JOURNAL OF THE ASSOCIATION FOR MATHEMATICAL RESEARCH

AMR
ASSOCIATION FOR MATHEMATICAL RESEARCH

VOLUME 2 : ISSUE 2

2024

JOURNAL OF THE ASSOCIATION FOR MATHEMATICAL RESEARCH

PUBLISHED BY

Association for Mathematical Research, amathr.org,
Davis, CA; Jenkintown PA.

JOURNAL OF THE ASSOCIATION FOR MATHEMATICAL RESEARCH, 2(2):154–208, 2024
jamathr.org

On Euler systems for motives and Heegner points

Takenori Kataoka * Takamichi Sano †

Received 14 Apr 2023; Revised 28 Jan 2024; Accepted 3 Apr 2024

Abstract:
We formulate an Iwasawa main conjecture for a higher rank Euler system for a general motive. We prove "one half" of the main conjecture under mild hypotheses. We also formulate a conjecture on "Darmon-type derivatives" of Euler systems and give an application to the Tamagawa number conjecture. Lastly, we specialize our general framework to the setting of Heegner points and give a natural interpretation of the Heegner point main conjecture in terms of rank two Euler systems.

Key words and phrases: Iwasawa main conjecture, Euler systems, Tamagawa number conjecture, Heegner points

1 Introduction

The aim of this paper is to study Iwasawa theory for motives in terms of higher rank Euler systems.

Let K be a number field and M a (pure) motive defined over K. Let p be an odd prime number and T a stable lattice of the p-adic étale realization of M. For simplicity, we assume that the coefficient ring $\mathcal{A}$ of T is the ring of integers of a finite extension of $\mathbb{Q}_p$ (e.g., $\mathcal{A} = \mathbb{Z}_p$). Let $T^*(1)$ denote the Kummer dual of T. Then we define the *basic rank* of T by

$$r = r_T := \mathrm{rank}_{\mathcal{A}} \left(\bigoplus_{v \in S_\infty(K)} H^0(K_v, T^*(1)) \right),$$

*Kataoka was partially supported by JSPS KAKENHI Grant Number 19J00763.
†Sano was partially supported by JSPS KAKENHI Grant Number 17K14171.

DOI: 10.56994/JAMR.002.002.001

where $S_\infty(K)$ denotes the set of all infinite places of K.

One expects that there is a canonical Euler system for T of rank r. For example, when $T = \mathbb{Z}_p(1)$, one sees that $r_{\mathbb{Z}_p(1)} = 1$ if and only if K is either $\mathbb{Q}$ or an imaginary quadratic field. When $K = \mathbb{Q}$, we have the cyclotomic unit Euler system. When K is an imaginary quadratic field, we have the elliptic unit Euler system. If $r_{\mathbb{Z}_p(1)} > 1$, then it is known that conjectural Rubin-Stark elements constitute an Euler system. When T is the p-adic Tate module of an elliptic curve over $\mathbb{Q}$ and $K = \mathbb{Q}$, then $r_T = 1$ and we have Kato's Euler system. In general, the validity of the equivariant Tamagawa number conjecture for the dual motive $M^*(1)$ implies the existence of a canonical higher rank Euler system for T of rank r which is related to leading terms of (complex) L-functions for $M^*(1)$ at $s = 0$.

In this paper, we study arithmetic properties of Euler systems in a general setting and give a new example of our general theory in the setting of Heegner points.

1.1 The Iwasawa main conjecture for motives

We first formulate an Iwasawa main conjecture for T. There are two types of the formulation: "with p-adic L-functions" and "without p-adic L-functions". In this paper, we study only the latter.

We sketch the formulation. Fix a finite set S of places of K which contains all infinite and p-adic places of K and all "bad" places for T. Let L/K be a finite abelian extension and K_∞/K a $\mathbb{Z}_p^d$-extension for some $d \geq 1$. We suppose that, for each $v \in S_\infty(K)$ that ramifies in L, we have $\mathrm{rank}_{\mathcal{A}}(H^0(K_v, T^*(1))) = \frac{1}{2}\mathrm{rank}_{\mathcal{A}}(T)$. We set $L_\infty := L \cdot K_\infty$ and $\Lambda := \mathcal{A}[[\mathrm{Gal}(L_\infty/K)]]$. Let $\mathbb{T} := T \otimes \Lambda$ be the Λ-adic deformation of T. Then one can construct a canonical map

$$\Theta : \det_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S}, \mathbb{T})) \to {\bigcap}_\Lambda^r H^1(\mathcal{O}_{K,S}, \mathbb{T}).$$

Here $\mathbf{R}\Gamma(\mathcal{O}_{K,S}, \mathbb{T})$ denotes the usual S-cohomology complex, $H^i(\mathcal{O}_{K,S}, \mathbb{T})$ its cohomology, and ${\bigcap}_\Lambda^r$ the r-th "exterior power bidual" over Λ (see §1.4 below). The Iwasawa main conjecture is formulated as follows.

Conjecture 1.1 (The Iwasawa main conjecture, see Conjecture 3.4)**.** *Suppose that a canonical rank $r(= r_T)$ Euler system $c_{L_\infty} \in {\bigcap}_\Lambda^r H^1(\mathcal{O}_{K,S}, \mathbb{T})$ is given. Then there exists a Λ-basis*

$$\mathfrak{z}_{L_\infty} \in \det_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S}, \mathbb{T}))$$

such that $\Theta(\mathfrak{z}_{L_\infty}) = c_{L_\infty}$.

When $T = \mathbb{Z}_p(1)$, a formulation of this form is given by Burns, Kurihara and the second author in [BKS17, Conj. 3.1 and Th. 3.4].

Let us consider the simplest "non-equivariant" case, i.e., $L = K$ and $d = 1$ (so that $L_\infty = K_\infty$ is a $\mathbb{Z}_p$-extension of K). Then one proves that Conjecture 1.1 is equivalent to the following "classical" formulation:

$$\mathrm{char}_\Lambda\left({\bigcap}_\Lambda^r H^1(\mathcal{O}_{K,S}, \mathbb{T})/\Lambda \cdot c_{K_\infty}\right) = \mathrm{char}_\Lambda(H^2(\mathcal{O}_{K,S}, \mathbb{T})),$$

where char denotes the characteristic ideal. (See Proposition 3.10.) We remark that the Iwasawa main conjecture of this form was studied earlier by Büyükboduk in [Büy09] and [Büy10].

One of the main results of this paper is to prove "one half" of the Iwasawa main conjecture for any given Euler system under some standard hypotheses.

Theorem 1.2 (see Theorem 3.17 for the precise statement)**.** *Assume $r = r_T \geq 1$, $p \geq 5$, and some mild hypotheses. Then, for any rank r Euler system c, there exists an element*

$$\mathfrak{z}_{L_\infty} \in \det_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S}, \mathbb{T}))$$

such that $\Theta(\mathfrak{z}_{L_\infty}) = c_{L_\infty}$.

In particular, when $L = K$ and $d = 1$, we have

$$\mathrm{char}_\Lambda\left(\bigcap\nolimits_\Lambda^r H^1(\mathcal{O}_{K,S}, \mathbb{T})/\Lambda \cdot c_{K_\infty}\right) \subset \mathrm{char}_\Lambda(H^2(\mathcal{O}_{K,S}, \mathbb{T})).$$

For the proof, we essentially use the theory of higher rank Euler, Kolyvagin, and Stark systems established by Burns, Sakamoto, and the second author [BSS18], and also an idea of the recent work by the first author in [Kat22].

1.2 Derivatives of Euler systems

Let $r = r_T$ be the basic rank of T and c an Euler system of rank r for T. Fix a finite set S of places of K as above. Under suitable assumptions, one can consider a “Darmon-type derivative" of c for a $\mathbb{Z}_p$-extension K_∞/K:

$$\kappa_\infty \in \bigwedge\nolimits_{\mathcal{A}}^r H^1(\mathcal{O}_{K,S}, T) \otimes_{\mathcal{A}} I^e/I^{e+1},$$

where we set $e := \mathrm{rank}_{\mathcal{A}}(H^2(\mathcal{O}_{K,S}, T))$ and $I := \ker(\mathcal{A}[[\mathrm{Gal}(K_\infty/K)]] \twoheadrightarrow \mathcal{A})$ (see §4.3). We formulate a conjecture which relates κ_∞ with the leading term $L_S^*(M^*(1), 0)$ of the S-truncated (complex) L-function for the dual motive $M^*(1)$ at $s = 0$. To do this we introduce an “(extended) special element" for T over K

$$\widetilde{\eta}_K \in \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathcal{A}}^{r+e} H^1(\mathcal{O}_{K,S}, T),$$

which is by definition related to the leading term (see Definition 2.11). This element is a natural generalization of the “Birch and Swinnerton-Dyer element" introduced by Burns, Kurihara, and the second author in [BKS24, Def. 2.4]. We naturally construct a “Bockstein regulator map"

$$\mathrm{Boc}_\infty : \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathcal{A}}^{r+e} H^1(\mathcal{O}_{K,S}, T) \to \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathcal{A}}^{r} H^1(\mathcal{O}_{K,S}, T) \otimes_{\mathcal{A}} I^e/I^{e+1},$$

and formulate the conjecture as follows.

Conjecture 1.3 (Conjecture 4.7)**.** *We have*

$$\kappa_\infty = (-1)^{re}\mathrm{Boc}_\infty(\widetilde{\eta}_K) \text{ in } \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathcal{A}}^{r} H^1(\mathcal{O}_{K,S}, T) \otimes_{\mathcal{A}} I^e/I^{e+1}.$$

This conjecture generalizes the “generalized Perrin-Riou conjecture" in [BKS24, Conj. 4.9] and also the “Iwasawa theoretic Mazur-Rubin-Sano conjecture" in [BKS17, Conj. 4.2].

We give a strategy for proving the Tamagawa number conjecture for $M^*(1)$ by using the Iwasawa main conjecture and Conjecture 1.3 (see Theorem 4.11). This result is a generalization of [BKS24, Th. 7.6].

1.3 Heegner points

To give a new example of our general theory, we study Heegner points in detail. Let E be an elliptic curve over $\mathbb{Q}$ and K an imaginary quadratic field satisfying the Heegner hypothesis for E (i.e., every prime divisor of the conductor of E splits in K). We consider the motive $M = h^1(E/K)(1)$. In this case, T is the p-adic Tate module of E and the coefficient ring is $\mathcal{A} = \mathbb{Z}_p$.

The Euler system of Heegner points is usually considered to be a rank one Euler system. However, it is known that it does not satisfy the natural definition of Euler systems given by Rubin [Rub00]. In this sense, it might be unnatural to regard the Heegner point Euler system as a rank one Euler system.

In this paper, we make the following observation: *it is natural to interpret the system of Heegner points as a rank two Euler system*. In fact, *the basic rank r_T is two* in this setting, since we have

$$\bigoplus_{v\in S_\infty(K)} H^0(K_v, T^*(1)) = H^0(\mathbb{C}, T^*(1)) = T^*(1)$$

and this is a free $\mathbb{Z}_p$-module of rank two.

Our idea of interpreting Heegner points as a rank two Euler system is as follows. We assume that E has good ordinary reduction at p. Take a finite set S of places of K as usual. Let K_∞/K be the anticyclotomic $\mathbb{Z}_p$-extension and set $\Lambda := \mathbb{Z}_p[[\mathrm{Gal}(K_\infty/K)]]$ and $\mathbb{T} := T \otimes \Lambda$. In the Selmer group $\mathrm{Sel}(\mathbb{T}) = \mathrm{Sel}(K, \mathbb{T})$, one has a Λ-adic Heegner point $y_\infty \in \mathrm{Sel}(\mathbb{T})$ (see [Cas17, §3.1] for example). Under some mild conditions, we construct an isomorphism

$$Q(\Lambda) \otimes_\Lambda {\bigcap}_\Lambda^2 H^1(\mathcal{O}_{K,S}, \mathbb{T}) \simeq Q(\Lambda) \otimes_\Lambda (\mathrm{Sel}(\mathbb{T}) \otimes_\Lambda \mathrm{Sel}(\mathbb{T})^\iota),$$

where $Q(\Lambda)$ denotes the quotient field of Λ and $(-)^\iota$ the module on which Λ acts via the natural involution. We define a “Heegner element"

$$z_\infty^{\mathrm{Hg}} \in Q(\Lambda) \otimes_\Lambda {\bigcap}_\Lambda^2 H^1(\mathcal{O}_{K,S}, \mathbb{T})$$

to be the element corresponding to

$$y_\infty \otimes y_\infty \in \mathrm{Sel}(\mathbb{T}) \otimes_\Lambda \mathrm{Sel}(\mathbb{T})^\iota$$

under the isomorphism above (see Definition 5.13).

We show that the “Heegner point main conjecture" of Perrin-Riou is equivalent to our formulation of the Iwasawa main conjecture for z_∞^{Hg}.

Theorem 1.4 (Theorem 5.17)**.** *The Heegner point main conjecture holds if and only if we have $z_\infty^{\mathrm{Hg}} \in {\bigcap}_\Lambda^2 H^1(\mathcal{O}_{K,S}, \mathbb{T})$ and an equality*

$$\mathrm{char}_\Lambda \left({\bigcap}_\Lambda^2 H^1(\mathcal{O}_{K,S}, \mathbb{T}) / \Lambda \cdot z_\infty^{\mathrm{Hg}} \right) = \mathrm{char}_\Lambda (H^2(\mathcal{O}_{K,S}, \mathbb{T})).$$

As an application, we give a formal construction of a rank two Euler system whose “K_∞-component" is z_∞^{Hg}.

Theorem 1.5 (Theorem 5.18)**.** *Assume the Heegner point main conjecture. Then there exists a rank two Euler system c such that* $c_{K_\infty} = z_\infty^{\mathrm{Hg}}$.

Remark 1.6. Burungale-Castella-Kim has recently proved the Heegner point main conjecture under mild hypotheses (see [BCK21, Th. A]). So Theorem 1.5 gives an unconditional construction of a rank two Euler system which is related to Heegner points. However, it should be noted that our construction is non-canonical: roughly speaking, the constructed Euler system is just a "lift" of z_∞^{Hg}. (The Heegner point main conjecture is assumed in order to ensure the existence of a lift.)

Remark 1.7. There is a non-Iwasawa theoretic version of Theorem 1.5: see Theorem 5.5. A Heegner element (over K) is more explicitly defined in this case without assuming that E has good ordinary reduction at p (see Definition 5.2).

Remark 1.8. Lei-Loeffler-Zerbes conjectured in [LLZ14, Conj. 8.2.6] that there should be a rank two Euler system which is related to the Euler system of Beilinson-Flach elements. Although a direct link is not clear to the present authors, our work may shed new light on their conjecture.

Lastly, we give an explicit interpretation of Conjecture 1.3 for the Heegner element z_∞^{Hg}. We assume the following:

(i) $E(K)[p] = 0$;

(ii) $\mathrm{rank}(E(\mathbb{Q})) \geq 1$ and $\mathrm{rank}(E^K(\mathbb{Q})) \geq 1$, where E^K denotes the quadratic twist of E by K;

(iii) $\#\text{Ш}(E/K)[p^\infty] < \infty$.

(See Hypothesis 5.22.) Note that the Heegner hypothesis and the validity of the parity conjecture imply that $\mathrm{rank}(E(K))$ is odd. So the condition (ii) implies $\mathrm{rank}(E(K)) \geq 3$. (One sees that Conjecture 1.3 is not interesting when $\mathrm{rank}(E(K)) = 1$.)

We set $e := \mathrm{rank}_{\mathbb{Z}_p}(H^2(\mathcal{O}_{K,S}, T))$ and $I := \ker(\Lambda \twoheadrightarrow \mathbb{Z}_p)$. Let

$$\kappa_\infty^{\mathrm{Hg}} \in \bigwedge\nolimits_{\mathbb{Z}_p}^{2} H^1(\mathcal{O}_{K,S}, T) \otimes_{\mathbb{Z}_p} I^e/I^{e+1}$$

be the Darmon-type derivative of z_∞^{Hg}. We define a canonical "anticyclotomic Bockstein regulator"

$$R_{K_\infty}^{\mathrm{Boc}} \in \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathbb{Z}_p}^{2} H^1(\mathcal{O}_{K,S}, T) \otimes_{\mathbb{Z}_p} I^e/I^{e+1}$$

as an analogue of the "(cyclotomic) Bockstein regulator" introduced by Burns, Kurihara, and the second author in [BKS24, Def. 4.11].

We prove that Conjecture 1.3 in this case is equivalent to the following explicit formula.

Conjecture 1.9 (see Proposition 5.26)**.** *We have*

$$\kappa_\infty^{\mathrm{Hg}} = \frac{L_S^*(E/K,1)\sqrt{|D_K|}}{\Omega_{E/K} \cdot R_{E/K}} \cdot R_{K_\infty}^{\mathrm{Boc}},$$

where $L_S^*(E/K,1)$ *denotes the leading term of the S-truncated L-function of* E/K *at* $s = 1$*,* D_K *the discriminant of K,* $\Omega_{E/K}$ *the Néron period, and* $R_{E/K}$ *the Néron-Tate regulator.*

According to the conjectural Birch-Swinnerton-Dyer formula, the analytic constant

$$\frac{L_S^*(E/K,1)\sqrt{|D_K|}}{\Omega_{E/K}\cdot R_{E/K}}$$

should be equal to the algebraic constant

$$\mathrm{Eul}_S\cdot\#\text{Ш}(E/K)[p^\infty]\cdot\mathrm{Tam}(E/K)$$

up to $\mathbb{Z}_p^\times$, where Eul_S denotes the product of Euler factors at primes in S (so that $\mathrm{Eul}_S\cdot L^*(E/K,1)=L_S^*(E/K,1)$) and $\mathrm{Tam}(E/K)$ the product of Tamagawa factors of E/K. We prove that an algebraic variant of Conjecture 1.9 follows from the Heegner point main conjecture up to $\mathbb{Z}_p^\times$.

Theorem 1.10 (Theorem 5.27)**.** *Assume the Heegner point main conjecture. Then there exists $u\in\mathbb{Z}_p^\times$ such that*

$$\kappa_\infty^{\mathrm{Hg}}=u\cdot\mathrm{Eul}_S\cdot\#\text{Ш}(E/K)[p^\infty]\cdot\mathrm{Tam}(E/K)\cdot R_{K_\infty}^{\mathrm{Boc}}.$$

In a forthcoming work, we show that Conjecture 1.9 (or rather its algebraic variant) implies the conjecture of Bertolini and Darmon [BeDa96, Conj. 4.5(1)] (see also [AgCa21, Conj. 3.6]). This gives further evidence for Conjecture 1.9.

Finally, we give a strategy for proving the Birch-Swinnerton-Dyer formula for E/K.

Theorem 1.11 (Theorem 5.29)**.** *If we assume*

- *the Heegner point main conjecture,*
- *Conjecture 1.9, and*
- $R_{K_\infty}^{\mathrm{Boc}}\neq 0$,

then the p-part of the Birch-Swinnerton-Dyer formula for E/K holds, i.e., there exists $u\in\mathbb{Z}_p^\times$ such that

$$L^*(E/K,1)=u\cdot\#\text{Ш}(E/K)[p^\infty]\cdot\mathrm{Tam}(E/K)\cdot\frac{1}{\sqrt{|D_K|}}\Omega_{E/K}\cdot R_{E/K}.$$

We remark that Theorems 1.10 and 1.11 are analogues of the results of Burns, Kurihara, and the second author [BKS24, Th. 7.3 and 7.6] respectively.

1.4 Notation

For a commutative ring R and an R-module N, we set $N^*:=\mathrm{Hom}_R(N,R)$. The R-torsion submodule of N is denoted by N_{tors}. The R-torsion-free quotient N/N_{tors} is denoted by N_{tf}. For a non-negative integer a, the a-th exterior power bidual of N over R is defined by

$$\bigcap\nolimits_R^a N:=\left(\bigwedge\nolimits_R^a(N^*)\right)^*.$$

For basic properties, see [BuSa21, Appendix A].

For an abelian group ($\mathbb{Z}$-module) A and a prime number p, we set

$$A[p] := \{a \in A \mid p \cdot a = 0\} \text{ and } A[p^\infty] := \{a \in A \mid p^n \cdot a = 0 \text{ for some } n\}.$$

Let K be a number field, which is regarded as a finite extension of $\mathbb{Q}$ inside a fixed algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$. We denote the absolute Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ by G_K. For each place v of K, we fix a place w of $\overline{\mathbb{Q}}$ lying above v. The decomposition group of w in G_K is identified with $\mathrm{Gal}(\overline{\mathbb{Q}}_w/K_v)$. In particular, we regard $\mathrm{Gal}(\overline{\mathbb{Q}}_w/K_v) \subset G_K$. Let K_v^{ur} be the maximal unramified extension of K_v inside $\overline{\mathbb{Q}}_w$ and $\mathrm{Fr}_v \in \mathrm{Gal}(K_v^{\mathrm{ur}}/K_v)$ the Frobenius element of v. We fix a lift of Fr_v in $\mathrm{Gal}(\overline{\mathbb{Q}}_w/K_v)$ and denote it also by the same symbol.

The set of all infinite (resp. p-adic) places of K is denoted by $S_\infty(K)$ (resp. $S_p(K)$).

For a Galois extension F/K, we often denote $\mathrm{Gal}(F/K)$ by $\mathcal{G}_F$. The set of finite places of K which ramify in F is denoted by $S_{\mathrm{ram}}(F/K)$. For a finite set S of places of K, we set

$$S_F := \{w : \text{a place of } F \mid \text{the place of } K \text{ lying under } w \text{ belongs to } S\}$$

and

$$S(F) := S \cup S_{\mathrm{ram}}(F/K).$$

We use some standard notations concerning Galois (étale) cohomology: $\mathbf{R}\Gamma(\mathcal{O}_{K,S}, -)$, $\mathbf{R}\Gamma_f(K_v, -)$, $\mathbf{R}\Gamma_{/f}(K_v, -)$, etc. For the definitions, see [BuSa21, §1.4] for example.

2 Euler systems for motives

In this section, we give a review of a general conjecture on Euler systems given in [BSS19, §4] (see Conjecture 2.6). This conjecture predicts what kind of Euler system should exist in a general setting of motives. In §2.5, we give a generalization of “Birch-Swinnerton-Dyer elements" introduced in [BKS24, §2.2], which will be used in §4.

2.1 The definition of Euler systems

Let K be a number field and $p > 2$ an odd prime number. Let M be a (pure) motive defined over K with coefficients in a finite dimensional semisimple commutative $\mathbb{Q}$-algebra R, which is necessarily a finite product of number fields. Let A be a finite extension of $\mathbb{Q}_p$ which arises as a component of $\mathbb{Q}_p \otimes_{\mathbb{Q}} R$ and $\mathcal{A}$ the ring of integers of A. Let $V_p(M)$ be the p-adic étale realization of M and set $V := A \otimes_{\mathbb{Q}_p \otimes_{\mathbb{Q}} R} V_p(M)$, which is a finite dimensional A-vector space endowed with a continuous action of G_K. Fix a G_K-stable lattice $T \subset V$, which is a free $\mathcal{A}$-module of finite rank. Let $T^*(1) := \mathrm{Hom}_{\mathcal{A}}(T, \mathcal{A}(1))$ be the Kummer dual of T. We set

$$Y_K(T^*(1)) := \bigoplus_{v \in S_\infty(K)} H^0(K_v, T^*(1)).$$

Note that, for each $v \in S_\infty(K)$, the $\mathcal{A}$-module $H^0(K_v, T^*(1))$ is a direct summand of $T^*(1)$, so in particular it is free of rank less than or equal to $\mathrm{rank}_{\mathcal{A}}(T^*(1)) = \mathrm{rank}_{\mathcal{A}}(T)$. Therefore, $Y_K(T^*(1))$ is also a free $\mathcal{A}$-module with an upper bound of the rank.

Definition 2.1. We define the *basic rank* of T by

$$r = r_T := \mathrm{rank}_{\mathcal{A}}(Y_K(T^*(1))).$$

We fix a finite set S of places of K such that

$$S_\infty(K) \cup S_p(K) \cup S_{\mathrm{ram}}(T) \subset S,$$

where $S_{\mathrm{ram}}(T)$ denotes the set of finite places of K at which T ramifies. For any $v \notin S$, we set

$$P_v(x) = P_v(x;T) := \det(1 - \mathrm{Fr}_v^{-1}x \mid T^*(1)) \in \mathcal{A}[x].$$

Let $\mathcal{K}/K$ be an abelian extension. Let $\Omega(\mathcal{K})$ be the set of finite subextensions F/K of $\mathcal{K}/K$. For each $F \in \Omega(\mathcal{K})$ we set

$$\mathcal{G}_F := \mathrm{Gal}(F/K)$$

and

$$S(F) := S \cup S_{\mathrm{ram}}(F/K).$$

We impose the following hypothesis on the extension $\mathcal{K}/K$: for each $F \in \Omega(\mathcal{K})$, the $\mathcal{A}[\mathcal{G}_F]$-module

$$Y_F(T^*(1)) := \bigoplus_{w \in S_\infty(F)} H^0(F_w, T^*(1))$$

is free of rank r_T. One sees that this condition is equivalent to the following:

For any $v \in S_\infty(K)$ which ramifies in $\mathcal{K}$, we have $\mathrm{rank}_{\mathcal{A}}(H^0(K_v, T^*(1))) = \dfrac{1}{2}\mathrm{rank}_{\mathcal{A}}(T)$.

For example, this hypothesis is satisfied as long as every $v \in S_\infty(K)$ splits completely in $\mathcal{K}$.

Let Σ be a finite set (possibly empty) of places of K which is disjoint from $S(F)$ for any $F \in \Omega(\mathcal{K})$. (This means that Σ is disjoint from S and every $v \in \Sigma$ is unramified in $\mathcal{K}$.) Following [BuSa21, §2.3], for any $F \in \Omega(\mathcal{K})$, we define the Σ-modified cohomology complex $\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{F,S(F)}, T)$ by the exact triangle

$$\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{F,S(F)}, T) \to \mathbf{R}\Gamma(\mathcal{O}_{F,S(F)}, T) \to \bigoplus_{w \in \Sigma_F} \mathbf{R}\Gamma_f(F_w, T) \to .$$

For any $F, F' \in \Omega(\mathcal{K})$ with $F \subset F'$, we write

$$\mathrm{Cor}^r_{F'/F} : {\bigcap}^r_{\mathcal{A}[\mathcal{G}_{F'}]} H^1_\Sigma(\mathcal{O}_{F',S(F')}, T) \to {\bigcap}^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S(F')}, T)$$

for the map induced by the corestriction map $\mathrm{Cor}_{F'/F} : H^1_\Sigma(\mathcal{O}_{F',S(F')}, T) \to H^1_\Sigma(\mathcal{O}_{F,S(F')}, T)$.

Definition 2.2. An *Euler system* of rank r for $(T, \mathcal{K})$ (with an implicit choice of S and Σ) is an element

$$c = (c_F)_F \in \prod_{F \in \Omega(\mathcal{K})} {\bigcap}^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S(F)}, T)$$

satisfying the following: for any $F, F' \in \Omega(\mathcal{K})$ with $F \subset F'$, we have

$$\mathrm{Cor}^r_{F'/F}(c_{F'}) = \left(\prod_{v \in S(F') \setminus S(F)} P_v(\mathrm{Fr}_v^{-1}) \right) c_F.$$

The set ($\mathcal{A}[[\mathrm{Gal}(\mathcal{K}/K)]]$-module) of Euler systems of rank r for $(T, \mathcal{K})$ is denoted by $\mathrm{ES}_r(T, \mathcal{K})$.

2.2 Conjectural Euler systems

In this subsection, we formulate an explicit conjecture concerning the existence of an Euler system for T which is related with L-functions of the motive $M^*(1)$, under the following hypothesis.

Hypothesis 2.3. For any $F \in \Omega(\mathcal{K})$, we have

(i) $H^0(F,T)=0$,

(ii) either Σ is non-empty or $H^1(\mathcal{O}_{F,S(F)},T)$ is $\mathcal{A}$-free, and

(iii) for any $w \in \Sigma_F$, we have $H^0(F_w,T)=0$.

Remark 2.4. If we assume Hypothesis 2.3(i), (iii) and that Σ is non-empty, then one easily sees that $H^1_\Sigma(\mathcal{O}_{F,S(F)},T)$ is $\mathcal{A}$-free. So Hypothesis 2.3(ii) can be replaced by

(ii′) $H^1_\Sigma(\mathcal{O}_{F,S(F)},T)$ is $\mathcal{A}$-free.

Remark 2.5. Hypothesis 2.3(iii) implies that $H^1_f(F_w,T)$ is finite for any $w \in \Sigma_F$. So in this case we have

$$\mathbb{Q}_p \otimes_{\mathbb{Z}_p} H^i_\Sigma(\mathcal{O}_{F,S(F)},T) = H^i(\mathcal{O}_{F,S(F)},V).$$

This identification will frequently be used.

We set some notations. Let $F \in \Omega(\mathcal{K})$. We write $\widehat{\mathcal{G}}_F$ for the set of $\overline{\mathbb{Q}}$-valued characters of $\mathcal{G}_F$. By the fixed embeddings $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$ and $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}_p$, each $\chi \in \widehat{\mathcal{G}}_F$ is regarded as both $\mathbb{C}$-valued and $\mathbb{C}_p$-valued. For $\chi \in \widehat{\mathcal{G}}_F$, we define the usual idempotent by $e_\chi := (\#\mathcal{G}_F)^{-1}\sum_{\sigma\in\mathcal{G}_F}\chi(\sigma)\sigma^{-1}$. We set

$$\Upsilon(T,F) := \{\chi \in \widehat{\mathcal{G}}_F \mid e_\chi(\mathbb{C}_p \otimes_{\mathbb{Z}_p} H^2(\mathcal{O}_{F,S(F)},T)) = 0\}$$

and

$$e_{T,F} := \sum_{\chi\in\Upsilon(T,F)} e_\chi \in A[\mathcal{G}_F].$$

Roughly speaking, $e_{T,F}$ is the “maximal" idempotent that annihilates $H^2(\mathcal{O}_{F,S(F)},T)$.

We recall the definition of motivic L-functions. To do this, we assume that the Euler factors have rational coefficients, i.e., $Q_v(x) := \det(1-\mathrm{Fr}_v^{-1}x \mid V_p(M^*(1)))$ belongs to $R[x]$ for any $v \notin S$.

