3–colored Khovanov cohomology can easily be calculated from [3, Section 4.4]. We summarize this in Table 1.

We observe that the tail is 3–periodic in quantum degrees $q = 2j+1$ starting from $j \geq 2$ with a homological shift by 4. Also, by simply looking at these groups we see

	i			
	1	2	3	4
$\mathrm{Kh}_{\mathrm{col}}^{i-4,-3}(U_3)$				$\mathbb{Z}$
$\mathrm{Kh}_{\mathrm{col}}^{i-4,-1}(U_3)$				$\mathbb{Z}$
$\mathrm{Kh}_{\mathrm{col}}^{i,1}(U_3)$		$\mathbb{Z}$		
$\mathrm{Kh}_{\mathrm{col}}^{i,3}(U_3)$			$\mathbb{Z}/2$	$\mathbb{Z}$
$\mathrm{Kh}_{\mathrm{col}}^{i+4j,6j+5}(U_3),\ j \geq 0$			$\mathbb{Z}$	$\mathbb{Z}$
$\mathrm{Kh}_{\mathrm{col}}^{i+4j,6j+1}(U_3),\ j \geq 1$	$\mathbb{Z}$	$\mathbb{Z}$		
$\mathrm{Kh}_{\mathrm{col}}^{i+4j,6j+3}(U_3),\ j \geq 1$	$\mathbb{Z}$		$\mathbb{Z}/2$	$\mathbb{Z}$

Table 1: The 3–colored Khovanov cohomology of the unknot

that except for quantum degrees $q = 6j + 3$ with $j \geq 0$ the stable homotopy types are wedges of spheres. In quantum degree $q = 3$ we have the nontrivial Steenrod Square coming from the torus knot $T_{4,3}$ first observed in [9], and which stably survives by [12].

The quantum degree $q = 9$ can be realized by the torus knot $T_{7,3}$, and computer calculations show a nontrivial Sq^2 in degree 5, with Sq^2 trivial in degree 6. The triviality in degree 6 indicates that the tail of the stable homotopy types is not 3–periodic, as the difference in 3–colored Khovanov cohomology in quantum degrees $q = 3$ and $q = 9$ comes from an extra generator in homological degree 0 killing the cocycle in degree 1, which survives in degree 5 for $q = 9$.

Computer calculations on $T_{8,3}$ show a trivial Sq^2 in degree 9, although this is not yet in the stable range for $q = 15$. Using a suitable diagram with a low number of tangles we have made computer calculations for $T_{13,3}$ which give evidence for the conjecture below:

Conjecture 4.1 *For $j \geq 1$ we have*

$$\mathcal{X}_{\mathrm{col}}^{12j-3}(U_3) = X(\eta 2, 8j-3) \vee S^{8j},$$
$$\mathcal{X}_{\mathrm{col}}^{12j+3}(U_3) = S^{8j+1} \vee X({}_2\eta, 8j+2).$$

Note that these two spaces are not stably homotopy equivalent, although they are Spanier–Whitehead dual when appropriately shifted. Following consideration of the Cooper–Krushkal projector P_3 explicitly described in [3] this conjecture is somewhat surprising. From P_3 the 3–periodicity follows immediately, so one may expect the same periodicity in the tail of the stable homotopy type.

Indeed, if one attempts a spacification based on lifting the Cooper–Krushkal projector P_3 to a framed flow category, one finds that the natural first attempt gives rise to 1–dimensional moduli spaces the framings of which also follow 3–periodicity. However, if one then pushes a little further to determine if one can genuinely lift P_3 to a CKffc, one runs into "ladybug matching" type problems which cannot all be solved simultaneously in a natural way, at least to the authors' eyes.