For $\chi \in \widehat{\mathcal{G}}_F$, the $S(F)$-truncated χ-twisted L-function of $M^*(1)$ is defined by

$$L_{S(F)}(M^*(1),\chi,s) := \prod_{v\notin S(F)} Q_v(\chi(\mathrm{Fr}_v)\mathrm{N}v^{-s})^{-1}.$$

This is a complex function which takes values in $\mathbb{C}\otimes_{\mathbb{Q}} R$ and converges if $\mathrm{Re}(s)$ is large enough. We assume that it is analytically continued to the whole complex plane. We define a Σ-modified version by

$$L_{S(F),\Sigma}(M^*(1),\chi,s) := \left(\prod_{v\in\Sigma} Q_v(\chi(\mathrm{Fr}_v)\mathrm{N}v^{1-s})\right) L_{S(F)}(M^*(1),\chi,s).$$

We then define the $\mathcal{G}_F$-equivariant $(S(F),\Sigma)$-modified L-function of $M^*(1)$ by

$$\theta_{F/K,S(F),\Sigma}(M^*(1),s) := \sum_{\chi\in\widehat{\mathcal{G}}_F} L_{S(F),\Sigma}(M^*(1),\chi^{-1},s)e_\chi,$$

which takes values in $\mathbb{C}\otimes_{\mathbb{Q}} R[\mathcal{G}_F]$.

We write $L^*_{S(F),\Sigma}(M^*(1),\chi,0) \in (\mathbb{C}\otimes_{\mathbb{Q}} R)^\times$ for the leading term of $L_{S(F),\Sigma}(M^*(1),\chi,s)$ at $s=0$, i.e., the leading coefficient in the Laurent expansion at $s=0$. We define the leading term of $\theta_{F/K,S(F),\Sigma}(M^*(1),s)$ by

$$\theta^*_{F/K,S(F),\Sigma}(M^*(1),0) := \sum_{\chi\in\widehat{\mathcal{G}}_F} L^*_{S(F),\Sigma}(M^*(1),\chi^{-1},0)e_\chi \in (\mathbb{C}\otimes_{\mathbb{Q}} R[\mathcal{G}_F])^\times.$$

We fix an embedding $\mathbb{C}\hookrightarrow\mathbb{C}_p$ and regard $\theta^*_{F/K,S(F),\Sigma}(M^*(1),0)$ as an element of $(\mathbb{C}_p\otimes_{\mathbb{Q}} R[\mathcal{G}_F])^\times$. By the natural projection $\mathbb{Q}_p\otimes_{\mathbb{Q}} R\twoheadrightarrow A$, we regard $\theta^*_{F/K,S(F),\Sigma}(M^*(1),0)$ as an element of $(\mathbb{C}_p\otimes_{\mathbb{Q}_p} A[\mathcal{G}_F])^\times = (\mathbb{C}_p\otimes_{\mathbb{Z}_p}\mathcal{A}[\mathcal{G}_F])^\times$.

In §2.4 below, we will define a canonical "period-regulator isomorphism"

$$\lambda_{T,F} : e_{T,F}\left(\mathbb{C}_p\otimes_{\mathbb{Z}_p}{\bigwedge}^r_{\mathcal{A}[\mathcal{G}_F]}H^1(\mathcal{O}_{F,S(F)},T)\right)\xrightarrow{\sim} e_{T,F}\left(\mathbb{C}_p\otimes_{\mathbb{Z}_p}{\bigwedge}^r_{\mathcal{A}[\mathcal{G}_F]}Y_F(T^*(1))^*\right),$$

where $Y_F(T^*(1))^* := \mathrm{Hom}_{\mathcal{A}}(Y_F(T^*(1)),\mathcal{A})$.

Conjecture 2.6. *Assume Hypothesis 2.3, and fix an $\mathcal{A}[[\mathrm{Gal}(\mathcal{K}/K)]]$-basis*

$$b=(b_F)_F\in\varprojlim_{F\in\Omega(\mathcal{K})}{\bigwedge}^r_{\mathcal{A}[\mathcal{G}_F]}Y_F(T^*(1))^*(\simeq\varprojlim_{F\in\Omega(\mathcal{K})}\mathcal{A}[\mathcal{G}_F]=\mathcal{A}[[\mathrm{Gal}(\mathcal{K}/K)]]).$$

Then there exists a unique Euler system

$$c=c(b)\in\mathrm{ES}_r(T,\mathcal{K})$$

satisfying the following properties.

(i) For every $F\in\Omega(\mathcal{K})$, we have

$$(1-e_{T,F})c_F=0,$$

i.e., $c_F\in{\bigcap}^r_{\mathcal{A}[\mathcal{G}_F]}H^1_\Sigma(\mathcal{O}_{F,S(F)},T)\subset{\bigwedge}^r_{\mathcal{A}[\mathcal{G}_F]}H^1(\mathcal{O}_{F,S(F)},V)$ belongs to $e_{T,F}{\bigwedge}^r_{\mathcal{A}[\mathcal{G}_F]}H^1(\mathcal{O}_{F,S(F)},V)$.

(ii) For every $F\in\Omega(\mathcal{K})$, we have

$$\lambda_{T,F}(c_F)=e_{T,F}\cdot\theta^*_{F/K,S(F),\Sigma}(M^*(1),0)\cdot b_F \text{ in } e_{T,F}\left(\mathbb{C}_p\otimes_{\mathbb{Z}_p}{\bigwedge}^r_{\mathcal{A}[\mathcal{G}_F]}Y_F(T^*(1))^*\right).$$

It is convenient to give the following definition.

Definition 2.7. We fix b as in Conjecture 2.6. For each $F \in \Omega(\mathcal{K})$, we define the *special element* for T by

$$\begin{aligned}\eta_F = \eta_{F/K,S(F),\Sigma}(T) &:= \lambda_{T,F}^{-1}\left(e_{T,F}\cdot\theta^*_{F/K,S(F),\Sigma}(M^*(1),0)\cdot b_F\right)\\ &\in e_{T,F}\left(\mathbb{C}_p\otimes_{\mathbb{Z}_p}{\bigwedge}^r_{\mathcal{A}[\mathcal{G}_F]}H^1(\mathcal{O}_{F,S(F)},T)\right).\end{aligned}$$

(This is called the "Bloch-Kato element" in [BSS19, Def. 4.10].)

Remark 2.8. One can show that the collection $(\eta_F)_{F\in\Omega(\mathcal{K})}$ satisfies the norm relation, i.e., we have

$$\mathrm{Cor}^r_{F'/F}(\eta_{F'}) = \left(\prod_{v\in S(F')\setminus S(F)} P_v(\mathrm{Fr}_v^{-1})\right)\eta_F$$

for any $F, F' \in \Omega(\mathcal{K})$ with $F \subset F'$. Also, η_F satisfies the properties (i) and (ii) in Conjecture 2.6 by definition. Thus Conjecture 2.6 is equivalent to the following assertion: for every $F \in \Omega(\mathcal{K})$, we have

$$\eta_F \in {\bigcap}^r_{\mathcal{A}[\mathcal{G}_F]}H^1_\Sigma(\mathcal{O}_{F,S(F)},T). \tag{2.2.1}$$

The conjecture of this form is given in [BSS19, Conj. 4.15].

Remark 2.9. By [BuSa21, Rem. 2.11 and Th. 2.18], one sees that the equivariant Tamagawa number conjecture for $(M^*(1)\otimes_K F, \mathcal{A}[\mathcal{G}_F])$ (see [BuFl01, Conj. 4]) implies (2.2.1) for $F \in \Omega(\mathcal{K})$. This gives theoretical evidence for Conjecture 2.6.

Remark 2.10. In this paper, we often assume the existence of a canonical Euler system and study its properties. So it may be reasonable to assume Conjecture 2.6 throughout. However, it turns out that *assuming Conjecture 2.6 is too strong*. For example, in the case of elliptic curves over $\mathbb{Q}$, we have a canonical Euler system called Kato's Euler system, but *it is still not known whether it satisfies the properties in Conjecture 2.6*. In fact, Conjecture 2.6 for Kato's Euler system is equivalent to a natural equivariant refinement of Perrin-Riou's conjecture, which has not yet been fully proved (see §2.3.2 below for the details). For this reason, we will only assume the existence of a canonical Euler system and propose conjectures for it.

2.3 Examples

Let us consider two special cases for which the conjectural Euler systems are more familiar. The first case is the (conjectural) system of Rubin-Stark elements over number fields, which specializes to the cyclotomic unit Euler system when $K = \mathbb{Q}$ and to the elliptic unit Euler system when K is an imaginary quadratic field. The second is Kato's Euler system for elliptic curves over $\mathbb{Q}$.

2.3.1 The Rubin-Stark Euler system

Consider the "$\mathbb{G}_m$ case", i.e., $M = h^0(K)(1)$, $R = \mathbb{Q}$, $A = \mathbb{Q}_p$, $\mathcal{A} = \mathbb{Z}_p$, and $T = \mathbb{Z}_p(1)$. Take $\mathcal{K}$ so that every $v \in S_\infty(K)$ splits completely in $\mathcal{K}$ and also choose Σ so that Hypothesis 2.3 is satisfied. In this case, we have

$$r = r_{\mathbb{Z}_p(1)} = \#S_\infty(K).$$

Also, for $F \in \Omega(\mathcal{K})$ and $\chi \in \widehat{\mathcal{G}}_F$ we have

$$\begin{aligned} L_{S(F),\Sigma}(M^*(1),\chi,s) &= L_{S(F),\Sigma}(\chi,s) \\ &:= \prod_{v\in\Sigma}(1-\chi(\mathrm{Fr}_v)\mathrm{N}v^{1-s}) \prod_{v\notin S(F)}(1-\chi(\mathrm{Fr}_v)\mathrm{N}v^{-s})^{-1}. \end{aligned}$$

This is the usual $(S(F),\Sigma)$-modified Artin L-function for χ (see [BKS16, §3.1] for example). So we have

$$\theta_{F/K,S(F),\Sigma}(M^*(1),s) = \theta_{F/K,S(F),\Sigma}(s) := \sum_{\chi\in\widehat{\mathcal{G}}_F} L_{S(F),\Sigma}(\chi^{-1},s)e_\chi.$$

For any finite set U of places of K, we set

$$Y_{F,U} := \bigoplus_{w\in U_F} \mathbb{Z}\cdot w \text{ and } X_{F,U} := \left\{ \sum_w a_w \cdot w \in Y_{F,U} \;\middle|\; \sum_w a_w = 0 \right\}.$$

Then $Y_F(T^*(1))^* = Y_F(\mathbb{Z}_p)^*$ is identified with $\mathbb{Z}_p \otimes_{\mathbb{Z}} Y_{F,S_\infty(K)}$. Let

$$\delta_F : \mathbb{R}\otimes_{\mathbb{Z}} \mathcal{O}_{F,S(F)}^\times \xrightarrow{\sim} \mathbb{R}\otimes_{\mathbb{Z}} X_{F,S(F)};\ u \mapsto -\sum_{w\in S(F)_F} \log|u|_w \cdot w$$

be the Dirichlet regulator. The period-regulator isomorphism in this case is defined by

$$\begin{aligned} \lambda_{\mathbb{Z}_p(1),F} : e_{\mathbb{Z}_p(1),F}\left(\mathbb{C}_p\otimes_{\mathbb{Z}_p}{\bigwedge}^r_{\mathbb{Z}_p[\mathcal{G}_F]} H^1(\mathcal{O}_{F,S(F)},\mathbb{Z}_p(1))\right) &\simeq e_{\mathbb{Z}_p(1),F}\left(\mathbb{C}_p\otimes_{\mathbb{Z}}{\bigwedge}^r_{\mathbb{Z}[\mathcal{G}_F]}\mathcal{O}_{F,S(F)}^\times\right) \\ &\overset{\delta_F}{\simeq} e_{\mathbb{Z}_p(1),F}\left(\mathbb{C}_p\otimes_{\mathbb{Z}}{\bigwedge}^r_{\mathbb{Z}[\mathcal{G}_F]}X_{F,S(F)}\right) \\ &\simeq e_{\mathbb{Z}_p(1),F}\left(\mathbb{C}_p\otimes_{\mathbb{Z}}{\bigwedge}^r_{\mathbb{Z}[\mathcal{G}_F]}Y_{F,S_\infty(K)}\right), \end{aligned}$$

where the first isomorphism is induced by the Kummer isomorphism $\mathbb{Z}_p\otimes_{\mathbb{Z}}\mathcal{O}_{F,S(F)}^\times \simeq H^1(\mathcal{O}_{F,S(F)},\mathbb{Z}_p(1))$, and the last isomorphism follows by the definition of $e_{\mathbb{Z}_p(1),F}$ and a canonical exact sequence

$$0 \to H^2(\mathcal{O}_{F,S(F)},\mathbb{Q}_p(1)) \to \mathbb{Q}_p\otimes_{\mathbb{Z}} X_{F,S(F)} \to \mathbb{Q}_p\otimes_{\mathbb{Z}} Y_{F,S_\infty(K)} \to 0$$

(see [Nek06, (9.2.1.2)] for example). One can choose a $\mathbb{Z}[\mathcal{G}_F]$-basis b_F of ${\bigwedge}^r_{\mathbb{Z}[\mathcal{G}_F]}Y_{F,S_\infty(K)}$ by fixing a labeling $S_\infty(K) = \{v_1,\dots,v_r\}$ and a place w_i of F lying above each v_i. Namely, one sets $b_F := w_1\wedge\cdots\wedge w_r$. The Rubin-Stark element (for $F/K, S(F), \Sigma, S_\infty(K)$) is defined by

$$\eta_F^{\mathrm{RS}} := \eta_{F/K,S(F),\Sigma}^{S_\infty(K)} := \lambda_{\mathbb{Z}_p(1),F}^{-1}(e_{\mathbb{Z}_p(1),F}\cdot\theta^*_{F/K,S(F),\Sigma}(0)\cdot b_F).$$

(See [BKS16, §5.1] for example.) This coincides with the special element $\eta_{F/K,S(F),\Sigma}(\mathbb{Z}_p(1))$ in Definition 2.7. The conjecture (2.2.1) is equivalent to the (p-part of the) Rubin-Stark conjecture (see [Rub96, Conj. B′] or [BKS16, Conj. 5.1]). Thus, if we assume the Rubin-Stark conjecture for all $F \in \Omega(\mathcal{K})$, then the conjectural Euler system in Conjecture 2.6 coincides with the Rubin-Stark Euler system

$$\eta^{\mathrm{RS}} := (\eta_F^{\mathrm{RS}})_F \in \mathrm{ES}_r(\mathbb{Z}_p(1),\mathcal{K}).$$

In particular, Conjecture 2.6 is true when $K = \mathbb{Q}$. (In this case, the Rubin-Stark Euler system is the cyclotomic unit Euler system.)

2.3.2 Kato's Euler system

Let E be an elliptic curve over $\mathbb{Q}$ and consider the case when $K=\mathbb{Q}$, $M=h^1(E)(1)$, $R=\mathbb{Q}$, $A=\mathbb{Q}_p$, $\mathcal{A}=\mathbb{Z}_p$, and T is a lattice of $V_p(E)$. In this case, we have

$$r=r_T=1.$$

We take $\mathcal{K}/\mathbb{Q}$ to be an abelian p-extension, Σ to be empty and assume Hypothesis 2.3. (If $T=T_p(E)$, then Hypothesis 2.3 is equivalent to $E(\mathbb{Q})[p]=0$.) For each $F\in\Omega(\mathcal{K})$, Kato [Kat04] constructed a "zeta element"

$$z_F^{\mathrm{Kato}}\in H^1(\mathcal{O}_{F,S(F)},V_p(E)).$$

(See [BSS19, Def. 6.8] for normalization. It depends on the choice of a $\mathbb{Z}_p[\mathcal{G}_F]$-basis $b_F\in Y_F(T^*(1))^*$.) If we assume the "integrality", i.e., $z_F^{\mathrm{Kato}}\in H^1(\mathcal{O}_{F,S(F)},T)$ for every $F\in\Omega(\mathcal{K})$, then we have

$$z^{\mathrm{Kato}}:=(z_F^{\mathrm{Kato}})_F\in \mathrm{ES}_1(T,\mathcal{K}).$$

(See [BSS19, Lem. 6.7].) This Euler system is called Kato's Euler system.

We shall describe the period-regulator isomorphism in this case. We prepare some notations. In the following, we assume $\#\text{Ш}(E/F)[p^\infty]<\infty$ for every $F\in\Omega(\mathcal{K})$. We abbreviate $\Upsilon(T,F)$ to $\Upsilon(F)$. For a non-negative integer i, we define

$$\Upsilon(F)_i^{\mathrm{an}}:=\{\chi\in\widehat{\mathcal{G}}_F\mid \mathrm{ord}_{s=1}L(E,\chi,s)=i\}$$

and

$$\Upsilon(F)_i^{\mathrm{alg}}:=\{\chi\in\widehat{\mathcal{G}}_F\mid \dim_{\mathbb{C}}(e_\chi(\mathbb{C}\otimes_{\mathbb{Z}}E(F)))=i\}.$$

Then one sees by [Kat04, Th. 14.2(2)] that

$$\Upsilon(F)_0^{\mathrm{an}}\subset\Upsilon(F)_0^{\mathrm{alg}}\subset\Upsilon(F)_0^{\mathrm{alg}}\cup\Upsilon(F)_1^{\mathrm{alg}}=\Upsilon(F).$$

(See [BSS19, Lem. 6.1(iii)] for the last equality.) We define the associated idempotent by

$$e_{F,i}^{\star}:=\sum_{\chi\in\Upsilon(F)_i^{\star}}e_\chi\in\mathbb{Q}[\mathcal{G}_F],$$

where $\star\in\{\mathrm{an},\mathrm{alg}\}$. Note that $e_{T,F}=e_{F,0}^{\mathrm{alg}}+e_{F,1}^{\mathrm{alg}}$.

We first describe the $e_{F,0}^{\mathrm{alg}}$-component of the period-regulator isomorphism. It is defined by the composition

$$\begin{aligned}\lambda_{T,F}:e_{F,0}^{\mathrm{alg}}\left(\mathbb{C}_p\otimes_{\mathbb{Z}_p}H^1(\mathcal{O}_{F,S(F)},T)\right)&\overset{\exp^*}{\simeq}e_{F,0}^{\mathrm{alg}}\left(\mathbb{C}_p\otimes_{\mathbb{Q}}\Gamma(E,\Omega^1_{E/F})\right)\\&\overset{\alpha}{\simeq}e_{F,0}^{\mathrm{alg}}\left(\mathbb{C}_p\otimes_{\mathbb{Q}}\left(\bigoplus_{\iota:F\hookrightarrow\mathbb{C}}H_1(E^\iota(\mathbb{C}),\mathbb{Q})\right)^+\right)^*\\&\overset{\beta}{\simeq}e_{F,0}^{\mathrm{alg}}\left(\mathbb{C}_p\otimes_{\mathbb{Z}_p}Y_F(T^*(1))^*\right),\end{aligned}$$

where the first isomorphism is induced by (the localization map and) the dual exponential map

$$\exp^* : \bigoplus_{w \in S_p(F)} H^1_{/f}(F_w, V) \simeq \bigoplus_{w \in S_p(F)} \mathbb{Q}_p \otimes_{\mathbb{Z}_p} E_1(F_w)^* \xrightarrow{\sim} \mathbb{Q}_p \otimes_{\mathbb{Q}} \Gamma(E, \Omega^1_{E/F}),$$

the second by the period map $\omega \mapsto (\gamma \mapsto \int_\gamma \omega)$, and the last by the comparison isomorphism

$$\left(\bigoplus_{\iota: F \hookrightarrow \mathbb{C}} H_1(E^\iota(\mathbb{C}), \mathbb{Q}_p) \right)^+ \simeq \mathbb{Q}_p \otimes_{\mathbb{Z}_p} Y_F(T^*(1)).$$

Next, we describe the $e^{\mathrm{alg}}_{F,1}$-component. It is defined by the composition

$$\begin{aligned} \lambda_{T,F} : e^{\mathrm{alg}}_{F,1} \left(\mathbb{C}_p \otimes_{\mathbb{Z}_p} H^1(\mathcal{O}_{F,S(F)}, T) \right) &\simeq e^{\mathrm{alg}}_{F,1} \left(\mathbb{C}_p \otimes_{\mathbb{Z}} E(F) \right) \\ &\simeq e^{\mathrm{alg}}_{F,1} \left(\mathbb{C}_p \otimes_{\mathbb{Z}} E(F) \right)^* \\ &\simeq e^{\mathrm{alg}}_{F,1} \left(\mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigoplus_{w \in S_p(F)} E_1(F_w) \right)^* \\ &\simeq e^{\mathrm{alg}}_{F,1} \left(\mathbb{C}_p \otimes_{\mathbb{Z}_p} Y_F(T^*(1))^* \right), \end{aligned}$$

where the first isomorphism is induced by the Kummer map $E(F) \to H^1(\mathcal{O}_{F,S(F)}, T_p(E))$ (see [BSS19, (21)]), the second by the Néron-Tate height pairing, the third by the localization map $E(F) \to E(F_w)$, and the last by $\beta \circ \alpha \circ \exp^*$.

We now relate Kato's zeta element z_F^{Kato} with the special element $\eta_F = \eta_{F/\mathbb{Q},S(F),\emptyset}(T)$ in Definition 2.7. By Kato's deep result [Kat04, Th. 6.6 and 9.7], we have

$$\lambda_{T,F}(e^{\mathrm{an}}_{F,0} \cdot z_F^{\mathrm{Kato}}) = e^{\mathrm{an}}_{F,0} \cdot \theta_{F/\mathbb{Q},S(F)}(E,1) \cdot b_F,$$

where $\theta_{F/\mathbb{Q},S(F)}(E,s) := \theta_{F/\mathbb{Q},S(F),\emptyset}(M^*(1), s-1) = \sum_{\chi \in \widehat{\mathcal{G}}_F} L_{S(F)}(E, \chi^{-1}, s) e_\chi$. So by the definition of η_F we have

$$e^{\mathrm{an}}_{F,0} \cdot z_F^{\mathrm{Kato}} = e^{\mathrm{an}}_{F,0} \cdot \eta_F.$$

It is natural to expect

$$z_F^{\mathrm{Kato}} = \eta_F.$$

(This is the conjecture made in [BSS19, Conj. 6.2].) We remark that the equality

$$z_{\mathbb{Q}}^{\mathrm{Kato}} = \eta_{\mathbb{Q}}$$

is equivalent to Perrin-Riou's conjecture [Per93] (see [BSS19, Prop. 6.5] or [BKS24, Prop. 2.10]).

2.4 The period-regulator isomorphism

In this subsection, we give a general definition of the period-regulator isomorphism

$$\lambda_{T,F} : e_{T,F}\left(\mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits^r_{\mathcal{A}[\mathcal{G}_F]} H^1(\mathcal{O}_{F,S(F)},T)\right) \xrightarrow{\sim} e_{T,F}\left(\mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits^r_{\mathcal{A}[\mathcal{G}_F]} Y_F(T^*(1))^*\right).$$

Let $\mathbf{R}\Gamma_{c,\Sigma}(\mathcal{O}_{F,S(F)},T^*(1))$ be the Σ-modified compactly supported cohomology complex defined in [BuSa21, §2.3.2] and set

$$C_{F,S(F),\Sigma}(T) := \mathbf{R}\operatorname{Hom}_{\mathbb{Z}_p}(\mathbf{R}\Gamma_{c,\Sigma}(\mathcal{O}_{F,S(F)},T^*(1)),\mathbb{Z}_p[-2]).$$

It is well-known that $C_{F,S(F),\Sigma}(T)$ is a perfect complex of $\mathcal{A}[\mathcal{G}_F]$-modules, acyclic outside degrees zero and one (under Hypothesis 2.3(i)), and the Euler characteristic is zero. By [BuSa21, Prop. 2.22], we have a canonical isomorphism

$$H^0(C_{F,S(F),\Sigma}(T)) \simeq H^1_\Sigma(\mathcal{O}_{F,S(F)},T)$$

and a canonical exact sequence

$$0 \to H^2_\Sigma(\mathcal{O}_{F,S(F)},T) \to H^1(C_{F,S(F),\Sigma}(T)) \to Y_F(T^*(1))^* \to 0.$$

Let

$$\vartheta_{T,F} : \mathbb{C}_p \otimes_{\mathbb{Z}_p} \det_{\mathcal{A}[\mathcal{G}_F]}(C_{F,S(F),\Sigma}(T)) \xrightarrow{\sim} \mathbb{C}_p \otimes_{\mathbb{Z}_p} \mathcal{A}[\mathcal{G}_F]$$

be the isomorphism used in the formulation of the equivariant Tamagawa number conjecture (see [BuFl01, §3.4]). We normalize $\vartheta_{T,F}$ so that the equivariant Tamagawa number conjecture for $(M^*(1)\otimes_K F,\mathcal{A}[\mathcal{G}_F])$ is equivalent to the equality

$$\vartheta_{T,F}(\det_{\mathcal{A}[\mathcal{G}_F]}(C_{F,S(F),\Sigma}(T))) = \mathcal{A}[\mathcal{G}_F]\cdot\theta^*_{F/K,S(F),\Sigma}(M^*(1),0).$$

The map $\vartheta_{T,F}$ induces an isomorphism

$$\mathbb{C}_p \otimes_{\mathbb{Q}_p} \det_{A[\mathcal{G}_F]}(\mathbb{Q}_p \otimes_{\mathbb{Z}_p} H^0(C_{F,S(F),\Sigma}(T))) \xrightarrow{\sim} \mathbb{C}_p \otimes_{\mathbb{Q}_p} \det_{A[\mathcal{G}_F]}(\mathbb{Q}_p \otimes_{\mathbb{Z}_p} H^1(C_{F,S(F),\Sigma}(T))),$$

which becomes

$$\mathbb{C}_p \otimes_{\mathbb{Q}_p} \det_{A[\mathcal{G}_F]}(H^1(\mathcal{O}_{F,S(F)},V)) \xrightarrow{\sim} \mathbb{C}_p \otimes_{\mathbb{Q}_p} \left(\det_{A[\mathcal{G}_F]}(H^2(\mathcal{O}_{F,S(F)},V)) \otimes_{\mathcal{A}[\mathcal{G}_F]} \bigwedge\nolimits^r_{\mathcal{A}[\mathcal{G}_F]} Y_F(T^*(1))^*\right). \tag{2.4.1}$$

We define $\lambda_{T,F}$ to be the isomorphism induced by this isomorphism (note that the idempotent $e_{T,F}$ kills $H^2(\mathcal{O}_{F,S(F)},V)$).

2.5 Extended special elements

Let $\eta_F = \eta_{F/K,S(F),\Sigma}(T)$ be the special element for $F \in \Omega(\mathcal{K})$ in Definition 2.7. For later use, we study connections between this element for $F = K$ and the Tamagawa number conjecture for $M^*(1)$.

First, note that by definition $\eta_K = \eta_{K/K,S,\Sigma}(T) \in \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge_{\mathcal{A}}^r H^1(\mathcal{O}_{K,S},T)$ can be zero. In fact, we have

$$\eta_K = 0 \Leftrightarrow H^2(\mathcal{O}_{K,S},V) \neq 0.$$

(This phenomenon can be regarded as a "trivial zero (or exceptional zero) phenomenon" for Euler systems. See Remark 2.14 below.) For this reason, we extend the definition of η_K so that it always becomes non-zero.

We set

$$e := \dim_A(H^2(\mathcal{O}_{K,S},V)).$$

We define the "extended period-regulator isomorphism"

$$\widetilde{\lambda}_{T,K} : \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathcal{A}}^{r+e} H^1(\mathcal{O}_{K,S},T) \xrightarrow{\sim} \mathbb{C}_p \otimes_{\mathbb{Z}_p} \left(\bigwedge\nolimits_{\mathcal{A}}^{e} H^2_{\Sigma}(\mathcal{O}_{K,S},T)_{\mathrm{tf}} \otimes_{\mathcal{A}} \bigwedge\nolimits_{\mathcal{A}}^{r} Y_K(T^*(1))^* \right)$$

to be the isomorphism induced by (2.4.1) (with $F = K$).

Definition 2.11. Assume $H^0(K,T) = 0$. Fix $\mathcal{A}$-bases $x \in \bigwedge_{\mathcal{A}}^e H^2_{\Sigma}(\mathcal{O}_{K,S},T)_{\mathrm{tf}}$ and $b \in \bigwedge_{\mathcal{A}}^r Y_K(T^*(1))^*$. We define the *extended special element* for T by

$$\widetilde{\eta}_K = \widetilde{\eta}_{K,S,\Sigma}(T) := \widetilde{\lambda}_{T,K}^{-1}(L^*_{S,\Sigma}(M^*(1),0) \cdot (x \otimes b)) \in \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathcal{A}}^{r+e} H^1(\mathcal{O}_{K,S},T).$$

Remark 2.12. By definition, $\widetilde{\eta}_K$ is always non-zero. When $e = 0$ (i.e., $H^2(\mathcal{O}_{K,S},V) = 0$), we have

$$\eta_K = \widetilde{\eta}_K \neq 0.$$

Remark 2.13. We can expect $H^2(\mathcal{O}_{K,S},V) = 0$ in many cases. For example, it is well-known that $H^2(\mathcal{O}_{K,S},\mathbb{Q}_p) = 0$ is equivalent to the Leopoldt conjecture for K. Soulé proved $H^2(\mathcal{O}_{K,S},\mathbb{Q}_p(j)) = 0$ for $j > 1$, and Schneider conjectured that $H^2(\mathcal{O}_{K,S},\mathbb{Q}_p(j)) = 0$ for $j \leq 0$ (see [NSW00, p. 641]). More generally, for a smooth projective scheme X/K, Jannsen conjectured that $H^2(\mathcal{O}_{K,S},H^i_{\text{ét}}(X \times_K \overline{\mathbb{Q}},\mathbb{Q}_p)(j)) = 0$ if $i+1 < j$ or $i+1 > 2j$ (see [Jan89, Conj. 1]). However, $H^2(\mathcal{O}_{K,S},V)$ can be non-zero in the cases considered in §2.3 (see also Remarks 2.14, 2.15 and 2.16 below).

Remark 2.14. Let χ be a non-trivial character of G_K of finite order and set $F := \overline{\mathbb{Q}}^{\ker \chi}$. If $T = \mathbb{Z}_p[\mathrm{im}\,\chi](1) \otimes \chi^{-1}$ and $S = S_\infty(K) \cup S_p(K) \cup S_{\mathrm{ram}}(F/K)$, then we have

$$H^2(\mathcal{O}_{K,S},V) \simeq e_\chi(\mathbb{Q}_p(\mathrm{im}\,\chi) \otimes_{\mathbb{Z}} X_{F,S_p(K)}),$$

where $X_{F,S_p(K)}$ is defined in §2.3.1. So in this case

$$H^2(\mathcal{O}_{K,S},V) = 0 \Leftrightarrow \chi(\mathrm{Fr}_v) \neq 1 \text{ for all } v \in S_p(K).$$

This is the usual "no trivial zeros" condition (see [BKS17, p. 1555]). However, when $T = T_p(E)$ with an elliptic curve E, the condition $H^2(\mathcal{O}_{K,S},V) \neq 0$ has nothing to do with the "exceptional zero" phenomenon in the sense of Mazur-Tate-Teitelbaum [MTT86].

Remark 2.15. Consider the $\mathbb{G}_m$ case (i.e., $M = h^0(K)(1)$ and $T = \mathbb{Z}_p(1)$). In this case, the extended special element is explicitly described as follows. First, note that there is a canonical exact sequence

$$0 \to \mathbb{Z}_p \otimes_{\mathbb{Z}} \mathrm{Cl}_S^{\Sigma}(K) \to H^2_{\Sigma}(\mathcal{O}_{K,S}, \mathbb{Z}_p(1)) \to \mathbb{Z}_p \otimes_{\mathbb{Z}} X_{K,S\setminus S_\infty(K)} \to 0,$$

where $\mathrm{Cl}_S^{\Sigma}(K)$ is the (S,Σ)-class group of K (see [BKS16, §1.7] for example) and $X_{K,S\setminus S_\infty(K)}$ is defined in §2.3.1. So we have

$$H^2_{\Sigma}(\mathcal{O}_{K,S}, \mathbb{Z}_p(1))_{\mathrm{tf}} \simeq \mathbb{Z}_p \otimes_{\mathbb{Z}} X_{K,S\setminus S_\infty(K)}$$

and

$$e = \#S - \#S_\infty(K) - 1.$$

We set $r := \#S_\infty(K)$ (so that $\#S = r+e+1$). We fix a labeling $S_\infty(K) = \{v_1, \dots, v_r\}$ and $S = \{v_0, v_1, \dots, v_{r+e}\}$. Recall that $Y_K(T^*(1))^*$ is identified with $\mathbb{Z}_p \otimes_{\mathbb{Z}} Y_{K,S_\infty(K)}$ (see §2.3.1). We define $\mathbb{Z}$-bases $x \in \bigwedge_{\mathbb{Z}}^e X_{K,S\setminus S_\infty(K)}$ and $b \in \bigwedge_{\mathbb{Z}}^r Y_{K,S_\infty(K)}$ by setting

$$x := (v_{r+1} - v_0) \wedge \cdots \wedge (v_{r+e} - v_0) \text{ and } b := v_1 \wedge \cdots \wedge v_r.$$

We assume Σ is chosen so that the (S,Σ)-unit group $\mathcal{O}_{K,S,\Sigma}^\times := \ker(\mathcal{O}_{K,S}^\times \to \bigoplus_{v\in\Sigma}(\mathcal{O}_K/v)^\times)$ is torsion-free, and let $\{u_1, \dots, u_{r+e}\}$ be a $\mathbb{Z}$-basis of $\mathcal{O}_{K,S,\Sigma}^\times$. Then one sees that the extended special element is described explicitly as

$$\widetilde{\eta}_{K,S,\Sigma}(\mathbb{Z}_p(1)) = \pm\#\mathrm{Cl}_S^{\Sigma}(K) \cdot u_1 \wedge \cdots \wedge u_{r+e} \in \mathbb{Z}_p \otimes_{\mathbb{Z}} \bigwedge\nolimits_{\mathbb{Z}}^{r+e} \mathcal{O}_{K,S,\Sigma}^\times \simeq \bigwedge\nolimits_{\mathbb{Z}_p}^{r+e} H^1_{\Sigma}(\mathcal{O}_{K,S}, \mathbb{Z}_p(1)).$$

(This also coincides with the Rubin-Stark element for $(K/K, S, \Sigma, S \setminus \{v_0\})$.) In fact, the extended period-regulator isomorphism in this case is induced by the Dirichlet regulator

$$\delta_K : \mathbb{R} \otimes_{\mathbb{Z}} \mathcal{O}_{K,S}^\times \xrightarrow{\sim} \mathbb{R} \otimes_{\mathbb{Z}} X_{K,S};\ u \mapsto -\sum_{v\in S} \log|u|_v \cdot v,$$

and the above description follows from the well-known class number formula

$$L^*_{S,\Sigma}(M^*(1), 0) = \zeta^*_{K,S,\Sigma}(0) = \pm\#\mathrm{Cl}_S^{\Sigma}(K) \cdot \det(\log|u_i|_{v_j})_{1\le i,j\le r+e},$$

where $\zeta_{K,S,\Sigma}(s)$ is the usual (S,Σ)-modified Dedekind zeta function for K.

Remark 2.16. Consider the case when $K = \mathbb{Q}$ and $M = h^1(E)(1)$ with an elliptic curve E over $\mathbb{Q}$. If we assume $\text{Ш}(E/\mathbb{Q})[p^\infty] < \infty$, then we have

$$e = \max\{0, \mathrm{rank}(E(\mathbb{Q})) - 1\}.$$

(See [BSS19, Lem. 6.1].) If $L(E,1) \neq 0$, then by the argument in §2.3.2 we have $e = 0$ and

$$\widetilde{\eta}_{\mathbb{Q}} = \eta_{\mathbb{Q}} = z_{\mathbb{Q}}^{\mathrm{Kato}}.$$

(We take Σ to be empty.) If $\mathrm{rank}(E(\mathbb{Q})) > 0$, then $\widetilde{\eta}_{\mathbb{Q}}$ coincides with the "Birch-Swinnerton-Dyer element" η_x^{BSD} defined in [BKS24, Def. 2.4] (by letting $b \in Y_{\mathbb{Q}}(T^*(1))^*$ be $e^+\delta(\xi)^*$ in loc. cit.).

The following is a generalization of [BKS24, Prop. 2.6].

Proposition 2.17. *Assume that $H^0(K,T)=0$ and that either Σ is non-empty or $H^1(\mathcal{O}_{K,S},T)$ is $\mathcal{A}$-free. Then the Tamagawa number conjecture for $M^*(1)$ (with coefficients in $\mathcal{A}$) holds if and only if we have an equality of $\mathcal{A}$-modules*

$$\mathcal{A}\cdot\widetilde{\eta}_K = \mathrm{Fitt}_{\mathcal{A}}(H^2_\Sigma(\mathcal{O}_{K,S},T)_{\mathrm{tors}})\cdot {\bigwedge}_{\mathcal{A}}^{r+e} H^1_\Sigma(\mathcal{O}_{K,S},T).$$

In particular, the Tamagawa number conjecture for $M^*(1)$ implies the "integrality" of $\widetilde{\eta}_K$:

$$\widetilde{\eta}_K \in {\bigwedge}_{\mathcal{A}}^{r+e} H^1_\Sigma(\mathcal{O}_{K,S},T).$$

Proof. Consider the following map:

$$\begin{aligned}
\det_{\mathcal{A}}(C_{K,S,\Sigma}(T)) &\hookrightarrow \mathbb{Q}_p\otimes_{\mathbb{Z}_p}\det_{\mathcal{A}}(C_{K,S,\Sigma}(T))\\
&\simeq \det_{\mathcal{A}}(H^1_\Sigma(\mathcal{O}_{K,S},V))\otimes_{\mathcal{A}}\det_{\mathcal{A}}^{-1}(H^2_\Sigma(\mathcal{O}_{K,S},V))\otimes_{\mathcal{A}}\det_{\mathcal{A}}^{-1}(Y_K(T^*(1))^*)\\
&= \mathbb{Q}_p\otimes_{\mathbb{Z}_p}\Big({\bigwedge}_{\mathcal{A}}^{r+e}H^1_\Sigma(\mathcal{O}_{K,S},T)\otimes_{\mathcal{A}}{\bigwedge}_{\mathcal{A}}^{e}H^2_\Sigma(\mathcal{O}_{K,S},T)^*\otimes_{\mathcal{A}}{\bigwedge}_{\mathcal{A}}^{r}Y_K(T^*(1))\Big)\\
&\simeq \mathbb{Q}_p\otimes_{\mathbb{Z}_p}{\bigwedge}_{\mathcal{A}}^{r+e}H^1_\Sigma(\mathcal{O}_{K,S},T),
\end{aligned}$$

where the last isomorphism is defined by using the fixed $\mathcal{A}$-bases of ${\bigwedge}_{\mathcal{A}}^{e}H^2_\Sigma(\mathcal{O}_{K,S},T)_{\mathrm{tf}}$ and ${\bigwedge}_{\mathcal{A}}^{r}Y_K(T^*(1))^*$. The proposition follows by noting that the image of this map is

$$\mathrm{Fitt}_{\mathcal{A}}(H^2_\Sigma(\mathcal{O}_{K,S},T)_{\mathrm{tors}})\cdot{\bigwedge}_{\mathcal{A}}^{r+e}H^1_\Sigma(\mathcal{O}_{K,S},T).$$

□

3 The Iwasawa main conjecture for motives

In this section, we study Iwasawa theory for motives. In §3.1, we give a formulation of the (equivariant) Iwasawa main conjecture (see Conjecture 3.4). In §3.2, we give another formulation in the "non-equivariant" case (see Conjecture 3.9). In §3.3, we state a theorem that gives us an approach to prove "one half" of the Iwasawa main conjecture under standard hypotheses (see Theorem 3.17). §3.4 is devoted to the proof of Theorem 3.17.

Throughout this section, we let K be a number field and $p>2$ an odd prime number. Let $\mathcal{A}$ be the ring of integers of a finite extension of $\mathbb{Q}_p$. Let T be a free $\mathcal{A}$-module of finite rank equipped with a continuous action of the absolute Galois group G_K of K which is unramified outside a finite set of places of K.

3.1 Formulation of the Iwasawa main conjecture

We fix the following data:

- L/K: a finite abelian extension in which all $v\in S_\infty(K)$ split completely;

- K_∞/K: a $\mathbb{Z}_p^d$-extension with $d \geq 1$;
- S: a finite set of places of K containing $S_\infty(K) \cup S_p(K) \cup S_{\mathrm{ram}}(L/K) \cup S_{\mathrm{ram}}(T)$;
- Σ: a finite set of places of K such that $S \cap \Sigma = \emptyset$.

We set some notations attached to these data. We set

$$L_\infty := L \cdot K_\infty,\ \mathcal{G}_\infty := \mathrm{Gal}(L_\infty/K) \text{ and } \Lambda = \Lambda_{L_\infty} := \mathcal{A}[[\mathcal{G}_\infty]].$$

We set

$$\mathbb{T} := T \otimes_{\mathcal{A}} \Lambda,$$

on which G_K acts by

$$\sigma \cdot (t \otimes \lambda) := \sigma t \otimes \lambda \overline{\sigma}^{-1} \quad (\sigma \in G_K,\ t \in T,\ \lambda \in \Lambda),$$

where $\overline{\sigma}^{-1} \in \mathcal{G}_\infty$ denotes the image of $\sigma^{-1} \in G_K$ under the natural surjection $G_K \twoheadrightarrow \mathcal{G}_\infty$.

We keep assuming Hypothesis 2.3, and let $r = r_T := \mathrm{rank}_{\mathcal{A}}(Y_K(T^*(1)))$ be the basic rank (see Definition 2.1). It is well-known that $\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S}, \mathbb{T})$ is a perfect complex of Λ-modules, which is acyclic outside degrees one and two, and that there is a non-canonical isomorphism

$$Q(\Lambda) \otimes_\Lambda H^1_\Sigma(\mathcal{O}_{K,S}, \mathbb{T}) \simeq Q(\Lambda) \otimes_\Lambda (H^2_\Sigma(\mathcal{O}_{K,S}, \mathbb{T}) \oplus \Lambda^r). \tag{3.1.1}$$

Here $Q(\Lambda)$ denotes the total quotient ring of Λ. In other words, the Euler characteristic of the complex $\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S}, \mathbb{T})$ is r. We now assume the following.

Hypothesis 3.1 (The weak Leopoldt conjecture)**.** $H^2_\Sigma(\mathcal{O}_{K,S}, \mathbb{T})$ is Λ-torsion, i.e.,

$$Q(\Lambda) \otimes_\Lambda H^2_\Sigma(\mathcal{O}_{K,S}, \mathbb{T}) = 0.$$

Remark 3.2. If K_∞/K is the cyclotomic $\mathbb{Z}_p$-extension, then it is expected that Hypothesis 3.1 is always satisfied (see [Per95, §1.3]). When $T = \mathbb{Z}_p(1)$, by a well-known theorem of Iwasawa, Hypothesis 3.1 is satisfied if no finite place of K splits completely in K_∞ (in particular, it is satisfied if K_∞/K is the cyclotomic $\mathbb{Z}_p$-extension). When $K = \mathbb{Q}$ and $T = T_p(E)$ with an elliptic curve E over $\mathbb{Q}$, Hypothesis 3.1 is proved by Kato [Kat04, Th. 12.4(1)]. For the case when K is imaginary quadratic and K_∞/K is the anticyclotomic $\mathbb{Z}_p$-extension, see Remark 5.15.

Under Hypothesis 3.1, we see by (3.1.1) that

$$Q(\Lambda) \otimes_\Lambda H^1_\Sigma(\mathcal{O}_{K,S}, \mathbb{T}) \simeq Q(\Lambda)^r.$$

In particular, we have a canonical isomorphism

$$Q(\Lambda) \otimes_\Lambda \mathrm{det}^{-1}_\Lambda(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S}, \mathbb{T})) \simeq Q(\Lambda) \otimes_\Lambda \bigwedge\nolimits^r_\Lambda H^1_\Sigma(\mathcal{O}_{K,S}, \mathbb{T}).$$

Since we have $Q(\Lambda) \otimes_\Lambda \bigcap^r_\Lambda H \simeq Q(\Lambda) \otimes_\Lambda \bigwedge^r_\Lambda H$ for any finitely generated Λ-module H, we obtain a canonical isomorphism

$$Q(\Lambda) \otimes_\Lambda \mathrm{det}^{-1}_\Lambda(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S}, \mathbb{T})) \simeq Q(\Lambda) \otimes_\Lambda \bigcap\nolimits^r_\Lambda H^1_\Sigma(\mathcal{O}_{K,S}, \mathbb{T}). \tag{3.1.2}$$

We need the following lemma proved by Sakamoto in [Sak23, Lem. B.15], which is used frequently in this paper.

Lemma 3.3. *Assume Hypothesis 2.3. Then there is a canonical isomorphism*

$$\bigcap_{\Lambda}^{r} H^1_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T}) \simeq \varprojlim_{F\in\Omega(L_\infty)} \bigcap_{\mathcal{A}[\mathcal{G}_F]}^{r} H^1_{\Sigma}(\mathcal{O}_{F,S},T).$$

(Recall that $\Omega(L_\infty)$ denotes the set of finite subextensions F/K of L_∞/K and $\mathcal{G}_F := \mathrm{Gal}(F/K)$.)

We shall now formulate the Iwasawa main conjecture. We assume that a certain canonical element

$$c_{L_\infty} \in \bigcap_{\Lambda}^{r} H^1_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T})$$

is given. By Lemma 3.3, this is equivalent to assuming that a canonical Euler system

$$c \in \mathrm{ES}_r(T,L_\infty) = \varprojlim_{F\in\Omega(L_\infty)} \bigcap_{\mathcal{A}[\mathcal{G}_F]}^{r} H^1_{\Sigma}(\mathcal{O}_{F,S},T)$$

is given. A reasonable way is to assume that T comes from a motive M as in §2 and assume Conjecture 2.6, but we do not need to assume it. For example, in the elliptic curve case, one can take c to be Kato's Euler system, although we do not know if it satisfies the properties (i) and (ii) in Conjecture 2.6 (see Remark 2.10 and §2.3.2).

Conjecture 3.4 (The Iwasawa main conjecture)**.** *Assume Hypotheses 2.3 and 3.1. Then there exists a (unique) Λ-basis*

$$\mathfrak{z}_{L_\infty} \in \det\nolimits_{\Lambda}^{-1}(\mathbf{R}\Gamma_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T}))$$

such that the isomorphism

$$Q(\Lambda)\otimes_{\Lambda}\det\nolimits_{\Lambda}^{-1}(\mathbf{R}\Gamma_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T})) \overset{(3.1.2)}{\simeq} Q(\Lambda)\otimes_{\Lambda}\bigcap_{\Lambda}^{r} H^1_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T})$$

sends $\mathfrak{z}_{L_\infty}$ to c_{L_∞}.

Remark 3.5. In the $\mathbb{G}_m$ case (i.e., $M = h^0(K)(1)$ and $c = \eta^{\mathrm{RS}}$: see §2.3.1), Conjecture 3.4 is equivalent to $\mathrm{IMC}(L_\infty/K,S,\Sigma)$ in [BKS17, Conj. 3.1] (see also [BKS17, Rem. 3.6]).

Remark 3.6. If $K = \mathbb{Q}$, $M = h^1(E)(1)$ with an elliptic curve E over $\mathbb{Q}$ and $c = z^{\mathrm{Kato}}$ (see §2.3.2), then Conjecture 3.4 in the case $L = K$ is equivalent to Kato's main conjecture [Kat04, Conj. 12.10] (see [BKS24, Rem. 7.2]). In the "equivariant" case (i.e., L is a general abelian extension of K), Conjecture 3.4 is studied by the first author in [Kat21] and [Kat22].

Remark 3.7. Using the language introduced in [BuSa21], we can rephrase Conjecture 3.4 as follows. The module of "vertical determinantal systems" introduced in loc. cit. for L_∞ is simply defined by

$$\mathrm{VS}(T,L_\infty) := \det\nolimits_{\Lambda}^{-1}(\mathbf{R}\Gamma_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T})).$$

Then by [BuSa21, Th. 2.18] there is a canonical map

$$\Theta_{T,L_\infty} : \mathrm{VS}(T,L_\infty) \to \mathrm{ES}_r(T,L_\infty)$$

which induces (3.1.2). (Namely, one can show that the image of $\det_{\Lambda}^{-1}(\mathbf{R}\Gamma_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T}))$ under the map (3.1.2) lies in $\bigcap_{\Lambda}^{r} H^1_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T})$.) Thus Conjecture 3.4 predicts the existence of a Λ-basis

$$\mathfrak{z}_{L_\infty} \in \mathrm{VS}(T, L_\infty)$$

such that

$$\Theta_{T,L_\infty}(\mathfrak{z}_{L_\infty}) = c_{L_\infty}.$$

We note that the map Θ_{T,L_∞} can be defined without assuming Hypothesis 3.1. However, if Hypothesis 3.1 is not satisfied, then one can show that Θ_{T,L_∞} is not injective.

Remark 3.8. Let $\mathcal{K}/K$ be an abelian extension as in §2. The observation in Remark 3.7 can be generalized naturally for $\mathcal{K}$: under Hypothesis 2.3, one can define an $\mathcal{A}[[\mathrm{Gal}(\mathcal{K}/K)]]$-module $\mathrm{VS}(T,\mathcal{K})$, which is free of rank one, and a canonical map

$$\Theta_{T,\mathcal{K}} : \mathrm{VS}(T,\mathcal{K}) \to \mathrm{ES}_r(T,\mathcal{K}).$$

It is natural to expect that, for a given canonical Euler system $c \in \mathrm{ES}_r(T,\mathcal{K})$, there exists an $\mathcal{A}[[\mathrm{Gal}(\mathcal{K}/K)]]$-basis

$$\mathfrak{z} \in \mathrm{VS}(T,\mathcal{K})$$

such that

$$\Theta_{T,\mathcal{K}}(\mathfrak{z}) = c.$$

This gives a generalization of Conjecture 3.4. (Namely, Conjecture 3.4 is the special case of this prediction for $\mathcal{K} = L_\infty$.) We note that, by the work of Burns-Greither [BuGr03], this generalization of Conjecture 3.4 is known to be true when $K = \mathbb{Q}$ and c is the cyclotomic unit Euler system (see §2.3.1 and [BDSS23, Lem. 5.2]).

3.2 The “non-equivariant" Iwasawa main conjecture

In this subsection, we give another formulation of Conjecture 3.4 in the “non-equivariant" case, i.e., when $L = K$. We moreover assume that K_∞/K is a $\mathbb{Z}_p$-extension, that is, $d = 1$. In this case, the Iwasawa algebra $\Lambda = \Lambda_{K_\infty} := \mathcal{A}[[\mathrm{Gal}(K_\infty/K)]]$ is a two-dimensional regular local ring.

As before, we assume that a canonical element (Euler system)

$$c_{K_\infty} \in \bigcap\nolimits_{\Lambda}^{r} H^1_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T}) \simeq \varprojlim_{n} \bigcap\nolimits_{\mathcal{A}[\mathcal{G}_{K_n}]}^{r} H^1_{\Sigma}(\mathcal{O}_{K_n,S},T) = \mathrm{ES}_r(T,K_\infty)$$

is given, where K_n denotes the n-th layer of the $\mathbb{Z}_p$-extension K_∞/K.

We propose the following.

Conjecture 3.9 (The non-equivariant Iwasawa main conjecture)**.** *Assume Hypotheses 2.3 and 3.1. Then we have*

$$\mathrm{char}_{\Lambda}\left(\bigcap\nolimits_{\Lambda}^{r} H^1_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T})/\Lambda \cdot c_{K_\infty}\right) = \mathrm{char}_{\Lambda}(H^2_{\Sigma}(\mathcal{O}_{K,S},\mathbb{T})).$$

Proposition 3.10. *Conjecture 3.4 for $L = K$ and $d = 1$ is equivalent to Conjecture 3.9.*

To prove this proposition, we need the following algebraic lemma.

Lemma 3.11. *Let Λ be a ring isomorphic to the formal power series ring $\mathcal{A}[[X]]$. Let $Q(\Lambda)$ be the quotient field of Λ. Let H be a finitely generated Λ-module and set $r := \dim_{Q(\Lambda)}(Q(\Lambda)\otimes_\Lambda H)$.*

(i) The Λ-module $\bigcap_\Lambda^r H$ is free of rank one and the natural map

$$\bigcap\nolimits_\Lambda^r H \to Q(\Lambda)\otimes_\Lambda \bigcap\nolimits_\Lambda^r H \simeq Q(\Lambda)\otimes_\Lambda \bigwedge\nolimits_\Lambda^r H$$

is injective.

(ii) The image of the canonical map

$$\det_\Lambda(H) \to Q(\Lambda)\otimes_\Lambda \det_\Lambda(H) \simeq Q(\Lambda)\otimes_\Lambda \bigwedge\nolimits_\Lambda^r H$$

coincides with

$$\mathrm{char}_\Lambda(H_{\mathrm{tors}})^{-1}\cdot \bigcap\nolimits_\Lambda^r H,$$

where H_{tors} is the Λ-torsion submodule of H, and we regard $\bigcap_\Lambda^r H \subset Q(\Lambda)\otimes_\Lambda \bigwedge_\Lambda^r H$ by (i).

Proof. Since we have $Q(\Lambda)\otimes_\Lambda \bigcap_\Lambda^r H \simeq Q(\Lambda)\otimes_\Lambda \bigwedge_\Lambda^r H$, we see that the Λ-rank of $\bigcap_\Lambda^r H$ is one. The Λ-module $\bigcap_\Lambda^r H$ is actually free, since it is reflexive by definition (see [NSW00, Cor. 5.1.3 and Prop. 5.1.9]). This proves (i).

To prove (ii), it is sufficient to show that for any height one prime $\mathfrak{p}$ of Λ we have

$$\mathrm{im}\left(\det_{\Lambda_\mathfrak{p}}(H_\mathfrak{p}) \to Q(\Lambda)\otimes_{\Lambda_\mathfrak{p}} \bigwedge\nolimits_{\Lambda_\mathfrak{p}}^r H_\mathfrak{p}\right) = \mathrm{Fitt}_{\Lambda_\mathfrak{p}}(H_{\mathfrak{p},\mathrm{tors}})^{-1}\cdot \left(\bigcap\nolimits_\Lambda^r H\right)_\mathfrak{p}.$$

Since $\mathrm{Hom}_{\Lambda_\mathfrak{p}}(H_\mathfrak{p},\Lambda_\mathfrak{p}) = \mathrm{Hom}_{\Lambda_\mathfrak{p}}(H_{\mathfrak{p},\mathrm{tf}},\Lambda_\mathfrak{p})$, we have an identification

$$\left(\bigcap\nolimits_\Lambda^r H\right)_\mathfrak{p} = \bigwedge\nolimits_{\Lambda_\mathfrak{p}}^r H_{\mathfrak{p},\mathrm{tf}} \quad \text{in} \quad Q(\Lambda)\otimes_{\Lambda_\mathfrak{p}} \bigwedge\nolimits_{\Lambda_\mathfrak{p}}^r H_\mathfrak{p}.$$

The claim follows from the following well-known fact: for a discrete valuation ring R and a finitely generated R-module M, we have

$$\mathrm{im}\left(\det_R(M) \to Q(R)\otimes_R \bigwedge\nolimits_R^s M\right) = \mathrm{Fitt}_R(M_{\mathrm{tors}})^{-1}\cdot \bigwedge\nolimits_R^s M_{\mathrm{tf}},$$

where $s := \dim_{Q(R)}(Q(R)\otimes_R M)$.

□

Proof of Proposition 3.9. Since Λ is regular, we have a canonical isomorphism

$$\det\nolimits_\Lambda^{-1}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},\mathbb{T})) \simeq \det\nolimits_\Lambda(H^1_\Sigma(\mathcal{O}_{K,S},\mathbb{T}))\otimes_\Lambda \det\nolimits_\Lambda^{-1}(H^2_\Sigma(\mathcal{O}_{K,S},\mathbb{T})).$$

By Lemma 3.11, under (3.1.2), this module corresponds to

$$\mathrm{char}_\Lambda(H^2_\Sigma(\mathcal{O}_{K,S},\mathbb{T}))\cdot \bigcap\nolimits_\Lambda^r H^1_\Sigma(\mathcal{O}_{K,S},\mathbb{T}).$$

The claim easily follows from this. □

3.3 Deduction of one half of the main conjecture

In this subsection, we state Theorem 3.17 below, which is one of the main results in this paper. It gives a general strategy to solve Conjecture 3.4. The proof will be given in the next subsection.

We keep the preceding notations. For simplicity, we assume $\Sigma = \emptyset$. Note that then, for each abelian extension $\mathcal{K}/K$, Hypothesis 2.3 is equivalent to $H^0(\mathcal{K}, T/p) = 0$. We will essentially consider the case where $\mathcal{K}$ contains L_∞ and $\mathcal{K}/L_\infty$ is a pro-p extension, and in that case $H^0(\mathcal{K}, T/p) = 0$ is equivalent to $H^0(L, T/p) = 0$, which will be a part of Hypothesis 3.13 below.

We give a list of hypotheses. Let $(-)^\vee$ denote the Pontryagin dual.

Hypothesis 3.12. For every $\mathfrak{q} \in S \setminus S_\infty(K)$, we have

$$H^0(L \otimes_K K_\mathfrak{q}, T^\vee(1)) = 0,$$

which is by the local duality equivalent to

$$H^2(L \otimes_K K_\mathfrak{q}, T) = 0.$$

Hypothesis 3.13 ([BSS18, Hyp. 3.3])**.** We have

$$H^0(L, T/p) = 0$$

and

$$H^0(L, (T/p)^\vee(1)) = 0.$$

For each integers $m \geq 1$ and $n \geq 0$, we put

$$R_{m,n} = \mathcal{A}/p^m[\mathrm{Gal}(L_n/K)],$$

where L_n is the n-th layer of L_∞/L (i.e., $\mathrm{Gal}(L_\infty/L_n) = \mathrm{Gal}(L_\infty/L)^{p^n}$ and $\mathrm{Gal}(L_n/L) \simeq (\mathbb{Z}/p^n)^d$). Then $R_{m,n}$ is a zero-dimensional Gorenstein ring, and we have $\Lambda = \varprojlim_{m,n} R_{m,n}$. We also put

$$T_{m,n} = T \otimes_\mathcal{A} R_{m,n},$$

which we regard as a Galois representation over $R_{m,n}$ in the same way as $\mathbb{T} = T \otimes \Lambda$.

As in [BSS18, §3.1.2], we use the following notation. We put

$$K_{p^m} = K(\mu_{p^m}, (\mathcal{O}_K^\times)^{1/p^m})K(1),$$

where $K(1)$ denotes the maximal p-extension in the Hilbert class field of K. For each m, n, let $K(T_{m,n})$ be the minimal Galois extension of K such that the action of G_K on $T_{m,n}$ factors through $\mathrm{Gal}(K(T_{m,n})/K)$. We put $K(T_{m,n})_{p^m} = K(T_{m,n})K_{p^m}$.

Hypothesis 3.14 ([BSS18, Hyp. 3.2])**.**

(i) The residual representation of T is irreducible as a representation of G_K.

(ii) For every $m \geq 1$ and $n \geq 0$, there exists $\tau \in G_{K_{p^m}}$ such that $T_{m,n}/(\tau-1)T_{m,n}$ is a free $R_{m,n}$-module of rank one.

(iii) For every $m \geq 1$, $n \geq 0$, we have

$$H^1(K(T_{m,n})_{p^m}/K, T_{m,n}) = 0$$

and

$$H^1(K(T_{m,n})_{p^m}/K, T_{m,n}^\vee(1)) = 0.$$

The following result gives a sufficient condition for Hypotheses 3.13 and 3.14 to be satisfied.

Proposition 3.15. *Put $a := \mathrm{rank}_{\mathcal{A}}(T)$ and suppose $a \geq 2$. Then Hypotheses 3.13 and 3.14 are satisfied if $p \geq 5$, $(a, p-1) \neq 1$, and the image of the Galois representation*

$$G_K \to \mathrm{Aut}_{\mathcal{A}}(T) \simeq \mathrm{GL}_a(\mathcal{A})$$

contains $\mathrm{SL}_a(\mathbb{Z}_p)$.