On the other hand, suppose that D is a tangle diagram in a disc with 6 ordered boundary points. This gives a cochain complex of free abelian groups $\langle D_{\boldsymbol{P}}\rangle$. Now consider the "reduced" subcomplex $\langle D_{\boldsymbol{P}}\rangle^{\mathrm{red}}$ of $\langle D_{\boldsymbol{P}}\rangle$ obtained by restricting to half the generators of $\langle D_{\boldsymbol{P}}\rangle$. Specifically, restrict to only those generators arising from a decoration by v_- of a chosen boundary point of D. In such a situation one can lift the cochain complex $\langle D_{\boldsymbol{P}}\rangle^{\mathrm{red}}$ to a framed flow category refining it. The ladybug matching problems no longer occur since we have thrown out enough objects of the flow category to kill them.

Unfortunately, this subcomplex is not really a very natural object to consider. The graded Euler characteristic is a renormalized version of the 3–colored Reshetikhin–Turaev invariant as discussed at the start of Section 3, but it is hard to motivate why one should consider this renormalization. Therefore we do not pursue this further here.

References

[1] **D Bar-Natan**, *Khovanov's homology for tangles and cobordisms*, Geom. Topol. 9 (2005) 1443–1499 MR

[2] **H J Baues**, **M Hennes**, *The homotopy classification of $(n-1)$–connected $(n+3)$–dimensional polyhedra, $n \geq 4$*, Topology 30 (1991) 373–408 MR

[3] **B Cooper**, **V Krushkal**, *Categorification of the Jones–Wenzl projectors*, Quantum Topol. 3 (2012) 139–180 MR

[4] **D Jones**, **A Lobb**, **D Schütz**, *An $\mathfrak{sl}_n$ stable homotopy type for matched diagrams*, preprint (2015) arXiv

[5] **M Khovanov**, *A categorification of the Jones polynomial*, Duke Math. J. 101 (2000) 359–426 MR

[6] **M Khovanov**, **L Rozansky**, *Matrix factorizations and link homology*, Fund. Math. 199 (2008) 1–91 MR

[7] **T Lawson**, **R Lipshitz**, **S Sarkar**, *Khovanov homotopy type, Burnside category, and products*, preprint (2015) arXiv

[8] **R Lipshitz**, **S Sarkar**, *A Khovanov stable homotopy type*, J. Amer. Math. Soc. 27 (2014) 983–1042 MR

[9] **R Lipshitz**, **S Sarkar**, *A Steenrod square on Khovanov homology*, J. Topol. 7 (2014) 817–848 MR

[10] **L Rozansky**, *An infinite torus braid yields a categorified Jones–Wenzl projector*, Fund. Math. 225 (2014) 305–326 MR

[11] **D Schütz**, *KnotJob*, software (2015) Available at `http://www.maths.dur.ac.uk/~dma0ds/knotjob.html`

[12] **M Willis**, *Stabilization of the Khovanov homotopy type of torus links*, preprint (2015) arXiv

[13] **M Willis**, *A colored Khovanov homotopy type for links, and its tail for the unknot*, preprint (2016) arXiv

AL and DS: *Department of Mathematical Sciences, Durham University*
Lower Mountjoy, Stockton Road, Durham, DH1 3LE, United Kingdom

PO: *Département de Mathématiques, Université du Québec à Montréal*
Montréal, Québec, H3C 3P8, Canada

`andrew.lobb@durham.ac.uk`, `patrick.orson@cirget.ca`,
`dirk.schuetz@durham.ac.uk`

Received: 27 April 2016 Revised: 12 August 2016

Guidelines for Authors

Submitting a paper to Algebraic & Geometric Topology

Papers must be submitted using the upload page at the AGT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

Preparing your article for Algebraic & Geometric Topology

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.

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 AGT is available on the AGT website.

`arXiv` papers

AGT papers are published simultaneously on the `arXiv`. If your paper has previously been deposited there, we shall need the `arXiv` password at acceptance time, to submit the published version to the `arXiv`.

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. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple, Mathematica or XFig. Many drawing tools such as Adobe Illustrator and Aldus FreeHand also produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, check with production by sending an email to agt@msp.warwick.ac.uk.

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.

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 17 Issue 2 (pages 645–1281) 2017

1472-2747(2017)17:2;1-U