Proof. This result will not be used in this paper, so the reader may skip the proof.

The basic idea is the same as [BSS19, Lem. 6.17(ii)]. We first observe the following claim: for any solvable extension M/K (i.e., a Galois extension whose Galois group is solvable), the image of the homomorphism

$$G_M \to \mathrm{Aut}_{\mathcal{A}}(T) \simeq \mathrm{GL}_a(\mathcal{A})$$

also contains $\mathrm{SL}_a(\mathbb{Z}_p)$. This follows from the fact that $\mathrm{SL}_a(\mathbb{Z}_p)$ is a perfect group since we assume $p \geq 5$.

Let π be a uniformizer of $\mathcal{A}$. Then the above claim implies that, for any solvable extension M/K, the image of the homomorphism

$$G_M \to \mathrm{Aut}_{\mathcal{A}}(T/\pi) \simeq \mathrm{GL}_a(\mathcal{A}/\pi)$$

contains $\mathrm{SL}_a(\mathbb{F}_p)$. It is easy to see that the action of $\mathrm{SL}_a(\mathbb{F}_p)$ on $(\mathcal{A}/\pi)^{\oplus a}$ is irreducible. From these facts, Hypothesis 3.13 and Hypothesis 3.14(i) follow immediately.

Let us show Hypothesis 3.14(ii). By the above claim, we can take an element $\tau \in G_{L_n K_{p^m}}$ such that τ acts on $T \simeq \mathcal{A}^{\oplus a}$ as the matrix

$$\begin{pmatrix} 1 & 1 & & & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & 1 \\ & & & & 1 \end{pmatrix}.$$

Then $T/(\tau-1)T$ is a free $\mathcal{A}$-module of rank one. Moreover, since $\tau \in G_{L_n}$, we have

$$T_{m,n}/(\tau-1)T_{m,n} \simeq T/(\tau-1)T \otimes_{\mathcal{A}} R_{m,n}.$$

Therefore, this τ satisfies the condition of Hypothesis 3.14(ii).

Finally let us show Hypothesis 3.14(iii). We have an injective map $H^1(K(T_{m,n})_{p^m}/K, T_{m,n}) \hookrightarrow H^1(L_nK(T_{m,n})_{p^m}/K, T_{m,n})$ by inflation. Also, we have the inflation-restriction exact sequence

$$\begin{aligned} H^1(L_nK_{p^m}/K, H^0(L_nK_{p^m}, T_{m,n})) &\to H^1(L_nK(T_{m,n})_{p^m}/K, T_{m,n}) \\ &\to H^1(L_nK(T_{m,n})_{p^m}/L_nK_{p^m}, T_{m,n}). \end{aligned}$$

The first term here vanishes because we have

$$H^0(L_nK_{p^m}, T_{m,n}) \simeq H^0(L_nK_{p^m}, T/p^m) \otimes_{\mathcal{A}/p^m} R_{m,n}$$

and the above irreducibility of the Galois representation implies that $H^0(L_nK_{p^m}, T/p^m) = 0$. Moreover, we can show that the last term of the sequence also vanishes in the following manner. The action of $\mathrm{Gal}(L_nK(T_{m,n})_{p^m}/L_nK_{p^m})$ on $T_{m,n}$ is presented by a homomorphism

$$\mathrm{Gal}(L_nK(T_{m,n})_{p^m}/L_nK_{p^m}) \hookrightarrow \mathrm{Aut}_{R_{m,n}}(T_{m,n}) \simeq \mathrm{GL}_a(R_{m,n}).$$

The first claim in this proof implies that the image of this homomorphism contains $\mathrm{SL}_a(\mathbb{Z}_p/p^m)$. By the assumption $(a, p-1) \neq 1$, there exists an element $\lambda \in \mathbb{Z}_p^\times$ such that $\lambda \neq 1$ and $\lambda^a = 1$. Then we may use the "center kills" argument for the scalar matrix $\lambda \in \mathrm{SL}_a(\mathbb{Z}_p/p^m) \subset \mathrm{GL}_a(R_{m,n})$ and obtain $H^1(L_nK(T_{m,n})_{p^m}/L_nK_{p^m}, T_{m,n}) = 0$ as claimed. □

Hypothesis 3.16 ([BSS18, Hyp. 6.11])**.** $\mathrm{Fr}_{\mathfrak{q}}^{p^k} - 1$ is injective on T for every finite place $\mathfrak{q} \notin S$ and $k \geq 0$.

Recall that, for an abelian extension $\mathcal{K}/K$, the module of Euler systems $\mathrm{ES}_r(T, \mathcal{K})$ is defined in Definition 2.2 (we take $\Sigma = \emptyset$). When $\mathcal{K} \supset L_\infty$, we can consider the composite map

$$\mathrm{ES}_r(T, \mathcal{K}) \to \mathrm{ES}_r(T, L_\infty) \simeq {\bigcap}_\Lambda^r H^1(\mathcal{O}_{K,S}, \mathbb{T}),$$

where the first map is the natural restriction map and the last isomorphism follows from Lemma 3.3. For each $c = (c_F)_F \in \mathrm{ES}_r(T, \mathcal{K})$, we write $c_{L_\infty} \in {\bigcap}_\Lambda^r H^1(\mathcal{O}_{K,S}, \mathbb{T})$ for the image of c.

The following is the main result of this section, which roughly says that "one half" of the Iwasawa main conjecture (Conjecture 3.4) holds for any Euler system $c \in \mathrm{ES}_r(T, \mathcal{K})$.

Theorem 3.17. *Let $\mathcal{K}$ be an abelian extension of K that contains L_∞ and $K(\mathfrak{q})$ for each finite place $\mathfrak{q} \notin S$, where $K(\mathfrak{q})$ denotes the maximal p-extension in the ray class field of K modulo $\mathfrak{q}$. Let us assume that Hypotheses 3.1, 3.12, 3.13, 3.14, and 3.16 hold. We assume $r \geq 1$ and $p \geq 5$. Then, for any Euler system*

$$c \in \mathrm{ES}_r(T, \mathcal{K}),$$

there exists an element

$$\mathfrak{z}_{L_\infty} \in \mathrm{det}_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S}, \mathbb{T}))$$

such that the isomorphism

$$Q(\Lambda) \otimes_\Lambda \mathrm{det}_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S}, \mathbb{T})) \overset{(3.1.2)}{\simeq} Q(\Lambda) \otimes_\Lambda {\bigcap}_\Lambda^r H^1(\mathcal{O}_{K,S}, \mathbb{T})$$

sends $\mathfrak{z}_{L_\infty}$ to c_{L_∞}.

3.4 Proof of Theorem 3.17

This subsection is devoted to the proof of Theorem 3.17. One of the main ingredients is the work [BSS18] by Burns, Sakamoto, and the second-named author. Another is the work [Kat22] by the first-named author.

3.4.1 Reformulation using basic elements

The first step is to reformulate Theorem 3.17 in terms of basic elements that are introduced in [Kat22]. It will be clear that the reformulation has a similar background to that in Remark 3.7.

Recall that the complex $\mathbf{R}\Gamma(\mathcal{O}_{K,S},\mathbb{T})$ is perfect over Λ. By the Euler-Poincare characteristic formula, the Euler characteristic of $\mathbf{R}\Gamma(\mathcal{O}_{K,S},\mathbb{T})$ equals to the basic rank $r = r_T$.

Definition 3.18 ([Kat22, Def. 3.1 and 3.2])**.** Suppose that $H^0(L,T/p) = 0$ holds. This implies that, for each $m \geq 1, n \geq 0$, the complex $\mathbf{R}\Gamma(\mathcal{O}_{K,S},T_{m,n})$ is acyclic outside degrees one and two. Then we have a natural homomorphism

$$\Pi_{m,n} : \det^{-1}_{R_{m,n}}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},T_{m,n})) \to {\bigcap}^r_{R_{m,n}} H^1(\mathcal{O}_{K,S},T_{m,n})$$

for each m,n (see [Kat22, Def. 3.1]). Then an element of ${\bigcap}^r_{R_{m,n}} H^1(\mathcal{O}_{K,S},T_{m,n})$ is called *basic* (resp. *primitive basic*) if it is the image of an element (resp. a basis) of $\det^{-1}_{R_{m,n}}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},T_{m,n}))$ under $\Pi_{m,n}$.

Note that, by Sakamoto [Sak23, Lem. B.15] as in Lemma 3.3, we have a natural isomorphism

$${\bigcap}^r_{\Lambda} H^1(\mathcal{O}_{K,S},\mathbb{T}) \simeq \varprojlim_{m,n} {\bigcap}^r_{R_{m,n}} H^1(\mathcal{O}_{K,S},T_{m,n}).$$

Then, taking the limit of $\Pi_{m,n}$, we obtain a natural homomorphism

$$\Pi : \det^{-1}_{\Lambda}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},\mathbb{T})) \to {\bigcap}^r_{\Lambda} H^1(\mathcal{O}_{K,S},\mathbb{T}).$$

In a similar way as above, we define the notion of (primitive) basic elements in ${\bigcap}^r_{\Lambda} H^1(\mathcal{O}_{K,S},\mathbb{T})$.

In fact, the homomorphism Π can be identified with Θ_{T,L_∞} in Remark 3.7. Therefore, we can reformulate Conjecture 3.4 as follows, respecting the spirit of [Kat22].

Conjecture 3.19. *The given canonical element $c_{L_\infty} \in {\bigcap}^r_{\Lambda} H^1(\mathcal{O}_{K,S},\mathbb{T})$ as in §3.1 is primitive basic for* $\mathbf{R}\Gamma(\mathcal{O}_{K,S},\mathbb{T})$.

If we assume Hypothesis 3.1 (the weak Leopoldt conjecture), then Conjectures 3.4 and 3.19 are equivalent. Moreover, as in Remark 3.7, the homomorphism Π is injective if and only if Hypothesis 3.1 holds. In the rest of this section, we do not assume Hypothesis 3.1. In fact, a general strategy to prove it is to use Theorem 3.20 below and the property of c_{L_∞} that its annihilator ideal vanishes.

In the same way as the equivalence between Conjectures 3.4 and 3.19, we can now restate Theorem 3.17 as follows.

Theorem 3.20. *Assume the same hypotheses as in Theorem 3.17 except for Hypothesis 3.1. Then every element in the image of the natural map*

$$\mathrm{ES}_r(T,\mathcal{K}) \to {\bigcap}_{\Lambda}^{r} H^1(\mathcal{O}_{K,S},\mathbb{T})$$

that sends c to c_{L_∞} is basic.

The rest of this subsection is devoted to the proof of Theorem 3.20. First let us observe that we may assume that $\mathcal{K}/L$ is a pro-p extension. For, writing $\mathcal{K}'$ for the maximal pro-p subextension of $\mathcal{K}/L$, the extension $\mathcal{K}'/K$ also satisfies the conditions in Theorem 3.17, and that the homomorphism $\mathrm{ES}_r(T,\mathcal{K}) \to \bigcap_{\Lambda}^{r} H^1(\mathcal{O}_{K,S},\mathbb{T})$ factors through $\mathrm{ES}_r(T,\mathcal{K}')$.

Let us outline the proof of Theorem 3.20. We assume that $\mathcal{K}/L$ is a pro-p extension. By [Kat22, Prop. 3.3], the theorem is equivalent to that, for each $m \geq 1$ and $n \geq 0$, every element in the image of

$$\mathrm{ES}_r(T,\mathcal{K}) \to \mathrm{ES}_r(T,L_\infty) \to {\bigcap}_{R_{m,n}}^{r} H^1(\mathcal{O}_{K,S},T_{m,n})$$

is basic. We shall prove this by constructing the following natural commutative diagram whose upper diagonal arrow is the last displayed map:

$$\begin{array}{ccc}
\mathrm{ES}_r(T,\mathcal{K}) & \searrow & \\
\downarrow{\scriptstyle \mathcal{D}_r} & & \\
\mathrm{KS}_r(T_{m,n},\mathcal{P}(T_{m,n})) & \longrightarrow & \bigcap_{R_{m,n}}^{r} H^1(\mathcal{O}_{K,S},T_{m,n}) \\
\uparrow{\scriptstyle \mathrm{Reg}_r\ \simeq} & \nearrow & \uparrow{\scriptstyle \Pi_{m,n}} \\
\mathrm{SS}_r(T_{m,n},\mathcal{P}(T_{m,n})) & \xrightarrow[\simeq]{} & \det_{R_{m,n}}^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},T_{m,n})).
\end{array} \tag{3.4.1}$$

Here KS_r denotes the module of Kolyvagin systems, SS_r the module of Stark systems, $\mathcal{D}_r$ the Kolyvagin derivative map, Reg_r the regulator map, and the other unnamed arrows are all natural maps (see the subsequent discussion for more detailed definitions). It is clear that this diagram proves the theorem. Basically, the lower triangle is (a direct generalization of) [Kat22, Th. 5.12]. The other triangles are obtained by applying main results of [BSS18].

Note that we have an advantage in using Π instead of Θ_{T,L_∞} in Remark 3.7; Π has obvious counterparts $\Pi_{m,n}$ for zero-dimensional ring $R_{m,n}$, for which the above diagram can be constructed.

In the rest of this subsection, fixing $m \geq 1$ and $n \geq 0$, we put

$$R = R_{m,n}, \qquad A = T_{m,n}.$$

(Note that A denoted the quotient field of $\mathcal{A}$ in §2, but it will never appear in the following and there will be no danger of confusion.)

3.4.2 Stark systems

Let us define the module of Stark systems and then construct the lower triangle of (3.4.1). We closely follow [Kat22, §5].

We set

$$\mathcal{P}(A) = \{\mathfrak{q} \notin S \mid \mathfrak{q} \text{ splits completely in } K_{p^m} \text{ and } A/(\mathrm{Fr}_{\mathfrak{q}} - 1)A \text{ is free of rank one over } R\}.$$

This coincides with the set $\mathcal{P}$ in [BSS18, §3.1.2]. Then for each $\mathfrak{q} \in \mathcal{P}(A)$, we have submodules

$$H^1_f(K_{\mathfrak{q}}, A), H^1_{\mathrm{tr}}(K_{\mathfrak{q}}, A) \subset H^1(K_{\mathfrak{q}}, A),$$

which are free of rank one, and $H^1(K_{\mathfrak{q}}, A) = H^1_f(K_{\mathfrak{q}}, A) \oplus H^1_{\mathrm{tr}}(K_{\mathfrak{q}}, A)$ (see [BSS18, §3.1.3]).

For a subset $\mathcal{Q} \subset \mathcal{P}(A)$, we define $\mathcal{N}(\mathcal{Q})$ as the set of square-free products of elements of $\mathcal{Q}$. For $\mathfrak{n} \in \mathcal{N}(\mathcal{P}(A))$, we write $\nu(\mathfrak{n})$ for the number of the prime divisors of $\mathfrak{n}$. By convention, for the unit ideal (1), we have $(1) \in \mathcal{N}(\mathcal{Q})$ for any $\mathcal{Q}$ and $\nu((1)) = 0$.

Definition 3.21 ([Kat22, Def. 5.2])**.** We say that $\mathfrak{n} \in \mathcal{N}(\mathcal{P}(A))$ is large (for A) if the natural localization map

$$H^1(\mathcal{O}_{K,S}, A^\vee(1)) \to \bigoplus_{\mathfrak{q} \mid \mathfrak{n}} H^1_f(K_{\mathfrak{q}}, A^\vee(1))$$

is injective.

For each $\mathfrak{n} \in \mathcal{N}(\mathcal{P}(A))$, let us put $S^{\mathfrak{n}} = S \cup \mathrm{prime}(\mathfrak{n})$, where $\mathrm{prime}(\mathfrak{n})$ denotes the set of prime divisors of $\mathfrak{n}$. Then, since we have an exact sequence

$$0 \to H^1(\mathcal{O}_{K,S}, A^\vee(1)) \to H^1(\mathcal{O}_{K,S^{\mathfrak{n}}}, A^\vee(1)) \to \bigoplus_{\mathfrak{q} \mid \mathfrak{n}} H^1_{/f}(K_{\mathfrak{q}}, A^\vee(1)),$$

we see that $\mathfrak{n}$ is large if and only if the map

$$H^1(\mathcal{O}_{K,S^{\mathfrak{n}}}, A^\vee(1)) \to \bigoplus_{\mathfrak{q} \mid \mathfrak{n}} H^1(K_{\mathfrak{q}}, A^\vee(1))$$

is injective.

The following is a consequence of the Chebotarev density theorem. We make essential use of Hypothesis 3.14 here.

Lemma 3.22 ([BSS18, Lem. 3.9])**.** *Suppose Hypothesis 3.14 holds. Then there exist infinitely many elements $\mathfrak{n} \in \mathcal{N}(\mathcal{P}(A))$ which are large for A.*

Proposition 3.23 ([Kat22, Prop. 5.8])**.** *Suppose that Hypotheses 3.12 and 3.13 hold. Let $\mathfrak{n} \in \mathcal{N}(\mathcal{P}(A))$ be large. Then $H^1(\mathcal{O}_{K,S^{\mathfrak{n}}}, A)$ is a free R-module of rank $r + \nu(\mathfrak{n})$ and we have a quasi-isomorphism*

$$\mathbf{R}\Gamma(\mathcal{O}_{K,S}, A) \simeq \left[H^1(\mathcal{O}_{K,S^{\mathfrak{n}}}, A) \to \bigoplus_{\mathfrak{q} \mid \mathfrak{n}} H^1_{/f}(K_{\mathfrak{q}}, A) \right], \tag{3.4.2}$$

where the right hand side is regarded as a complex concentrated in degrees one and two.

Proof. We shall first show that $H^1(\mathcal{O}_{K,S^{\mathfrak{n}}},A)$ is free of rank $r+\nu(\mathfrak{n})$. By Hypothesis 3.13, we have $H^0(\mathcal{O}_{K,S^{\mathfrak{n}}},A)=0$. Since $\mathbf{R}\Gamma(\mathcal{O}_{K,S^{\mathfrak{n}}},A)$ is a perfect complex with Euler characteristic r, it is enough to show that $H^2(\mathcal{O}_{K,S^{\mathfrak{n}}},A)$ is free of rank $\nu(\mathfrak{n})$.

We use the Poitou-Tate duality:

$$\begin{aligned} H^1(\mathcal{O}_{K,S^{\mathfrak{n}}},A) \to \bigoplus_{\mathfrak{q}\in S^{\mathfrak{n}}\setminus S_\infty(K)} H^1(K_{\mathfrak{q}},A) \xrightarrow{*} H^1(\mathcal{O}_{K,S^{\mathfrak{n}}},A^\vee(1))^\vee \\ \to H^2(\mathcal{O}_{K,S^{\mathfrak{n}}},A) \to \bigoplus_{\mathfrak{q}\in S^{\mathfrak{n}}\setminus S_\infty(K)} H^2(K_{\mathfrak{q}},A) \to H^0(\mathcal{O}_{K,S^{\mathfrak{n}}},A^\vee(1))^\vee. \end{aligned}$$

By the assumption that $\mathfrak{n}$ is large, the map with $*$ is surjective. Also, by Hypothesis 3.13, the last term vanishes. Hence, by Hypothesis 3.12, we obtain an isomorphism

$$H^2(\mathcal{O}_{K,S^{\mathfrak{n}}},A) \xrightarrow{\sim} \bigoplus_{\mathfrak{q}\mid\mathfrak{n}} H^2(K_{\mathfrak{q}},A), \tag{3.4.3}$$

which shows that $H^2(\mathcal{O}_{K,S^{\mathfrak{n}}},A)$ is free of rank $\nu(\mathfrak{n})$.

To complete the proof, we need to show the quasi-isomorphism (3.4.2). But this follows immediately from (3.4.3) and the natural long exact sequence

$$\begin{aligned} 0 \to H^1(\mathcal{O}_{K,S},A) \to H^1(\mathcal{O}_{K,S^{\mathfrak{n}}},A) \to \bigoplus_{\mathfrak{q}\mid\mathfrak{n}} H^1_{/f}(K_{\mathfrak{q}},A) \\ \to H^2(\mathcal{O}_{K,S},A) \to H^2(\mathcal{O}_{K,S^{\mathfrak{n}}},A) \to \bigoplus_{\mathfrak{q}\mid\mathfrak{n}} H^2(K_{\mathfrak{q}},A). \end{aligned}$$

□

We shall review the definition of the module of Stark systems.

Definition 3.24 ([BuSa21, §3.2], [BSS18, §4.1], [Kat22, Def. 5.4 and 5.5])**.** Let $\mathcal{Q} \subset \mathcal{P}(A)$ be a subset. For each $\mathfrak{n} \in \mathcal{N}(\mathcal{Q})$, we put

$$X^r_{\mathfrak{n}}(A) = \left(\bigcap\nolimits_R^{r+\nu(\mathfrak{n})} H^1(\mathcal{O}_{K,S^{\mathfrak{n}}},A)\right) \otimes_R \det\nolimits_R^{-1}\left(\bigoplus_{\mathfrak{q}\mid\mathfrak{n}} H^1_{/f}(K_{\mathfrak{q}},A)\right).$$

For each $\mathfrak{n} \mid \mathfrak{n}'$, by the exact sequence

$$0 \to H^1(\mathcal{O}_{K,S^{\mathfrak{n}}},A) \to H^1(\mathcal{O}_{K,S^{\mathfrak{n}'}},A) \to \bigoplus_{\mathfrak{q}\mid\mathfrak{n}'/\mathfrak{n}} H^1_{/f}(K_{\mathfrak{q}},A),$$

we have a natural homomorphism $X^r_{\mathfrak{n}'}(A) \to X^r_{\mathfrak{n}}(A)$. With respect to these transition maps, we define

$$\mathrm{SS}_r(A,\mathcal{Q}) = \varprojlim_{\mathfrak{n}\in\mathcal{N}(\mathcal{Q})} X^r_{\mathfrak{n}}(A).$$

We shall obtain the lower triangle in the diagram (3.4.1).

Theorem 3.25 ([Kat22, Th. 5.12])**.** *Suppose that Hypotheses 3.12, 3.13, and 3.14 hold. Then we have a commutative diagram*

$$\begin{array}{ccc} \mathrm{SS}_r(A,\mathcal{P}(A)) & \longrightarrow & X^r_{(1)}(A) \\ \downarrow \simeq & & \| \\ \det_R^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},A)) & \xrightarrow[\Pi_{m,n}]{} & \bigcap_R^r H^1(\mathcal{O}_{K,S},A). \end{array}$$

Here, (1) denotes the unit ideal. The upper horizontal map is the canonical projection. The right vertical equality is by definition.

Proof. The left vertical isomorphism follows from Lemma 3.22 and Proposition 3.23. The commutativity is a consequence of the actual constructions of maps. □

3.4.3 Kolyvagin systems

We introduce the module of Kolyvagin systems and construct the other two triangles of (3.4.1).

First we introduce a Selmer structure $\mathcal{F}$ on the Galois representation $A = T_{m,n}$ (see [BSS18, §3.1.1] for a general definition of Selmer structures) by

$$H^1_{\mathcal{F}}(K_{\mathfrak{q}},A) = \begin{cases} H^1(K_{\mathfrak{q}},A) & (\mathfrak{q} \in S \setminus S_\infty(K)) \\ H^1_f(K_{\mathfrak{q}},A) & (\mathfrak{q} \notin S). \end{cases}$$

Then the Selmer group $H^1_{\mathcal{F}}(K,A)$ is identified with $H^1(\mathcal{O}_{K,S},A)$. More generally, for $\mathfrak{n} \in \mathcal{N}(\mathcal{P}(A))$, we have

$$H^1_{\mathcal{F}^{\mathfrak{n}}}(K,A) = H^1(\mathcal{O}_{K,S^{\mathfrak{n}}},A)$$

and

$$H^1_{\mathcal{F}(\mathfrak{n})}(K,A) = \mathrm{Ker}\left(H^1(\mathcal{O}_{K,S^{\mathfrak{n}}},A) \to \bigoplus_{\mathfrak{q}|\mathfrak{n}} H^1_{/\mathrm{tr}}(K_{\mathfrak{q}},A)\right)$$

(see [BSS18, §3.1.3] for the definitions of the modified Selmer structures $\mathcal{F}^{\mathfrak{n}}$ and $\mathcal{F}(\mathfrak{n})$).

For each $\mathfrak{q} \in \mathcal{P}(A)$, we put $G_{\mathfrak{q}} = \mathrm{Gal}(K(\mathfrak{q})/K(1))$ (recall that $K(\mathfrak{q})$ denotes the maximal p-extension in the $\mathfrak{q}$-ray class field of K). More generally, for $\mathfrak{n} \in \mathcal{N}(\mathcal{P}(A))$, we put $G_{\mathfrak{n}} = \bigotimes_{\mathfrak{q}|\mathfrak{n}} G_{\mathfrak{q}}$.

Definition 3.26 ([BuSa21, Def. 4.1], [BSS18, §5.1])**.** Let $\mathcal{Q} \subset \mathcal{P}(A)$ be a subset. We define a Kolyvagin system

$$\kappa = (\kappa_{\mathfrak{n}})_{\mathfrak{n}} \in \prod_{\mathfrak{n} \in \mathcal{N}(\mathcal{Q})} \bigcap_R^r H^1_{\mathcal{F}(\mathfrak{n})}(K,A) \otimes G_{\mathfrak{n}}$$

by requiring the "finite-singular relation" (we do not recall the precise definition). Let $\mathrm{KS}_r(A,\mathcal{Q})$ denote the module of Kolyvagin systems.

Now we consider the middle triangle of (3.4.1).

Definition 3.27 ([BuSa21, §4.2], [BSS18, §5.2])**.** Let $\mathcal{Q} \subset \mathcal{P}(A)$ be a subset. We define the regulator map

$$\mathrm{Reg}_r : \mathrm{SS}_r(A, \mathcal{Q}) \to \mathrm{KS}_r(A, \mathcal{Q})$$

as follows. For each $\mathfrak{n} \in \mathcal{N}(\mathcal{Q})$, by the exact sequence defining $H^1_{\mathcal{F}(\mathfrak{n})}(K, A)$, we have a natural map

$$\bigcap\nolimits_R^{r+\nu(\mathfrak{n})} H^1(\mathcal{O}_{K,S^{\mathfrak{n}}}, A) \to \bigcap\nolimits_R^{r} H^1_{\mathcal{F}(\mathfrak{n})}(K, A) \otimes \bigotimes_{\mathfrak{q} \mid \mathfrak{n}} H^1_{/\mathrm{tr}}(K_{\mathfrak{q}}, A).$$

By combining with the "finite-singular comparison map," we then obtain a map

$$X^r_{\mathfrak{n}}(A) \to \bigcap\nolimits_R^{r} H^1_{\mathcal{F}(\mathfrak{n})}(K, A) \otimes G_{\mathfrak{n}}.$$

These maps for various $\mathfrak{n}$ define Reg_r.

By the construction, we obtain the middle commutative triangle in (3.4.1). Then we have to show that Reg_r is an isomorphism:

Theorem 3.28 ([BSS18, Th. 5.2(i)])**.** *Suppose that Hypotheses 3.12, 3.13, and 3.14 hold. Suppose $p \geq 5$. Then the regulator map*

$$\mathrm{Reg}_r : \mathrm{SS}_r(A, \mathcal{P}(A)) \xrightarrow{\sim} \mathrm{KS}_r(A, \mathcal{P}(A))$$

is an isomorphism.

Proof. In order to apply [BSS18, Th. 5.2(i)], we have to check [BSS18, Hyp. 4.2], i.e., that there exist infinitely many elements $\mathfrak{n} \in \mathcal{N}(\mathcal{P}(A))$ such that

$$H^1_{(\mathcal{F}^*)_{\mathfrak{n}}}(K, A^\vee(1)) = 0$$

and that $H^1_{\mathcal{F}^{\mathfrak{n}}}(K, A)$ is free of rank $r + \nu(\mathfrak{n})$ over R. By Lemma 3.22, it is enough to show that every $\mathfrak{n} \in \mathcal{N}(\mathcal{P}(A))$ which is large for A (in the sense of Definition 3.21) satisfies the conditions. By Proposition 3.23, the module $H^1_{\mathcal{F}^{\mathfrak{n}}}(K, A) = H^1(\mathcal{O}_{K,S^{\mathfrak{n}}}, A)$ is actually free of rank $r + \nu(\mathfrak{n})$. Moreover, since $H^1_{(\mathcal{F}^*)_{\mathfrak{n}}}(K, A^\vee(1))$ is the kernel of

$$H^1(\mathcal{O}_{K,S}, A^\vee(1)) \to \bigoplus_{\mathfrak{q} \in S^{\mathfrak{n}} \setminus S_\infty(K)} H^1(K_{\mathfrak{q}}, A^\vee(1)),$$

it vanishes as $\mathfrak{n}$ is large. □

Finally we consider the upper triangle of (3.4.1).

Theorem 3.29 ([BSS18, Cor. 6.13])**.** *Let $\mathcal{K}$ be an abelian extension of K such that $\mathcal{K} \supset L_\infty$ and $\mathcal{K}/L$ is a pro-p extension. Suppose that $\mathcal{K} \supset K(\mathfrak{q})$ for every $\mathfrak{q} \in \mathcal{P}(A)$. Suppose that $H^0(L, T/p) = 0$ and Hypothesis 3.16 hold. Then we have a natural homomorphism*

$$\mathcal{D}_r : \mathrm{ES}_r(T, \mathcal{K}) \to \mathrm{KS}_r(A, \mathcal{P}(A)),$$

called the derivative operator, such that we have a commutative diagram

$$\begin{array}{ccc} \mathrm{ES}_r(T,\mathcal{K}) & & \\ \Big\downarrow{\scriptstyle \mathcal{D}_r} & \searrow & \\ \mathrm{KS}_r(A,\mathcal{P}(A)) & \longrightarrow & \bigcap_R^r H^1(\mathcal{O}_{K,S},A), \end{array}$$

where the diagonal and the horizontal maps are the natural projection maps.

Proof. We only have to apply [BSS18, Cor. 6.13], taking the following remarks into account. In the corollary, the Selmer structure is taken as the canonical one $\mathcal{F}_{\mathrm{can}}$, but that does not matter since our Selmer structure $\mathcal{F}$ is larger. We also removed the latter condition in [BSS18, Hyp. 6.7], that is, "$\mathcal{K}$ contains a $\mathbb{Z}_p^d$-extension of K ($d \geq 1$) in which no finite place of K splits completely." For, this condition is only used in order to check that the Kolyvagin derivative of an Euler system satisfies the unramified condition outside S^{n}, which is already assumed in our definition of Euler systems. □

Thus we have the upper triangle in (3.4.1). This completes the construction of the commutative diagram (3.4.1), so we have also finished proving Theorem 3.20 and Theorem 3.17.

4 Derivatives of Euler systems

In this section, we generalize several conjectures and results in [BKS24], where the case of elliptic curves over $\mathbb{Q}$ is considered, to a general motive.

We use notations in §2. Throughout this section, we assume Hypothesis 2.3.

4.1 Bockstein maps

As in §2.5, we set

$$e := \dim_A(H^2(\mathcal{O}_{K,S},V)) = \mathrm{rank}_{\mathcal{A}}(H^2_\Sigma(\mathcal{O}_{K,S},T)_{\mathrm{tf}})$$

and fix an $\mathcal{A}$-basis $x \in \bigwedge_{\mathcal{A}}^e H^2_\Sigma(\mathcal{O}_{K,S},T)_{\mathrm{tf}}$. In this subsection, we define a "Bockstein regulator map"

$$\mathrm{Boc}_F = \mathrm{Boc}_{T,F,x} : \bigwedge\nolimits_{\mathcal{A}}^{r+e} H^1_\Sigma(\mathcal{O}_{K,S},T) \to \bigwedge\nolimits_{\mathcal{A}}^{r} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} I_F^e/I_F^{e+1}$$

for each $F \in \Omega(\mathcal{K})$, where I_F denotes the augmentation ideal of $\mathcal{A}[\mathcal{G}_F]$:

$$I_F := \ker(\mathcal{A}[\mathcal{G}_F] \twoheadrightarrow \mathcal{A}).$$

The definition of the map is more or less the same as that in [BKS24, §2.3].

Let

$$\beta = \beta_F : H^1_\Sigma(\mathcal{O}_{K,S},T) \to H^2_\Sigma(\mathcal{O}_{K,S},T)_{\mathrm{tf}} \otimes_{\mathcal{A}} I_F/I_F^2$$

be the Bockstein map, i.e., the map induced by the connecting homomorphism of the natural exact triangle

$$\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},T) \otimes^{\mathbf{L}}_{\mathcal{A}} I_F/I_F^2 \to \mathbf{R}\Gamma_\Sigma(\mathcal{O}_{F,S},T) \otimes^{\mathbf{L}}_{\mathcal{A}[\mathcal{G}_F]} \mathcal{A}[\mathcal{G}_F]/I_F^2 \to \mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},T) \to .$$

Write $x = x_1 \wedge \cdots \wedge x_e$ and define $\beta_i : H^1_\Sigma(\mathcal{O}_{K,S}, T) \to I_F/I_F^2$ by

$$\beta(a) = \sum_{i=1}^{e} x_i \otimes \beta_i(a).$$

The Bockstein regulator map is now defined by

$$\mathrm{Boc}_F(a_1 \wedge \cdots \wedge a_{r+e}) = \sum_{\sigma} \mathrm{sgn}(\sigma) a_{\sigma(e+1)} \wedge \cdots \wedge a_{\sigma(e+r)} \otimes \det(\beta_i(a_{\sigma(j)}))_{1 \le i,j \le e},$$

where σ runs over the elements of the symmetric group S_{r+e} such that $\sigma(1) < \cdots < \sigma(e)$ and $\sigma(e+1) < \cdots < \sigma(e+r)$. Note that Boc_F depends on the choice of x, but is independent of $x_1, \ldots, x_e$.

4.2 Derivatives and extended special elements

We now assume that a canonical Euler system

$$c \in \mathrm{ES}_r(T, \mathcal{K})$$

is given. In this subsection, we formulate a conjecture which relates "derivatives" of c with the extended special element $\widetilde{\eta}_K$ defined in §2.5. (Here "derivatives" are different from Kolyvagin's derivatives, which are considered in the usual Euler system argument [Rub00] as in Theorem 3.29, but similar to those considered by Darmon [Dar92].) This conjecture is a generalization of the "generalized Perrin-Riou conjecture" for Kato's Euler system formulated in [BKS24, Conj. 2.12].

As in Conjecture 2.6, we fix an $\mathcal{A}[[\mathrm{Gal}(\mathcal{K}/K)]]$-basis

$$b = (b_F)_F \in \varprojlim_{F \in \Omega(\mathcal{K})} {\bigwedge}^r_{\mathcal{A}[\mathcal{G}_F]} Y_F(T^*(1))^*.$$

(As Conjecture 2.6 suggests, the Euler system c should depend on the choice of b.)

We fix $F \in \Omega(\mathcal{K})$, which is unramified outside S (so that $S(F) = S$). Let

$$\begin{aligned} \iota_F : {\bigwedge}^r_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S}, T) \otimes_{\mathcal{A}} I_F^e/I_F^{e+1} &\hookrightarrow {\bigcap}^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S}, T) \otimes_{\mathcal{A}} I_F^e/I_F^{e+1} \\ &\subset {\bigcap}^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S}, T) \otimes_{\mathcal{A}} \mathcal{A}[\mathcal{G}_F]/I_F^{e+1} \end{aligned}$$

be the canonical injection defined in the same way as [San14, Lem. 2.11]. (Note that ${\bigwedge}^r_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S}, T) = {\bigcap}^r_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S}, T)$ since the $\mathcal{A}$-module $H^1_\Sigma(\mathcal{O}_{K,S}, T)$ is free.)

Let

$$\widetilde{\eta}_K = \widetilde{\eta}_{K,S,\Sigma}(T) \in \mathbb{C}_p \otimes_{\mathbb{Z}_p} {\bigwedge}^{r+e}_{\mathcal{A}} H^1(\mathcal{O}_{K,S}, T)$$

be the extended special element in Definition 2.11 (which depends on the fixed $\mathcal{A}$-bases $x \in {\bigwedge}^e_{\mathcal{A}} H^2_\Sigma(\mathcal{O}_{K,S}, T)_{\mathrm{tf}}$ and $b_K \in {\bigwedge}^r_{\mathcal{A}} Y_K(T^*(1))^*$).

The following conjecture predicts a precise relation between the element

$$\mathcal{N}_{F/K}(c_F) := \sum_{\sigma \in \mathcal{G}_F} \sigma c_F \otimes \sigma^{-1} \in {\bigcap}^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S}, T) \otimes_{\mathcal{A}} \mathcal{A}[\mathcal{G}_F]$$

and $\widetilde{\eta}_K$.

Conjecture 4.1. *Assume Hypothesis 2.3 for F.*

(i) We have

$$\mathcal{N}_{F/K}(c_F) \in \bigcap\nolimits^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S},T) \otimes_{\mathcal{A}} I_F^e.$$

(ii) Assume the integrality of $\widetilde{\eta}_K$, i.e.,

$$\widetilde{\eta}_K \in \bigwedge\nolimits^{r+e}_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T).$$

(This is a consequence of the Tamagawa number conjecture: see Proposition 2.17.) Then we have

$$\mathcal{N}_{F/K}(c_F) = (-1)^{re}\iota_F\left(\mathrm{Boc}_F(\widetilde{\eta}_K)\right) \textit{ in } \bigcap\nolimits^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S},T) \otimes_{\mathcal{A}} I_F^e/I_F^{e+1},$$

where

$$\mathrm{Boc}_F = \mathrm{Boc}_{T,F,x} : \bigwedge\nolimits^{r+e}_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) \to \bigwedge\nolimits^r_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} I_F^e/I_F^{e+1}$$

is the Bockstein regulator map defined in §4.1.

Remark 4.2. Conjecture 4.1(ii) in particular predicts the existence of a unique element

$$\kappa_F \in \bigwedge\nolimits^r_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} I_F^e/I_F^{e+1}$$

such that

$$\iota_F(\kappa_F) = \mathcal{N}_{F/K}(c_F) \text{ in } \bigcap\nolimits^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S},T) \otimes_{\mathcal{A}} \mathcal{A}[\mathcal{G}_F]/I_F^{e+1}.$$

We call this conjectural element the *Darmon derivative* of the Euler system c_F. Conjecture 4.1(ii) is then equivalent to the equality

$$\kappa_F = (-1)^{re}\mathrm{Boc}_F(\widetilde{\eta}_K).$$

Remark 4.3. When $e = 0$ (i.e., $H^2(\mathcal{O}_{K,S},V) = 0$), Conjecture 4.1(i) is trivially true and (ii) is equivalent to the equality

$$c_K = \eta_K,$$

where $\eta_K \in \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge^r_{\mathcal{A}} H^1(\mathcal{O}_{K,S},T)$ is the special element in Definition 2.7 (in this case we do not need to assume the integrality). If $K = \mathbb{Q}$, $M = h^1(E)(1)$ with an elliptic curve E over $\mathbb{Q}$ and $c = z^{\mathrm{Kato}}$, then Conjecture 4.1 is equivalent to [BKS24, Conj. 2.12] and the equality $z^{\mathrm{Kato}}_{\mathbb{Q}} = \eta_{\mathbb{Q}}$ is equivalent to Perrin-Riou's conjecture (see §2.3.2).

Remark 4.4. In the $\mathbb{G}_m$ case (i.e., $M = h^0(K)(1)$ and $c = \eta^{\mathrm{RS}}$: see §2.3.1), Conjecture 4.1 is equivalent to the (p-part of the) "Mazur-Rubin-Sano conjecture" formulated in [MaRu16, Conj. 5.2] and [San14, Conj. 3]. More precisely, Conjecture 4.1 is equivalent to $\mathrm{MRS}(F/K/K,S,\Sigma,S_\infty(K),S\setminus\{v_0\})_p$ in [BKS17, Conj. 4.2], where we choose a finite place $v_0 \in S$. (This choice corresponds to the choice of a $\mathbb{Z}_p$-basis $x \in \bigwedge^e_{\mathbb{Z}_p} H^2_\Sigma(\mathcal{O}_{K,S},\mathbb{Z}_p(1))_{\mathrm{tf}}$. See Remark 2.15.)

Remark 4.5. One can show that the equivariant Tamagawa number conjecture for $(M \otimes_K F, \mathcal{A}[\mathcal{G}_F])$ (see [BuFl01, Conj. 4]) implies Conjecture 4.1 when c_F is the special element η_F in Definition 2.7 (see also Remark 2.9).

Concerning the existence of Darmon derivatives in Remark 4.2, we have the following result.

Proposition 4.6. *Let L_∞/K be an extension as in §3.1 and $c_{L_\infty} \in \bigcap_\Lambda^r H^1_\Sigma(\mathcal{O}_{K,S},\mathbb{T})$ an Euler system. Assume Hypothesis 2.3 for $\mathcal{K} = L_\infty$. If there exists an element $\mathfrak{z}_{L_\infty} \in \det_\Lambda^{-1}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},\mathbb{T}))$ such that the map (3.1.2) sends $\mathfrak{z}_{L_\infty}$ to c_{L_∞}, then the Darmon derivative κ_F of c_F exists for any $F \in \Omega(L_\infty)$.*

Proof. Let

$$\mathfrak{z}_F \in \det^{-1}_{\mathcal{A}[\mathcal{G}_F]}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{F,S},T))$$

be the image of $\mathfrak{z}_{L_\infty}$ under the natural surjection $\det_\Lambda^{-1}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},\mathbb{T})) \twoheadrightarrow \det^{-1}_{\mathcal{A}[\mathcal{G}_F]}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{F,S},T))$.

We have the following commutative diagram:

$$\begin{array}{ccccc}
\det^{-1}_{\mathcal{A}[\mathcal{G}_F]}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{F,S},T)) & \xrightarrow{\Theta_{T,F}} & \bigcap^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S},T) & \xrightarrow{\mathcal{N}_{F/K}} & \bigcap^r_{\mathcal{A}[\mathcal{G}_F]} H^1_\Sigma(\mathcal{O}_{F,S},T) \otimes_{\mathcal{A}} \mathcal{A}[\mathcal{G}_F]/I_F^{e+1} \\
\downarrow & & & & \uparrow \iota_F \\
\det^{-1}_{\mathcal{A}}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},T)) & \xrightarrow[\Theta_x]{} & \bigwedge^{r+e}_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) & \xrightarrow[(-1)^{re}\mathrm{Boc}_F]{} & \bigwedge^r_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} I_F^e/I_F^{e+1}.
\end{array} \tag{4.2.1}$$

Here $\Theta_{T,F}$ is the F-component of the map Θ_{T,L_∞} mentioned in Remark 3.7, and Θ_x is the following map:

$$\begin{aligned}
\det^{-1}_{\mathcal{A}}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},T)) &\simeq \det_{\mathcal{A}}(H^1_\Sigma(\mathcal{O}_{K,S},T)) \otimes_{\mathcal{A}} \det^{-1}_{\mathcal{A}}(H^2_\Sigma(\mathcal{O}_{K,S},T)) \\
&\simeq \mathrm{Fitt}_{\mathcal{A}}(H^2_\Sigma(\mathcal{O}_{K,S},T)_{\mathrm{tors}}) \cdot \bigwedge^{r+e}_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} \bigwedge^{e}_{\mathcal{A}} H^2_\Sigma(\mathcal{O}_{K,S},T)^*_{\mathrm{tf}} \\
&\simeq \mathrm{Fitt}_{\mathcal{A}}(H^2_\Sigma(\mathcal{O}_{K,S},T)_{\mathrm{tors}}) \cdot \bigwedge^{r+e}_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) \\
&\subset \bigwedge^{r+e}_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T),
\end{aligned}$$

where the third isomorphism is defined by using the basis $x \in \bigwedge^e_{\mathcal{A}} H^2_\Sigma(\mathcal{O}_{K,S},T)_{\mathrm{tf}}$. The commutativity of the diagram is proved in the same way as [BKS16, Lem. 5.22] or [BKS24, Th. 7.8].

Since we have $\Theta_{T,F}(\mathfrak{z}_F) = c_F$, the commutative diagram implies that the element $\mathcal{N}_{F/K}(c_F)$ lies in the image of ι_F. This shows the existence of the Darmon derivative. □

4.3 An Iwasawa theoretic version

As in [BKS24, §4.3], we can formulate a natural Iwasawa theoretic version of Conjecture 4.1.

Let K_∞/K be a $\mathbb{Z}_p$-extension and consider the case $\mathcal{K} = K_\infty$. We keep assuming Hypothesis 2.3. Note that each $F \in \Omega(K_\infty)$ is of the form K_n (the n-th layer) for some n. We set

$$I_n := I_{K_n} \text{ and } I := \ker(\mathcal{A}[[\mathrm{Gal}(K_\infty/K)]] \twoheadrightarrow \mathcal{A}) \simeq \varprojlim_n I_n.$$

We define the Bockstein regulator map for K_∞ by

$$\begin{aligned}
\mathrm{Boc}_\infty := \varprojlim_n \mathrm{Boc}_{K_n} : \bigwedge^{r+e}_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) &\to \bigwedge^{r}_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} \varprojlim_n I_n^e/I_n^{e+1} \\
&\simeq \bigwedge^{r}_{\mathcal{A}} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} I^e/I^{e+1}.
\end{aligned}$$

This map induces

$$\mathrm{Boc}_\infty : \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathcal{A}}^{r+e} H^1_\Sigma(\mathcal{O}_{K,S},T) \to \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathcal{A}}^{r} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} I^e/I^{e+1}.$$

(Note that $I^e/I^{e+1} \simeq \mathcal{A}$ and so it does not vanish after taking $\mathbb{C}_p \otimes_{\mathbb{Z}_p} -$.)

For a given canonical Euler system $c \in \mathrm{ES}_r(T, K_\infty)$, we assume the existence of the Darmon derivative of c

$$\kappa_n := \kappa_{K_n} \in \bigwedge\nolimits_{\mathcal{A}}^{r} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} I_n^e/I_n^{e+1}$$

in Remark 4.2 for every n. Then one sees that $(\kappa_n)_n$ is an inverse system and so we can define the limit

$$\kappa_\infty := \varprojlim_n \kappa_n \in \bigwedge\nolimits_{\mathcal{A}}^{r} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} \varprojlim_n I_n^e/I_n^{e+1} \simeq \bigwedge\nolimits_{\mathcal{A}}^{r} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} I^e/I^{e+1}.$$

We now propose the following conjecture.

Conjecture 4.7. *We have*

$$\kappa_\infty = (-1)^{re}\mathrm{Boc}_\infty(\widetilde{\eta}_K) \text{ in } \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathcal{A}}^{r} H^1_\Sigma(\mathcal{O}_{K,S},T) \otimes_{\mathcal{A}} I^e/I^{e+1}.$$

Remark 4.8. Unlike Conjecture 4.1, we do not need to assume the integrality of $\widetilde{\eta}_K$ in Conjecture 4.7. This is one of the advantages of taking limits.

Remark 4.9. Conjecture 4.7 is a generalization of [BKS17, Conj. 4.2] and [BKS24, Conj. 4.9]. In particular, by [BKS17, Th. 4.9] and [BKS24, Cor. 6.7], Conjecture 4.7 constitutes a generalization of both the Gross-Stark conjecture [Gro82] and the p-adic Birch-Swinnerton-Dyer conjecture [MTT86].

We shall show that Conjectures 4.1 and 4.7 are equivalent under suitable assumptions.

Proposition 4.10. *Let L_∞/K be an extension as in §3.1 (with $d = 1$) and $c_{L_\infty} \in \bigcap_\Lambda^r H^1_\Sigma(\mathcal{O}_{K,S},\mathbb{T})$ an Euler system. We assume Hypothesis 2.3 for $\mathcal{K} = L_\infty$. We also assume the following.*

(i) There exists an element $\mathfrak{z}_{L_\infty} \in \det_\Lambda^{-1}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},\mathbb{T}))$ such that the map (3.1.2) sends $\mathfrak{z}_{L_\infty}$ to c_{L_∞}.

(ii) The map Boc_∞ is non-zero.

Then Conjecture 4.1 for any $F \in \Omega(L_\infty)$ is equivalent to Conjecture 4.7.

Proof. It is obvious that Conjecture 4.1 (for K_n for all n) implies Conjecture 4.7, so we shall prove the converse.

Let

$$\mathfrak{z}_K \in \det_{\mathcal{A}}^{-1}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},T))$$

be the image of $\mathfrak{z}_{L_\infty}$. Let

$$\Theta_x : \det_{\mathcal{A}}^{-1}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S},T)) \to \bigwedge\nolimits_{\mathcal{A}}^{r+e} H^1_\Sigma(\mathcal{O}_{K,S},T)$$

be the map defined in the proof of Proposition 4.6. By the commutative diagram (4.2.1) (for $F = K_n$ for all n), we see that the Darmon derivative of c_{K_∞} exists and it is given by

$$\kappa_\infty = (-1)^{re}\mathrm{Boc}_\infty(\Theta_x(\mathfrak{z}_K)).$$

Since Boc_∞ is non-zero by assumption, we see that Conjecture 4.7 is equivalent to the equality

$$\Theta_x(\mathfrak{z}_K) = \widetilde{\eta}_K. \tag{4.3.1}$$

Note that, since $\Theta_x(\mathfrak{z}_K)$ lies in $\bigwedge_{\mathcal{A}}^{r+e} H^1_\Sigma(\mathcal{O}_{K,S}, T)$, this equality implies the integrality of $\widetilde{\eta}_K$, which is assumed in Conjecture 4.1(ii).

Let $F \in \Omega(L_\infty)$. By the proof of Proposition 4.6, the Darmon derivative of c_F is given by

$$\kappa_F = (-1)^{re}\mathrm{Boc}_F(\Theta_x(\mathfrak{z}_K)).$$

So Conjecture 4.1 for F is equivalent to the equality

$$\mathrm{Boc}_F(\Theta_x(\mathfrak{z}_K)) = \mathrm{Boc}_F(\widetilde{\eta}_K).$$

This is obviously implied by (4.3.1). Thus we have proved that Conjecture 4.7 implies Conjecture 4.1 for any $F \in \Omega(L_\infty)$. □

4.4 A strategy for proving the Tamagawa number conjecture

As in the previous subsection, we consider the case $\mathcal{K} = K_\infty$.

Theorem 4.11. *Assume Hypotheses 2.3 and 3.1. If we also assume*

- *Conjecture 3.9 (the non-equivariant Iwasawa main conjecture) for c,*
- *Conjecture 4.7 for c, and*
- *the map* Boc_∞ *is non-zero,*

then the Tamagawa number conjecture for $M^*(1)$ (with coefficients in $\mathcal{A}$) is true.

Proof. Conjecture 3.9 implies the existence of a Λ-basis $\mathfrak{z}_{K_\infty} \in \det_\Lambda^{-1}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S}, \mathbb{T}))$ such that the map (3.1.2) sends $\mathfrak{z}_{K_\infty}$ to c_{K_∞}. Let

$$\mathfrak{z}_K \in \det_{\mathcal{A}}^{-1}(\mathbf{R}\Gamma_\Sigma(\mathcal{O}_{K,S}, T))$$

be the image of $\mathfrak{z}_{K_\infty}$, which is an $\mathcal{A}$-basis. By Proposition 2.17, it is sufficient to prove that

$$\Theta_x(\mathfrak{z}_K) = \widetilde{\eta}_K,$$

where Θ_x is the map in the diagram (4.2.1). Since Boc_∞ is non-zero by assumption, this is implied by Conjecture 4.7 (as in the proof of Proposition 4.10). So we have completed the proof. □

Remark 4.12. Theorem 4.11 is a direct generalization of [BKS24, Th. 7.6], where a strategy for proving the Birch-Swinnerton-Dyer formula for an elliptic curve over $\mathbb{Q}$ is given.

Remark 4.13. In the $\mathbb{G}_m$ case, a natural equivariant version of Theorem 4.11 is given in [BKS17, Th. 5.2]. Since the main aim of this paper is to study a general motive, we do not give an equivariant generalization of Theorem 4.11 in this paper.

5 Heegner points

The aim of this section is to study how the theory of Heegner points fits in the general framework given in earlier sections. In particular, we study relations between Perrin-Riou's "Heegner point main conjecture" and the (non-equivariant) Iwasawa main conjecture in Conjecture 3.9.

In this section, we consider an imaginary quadratic base field K and the motive $M = h^1(E/K)(1)$ (with coefficients $R = \mathbb{Q}$, $A = \mathbb{Q}_p$ and $\mathcal{A} = \mathbb{Z}_p$), where E is an elliptic curve over $\mathbb{Q}$ such that K satisfies the Heegner hypothesis for E (i.e., every prime divisor of the conductor of E splits in K). Let $T := T_p(E)$ and $V := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} T$. In this case, note that

$$Y_K(T^*(1)) := \bigoplus_{v \in S_\infty(K)} H^0(K_v, T^*(1)) = H^0(\mathbb{C}, T^*(1)) = T^*(1)$$

and so

$$\mathrm{rank}_{\mathbb{Z}_p}(Y_K(T^*(1))) = 2,$$

i.e., the basic rank is two (see Definition 2.1). So Conjecture 2.6 suggests that there should be a canonical Euler system of rank two in this setting. For this reason, we expect that *Heegner points are related with rank two Euler systems*.

5.1 Heegner elements and Euler systems of rank two

In this subsection, we introduce "Heegner elements", which lie in the second exterior power of H^1, and relate them with special elements in Definition 2.7 (see Proposition 5.3). As an application, we construct a rank two Euler system which is related with Heegner points in the case of analytic rank one (see Theorem 5.5).

We fix a finite set S of places of K containing $\{\infty\} \cup S_p(K) \cup S_{\mathrm{ram}}(T)$. (Here $S_\infty(K) = \{\infty\}$.) We need the following lemma.

Lemma 5.1. *Assume* $\mathrm{rank}(E(K)) = 1$ *and* $\#\text{Ш}(E/K)[p^\infty] < \infty$. *Then* $H^2(\mathcal{O}_{K,S}, V) = 0$ *and there exists a canonical exact sequence*

$$0 \to \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K) \to H^1(\mathcal{O}_{K,S}, V) \to \mathbb{Q}_p \otimes_{\mathbb{Q}} \Gamma(E, \Omega^1_{E/K}) \to \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K)^* \to 0.$$

Proof. We have a canonical exact triangle

$$\mathbf{R}\Gamma_f(K, V) \to \mathbf{R}\Gamma(\mathcal{O}_{K,S}, V) \to \bigoplus_{v \in S} \mathbf{R}\Gamma_{/f}(K_v, V) \to .$$

(See [BuFl01, p. 522] for example.) If $v \notin S_p(K)$, we know that $\mathbf{R}\Gamma_{/f}(K_v, V)$ is acyclic. So we obtain a long exact sequence

$$0 \to H^1_f(K, V) \to H^1(\mathcal{O}_{K,S}, V) \to H^1_{/f}(K_p, V) \to H^2_f(K, V) \to H^2(\mathcal{O}_{K,S}, V) \to 0.$$

Since we assume the finiteness of $\text{Ш}(E/K)[p^\infty]$, we have

$$H^1_f(K, V) = \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K) \text{ and } H^2_f(K, V) = \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K)^*.$$

Also, we have $H^1_{/f}(K_p,V) = \mathbb{Q}_p \otimes_{\mathbb{Z}_p} (\varprojlim_n E(K_p)/p^n)^*$, which is canonically isomorphic to $\mathbb{Q}_p \otimes_{\mathbb{Q}} \Gamma(E,\Omega^1_{E/K})$ via the dual exponential map. Hence we have an exact sequence

$$0 \to \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K) \to H^1(\mathcal{O}_{K,S},V) \to \mathbb{Q}_p \otimes_{\mathbb{Q}} \Gamma(E,\Omega^1_{E/K}) \to \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K)^* \to H^2(\mathcal{O}_{K,S},V) \to 0. \quad (5.1.1)$$

We shall show that $H^2(\mathcal{O}_{K,S},V) = 0$. Since we assume $\mathrm{rank}(E(K)) = 1$, we see that the map $\mathbb{Q}_p \otimes_{\mathbb{Z}} E(K) \to \mathbb{Q}_p \otimes_{\mathbb{Z}_p} (\varprojlim_n E(K_p)/p^n)$ is injective. This implies that the map $\mathbb{Q}_p \otimes_{\mathbb{Q}} \Gamma(E,\Omega^1_{E/K}) \to \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K)^*$ is surjective, so we have $H^2(\mathcal{O}_{K,S},V) = 0$. This completes the proof. □

Throughout this subsection, we assume $\mathrm{rank}(E(K)) = 1$ and $\#Ш(E/K)[p^\infty] < \infty$. Then, by Lemma 5.1, we obtain a canonical isomorphism

$$\det_{\mathbb{Q}_p}(H^1(\mathcal{O}_{K,S},V)) \simeq \det_{\mathbb{Q}_p}(\mathbb{Q}_p \otimes_{\mathbb{Z}} E(K)) \otimes_{\mathbb{Q}_p} \det_{\mathbb{Q}_p}(\mathbb{Q}_p \otimes_{\mathbb{Q}} \Gamma(E,\Omega^1_{E/K})) \otimes_{\mathbb{Q}_p} \det^{-1}_{\mathbb{Q}_p}(\mathbb{Q}_p \otimes_{\mathbb{Z}} E(K)^*),$$

i.e.,

$$\bigwedge^2_{\mathbb{Q}_p} H^1(\mathcal{O}_{K,S},V) \simeq \mathbb{Q}_p \otimes_{\mathbb{Q}} \left(E(K) \otimes_{\mathbb{Z}} E(K) \otimes_{\mathbb{Z}} \bigwedge^2_{\mathbb{Q}} \Gamma(E,\Omega^1_{E/K})\right). \quad (5.1.2)$$

Now we fix a modular parametrization $\phi : X_0(N) \to E$ and let

$$y_K \in E(K)$$

be the associated Heegner point. Let c_ϕ be the Manin constant and set $u_K := \#\mathcal{O}_K^\times/2$. We fix Néron differentials ω and ω^K of $E/\mathbb{Q}$ and $E^K/\mathbb{Q}$ respectively, where E^K be the quadratic twist of E by K. Then $\{\omega, \omega^K\}$ is a $\mathbb{Q}$-basis of $\Gamma(E,\Omega^1_{E/K})$. Let $\mathrm{Eul}_S \in \mathbb{Q}^\times$ be the product of Euler factors at $v \in S \setminus \{\infty\}$ satisfying

$$\mathrm{Eul}_S \cdot L^*(E/K,1) = L^*_S(E/K,1).$$

We now define the Heegner element.

Definition 5.2. Assume $\mathrm{rank}(E(K)) = 1$ and $\#Ш(E/K)[p^\infty] < \infty$. Then we define the *Heegner element* for E/K

$$z_K^{\mathrm{Hg}} \in \bigwedge^2_{\mathbb{Q}_p} H^1(\mathcal{O}_{K,S},V)$$

as the element corresponding to

$$\mathrm{Eul}_S \cdot (u_K c_\phi)^{-2} \otimes y_K \otimes y_K \otimes (\omega \wedge \omega^K) \in \mathbb{Q}_p \otimes_{\mathbb{Q}} \left(E(K) \otimes_{\mathbb{Z}} E(K) \otimes_{\mathbb{Z}} \bigwedge^2_{\mathbb{Q}} \Gamma(E,\Omega^1_{E/K})\right)$$

under the isomorphism (5.1.2).

We choose a $\mathbb{Z}_p$-basis

$$b_K \in \bigwedge^2_{\mathbb{Z}_p} Y_K(T^*(1))^* = \bigwedge^2_{\mathbb{Z}_p} T(-1) \quad (5.1.3)$$

in the following way. Let $\Omega_{E/K}$ be the Néron period of E/K. We first take a $\mathbb{Z}_{(p)}$-basis

$$\gamma^* \in \bigwedge^2_{\mathbb{Z}_{(p)}} H_1(E(\mathbb{C}),\mathbb{Z}_{(p)})^*$$

so that the period map

$$\mathbb{R}\otimes_{\mathbb{Q}}\bigwedge\nolimits_{\mathbb{Q}}^{2}\Gamma(E,\Omega^1_{E/K}) \xrightarrow{\sim} \mathbb{R}\otimes_{\mathbb{Q}}\bigwedge\nolimits_{\mathbb{Q}}^{2}H_1(E(\mathbb{C}),\mathbb{Q})^* \tag{5.1.4}$$

sends $\omega\wedge\omega^K$ to

$$\frac{1}{\sqrt{|D_K|}}\Omega_{E/K}\cdot\gamma^*.$$

(This is possible, since $\frac{1}{\sqrt{|D_K|}}\Omega_{E/K}$ coincides with $\Omega_E^+\Omega_{E^K}^+$ up to 2-power, where Ω_E^+ and $\Omega_{E^K}^+$ denote the real periods of $E/\mathbb{Q}$ and $E^K/\mathbb{Q}$ respectively. See [GrZa86, p. 312].) We then define b_K to be the image of γ^* under the comparison isomorphism

$$\bigwedge\nolimits_{\mathbb{Z}_p}^{2}H_1(E(\mathbb{C}),\mathbb{Z}_p)^*\simeq\bigwedge\nolimits_{\mathbb{Z}_p}^{2}T(-1). \tag{5.1.5}$$

Let

$$\eta_K=\eta_{K/K,S,\emptyset}(T)\in\mathbb{C}_p\otimes_{\mathbb{Z}_p}\bigwedge\nolimits_{\mathbb{Z}_p}^{2}H^1(\mathcal{O}_{K,S},T)$$

be the special element defined by using b_K (see Definition 2.7). A relation between z_K^{Hg} and η_K is given as follows.

Proposition 5.3. *Assume* $\mathrm{ord}_{s=1}L(E/K,s)=1$ *(which implies* $\mathrm{rank}(E(K))=1$ *and* $\#\text{Ш}(E/K)<\infty$ *by the well-known theorem of Gross-Zagier-Kolyvagin). Then we have*

$$z_K^{\mathrm{Hg}}=\eta_K.$$

In particular, η_K *lies in* $\bigwedge_{\mathbb{Q}_p}^2H^1(\mathcal{O}_{K,S},V)$.

Proof. Let

$$\lambda_{T,K}:\mathbb{C}_p\otimes_{\mathbb{Z}_p}\bigwedge\nolimits_{\mathbb{Z}_p}^{2}H^1(\mathcal{O}_{K,S},T)\xrightarrow{\sim}\mathbb{C}_p\otimes_{\mathbb{Z}_p}\bigwedge\nolimits_{\mathbb{Z}_p}^{2}Y_K(T^*(1))^*$$

be the period-regulator isomorphism defined in §2.4. By definition, the special element η_K is characterized by

$$\lambda_{T,K}(\eta_K)=L_S^*(E/K,1)\cdot b_K.$$

So it is sufficient to show that

$$\lambda_{T,K}(z_K^{\mathrm{Hg}})=L_S^*(E/K,1)\cdot b_K.$$

Under the assumption, one checks that $\lambda_{T,K}$ coincides with the following composition map:

$$\begin{aligned}\mathbb{C}_p\otimes_{\mathbb{Z}_p}\bigwedge\nolimits_{\mathbb{Z}_p}^{2}H^1(\mathcal{O}_{K,S},T) &\overset{(5.1.2)}{\simeq} \mathbb{C}_p\otimes_{\mathbb{Q}}\left(E(K)\otimes_{\mathbb{Z}}E(K)\otimes_{\mathbb{Z}}\bigwedge\nolimits_{\mathbb{Q}}^{2}\Gamma(E,\Omega^1_{E/K})\right)\\ &\simeq \mathbb{C}_p\otimes_{\mathbb{Q}}\bigwedge\nolimits_{\mathbb{Q}}^{2}\Gamma(E,\Omega^1_{E/K})\\ &\overset{(5.1.4)}{\simeq} \mathbb{C}_p\otimes_{\mathbb{Q}}\bigwedge\nolimits_{\mathbb{Q}}^{2}H_1(E(\mathbb{C}),\mathbb{Q})^*\\ &\overset{(5.1.5)}{\simeq} \mathbb{C}_p\otimes_{\mathbb{Z}_p}\bigwedge\nolimits_{\mathbb{Z}_p}^{2}T(-1)=\mathbb{C}_p\otimes_{\mathbb{Z}_p}\bigwedge\nolimits_{\mathbb{Z}_p}^{2}Y_K(T^*(1))^*,\end{aligned}$$

where the second isomorphism is induced by the Néron-Tate height pairing

$$\langle -,-\rangle_{\infty,K} : E(K)\times E(K)\to\mathbb{R}.$$

By the definition of z_K^{Hg}, we have

$$\lambda_{T,K}(z_K^{\mathrm{Hg}}) = \mathrm{Eul}_S\cdot(u_K c_\phi)^{-2}\cdot\langle y_K,y_K\rangle_{\infty,K}\cdot\frac{1}{\sqrt{|D_K|}}\cdot\Omega_{E/K}\cdot b_K.$$

By the Gross-Zagier formula [GrZa86], we know that

$$L'(E/K,1) = (u_K c_\phi)^{-2}\cdot\langle y_K,y_K\rangle_{\infty,K}\cdot\frac{1}{\sqrt{|D_K|}}\cdot\Omega_{E/K},$$

so we have

$$\lambda_{T,K}(z_K^{\mathrm{Hg}}) = \mathrm{Eul}_S\cdot L'(E/K,1)\cdot b_K = L_S^*(E/K,1)\cdot b_K.$$

This proves the proposition. □

Let $\mathrm{Tam}(E/K)$ be the product of Tamagawa factors and $R_{E/K}$ the Néron-Tate regulator for E/K. Recall that the p-part of the Birch-Swinnerton-Dyer formula for E/K predicts the equality (in $\mathbb{C}_p$)

$$\mathbb{Z}_p\cdot L^*(E/K,1) = \mathbb{Z}_p\cdot\frac{\#\text{Ш}(E/K)\cdot\mathrm{Tam}(E/K)}{\#E(K)_{\mathrm{tors}}^2}\cdot\frac{1}{\sqrt{|D_K|}}\Omega_{E/K}\cdot R_{E/K}.$$

Proposition 5.4. *Assume* $\mathrm{ord}_{s=1}L(E/K,s)=1$ *and* $E(K)[p]=0$. *Then the p-part of the Birch-Swinnerton-Dyer formula for E/K holds if and only if we have an equality of* $\mathbb{Z}_p$*-modules*

$$\mathbb{Z}_p\cdot z_K^{\mathrm{Hg}} = \#H^2(\mathcal{O}_{K,S},T)\cdot{\bigwedge}_{\mathbb{Z}_p}^2 H^1(\mathcal{O}_{K,S},T).$$

In particular, the p-part of the Birch-Swinnerton-Dyer formula for E/K implies the "integrality" of z_K^{Hg}*:*

$$z_K^{\mathrm{Hg}}\in{\bigwedge}_{\mathbb{Z}_p}^2 H^1(\mathcal{O}_{K,S},T).$$

Proof. Note that the assumption $E(K)[p]=0$ implies that $H^1(\mathcal{O}_{K,S},T)$ is $\mathbb{Z}_p$-free. Since the p-part of the Birch-Swinnerton-Dyer formula for E/K is equivalent to the Tamagawa number conjecture for $h^1(E/K)(1)$ (with coefficients in $\mathbb{Z}_p$), the proposition follows immediately from Propositions 2.17 and 5.3. □

The following result gives a connection between Heegner points and rank two Euler systems.

Theorem 5.5. *Assume* $\mathrm{ord}_{s=1}L(E/K,s)=1$, $E(K)[p]=0$ *and the p-part of the Birch-Swinnerton-Dyer formula for E/K holds. Then for any abelian p-extension* $\mathcal{K}/K$ *there exists a rank two Euler system* $c\in\mathrm{ES}_2(T,\mathcal{K})$ *such that*

$$c_K = z_K^{\mathrm{Hg}}.$$

Proof. Note that we have a canonical isomorphism

$$\Theta_{T,K} : \det^{-1}_{\mathbb{Z}_p}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},T)) \simeq \#H^2(\mathcal{O}_{K,S},T) \cdot {\bigwedge}^2_{\mathbb{Z}_p} H^1(\mathcal{O}_{K,S},T).$$

Since we assume the p-part of the Birch-Swinnerton-Dyer formula, Proposition 5.4 implies that there is a basis $\mathfrak{z}_K^{\mathrm{Hg}} \in \det^{-1}_{\mathbb{Z}_p}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},T))$ such that $\Theta_{T,K}(\mathfrak{z}_K^{\mathrm{Hg}}) = z_K^{\mathrm{Hg}}$. Let $\mathrm{VS}(T,\mathcal{K})$ be the module mentioned in Remark 3.8, which is defined to be a certain inverse limit $\varprojlim_{F\in\Omega(\mathcal{K})} \det^{-1}_{\mathbb{Z}_p[\mathcal{G}_F]}(\mathbf{R}\Gamma(\mathcal{O}_{F,S(F)},T))$ with surjective transition maps. Then we have a commutative diagram

$$\begin{array}{ccc} \mathrm{VS}(T,\mathcal{K}) & \xrightarrow{\Theta_{T,\mathcal{K}}} & \mathrm{ES}_2(T,\mathcal{K}) \\ \downarrow & & \downarrow \\ \det^{-1}_{\mathbb{Z}_p}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},T)) & \xrightarrow[\Theta_{T,K}]{} & {\bigwedge}^2_{\mathbb{Z}_p} H^1(\mathcal{O}_{K,S},T), \end{array}$$

where the left vertical surjection is the natural projection map and the right vertical arrow sends c to c_K. Take a lift

$$\mathfrak{z} \in \mathrm{VS}(T,\mathcal{K})$$

of $\mathfrak{z}_K^{\mathrm{Hg}} \in \det^{-1}_{\mathbb{Z}_p}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},T))$ and put

$$c := \Theta_{T,\mathcal{K}}(\mathfrak{z}) \in \mathrm{ES}_2(T,\mathcal{K}).$$

Then this Euler system has the desired property. □

Remark 5.6. The Euler system constructed in Theorem 5.5 is not canonical. A canonical rank two Euler system should be constructed *directly* from Heegner points over ring class fields (so that it satisfies the properties (i) and (ii) in Conjecture 2.6).

5.2 The Heegner point main conjecture

In this subsection, we relate our formulation of the Iwasawa main conjecture (Conjecture 3.9) with the Heegner point main conjecture formulated by Perrin-Riou [Per87] (which has been studied in many works including [Ber95], [How04], [Wan21], [Cas17] and [BCK21]).

In the following, *we assume E has good ordinary reduction at p and $K/\mathbb{Q}$ is unramified at p.*

5.2.1 Formulation of the Heegner point main conjecture

We first review the formulation of the Heegner point main conjecture. Let K_∞/K be the anticyclotomic $\mathbb{Z}_p$-extension and K_n its n-th layer. We set

$$\Gamma_n := \mathrm{Gal}(K_n/K),\ \Gamma := \mathrm{Gal}(K_\infty/K) \text{ and } \Lambda := \mathbb{Z}_p[[\Gamma]] \simeq \varprojlim_n \mathbb{Z}_p[\Gamma_n].$$

We also set

$$\mathbb{T} := \varprojlim_n \mathrm{Ind}_{K_n/K}(T),\ W := V/T \text{ and } \mathbb{W} := \varinjlim_n \mathrm{Ind}_{K_n/K}(W).$$

For $X \in \{\mathbb{T}, \mathbb{W}\}$, we define a Λ-adic Selmer group $\mathrm{Sel}(X)$ as follows. Note first that, since E has good ordinary reduction at each $v \in S_p(K)$, we have a natural filtration $F^+X \subset X$ (as G_{K_v}-modules). We set $F^-X := X/F^+X$ and define

$$\mathrm{Sel}(X) := \ker\left(H^1(\mathcal{O}_{K,S}, X) \to \bigoplus_{v \in S_p(K)} H^1(K_v, F^-X) \oplus \bigoplus_{v \in S \setminus (S_\infty(K) \cup S_p(K))} H^1(K_v^{\mathrm{ur}}, X)\right),$$

where K_v^{ur} denotes the maximal unramified extension of K_v.

We set

$$\mathrm{Sel}(\mathbb{W})^\vee := \mathrm{Hom}_{\mathbb{Z}_p}(\mathrm{Sel}(\mathbb{W}), \mathbb{Q}_p/\mathbb{Z}_p) \text{ and } Ш_\infty := (\mathrm{Sel}(\mathbb{W})^\vee)_{\mathrm{tors}}.$$

Let

$$y_\infty \in \mathrm{Sel}(\mathbb{T})$$

be the Λ-adic Heegner class, which is denoted by $\mathbf{z}_f$ in [Cas17, §3.1]. By [Cor02] and [CoVa07], we know that y_∞ is non-torsion.

Lastly, let $\iota : \Lambda \to \Lambda;\ a \mapsto a^\iota$ denote the involution induced by $\Gamma \to \Gamma;\ \gamma \mapsto \gamma^{-1}$.

Conjecture 5.7 (The Heegner point main conjecture)**.** *We have*

$$\mathrm{char}_\Lambda(\mathrm{Sel}(\mathbb{T})/\Lambda \cdot y_\infty) \cdot \mathrm{char}_\Lambda(\mathrm{Sel}(\mathbb{T})/\Lambda \cdot y_\infty)^\iota = \mathrm{char}_\Lambda(Ш_\infty).$$

Remark 5.8. Building on works by many people (including Howard [How06] and Zhang [Zha14]), Burungale-Castella-Kim has recently proved Conjecture 5.7 under mild hypotheses (see [BCK21, Th. A]).

5.2.2 Selmer complexes

For later purpose, we give another formulation of the Heegner point main conjecture by using Nekovář's Selmer complexes [Nek06].

Let

$$\widetilde{\mathbf{R}\Gamma}_f(\mathbb{T}) := \widetilde{\mathbf{R}\Gamma}_{f,\mathrm{Iw}}(K_\infty/K, T) \text{ and } \widetilde{\mathbf{R}\Gamma}_f(\mathbb{W}) := \widetilde{\mathbf{R}\Gamma}_f(K_S/K_\infty, W)$$

be Selmer complexes defined in [Nek06, (8.8.5)]. We write $\widetilde{H}^i_f(-)$ for $H^i(\widetilde{\mathbf{R}\Gamma}_f(-))$.

Lemma 5.9.

(i) Let $X \in \{\mathbb{T}, \mathbb{W}\}$. There is a canonical exact triangle

$$\widetilde{\mathbf{R}\Gamma}_f(X) \to \mathbf{R}\Gamma(\mathcal{O}_{K,S}, X) \to \bigoplus_{v \in S_p(K)} \mathbf{R}\Gamma(K_v, F^-X) \oplus \bigoplus_{v \in S \setminus (\{\infty\} \cup S_p(K))} U_v^-(X),$$

where $U_v^-(X)$ is defined by

$$U_v^-(X) := \mathrm{Cone}\left(\mathbf{R}\Gamma(K_v^{\mathrm{ur}}/K_v, X^{G_{K_v^{\mathrm{ur}}}}) \to \mathbf{R}\Gamma(K_v, X)\right).$$

(ii) We have

$$\widetilde{H}^1_f(\mathbb{T}) = \mathrm{Sel}(\mathbb{T}).$$

(iii) There is a canonical surjection

$$\widetilde{H}^1_f(\mathbb{W}) \twoheadrightarrow \mathrm{Sel}(\mathbb{W})$$

with finite kernel.

(iv) There is a canonical isomorphism

$$\widetilde{H}^i_f(\mathbb{T}) \simeq (\widetilde{H}^{3-i}_f(\mathbb{W})^{\iota})^{\vee},$$

where we write $(-)^{\iota}$ for the module on which Λ acts via the involution ι and $(-)^{\vee}$ for the Pontryagin dual.

(v) There is a canonical map

$$\widetilde{H}^2_f(\mathbb{T}) \to \mathrm{Hom}_{\Lambda}(\widetilde{H}^1_f(\mathbb{T}), \Lambda)^{\iota}$$

such that

- *the cokernel is pseudo-null,*
- *the kernel is pseudo-isomorphic to $Ш^{\iota}_{\infty}$.*

Proof. Claim (i) follows from [Nek06, (6.1.3.2)]. Claim (ii) follows from claim (i) by noting that $H^0(K_v, F^-\mathbb{T}) = 0$ for $v \in S_p(K)$ and $H^0(U^-_v(\mathbb{T})) = 0$ for $v \notin \{\infty\} \cup S_p(K)$. Claim (iii) again follows from claim (i) by noting that $H^0(K_v, F^-\mathbb{W})$ is finite for $v \in S_p(K)$ and $H^0(U^-_v(\mathbb{W})) = 0$ for $v \notin \{\infty\} \cup S_p(K)$. Claim (iv) follows from the duality [Nek06, (8.9.6.2)] and the natural identification $W = T^{\vee}(1)$ induced by the Weil pairing.

We prove claim (v). By [Nek06, Th. 8.9.9], we have a short exact sequence

$$0 \to \widetilde{H}^2_f(\mathbb{T})_{\mathrm{tors}} \to \widetilde{H}^2_f(\mathbb{T}) \to \mathrm{Hom}_{\Lambda}(\widetilde{H}^1_f(\mathbb{T}), \Lambda)^{\iota} \to 0$$

modulo pseudo-null. By (iii) and (iv), we see that $\widetilde{H}^2_f(\mathbb{T})_{\mathrm{tors}}$ is pseudo-isomorphic to $Ш^{\iota}_{\infty}$. This proves the claim. □

Remark 5.10. Combining Lemma 5.9(ii), (iii), (iv) and (v), we obtain a canonical isomorphism

$$Q(\Lambda) \otimes_{\Lambda} \mathrm{Sel}(\mathbb{T}) \simeq \mathrm{Hom}_{\Lambda}(\mathrm{Sel}(\mathbb{W})^{\vee}, Q(\Lambda)),$$

where $Q(\Lambda)$ denotes the quotient field of Λ. In particular, the Λ-ranks of $\mathrm{Sel}(\mathbb{T})$ and $\mathrm{Sel}(\mathbb{W})^{\vee}$ are the same.

We also know the following.

Proposition 5.11. $\mathrm{Sel}(\mathbb{T})$ *is a free Λ-module of rank one.*

Proof. This is proved in [Nek01, Lem. 2.3] (see also [Nek06, Prop. 9.6.7.5]). □

We define a canonical isomorphism

$$Q(\Lambda)\otimes_\Lambda \det_\Lambda^{-1}(\widetilde{\mathbf{R}\Gamma}_f(\mathbb{T})) \simeq Q(\Lambda)\otimes_\Lambda \mathrm{Sel}(\mathbb{T})\otimes_\Lambda \mathrm{Sel}(\mathbb{T})^\iota \tag{5.2.1}$$

by the composition

$$\begin{aligned} Q(\Lambda)\otimes_\Lambda \det_\Lambda^{-1}(\widetilde{\mathbf{R}\Gamma}_f(\mathbb{T})) &\simeq Q(\Lambda)\otimes_\Lambda \det_\Lambda(\widetilde{H}^1_f(\mathbb{T}))\otimes_\Lambda \det_\Lambda^{-1}(\widetilde{H}^2_f(\mathbb{T})) \\ &\simeq Q(\Lambda)\otimes_\Lambda \det_\Lambda(\mathrm{Sel}(\mathbb{T}))\otimes_\Lambda \det_\Lambda(\mathrm{Sel}(\mathbb{T})^\iota) \\ &\simeq Q(\Lambda)\otimes_\Lambda \mathrm{Sel}(\mathbb{T})\otimes_\Lambda \mathrm{Sel}(\mathbb{T})^\iota, \end{aligned}$$

where the first isomorphism follows by noting that $\widetilde{H}^0_f(\mathbb{T})=0$ and $\widetilde{H}^3_f(\mathbb{T})=(\widetilde{H}^0_f(\mathbb{W})^\iota)^\vee$ is finite (which follows from the fact that $E(K_\infty)[p^\infty]$ is finite, see [Nek01, Lem. 2.1(v)]), the second by Lemma 5.9(ii) and (v), and the last by Proposition 5.11.

Proposition 5.12. *Conjecture 5.7 holds if and only if there is a Λ-basis*

$$\widetilde{\mathfrak{z}}_\infty \in \det_\Lambda^{-1}(\widetilde{\mathbf{R}\Gamma}_f(\mathbb{T}))$$

such that the map (5.2.1) sends $\widetilde{\mathfrak{z}}_\infty$ to $y_\infty\otimes y_\infty$.

Proof. By Lemma 5.9(ii) and (v), we have a canonical isomorphism

$$\det_\Lambda^{-1}(\widetilde{H}^2_f(\mathbb{T})) \simeq \det_\Lambda^{-1}(Ш^\iota_\infty)\otimes_\Lambda \det_\Lambda(\mathrm{Sel}(\mathbb{T})^\iota).$$

Since we have $\det_\Lambda^{-1}(Ш^\iota_\infty)\simeq \mathrm{char}_\Lambda(Ш_\infty)^\iota$ (see the argument in the proof of Lemma 3.11(ii)), we see that the image of $\det_\Lambda^{-1}(\widetilde{\mathbf{R}\Gamma}_f(\mathbb{T}))$ under the map (5.2.1) coincides with

$$\mathrm{char}_\Lambda(Ш_\infty)^\iota\cdot \mathrm{Sel}(\mathbb{T})\otimes_\Lambda \mathrm{Sel}(\mathbb{T})^\iota.$$

The claim follows easily from this. □

5.2.3 Λ-adic Heegner elements

To simplify the notation, we set

$$\mathbb{H}^i := H^i(\mathcal{O}_{K,S},\mathbb{T}).$$

We shall define a "Λ-adic Heegner element"

$$z^{\mathrm{Hg}}_\infty \in Q(\Lambda)\otimes_\Lambda {\bigcap}^2_\Lambda \mathbb{H}^1$$

and compare the Iwasawa main conjecture (Conjecture 3.9) for z^{Hg}_∞ with the Heegner point main conjecture.

We fix an isomorphism

$$\det_\Lambda\left(\bigoplus_{v\in S_p(K)} \mathbf{R}\Gamma(K_v,F^-\mathbb{T})\oplus \bigoplus_{v\in S\setminus(\{\infty\}\cup S_p(K))} U^-_v(\mathbb{T})\right)\simeq\Lambda. \tag{5.2.2}$$

Then the exact triangle in Lemma 5.9(i) induces an isomorphism

$$\det_\Lambda^{-1}(\widetilde{\mathbf{R}\Gamma}_f(\mathbb{T}))\simeq \det_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},\mathbb{T})). \tag{5.2.3}$$

Definition 5.13. Assume Hypothesis 3.1 (i.e., $\mathbb{H}^2$ is Λ-torsion). We define the Λ*-adic Heegner element*

$$z_\infty^{\mathrm{Hg}} \in Q(\Lambda) \otimes_\Lambda \bigcap\nolimits_\Lambda^2 \mathbb{H}^1$$

as the image of

$$y_\infty \otimes y_\infty \in \mathrm{Sel}(\mathbb{T}) \otimes_\Lambda \mathrm{Sel}(\mathbb{T})^\iota$$

under the map

$$\begin{aligned} \mathrm{Sel}(\mathbb{T}) \otimes_\Lambda \mathrm{Sel}(\mathbb{T})^\iota &\overset{(5.2.1)}{\to} Q(\Lambda) \otimes_\Lambda \det_\Lambda^{-1}(\widetilde{\mathbf{R}\Gamma}_f(\mathbb{T})) \\ &\overset{(5.2.3)}{\simeq} Q(\Lambda) \otimes_\Lambda \det_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S}, \mathbb{T})) \\ &\overset{(3.1.2)}{\simeq} Q(\Lambda) \otimes_\Lambda \bigcap\nolimits_\Lambda^2 \mathbb{H}^1. \end{aligned}$$

Remark 5.14. The element z_∞^{Hg} is canonical up to $\Lambda^\times$ (it depends on the choice of an isomorphism (5.2.2)).

Remark 5.15. Note that Hypothesis 3.1 is equivalent to

$$H^2(\mathcal{O}_{K,S}, \mathbb{W}) = \varinjlim_n H^2(\mathcal{O}_{K_n,S}, E[p^\infty]) = 0.$$

(See [Per95, Prop. 1.3.2] or [Nek06, Lem. 9.1.5].) The vanishing of $H^2(\mathcal{O}_{K,S}, \mathbb{W})$ is proved by Bertolini [Ber01, Th. 5.4] under mild assumptions. Thus Hypothesis 3.1 is known to hold under mild assumptions.

The following is our formulation of the non-equivariant Iwasawa main conjecture (Conjecture 3.9) in the present setting.

Conjecture 5.16 (The Iwasawa main conjecture for z_∞^{Hg})**.** *Assume Hypothesis 3.1. Then we have* $z_\infty^{\mathrm{Hg}} \in \bigcap_\Lambda^2 \mathbb{H}^1$ *and*

$$\mathrm{char}_\Lambda\left(\bigcap\nolimits_\Lambda^2 \mathbb{H}^1 / \Lambda \cdot z_\infty^{\mathrm{Hg}}\right) = \mathrm{char}_\Lambda(\mathbb{H}^2).$$

Theorem 5.17. *Assume Hypothesis 3.1. Then the Heegner point main conjecture (Conjecture 5.7) is equivalent to Conjecture 5.16.*

Proof. This follows immediately from Propositions 5.12 and 3.10. □

The following is an Iwasawa theoretic analogue of Theorem 5.5.

Theorem 5.18. *Assume Hypothesis 3.1 and the Heegner point main conjecture (Conjecture 5.7). Then, for any abelian p-extension $\mathcal{K}/K$ containing K_∞, there exists a rank two Euler system $c \in \mathrm{ES}_2(T, \mathcal{K})$ such that*

$$\varprojlim_n c_{K_n} = z_\infty^{\mathrm{Hg}} \text{ in } \varprojlim_n \bigcap\nolimits_{\mathbb{Z}_p[\Gamma_n]}^2 H^1(\mathcal{O}_{K_n,S}, T) \simeq \bigcap\nolimits_\Lambda^2 \mathbb{H}^1.$$

Proof. The assumed validity of the Heegner point main conjecture implies the existence of a Λ-basis

$$\mathfrak{z}_\infty^{\mathrm{Hg}} \in \det_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S}, \mathbb{T}))$$

such that the map

$$\Theta_{T,K_\infty} : \det_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},\mathbb{T})) \to \bigcap_\Lambda^2 \mathbb{H}^1$$

induced by (3.1.2) sends $\mathfrak{z}_\infty^{\mathrm{Hg}}$ to z_∞^{Hg}. Similarly to the proof of Theorem 5.5, we have a commutative diagram

$$\begin{array}{ccc} \mathrm{VS}(T,\mathcal{K}) & \xrightarrow{\Theta_{T,\mathcal{K}}} & \mathrm{ES}_2(T,\mathcal{K}) \\ \downarrow & & \downarrow \\ \det_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},\mathbb{T})) & \xrightarrow[\Theta_{T,K_\infty}]{} & \bigcap_\Lambda^2 \mathbb{H}^1, \end{array}$$

where the right vertical arrow sends c to $\varprojlim_n c_{K_n}$. Take a lift

$$\mathfrak{z} \in \mathrm{VS}(T,\mathcal{K})$$

of $\mathfrak{z}_\infty^{\mathrm{Hg}} \in \det_\Lambda^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},\mathbb{T}))$ and put

$$c := \Theta_{T,\mathcal{K}}(\mathfrak{z}) \in \mathrm{ES}_2(T,\mathcal{K}).$$

Then this Euler system has the desired property. □

Remark 5.19. We expect that the element z_∞^{Hg} coincides with $\varprojlim_n \eta_{K_n}$ up to normalization, where

$$\eta_{K_n} = \eta_{K_n/K,S,\emptyset}(T) \in \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathbb{Z}_p[\Gamma_n]}^2 H^1(\mathcal{O}_{K_n,S},T)$$

is the special element in Definition 2.7. Since $\varprojlim_n \eta_{K_n}$ is defined even in the supersingular and bad reduction cases, we expect there is a construction of z_∞^{Hg} without assuming E has good ordinary reduction at p.

5.3 Derivatives of Heegner elements

In this subsection, we study Conjecture 4.7 for the Λ-adic Heegner element z_∞^{Hg} in Definition 5.13. We keep assuming that E has good ordinary reduction at p. *In this subsection, we always assume Hypothesis 3.1 (i.e., $\mathbb{H}^2$ is Λ-torsion).*

We set $e := \dim_{\mathbb{Q}_p}(H^2(\mathcal{O}_{K,S},V))$. We also set

$$I_n := \ker(\mathbb{Z}_p[\Gamma_n] \twoheadrightarrow \mathbb{Z}_p) \text{ and } I := \ker(\Lambda \twoheadrightarrow \mathbb{Z}_p) \simeq \varprojlim_n I_n.$$

Let

$$\begin{aligned} \iota_n := \iota_{K_n} : \bigwedge\nolimits_{\mathbb{Z}_p}^2 H^1(\mathcal{O}_{K,S},T) \otimes_{\mathbb{Z}_p} I_n^e/I_n^{e+1} & \hookrightarrow \bigcap\nolimits_{\mathbb{Z}_p[\Gamma_n]}^2 H^1(\mathcal{O}_{K_n,S},T) \otimes_{\mathbb{Z}_p} I_n^e/I_n^{e+1} \\ & \subset \bigcap\nolimits_{\mathbb{Z}_p[\Gamma_n]}^2 H^1(\mathcal{O}_{K_n,S},T) \otimes_{\mathbb{Z}_p} \mathbb{Z}_p[\Gamma_n]/I_n^{e+1} \end{aligned}$$

be the canonical injection in §4.2.

Proposition 5.20. *Assume the Heegner point main conjecture (Conjecture 5.7). Then there exists the Darmon derivative of* z_∞^{Hg}*:*

$$\kappa_\infty^{\mathrm{Hg}} = \varprojlim_n \kappa_n^{\mathrm{Hg}} \in \bigwedge^2_{\mathbb{Z}_p} H^1(\mathcal{O}_{K,S},T) \otimes_{\mathbb{Z}_p} \varprojlim_n I_n^e/I_n^{e+1} = \bigwedge^2_{\mathbb{Z}_p} H^1(\mathcal{O}_{K,S},T) \otimes_{\mathbb{Z}_p} I^e/I^{e+1},$$

i.e., the unique element satisfying

$$\iota_n(\kappa_n^{\mathrm{Hg}}) = \sum_{\sigma\in\Gamma_n} \sigma z_n^{\mathrm{Hg}} \otimes \sigma^{-1} \text{ in } \bigcap^2_{\mathbb{Z}_p[\Gamma_n]} H^1(\mathcal{O}_{K_n,S},T) \otimes_{\mathbb{Z}_p} \mathbb{Z}_p[\Gamma_n]/I_n^{e+1}$$

for every n. (Note that Theorem 5.17 implies that z_∞^{Hg} lies in $\bigcap^2_\Lambda \mathbb{H}^1 = \varprojlim_n \bigcap^2_{\mathbb{Z}_p[\Gamma_n]} H^1(\mathcal{O}_{K_n,S},T)$, and $z_n^{\mathrm{Hg}} \in \bigcap^2_{\mathbb{Z}_p[\Gamma_n]} H^1(\mathcal{O}_{K_n,S},T)$ denotes the element such that $z_\infty^{\mathrm{Hg}} = \varprojlim_n z_n^{\mathrm{Hg}}$.)

Proof. This follows from Theorem 5.17 and Proposition 4.6. □

Remark 5.21. Since z_∞^{Hg} is canonical up to $\Lambda^\times$, its Darmon derivative $\kappa_\infty^{\mathrm{Hg}}$ is canonical up to $\mathbb{Z}_p^\times$.

In the following, we assume the following hypothesis. (Recall that $E^K/\mathbb{Q}$ denotes the quadratic twist of $E/\mathbb{Q}$ by K.)

Hypothesis 5.22.

(i) $E(K)[p] = 0$;

(ii) $r^+ := \mathrm{rank}(E(\mathbb{Q})) > 0$ and $r^- := \mathrm{rank}(E^K(\mathbb{Q})) > 0$ (in particular, $\mathrm{rank}(E(K)) \geq 2$);

(iii) $\#\text{Ш}(E/K)[p^\infty] < \infty$.

Lemma 5.23. *Assume Hypothesis 5.22.*

(i) We have canonical isomorphisms

$$H^1(\mathcal{O}_{K,S},V) \simeq \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K) \simeq \mathbb{Q}_p \otimes_{\mathbb{Z}} (E(\mathbb{Q}) \oplus E^K(\mathbb{Q})).$$

(ii) We have a canonical exact sequence

$$0 \to \mathbb{Q}_p \otimes_{\mathbb{Q}} \Gamma(E,\Omega^1_{E/K}) \to \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K)^* \to H^2(\mathcal{O}_{K,S},V) \to 0. \tag{5.3.1}$$

In particular, we have

$$e := \dim_{\mathbb{Q}_p}(H^2(\mathcal{O}_{K,S},V)) = r^+ + r^- - 2.$$

Proof. By (5.1.1), it is sufficient to prove that the map $\mathbb{Q}_p \otimes_{\mathbb{Q}} \Gamma(E,\Omega^1_{E/K}) \to \mathbb{Q}_p \otimes_{\mathbb{Z}} E(K)^*$ is injective, or equivalently, the localization map $\mathbb{Q}_p \otimes_{\mathbb{Z}} E(K) \to \mathbb{Q}_p \otimes_{\mathbb{Z}_p} (\varprojlim_n E(K_p)/p^n)$ is surjective. Since we have

$$\mathbb{Q}_p \otimes_{\mathbb{Z}} E(K) = \mathbb{Q}_p \otimes_{\mathbb{Z}} (E(\mathbb{Q}) \oplus E^K(\mathbb{Q}))$$

and

$$\mathbb{Q}_p \otimes_{\mathbb{Z}_p} (\varprojlim_n E(K_p)/p^n) = \mathbb{Q}_p \otimes_{\mathbb{Z}_p} (\varprojlim_n E(\mathbb{Q}_p)/p^n \oplus \varprojlim_n E^K(\mathbb{Q}_p)/p^n),$$

it is sufficient to prove the surjectivity of $\mathbb{Q}_p \otimes_{\mathbb{Z}} A(\mathbb{Q}) \to \mathbb{Q}_p \otimes_{\mathbb{Z}_p} (\varprojlim_n A(\mathbb{Q}_p)/p^n)$ for $A \in \{E, E^K\}$. However, this is true by Hypothesis 5.22(ii). □

Remark 5.24. The Heegner hypothesis implies that $\mathrm{ord}_{s=1}L(E/K,s)$ is odd, and so by the parity conjecture [Nek01] we know that $\mathrm{rank}(E(K))$ is also odd. So by Lemma 5.23(ii) we have $e>0$ under Hypothesis 5.22.

We shall define a canonical "anticyclotomic Bockstein regulator"

$$R^{\mathrm{Boc}}_{K_\infty} \in {\bigwedge}^2_{\mathbb{Q}_p} H^1(\mathcal{O}_{K,S},V)\otimes_{\mathbb{Z}_p} I^e/I^{e+1}$$

as follows.

We fix a $\mathbb{Z}_p$-basis $x\in \bigwedge^e_{\mathbb{Z}_p} H^2(\mathcal{O}_{K,S},T)_{\mathrm{tf}}$, and let

$$\mathrm{Boc}_\infty = \varprojlim_n \mathrm{Boc}_{T,K_n,x} : {\bigwedge}^{e+2}_{\mathbb{Q}_p} H^1(\mathcal{O}_{K,S},V) \to {\bigwedge}^2_{\mathbb{Q}_p} H^1(\mathcal{O}_{K,S},V)\otimes_{\mathbb{Z}_p} I^e/I^{e+1}$$

be the limit of Bockstein regulator maps defined in §4.1. By Lemma 5.23, we have $\mathrm{rank}(E(K)) = r^+ + r^- = e+2$ and we fix a $\mathbb{Z}$-basis $\{P_1,\dots,P_{e+2}\}$ of $E(K)_{\mathrm{tf}}$, which is regarded as a $\mathbb{Q}_p$-basis of $H^1(\mathcal{O}_{K,S},V)$ by identifying $\mathbb{Q}_p\otimes_{\mathbb{Z}} E(K) = H^1(\mathcal{O}_{K,S},V)$. Let $\{P_1^*,\dots,P_{e+2}^*\}$ be the basis of $H^1(\mathcal{O}_{K,S},V)^*$ which is dual to $\{P_1,\dots,P_{e+2}\}$.

Recall that we fixed Néron differentials ω and ω^K of $E/\mathbb{Q}$ and $E^K/\mathbb{Q}$ respectively. Then $\{\omega,\omega^K\}$ is a $\mathbb{Q}$-basis of $\Gamma(E,\Omega^1_{E/K})$ and so we can identify $\Gamma(E,\Omega^1_{E/K}) = \mathbb{Q}^2$. By the exact sequence (5.3.1), we get an isomorphism

$$ {\bigwedge}^e_{\mathbb{Q}_p} H^2(\mathcal{O}_{K,S},V) \simeq {\bigwedge}^{e+2}_{\mathbb{Q}_p} (\mathbb{Q}_p\otimes_{\mathbb{Z}} E(K))^* = {\bigwedge}^{e+2}_{\mathbb{Q}_p} H^1(\mathcal{O}_{K,S},V)^*. \tag{5.3.2}$$

Definition 5.25. Assume Hypothesis 5.22. We define the *anticyclotomic Bockstein regulator* by

$$R^{\mathrm{Boc}}_{K_\infty} := C_x\cdot \mathrm{Boc}_\infty(P_1\wedge\cdots\wedge P_{e+2}) \in {\bigwedge}^2_{\mathbb{Q}_p} H^1(\mathcal{O}_{K,S},V)\otimes_{\mathbb{Z}_p} I^e/I^{e+1},$$

where $C_x\in\mathbb{Q}_p^\times$ is the element satisfying

$$x \overset{(5.3.2)}{\mapsto} C_x\cdot P_1^*\wedge\cdots\wedge P_{e+2}^*.$$

(One checks that $R^{\mathrm{Boc}}_{K_\infty}$ is independent of the choice of x and $\{P_1,\dots,P_{e+2}\}$.)

Recall that Conjecture 4.7 for z^{Hg}_∞ predicts the equality

$$\kappa^{\mathrm{Hg}}_\infty = \mathrm{Boc}_\infty(\widetilde{\eta}_K) \text{ in } \mathbb{C}_p\otimes_{\mathbb{Z}_p} {\bigwedge}^2_{\mathbb{Z}_p} H^1(\mathcal{O}_{K,S},T)\otimes_{\mathbb{Z}_p} I^e/I^{e+1},$$

where $\widetilde{\eta}_K = \widetilde{\eta}_{K,S,\emptyset}(T) \in \mathbb{C}_p\otimes_{\mathbb{Z}_p} \bigwedge^{e+2}_{\mathbb{Z}_p} H^1(\mathcal{O}_{K,S},T)$ is the extended special element in Definition 2.11 (with respect to $b_K\in\bigwedge^2_{\mathbb{Z}_p} Y_K(T^*(1))^*$ in (5.1.3) and $x\in\bigwedge^e_{\mathbb{Z}_p} H^2(\mathcal{O}_{K,S},T)_{\mathrm{tf}}$ fixed above).

Proposition 5.26. *Assume Hypothesis 5.22. Then Conjecture 4.7 for z^{Hg}_∞ holds if and only if we have an equality*

$$\kappa^{\mathrm{Hg}}_\infty = \frac{L^*_S(E/K,1)\sqrt{|D_K|}}{\Omega_{E/K}\cdot R_{E/K}}\cdot R^{\mathrm{Boc}}_{K_\infty} \text{ in } \mathbb{C}_p\otimes_{\mathbb{Z}_p} {\bigwedge}^2_{\mathbb{Z}_p} H^1(\mathcal{O}_{K,S},T)\otimes_{\mathbb{Z}_p} I^e/I^{e+1}.$$

Proof. It is sufficient to prove that

$$\widetilde{\eta}_K = \frac{L_S^*(E/K,1)\sqrt{|D_K|}}{\Omega_{E/K}\cdot R_{E/K}}\cdot C_x \cdot P_1 \wedge \cdots \wedge P_{e+2}.$$

Let

$$\widetilde{\lambda}_{T,K} : \mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathbb{Z}_p}^{e+2} H^1(\mathcal{O}_{K,S},T) \xrightarrow{\sim} \mathbb{C}_p \otimes_{\mathbb{Z}_p} \left(\bigwedge\nolimits_{\mathbb{Z}_p}^{e} H^2(\mathcal{O}_{K,S},T)_{\mathrm{tf}} \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathbb{Z}_p}^{2} Y_K(T^*(1))^*\right)$$

be the extended period-regulator isomorphism defined in §2.5. Since $\widetilde{\eta}_K$ is characterized by $\widetilde{\lambda}_{T,K}(\widetilde{\eta}_K) = L_S^*(E/K,1)\cdot(x\otimes b_K)$, it is sufficient to prove that

$$\widetilde{\lambda}_{T,K}(C_x\cdot P_1\wedge\cdots\wedge P_{e+2}) = \frac{\Omega_{E/K}\cdot R_{E/K}}{\sqrt{|D_K|}}\cdot(x\otimes b_K).$$

One checks that $\widetilde{\lambda}_{T,K}$ is explicitly given by the following composition map:

$$\begin{aligned}
\mathbb{C}_p \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathbb{Z}_p}^{e+2} H^1(\mathcal{O}_{K,S},T) \quad &= \quad \mathbb{C}_p \otimes_{\mathbb{Z}} \bigwedge\nolimits_{\mathbb{Z}}^{e+2} E(K)\\
&\simeq \quad \mathbb{C}_p \otimes_{\mathbb{Z}} \bigwedge\nolimits_{\mathbb{Z}}^{e+2} E(K)^*\\
&\overset{(5.3.1)}{\simeq} \quad \mathbb{C}_p \otimes_{\mathbb{Q}_p} \left(\bigwedge\nolimits_{\mathbb{Q}_p}^{e} H^2(\mathcal{O}_{K,S},V) \otimes_{\mathbb{Q}} \bigwedge\nolimits_{\mathbb{Q}}^{2} \Gamma(E,\Omega^1_{E/K})\right)\\
&\overset{(5.1.4)}{\simeq} \quad \mathbb{C}_p \otimes_{\mathbb{Q}_p} \left(\bigwedge\nolimits_{\mathbb{Q}_p}^{e} H^2(\mathcal{O}_{K,S},V) \otimes_{\mathbb{Q}} \bigwedge\nolimits_{\mathbb{Q}}^{2} H_1(E(\mathbb{C}),\mathbb{Q})^*\right)\\
&\overset{(5.1.5)}{\simeq} \quad \mathbb{C}_p \otimes_{\mathbb{Z}_p} \left(\bigwedge\nolimits_{\mathbb{Z}_p}^{e} H^2(\mathcal{O}_{K,S},T)_{\mathrm{tf}} \otimes_{\mathbb{Z}_p} \bigwedge\nolimits_{\mathbb{Z}_p}^{2} Y_K(T^*(1))^*\right),
\end{aligned}$$

where the first isomorphism is induced by the Néron-Tate height pairing. The claim follows easily from this description. □

The following result is an analogue of [BKS24, Th. 7.3]. (Recall that Eul_S is the product of Euler factors at $v\in S\setminus\{\infty\}$ satisfying $\mathrm{Eul}_S\cdot L^*(E/K,1) = L_S^*(E/K,1)$.)

Theorem 5.27. *Assume Hypothesis 5.22 and the Heegner point main conjecture (Conjecture 5.7). Then we have*

$$\mathbb{Z}_p\cdot\kappa_\infty^{\mathrm{Hg}} = \mathbb{Z}_p\cdot\mathrm{Eul}_S\cdot\#\text{Ш}(E/K)[p^\infty]\cdot\mathrm{Tam}(E/K)\cdot R_{K_\infty}^{\mathrm{Boc}} \text{ in } \bigwedge\nolimits_{\mathbb{Z}_p}^{2} H^1(\mathcal{O}_{K,S},T)\otimes_{\mathbb{Z}_p} I^e/I^{e+1}.$$

Proof. The Heegner point main conjecture implies the existence of a Λ-basis $\mathfrak{z}_\infty^{\mathrm{Hg}}$ such that the map (3.1.2) sends $\mathfrak{z}_\infty^{\mathrm{Hg}}$ to z_∞^{Hg}. Let

$$\mathfrak{z}_K^{\mathrm{Hg}} \in \det\nolimits_{\mathbb{Z}_p}^{-1}(\mathbf{R}\Gamma(\mathcal{O}_{K,S},T))$$

be the image of $\mathfrak{z}_\infty^{\mathrm{Hg}}$, which is a $\mathbb{Z}_p$-basis. By the commutative diagram (4.2.1) (for $F=K_n$), we have

$$\kappa_\infty^{\mathrm{Hg}} = \mathrm{Boc}_\infty(\Theta_x(\mathfrak{z}_K^{\mathrm{Hg}})).$$

So it is sufficient to prove that

$$\mathbb{Z}_p \cdot \Theta_x(\mathfrak{z}_K^{\mathrm{Hg}}) = \mathbb{Z}_p \cdot \mathrm{Eul}_S \cdot \#\text{Ш}(E/K)[p^\infty] \cdot \mathrm{Tam}(E/K) \cdot C_x \cdot P_1 \wedge \cdots \wedge P_{e+2}.$$

By the definition of Θ_x, we have

$$\mathbb{Z}_p \cdot \Theta_x(\mathfrak{z}_K^{\mathrm{Hg}}) = \mathbb{Z}_p \cdot \#H^2(\mathcal{O}_{K,S}, T)_{\mathrm{tors}} \cdot P_1 \wedge \cdots \wedge P_{e+2},$$

so it is sufficient to prove that

$$\mathbb{Z}_p \cdot \#H^2(\mathcal{O}_{K,S}, T)_{\mathrm{tors}} = \mathbb{Z}_p \cdot \mathrm{Eul}_S \cdot \#\text{Ш}(E/K)[p^\infty] \cdot \mathrm{Tam}(E/K) \cdot C_x.$$

It is not difficult to prove this equality, and we leave it as an exercise for the reader. □

Remark 5.28. In a forthcoming work, we prove that the formula in Theorem 5.27 implies the conjecture of Bertolini-Darmon (see [BeDa96, Conj. 4.5(1)] or [AgCa21, Conj. 3.6]) up to $\mathbb{Z}_p^\times$. Furthermore, we formulate a refinement of Conjecture 4.7 so that it essentially implies [AgCa21, Conj. 3.11].

The following result is an analogue of [BKS24, Th. 7.6] (and a special case of Theorem 4.11).

Theorem 5.29. *Assume Hypothesis 5.22. If we also assume*

- *Conjecture 5.7 (the Heegner point main conjecture),*
- *Conjecture 4.7 for z_∞^{Hg}, and*
- $R_{K_\infty}^{\mathrm{Boc}} \neq 0$*,*

then the p-part of the Birch-Swinnerton-Dyer formula for E/K holds, i.e.,

$$\mathbb{Z}_p \cdot L^*(E/K,1) = \mathbb{Z}_p \cdot \#\text{Ш}(E/K)[p^\infty] \cdot \mathrm{Tam}(E/K) \cdot \frac{1}{\sqrt{|D_K|}} \Omega_{E/K} \cdot R_{E/K}.$$

Proof. This is a consequence of Theorems 4.11 and 5.17, but we can also deduce it directly by combining Proposition 5.26 and Theorem 5.27. □

Acknowledgments

The authors would like to thank Francesc Castella for reading an earlier version of this paper and giving them helpful comments.

References

[AgCa21] A. Agboola, F. Castella, On anticyclotomic variants of the p-adic Birch and Swinnerton-Dyer conjecture, J. Théor. Nombres Bordeaux (Iwasawa 2019 special issue) **33** no.3.1 (2021) 629-658. 159, 204

[Ber95] M. Bertolini, Selmer groups and Heegner points in anticyclotomic $\mathbb{Z}_p$-extensions, Compositio Math. **99** no.2 (1995) 153-182. 195

[Ber01] M. Bertolini, Iwasawa theory for elliptic curves over imaginary quadratic fields, J. Th. Nombres Bordeaux **13** no.1 (2001) 1-25. 199

[BeDa96] M. Bertolini, H. Darmon, Heegner points on Mumford-Tate curves, Invent. Math. **126** no.3 (1996) 413-456. 159, 204

[BDSS23] D. Burns, A. Daoud, T. Sano, S. Seo, On Euler systems for the multiplicative group over general number fields, Publicacions Matemàtiques **67** no.1 (2023) 89-126. 174

[BuFl01] D. Burns, M. Flach, Tamagawa numbers for motives with (non-commutative) coefficients, Doc. Math. **6** (2001) 501-570. 164, 168, 187, 191

[BuGr03] D. Burns, C. Greither, On the Equivariant Tamagawa Number Conjecture for Tate motives, Invent. math. **153** (2003) 303-359. 174

[BKS16] D. Burns, M. Kurihara, T. Sano, On zeta elements for $\mathbb{G}_m$, Doc. Math. **21** (2016) 555-626. 165, 170, 188

[BKS17] D. Burns, M. Kurihara, T. Sano, On Iwasawa theory, zeta elements for $\mathbb{G}_m$ and the equivariant Tamagawa number conjecture, Algebra & Number Theory **11** (2017) 1527-1571. 155, 156, 169, 173, 187, 189, 190

[BKS24] D. Burns, M. Kurihara, T. Sano, On derivatives of Kato's Euler system for elliptic curves, J. Math. Soc. Japan **76** no.3 (2024) 855-919. 156, 158, 159, 160, 167, 170, 171, 173, 185, 186, 187, 188, 189, 190, 203, 204

[BSS18] D. Burns, R. Sakamoto, T. Sano, On the theory of higher rank Euler, Kolyvagin and Stark systems II, submitted, arXiv:1805.08448. 156, 176, 178, 179, 180, 181, 182, 183, 184, 185

[BSS19] D. Burns, R. Sakamoto, T. Sano, On the theory of higher rank Euler, Kolyvagin and Stark systems III, preprint, arXiv:1902.07002. 160, 164, 166, 167, 170, 177

[BuSa21] D. Burns, T. Sano, On the theory of higher rank Euler, Kolyvagin and Stark systems, Int. Math. Res. Not. **2021** no.13 (2021) 10118-10206. 159, 160, 161, 164, 168, 173, 182, 183, 184

[BCK21] A. Burungale, F. Castella, C.-H. Kim, A proof of Perrin-Riou's Heegner point main conjecture, Algebra & Number Theory **15** no.7 (2021) 1627-1653. 158, 195, 196

[Büy09] K. Büyükboduk, Stark units and the main conjectures for totally real fields, Compos. Math. **145** no.5 (2009)1163-1195. 155

[Büy10] K. Büyükboduk, On Euler systems of rank and their Kolyvagin systems, Indiana Univ. Math. J. **59** no.4 (2010) 1277-1332. 155

[Cas17] F. Castella, p-adic heights of Heegner points and Beilinson-Flach classes, Journal of the London Mathematical Society **96** no.1 (2017) 156-180. 157, 195, 196

[Cor02] C. Cornut, Mazur's conjecture on higher Heegner points, Invent. Math. **148** no.3 (2002) 495-523. 196

[CoVa07] C. Cornut, V. Vatsal, Nontriviality of Rankin-Selberg L-functions and CM points, L-functions and Galois representations, London Math. Soc. Lecture Note Ser., vol. 320, Cambridge Univ. Press, Cambridge (2007) 121-186. 196

[Dar92] H. Darmon, A refined conjecture of Mazur-Tate type for Heegner points, Invent. math. **110** no. 1 (1992) 123-146. 186

[Gro82] B. H. Gross, On p-adic L-series at $s = 0$, J. Fac. Sci. Univ. Tokyo, Sect. IA, **28** (1982) 979-994. 189

[GrZa86] B. Gross, D. Zagier, Heegner points and derivatives of L-series, Invent. Math. **84** no.2 (1986) 225-320. 193, 194

[How04] B. Howard, The Heegner point Kolyvagin system, Compositio Mathematica **140** no.6 (2004) 1439-1472. 195

[How06] B. Howard, Bipartite Euler systems, J. Reine Angew. Math. **597** (2006) 1-25. 196

[Jan89] U. Jannsen, On the ℓ-adic cohomology of varieties over number fields and its Galois cohomology, In: Ihara Y., Ribet K., Serre J.-P. (eds) Galois Groups over $\mathbb{Q}$. Mathematical Sciences Research Institute Publications, vol 16. Springer, New York, NY (1989) 315-360. 169

[Kat21] T. Kataoka, Equivariant Iwasawa theory for elliptic curves, Mathematische Zeitschrift **298** (2021) 1653-1725. 173

[Kat22] T. Kataoka, Stark systems and equivariant main conjectures, Osaka J. Math. **59** no.2 (2022) 417-452. 156, 173, 179, 180, 181, 182, 183

[Kat04] K. Kato, p-adic Hodge theory and values of zeta functions of modular forms, Astérisque, (295):ix, 117-290, 2004. Cohomologies p-adiques et applications arithmétiques. III. 166, 167, 172, 173

[LLZ14] A. Lei, D. Loeffler, S. Zerbes, Euler systems for Rankin-Selberg convolutions of modular forms, Ann. Math. **180** no.2 (2014) 653-771. 158

[MaRu16] B. Mazur, K. Rubin, Refined class number formulas for $\mathbb{G}_m$, J. Th. Nombres Bordeaux **28** (2016) 185-211. 187

[MTT86] B. Mazur, J. Tate, J. Teitelbaum, On p-adic analogues of the conjectures of Birch and Swinnerton-Dyer, Invent. math. **84**(1) (1986) 1-48. 169, 189

[Nek01] J. Nekovář, On the parity of ranks of Selmer groups. II, C. R. Acad. Sci. Paris Sér. I Math. **332** no.2 (2001) 99-104. 197, 198, 202

[Nek06] J. Nekovář, Selmer complexes, Astérisque **310** (2006). 165, 196, 197, 199

[NSW00] J. Neukirch, A. Schmidt, K. Wingberg, Cohomology of number fields, Springer Verlag, 2000. 169, 175

[Per87] B. Perrin-Riou, Fonctions L p-adiques, théorie d'Iwasawa et points de Heegner, Bulletin de la Société Mathématique de France **115** no.4 (1987) 399-456. 195

[Per93] B. Perrin-Riou, Fonctions L p-adiques d'une courbe elliptique et points rationnels, Ann. Inst. Fourier (Grenoble) **43** no.4 (1993) 945-995. 167

[Per95] B. Perrin-Riou, Fonctions L p-adiques des représentations p-adiques, Astérisque (229) 1995. 172, 199

[Rub96] K. Rubin, A Stark Conjecture 'over $\mathbb{Z}$' for abelian L-functions with multiple zeros, Ann. Inst. Fourier **46** (1996) 33-62. 165

[Rub00] K. Rubin, Euler systems, Annals of Math. Studies **147**, Princeton Univ. Press, 2000. 157, 186

[Sak23] R. Sakamoto, A higher rank Euler system for the multiplicative group over a totally real field, American J. Math. **145** no.1 (2023) 65-108. 172, 179

[San14] T. Sano, Refined abelian Stark conjectures and the equivariant leading term conjecture of Burns, Compositio Math. **150** (2014) 1809-1835. 186, 187

[Wan21] X. Wan, Heegner point Kolyvagin system and Iwasawa main conjecture, Acta Mathematica Sinica, English Series **37** (2021) 104-120. 195

[Zha14] W. Zhang, Selmer groups and the indivisibility of Heegner points, Camb. J. Math. **2** no.2 (2014) 191-253. 196

AUTHORS

Takenori Kataoka
Tokyo University of Science
Department of Mathematics
Faculty of Science Division II
1-3 Kagurazaka
Shinjuku-ku, Tokyo 162-8601, Japan

tkataoka@rs.tus.ac.jp

Takamichi Sano
Osaka Metropolitan University
Department of Mathematics
3-3-138 Sugimoto
Sumiyoshi-ku, Osaka 558-8585, Japan
tsano@omu.ac.jp

Journal of the Association for Mathematical Research, 2(2):209–335, 2024
jamathr.org

Generating the Goeritz group of S^3

Martin Scharlemann

Received 29 Nov 2022; Revised 15 Feb 2024; Accepted 13 Mar 2024

Abstract: In 1980 J. Powell [Po] proposed that five specific elements sufficed to generate the Goeritz group for any genus Heegaard splitting of S^3. Here we prove that a natural expansion of Powell's proposed generators, to include all eyeglass twists and all topological conjugates of Powell's generators, does suffice.

Key words and phrases: Geometric Structures on 3-Manifolds, Goeritz group

1 Introduction

Suppose that M is a closed orientable 3-manifold and $M = A \cup_T B$ is a Heegaard splitting of M. Following [JM] the *Goeritz group* $G(M,T)$ is the group of isotopy classes of diffeomorphisms $(M,T) \to (M,T)$ for which the induced diffeomorphism $M \to M$ is isotopic to the identity. An element of the Goeritz group can be viewed as the final result of a (possibly non-unique) loop $T_\theta, 0 \leq \theta \leq 2\pi$ of embeddings of T in M, that is an element in $\pi_1(\text{Diff}(M)/\text{Diff}(M,T))$ [JM, Theorem 1]. That is the viewpoint we will take.

Little is known about the Goeritz group, even in the case that $M = S^3$ and the Heegaard splittings are fully described [Wa]. In [Po] J. Powell proposed (indeed he believed he had proven) that five specific isotopies generate the Goeritz group $G(S^3,T)$ of the 3-sphere. (In [Sc2] one is found to be redundant, and so the number is reduced to four.) Figures 1 to 4 give important examples of Goeritz elements acting on the standard Heegaard splitting of S^3.

Powell's conjecture has been verified for genus ≤ 3 splittings of S^3 [FS1], but even the question of whether $G(S^3,T)$ is finitely generated remains open when genus$(T) \geq 4$. Here we define a generalization, shown in Figure 4 and called an *eyeglass twist*, of Powell's proposed generator D_θ. We ultimately show (Corollary 18.4) that if D_θ is replaced by the collection of all eyeglass twists, and we include also all topological conjugates of the other three Powell generators (i. e. conjugates by elements of $G(S^3,T)$), the subgroup $\mathcal{E} \subset G(S^3,T)$ thereby generated is all of $G(S^3,T)$.

DOI: 10.56994/JAMR.002.002.002

In [Sc3] we show that this leads to the following observation: Suppose T is the standard genus $g+1$ Heegaard surface in S^3. Any element of $G(S^3, T)$ that acts trivially on a standard genus 1 summand of T is a consequence of the Powell generators acting on T. In other words, the Powell Conjecture is true, stably.

The strategy for the proof of Corollary 18.4 is rather simple: Suppose T is the standard genus g Heegaard splitting of S^3, and $\tau \in G(S^3, T)$. In [FS1] it is shown that there is a 'cycle of weak reductions' capturing τ. That is, given a loop of embeddings $T_\theta, 0 \leq \theta \leq 2\pi$ of T in S^3 that represents τ, there is a natural topological way to extract, for generic θ, a pair of properly embedded disjoint essential disks $a_\theta \subset A, b_\theta \subset B$ that thereby weakly reduce T_θ. Moreover, at the finite number of non-generic points, the chosen weakly reducing pair does not change much. Since the method of choosing the pair (a_θ, b_θ) is topological, it follows relatively easily that $(a_{2\pi}, b_{2\pi}) = (\tau(a_0), \tau(b_0))$. A landmark result of Casson and Gordon [CG] shows that a weakly reducing pair gives rise to a reducing sphere for T; one might naturally hope that one could track such a reducing sphere K around θ, as was done for the weakly reducing pairs, and thereby be able to meaningfully compare K with $\tau(K)$ and so understand the action of τ.

Sadly, the transition from a weakly reducing pair (a, b) to a reducing sphere involves much choice, so there is no naturally derived reducing sphere K for T as hoped for above. The program here is to find one, via this natural topological method: Let $F \subset S^3$ be the surface obtained from T by weakly reducing along the pair (a, b). There is a natural and oft-used way to sweep out S^3 by level 2-spheres S_s and, given F, a natural value of s to pick: a value so that the genus of the part of F lying below S_s matches the genus of the part above. Could S_s be turned into a viable, topologically defined reducing sphere for T that could play the role of K above?

It turns out that although S_s itself may not play that role, it does give a recipe for weak reduction that is robust: it doesn't change abruptly as T moves through S^3. And as the weak reduction proceeds (as guided by S_s) and F becomes more complicated, its complementary components in S^3 (called the *chambers*) are more likely to contain reducible components. Reducing spheres in these reducible components naturally cut off summands of the original Heegaard surface, and these (inductively) determine an isotopy to the standard picture that is unique, up to action by $\mathcal{E}$. If no complementary component becomes reducible, then at the end of the process we fall back on S_s as the required reducing sphere for T (see the second bullet in the statement of Proposition 10.2).

One can think of the progression Proposition 3.7 → Corollary 3.9 → Corollary 8.4 → Corollary 8.9 → Corollary 17.4 → Corollary 18.4 as guideposts for the argument.

Although the program is easy to describe, the technical difficulties encountered below are complex since, in effect, they will involve 4 dimensions of sweep-outs. Much of the machinery is new, and then so is the terminology. We call the attention of the reader to the Index at the end of the paper for a guide to this new terminology. A plausible strategy for a proof is presented in Section 3, including a rough overview of the final proof in Subsection 3.3. Some of the argument is not restricted to $M = S^3$, so perhaps some of the methodology can be useful in understanding stabilized Heegaard splittings of other 3-manifolds as well.

General Remarks: All manifolds will be orientable and, unless obviously not, compact. We mostly work in the TOP category (locally flat embeddings, homeomorphisms) to avoid discussion of corner rounding, etc. But at several points (see especially Section 17) we need results from smooth topology. At those points we will work in the DIFF category. In these dimensions (two and three) the difference in categories is immaterial, see for example [Moi]. Unexplained notation in the statement of a Lemma, Proposition, etc. may well be found in the discussion that immediately precedes it.

2 The eyeglass subgroup

Powell's proposed generators are expressed in terms of *standard genus 1 summands* in a canonical picture of a genus g Heegaard surface in S^3. In [FS1, Section 2] some consequences of these Powell generators are presented that include the Powell generators themselves, but also include more general moves on the standard genus 1 summands. These more general moves would make sense on any stabilized Heegaard splitting of any closed orientable manifold, if we drop the requirement that the summands are 'standard', which is not meaningful in the context of arbitrary stabilized splittings. Below is a brief description of those generalized moves; more detail can be found in [FS1]. In addition, note that the last move described (an eyeglass twist) is a significant generalization of Powell's move D_θ.

For a Heegaard splitting $M = A \cup_T B$ define a *bubble* for T to be a 3-ball $\mathfrak{b}$ in M whose boundary intersects T in a single essential circle. Via Waldhausen [Wa] we know that any bubble is the boundary sum of a collection of genus 1 bubbles.

1. A *bubble move* is an isotopy of $\mathfrak{b}$ along a closed path in $T - \mathfrak{b}$ that begins and ends at $\mathfrak{b}$, returning $(\mathfrak{b}, \mathfrak{b} \cap T)$ to itself. See Figure 1. Note that if the closed path is null-homotopic in M, which is automatic if $M = S^3$, the bubble move is isotopic to the identity on M, so it represents an element of $G(M, T)$.

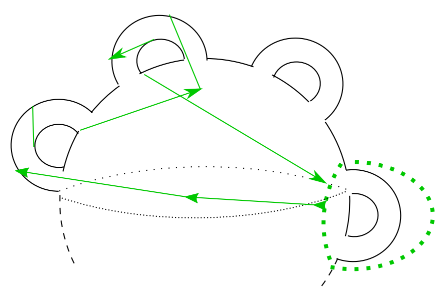

Figure 1: A bubble move

2. Let $\mathfrak{b}$ be a genus 1 bubble. A *flip* is the homeomorphism $(\mathfrak{b}, \mathfrak{b} \cap T) \to (\mathfrak{b}, \mathfrak{b} \cap T)$ that reverses orientation of both the meridian and longitude of the summand, as shown in Figure 2. Powell's [Po] label for a flip on the first standard summand is D_ω. In a genus 1 splitting, where T is a torus, regard the hyperelliptic involution $(S^3, T) \to (S^3, T)$ as (a degenerate case of) a flip.

3. Let $\mathfrak{b}_1, \mathfrak{b}_2$ be disjoint genus 1 bubbles, and let $v \subset T - (\mathfrak{b}_1 \cup \mathfrak{b}_2)$ be an arc connecting them. Let $\mathfrak{b}$ be the genus 2 bubble obtained by tubing together the genus 1 bubbles along v. A *bubble exchange* exchanges the two genus 1 bubbles within B, as shown in Figure 3. Powell's label for a bubble exchange on the first two standard summands is $D_{\eta_{12}}$.

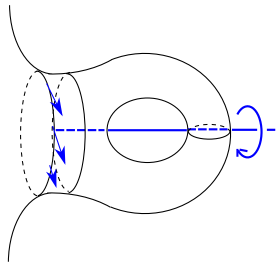

Figure 2: A flip

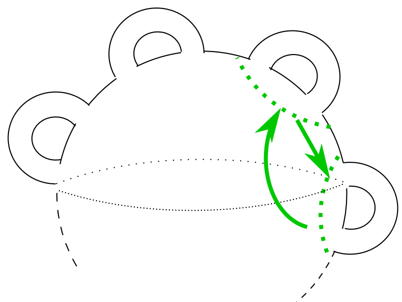

Figure 3: A bubble exchange

4. An *eyeglass* is the union of two disks, ℓ_a, ℓ_b (the *lenses*) with an arc v (the *bridge*) connecting their boundaries. Suppose an eyeglass η is embedded in M so that the 1-skeleton of η (called the *frame*) lies in T, one lens is properly embedded in A, and the other lens is properly embedded in B. The embedded η defines a natural automorphism $(M,T) \to (M,T)$, as illustrated in Figure 4, called an *eyeglass twist*. Powell's generator D_θ is an example, but topologically special because the lenses are primitive disks in each handlebody.

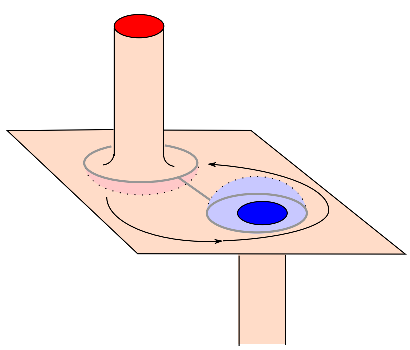

Figure 4: An eyeglass twist

Let Mod(M,T) be the group of path components of Diff(M,T); it is the Goeritz group $G(M,T)$ when, as is the case with S^3, Mod(M) is trivial [JM].

Definition 2.1. *Suppose $M = A \cup_T B$ is a Heegaard splitting. The subgroup of* $\mathrm{Mod}(M,T)$ *generated by the four types of moves just described is called the* eyeglass group $\mathcal{E} \subset \mathrm{Mod}(M,T)$*. Any finite composition of these eyeglass generators will be called an* eyeglass move.

When $M = S^3$ call the subgroup of $G(S^3,T) = \mathrm{Mod}(S^3,T)$ generated by Powell's proposed generators the Powell group*. Any finite composition of these generators will be called a* Powell move.

Remarks: It is shown in [FS1, Section 2] that the Powell group can also be described as the subgroup of $G(S^3,T)$ generated by D_θ together with those flips, bubble moves, and bubble exchanges that act on the standard genus 1 bubbles, not on arbitrary genus 1 bubbles.

It follows that any Powell move (indeed any topological conjugate of a Powell move) is an eyeglass move. Note also that whereas the Powell group is not known to be normal in $G(S^3,T)$, the eyeglass group is normal, since any topological conjugate of a generator is a generator.

The Powell Conjecture is that the Powell group is the entire Goeritz group $G(S^3,T)$; we will eventually show the weaker result that the eyeglass group is the entire Goeritz group. If one could show that each eyeglass twist and any topological conjugate of a Powell generator is in the Powell group, the Powell Conjecture would follow. This seems unlikely.

Let $M = A \cup_T B$ be a Heegaard splitting; we briefly review terminology, and make some elementary observations:

Definition 2.2. *A sphere S in M is* aligned *with the Heegaard splitting if $S \cap T$ is at most one circle. Similarly, a properly embedded disk $(D,\partial D) \subset (M,\partial M)$ is aligned with the splitting if $D \cap T$ is at most one circle, and, if it is one circle, the annulus component of $D - T$ is a spanning annulus in the compression body in which it lies. See [Sc1], [FS2].*

For example, the boundary of a bubble $\mathfrak{b}$ is an aligned sphere.

Suppose disjoint spheres S and S' are aligned, and α is a properly embedded arc in T so that α intersects S exactly in one end of α and intersects S' only in the other end of α. A thin tubular neighborhood $N \cong D^2 \times I$ of α will intersect $S \cup S'$ in the two disks $D^2 \times \{0,1\}$. Delete those disks from $S \cup S'$ and glue on the annulus $\partial D^2 \times I$.

Definition 2.3. *Call the resulting aligned sphere the* tube sum *of S and S' along α.*

The same construction on S and a disjoint aligned disk D results in another aligned disk, called the tube sum of D and S.

If $\mathfrak{b}$ and $\mathfrak{b}'$ are disjoint bubbles, say of genus p and q then the tube sum of $\partial\mathfrak{b}$ and $\partial\mathfrak{b}'$ naturally bound a genus $p+q$ bubble, called the tube sum of $\mathfrak{b}$ and $\mathfrak{b}'$.

Since a bubble move is an eyeglass move, the tube sum of a sphere or disk with a bubble, and, in particular, the tube sum of two bubbles, is well-defined up to eyeglass moves; it does not depend on the choice of α.

Definition 2.4. *For D an aligned disk in M and $\mathfrak{b}$ a disjoint bubble, the tube sum of D and $\mathfrak{b}$ is an aligned disk D' properly isotopic to D in M. Replacing D with D' is called a* bubble pass *of $\mathfrak{b}$ through D.*

Lemma 2.5. *Suppose D is an aligned* non-separating *disk in M, $\mathfrak{b}$ is a disjoint bubble, and D' is the disk obtained from D by a* bubble pass *of $\mathfrak{b}$ through D. Then there is an eyeglass move which isotopes D to D' in M.*

Proof. Let α be the arc between D and $\partial\mathfrak{b}$ along which the bubble pass is made, and let D' be the resulting aligned disk. Since D is assumed non-separating, there is another arc $\beta \subset T$ disjoint from α so that $\partial\beta = \partial\alpha$ but the end of β at D is on the opposite side of D as the end of α at D. Thus the union $\gamma = \alpha \cup \beta$ is a path in $T - \mathfrak{b}$ that passes exactly once through D. Then a bubble move of $\mathfrak{b}$ via γ will isotope D to D' as required. (See Figure 5. The bubble move of $\mathfrak{b}$ to itself that carries D to D' is along the concatenation loop $\beta\alpha$.) □

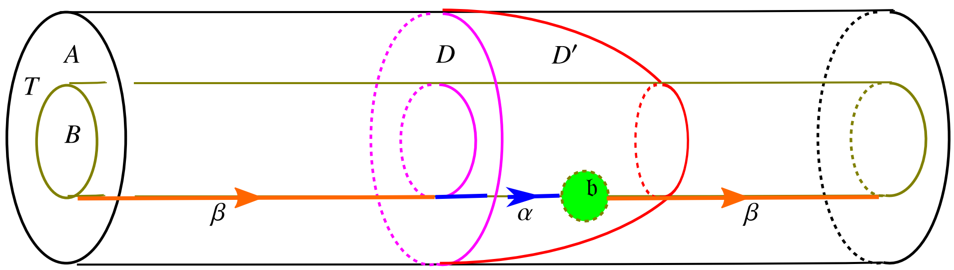

Figure 5: A bubble move of $\mathfrak{b}$ to itself that carries D to D'

3 The search for a plausible strategy

The proof that $G(S^3, T) = \mathcal{E}$ will be by induction on the genus of T; it is known for genus ≤ 3 by [FS1]. In this section we lay some groundwork and describe (see Corollary 3.12) a special circumstance that makes the inductive step straightforward. The goal of the rest of the long and technical argument in Sections 4 through 17 is simply to show that conditions sufficiently similar to this special circumstance always arise.

3.1 A technical lemma

The following technical lemma is needed almost immediately for an inductive step:

Let (S^3, T) be a genus $g \geq 1$ Heegaard splitting and B_p be a small ball around a point $p \in T$ that intersects T in a single disk D_p. Let B_- be the 3-ball that remains when $\text{int}(B_p)$ is removed, and $T_- = T \cap B_- = T - \text{int}(D_p)$. T_- is once-punctured genus g surface properly embedded in B_-. Let $\text{Mod}(B_-, T_-)$ be the group of path components of $\text{Diff}(B_-, T_-)$. (We do not require that a diffeomorphism $f : (B_-, T_-) \to (B_-, T_-) \in \text{Diff}(B_-, T_-)$ be the identity on $\partial B_p = \partial B_-$.)

Echoing the definition of bubble move above, define a *B_p-bubble move* $\beta_p : (S^3, T) \to (S^3, T)$ to be the result of an isotopy of (B_p, D_p) along a closed path in T_- that returns (B_p, D_p) to itself. It is understood that D_p remains in T throughout the isotopy, so T is preserved throughout the isotopy. This means that the final map $\beta_p : (S^3, T) \to (S^3, T)$ still represents the identity in $\text{Mod}(S^3, T)$. However, the restriction $\beta_p | B_- : (B_-, T_-) \to (B_-, T_-)$ may not be isotopic to the identity and so may represent a non-trivial element of $\text{Mod}(B_-, T_-)$

Define $\mathcal{E}_- \subset \text{Mod}(B_-, T_-)$ to be the subgroup generated by B_p-bubble moves, together with the flips, bubble moves, bubble exchanges and eyeglass twists that generate $\mathcal{E}$ and whose associated paths, bubbles, lenses and bridges are disjoint from B_p, so they all lie in B_-.

Lemma 3.1. *If* $\text{Mod}(S^3, T) = \mathcal{E}$ *then* $\text{Mod}(B_-, T_-) = \mathcal{E}_-$.

This is a natural statement, and is relevant here because $G(S^3, T) = \text{Mod}(S^3, T)$.

Proof. Let $h_- : (B_-, T_-) \to (B_-, T_-)$ represents an element of $\text{Mod}(B_-, T_-)$; after an isotopy we can assume that h_- is the identity near $\partial B_p = \partial B_-$. Extend h_- to a diffeomorphism $h : (S^3, T) \to (S^3, T)$ representing a class in $\text{Mod}(S^3, T)$ by setting $h|B_p$ to be the identity. Under the assumption $\text{Mod}(S^3, T) = \mathcal{E}$ there is an eyeglass move $h' : (S^3, T) \to (S^3, T)$ so that $h'h^{-1} : (S^3, T) \to (S^3, T)$ is isotopic to the identity. We may as well take the generators of h' to lie in B_-, since each of these generators has support near a 1-complex in T and so, by general position, can be chosen to be disjoint from B_p. For example, for an eyeglass twist choose the frame of the eyeglass disjoint from B_p and it becomes a frame in B_-; in a bubble move, regard the bubble as a thin neighborhood in T of the bubble's 1-skeleton, which we can also take to be disjoint from B_p. Once this is done, the generators give a diffeomorphism $h'_- : (B_-, T_-) \to (B_-, T_-)$ representing an element of $\mathcal{E}_-$. (There is no claim that h'_- is well-defined. In fact a particular choice of h_- depends on how the supporting 1-complexes of the generators of h' have been isotoped away from B_p.)

By construction, the diffeomorphisms $h'h^{-1}$ and $id_{(S^3,T)}$, both of which are the identity on B_p, are isotopic, but such an isotopy might well move B_p, so it does not induce an isotopy from $h'_- h_-^{-1} : (B_-, T_-) \to (B_-, T_-)$ to the identity. Our intention is to fix this, by suitably altering h'. Denote by θ_t the isotopy from $\theta_0 = h'h^{-1} : (S^3, T) \to (S^3, T)$ to $\theta_1 = id_{(S^3,T)}$ and view θ_t as a path in $\text{Diff}(S^3, T)$. The isotopy θ_t determines a p-based loop $\alpha : [0, 1] \to T$ in T via $\alpha(t) = \theta_t(p)$. Now alter h' to $h'' : (S^3, T) \to (S^3, T)$ by post-composing with a B_p-bubble move along the loop $\overline{\alpha}$ defined by $\overline{\alpha}(t) = \alpha(1 - t)$. By definition of $\mathcal{E}_-$, the B_p-bubble move $h''_- : (B_-, T_-) \to (B_-, T_-)$ still lies in $\mathcal{E}_-$. Furthermore, the maps h' and h'' are isotopic as diffeomorphisms on (S^3, T), but the isotopy θ_t is replaced by an isotopy θ'_t from $h''h^{-1}$ to the identity for which the loop $\theta'_t(p)$ in T becomes the concatenation $\alpha\overline{\alpha}$, and so is nullhomotopic in T.

Consider the evaluation map $e_p : \text{Diff}(S^3, T) \to T$ given by $e_p(f) = f(p)$. We can regard T as the space of embeddings $\text{Emb}(p, T)$ and deduce from [Pa] that e_p is a fibration. In particular, the null-homotopy of $\alpha\overline{\alpha}$ to $p \in T$ just described lifts to a homotopy rel end points from the arc θ'_t in $\text{Diff}(S^3, T)$ to an arc in $\text{Diff}(S^3, T)$ that lies entirely over p. This arc then defines an isotopy θ''_t from $h''h^{-1} : (S^3, T) \to (S^3, T)$ to the identity, but now an isotopy in which the point p is fixed throughout. That is, for every $0 \le t \le 1$ $\theta''_t(p) = p$.

Finally, we will alter below the isotopy θ'' so that for every $0 \le t \le 1$, $\theta''_t(B_p) = B_p$. This will imply that the restriction $\theta''_- = \theta''|(B_-, T_-)$ is an isotopy from $h''_- h_-^{-1} : (B_-, T_-) \to (B_-, T_-)$ to the identity, so h''_- and h_- represent the same element of $\text{Mod}(B_-, T_-)$; since $h'' \in \mathcal{E}_-$ this will conclude the proof.

Here then is a brief sketch of how, using standard tools of differential topology [GP], the isotopy θ'' can be altered so that not just p, but all of B_p is mapped diffeomorphically to itself during the isotopy. The first step is to observe that the derivative of any diffeomorphism $f : (S^3, T, p) \to (S^3, T, p)$ determines a linear automorphism Df_p of the tangent space of S^3 at p which also sends the tangent space of T at p to itself. The space of such linear automorphisms deformation retracts to the subspace in which the automorphism is orthogonal. (Such a deformation retraction can be easily constructed, for example, from the proof of [Sp, Theorem 4.2].) Via this deformation retraction, continuously alter θ'' so that for any $0 \le t \le 1$ the derivative $D(\theta''_t)_p$ is orthogonal. Once this is done, θ'' can be further altered near p so that each map θ''_t becomes itself the orthogonal map $D(\theta''_t)_p$ near p. In particular it takes a small ball around

p diffeomorphically to itself while continuing to preserve the Heegaard surface T . The final step is to deform θ'' so that the ball on which this is true contains the original B_p.

(Aside: when $genus(T) = 1$, B_p-bubble moves are not needed, per [EE, Theorem 1b].) □

A bubble exchange between genus 1 bubbles $\mathfrak{b}_1$ and $\mathfrak{b}_2$, as shown in Figure 3, takes place in the neighborhood of $\mathfrak{b}_1 \cup \mathfrak{b}_2 \cup v$, where v is an embedded arc in T that connects the bubbles. A more general operation, which we will call a *generalized bubble exchange*, isotopes $\mathfrak{b}_1$ to $\mathfrak{b}_2$ along an arc $v_1 \subset T$ while simultaneously isotoping $\mathfrak{b}_2$ to $\mathfrak{b}_1$ along an arc $v_2 \subset T$ disjoint from v_1. A generalized bubble exchange is a simple bubble exchange if the arcs v_1, v_2 are parallel in T.

Lemma 3.2. *Any generalized bubble exchange is an eyeglass move.*

Proof. A generalized exchange of bubbles $\mathfrak{b}_1$ and $\mathfrak{b}_2$ using arcs v_1 and v_2 (first column of Figure 6) can be written (see second column of Figure 6) as a composition of the simple bubble exchange along v_1 and a bubble move of $\mathfrak{b}_2$ along the closed path $v_1 \cup v_2$. Each of the latter is an eyeglass move. □

Following Lemma 3.2 we will use the term "bubble exchange" to include generalized bubble exchange, unless the distinction is important.

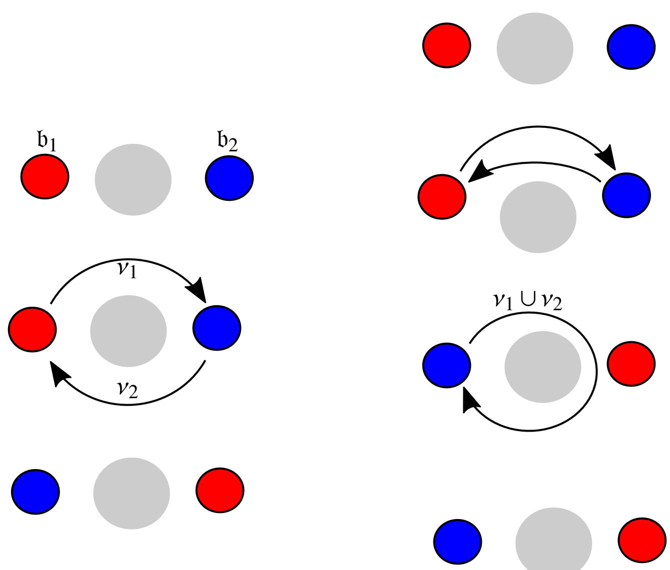

Figure 6: Generalized exchange as an eyeglass move

3.2 Resuming the search for a strategy

We begin by setting up a standard picture for the genus g Heegaard splitting of S^3 that is somewhat different than the one used by Powell. Let $T_g \subset S^3$ be the standard genus g Heegaard surface in S^3, dividing S^3 into the genus g handlebodies A_g and B_g.

Let $\{c_1, ..., c_{g-1}\}$ be the disjoint separating circles on T_g shown in Figure 7, with each c_i separating the first i standard summands from the last $g-i$ standard summands. Note that each c_i bounds a disk in both A and B and so defines a reducing sphere S_i for T_g. Let $\mathfrak{b}_i$ be the genus i Heegaard split 3-ball component of $S^3 - S_i$ containing S_1. Both $\mathfrak{b}_i$ and its complement $S^3 - \mathfrak{b}_i$ are bubbles.

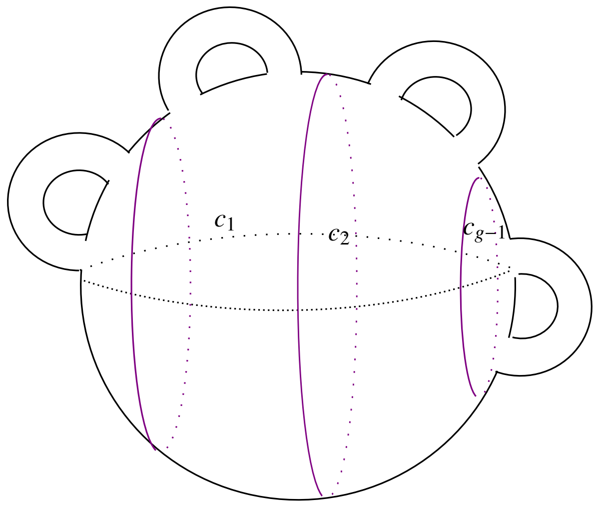

Figure 7: The standard Heegaard surface $T_g \subset S^3$

It is easy to construct a sequence of (generalized) bubble exchanges (and perhaps a flip) that ultimately moves each bubble $\mathfrak{b}_i, 1 \leq i \leq g-1$ to the bubble $S^3 - \mathfrak{b}_{g-i}$. The sequence is best seen in the elaboration Figure 8 of Figure 7 that we now describe. The unit sphere S in S^3 is shown intersecting the x-y-plane, with the z-axis coming out of the page. Denote by C the unit circle in which S intersects the x-y-plane. For $g = 2n$ or $2n-1$ (depending on the parity of g) consider n planes $Q_1, ..., Q_n$ parallel to the x-z plane (horizontal planes in the figure) starting at Q_1 the x-z plane itself and ascending from there. Each $Q_i, 1 \leq i \leq n-1$ intersects S in a circle d_i; $d_n = S \cap Q_n$ is either a circle or a point, depending on the parity of g. Place a genus 1 bubble (shown in green in the figure) on S at each point of $d_i \cap C$. The curves c_i that separate the bubbles, as shown in Figure 7 may be taken to be the intersection of vertical planes (that is, planes parallel to the y-z-plane) $P_1, ..., P_{g-1}$. These planes are distributed symmetrically across the y-z plane and are shown in red in Figure 8. In this picture, simple π-rotation around the y-axis will take each c_i to c_{g-i} and each genus i bubble $\mathfrak{b}_i$ to the bubble $S^3 - \mathfrak{b}_{g-i}$, as we seek.

It remains to show that (up to pairwise isotopy of (S^3, T_g)) this π rotation, which we will denote h_ρ, can be accomplished by generalized bubble exchanges and perhaps a flip. This is easy to see; for each $1 \leq i \leq n$ do a generalized bubble exchange between the two bubbles that lie in d_i (or a flip on the single bubble in d_n if g is odd) using the subarc of d_i having z positive (the front face of the sphere S) for one arc of the exchange and the subarc of d_i on the back face of S (z negative) for the other arc. Since each generalized bubble exchange (and the flip) is an eyeglass move, h_ρ is an eyeglass move. (In fact, since the bubbles are standard, h_ρ is a Powell move.)

Suppose $T \subset S^3$ is a genus g Heegaard surface, and $h, h' : (S^3, T) \to (S^3, T_g)$ are two orientation-

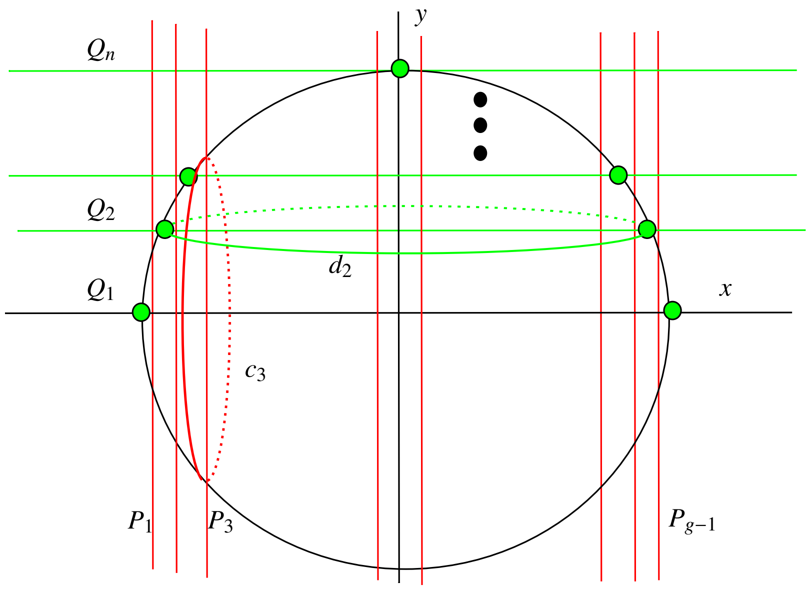

Figure 8: h_ρ as sequence of bubble exchanges (g odd)

preserving homeomorphisms.

Definition 3.3. *h, h' are* eyeglass equivalent (written $h \sim h'$) if the composition $h'h^{-1} : (S^3, T_g) \to (S^3, T_g)$ is isotopic in (S^3, T_g) to an eyeglass move.

For example, the argument above shows that h_ρ is eyeglass equivalent to the identity.

Throughout the remainder of this section we continue to make this inductive assumption:

Assumption 3.4. $G(S^3, T') = \mathcal{E}$ *whenever* genus$(T') \leq g-1$.

Lemma 3.5. *Suppose S is a reducing sphere for a genus g Heegaard splitting (S^3, T). Then there is an orientation preserving homeomorphism $h_S : (S^3, T) \to (S^3, T_g)$ so that $h_S(S) \in \{S_i, i = 1, ..., g-1\}$.*

Suppose that S' is a reducing sphere for (S^3, T) that is disjoint from S and $h_{S'} : (S^3, T) \to (S^3, T_g)$ is similarly defined, so that $h_{S'}(S') \in \{S_i, i = 1, ..., g-1\}$. Under Assumption 3.4, the homeomorphisms h_S and $h_{S'}$ are eyeglass equivalent. In particular h_S is well-defined up to eyeglass equivalence.

Proof. The first statement follows almost immediately from [Wa], as we now describe. S divides T into surfaces $T_\pm$ of genus $g_\pm \geq 1$ respectively, with $g_+ + g_- = g$. Choose any orientation-preserving homeomorphism $f : S^3 \to S^3$ that carries S to S_{g_+} and the complementary ball component of S that contains T_+ to the ball $\mathfrak{b}_{g_+}$. Adjust so that the circle $T \cap S$ is sent to $\mathfrak{c}_{g_+} \subset S_{g_+}$. Then $f(T_+)$ and $T_g \cap \mathfrak{b}_{g_+}$ each determine a genus g_+ Heegaard splitting of $\mathfrak{b}_{g_+}$ and [Wa] then implies that the surfaces are isotopic rel $\mathfrak{c}_{g_+}$. Similarly isotope $f(T_-)$ to $T_g \cap (S^3 - \mathfrak{b}_{g_+})$ in $(S^3 - \mathfrak{b}_{g_+})$. The resulting homeomorphism h_S takes S to S_{g_+} and T to T_g, as required.

Note that this construction subtly depends on a choice: if we had reversed the labels $g_\pm$ then the requirement that T_+ be sent to $\mathfrak{b}_{g_+}$ would have sent S to S_{g_-} instead of S_{g_+}. But this could also be accomplished by composing with the homeomorphism h_ρ defined just before this lemma, and we have shown that $h_\rho \sim id_{(S^3,T_g)}$. So the choice of labelling makes no difference in the construction of h_S, up to eyeglass equivalence.

Proceeding then with the second statement, we first claim that h_S is well-defined up to eyeglass equivalence. To that end, suppose another orientation preserving homeomorphism $h' : (S^3, T) \to (S^3, T_g)$ has $h'(S) \in \{S_i, i = 1, ..., g-1\}$. Using possibly h_ρ as above, we may as well assume $h_S h'^{-1}(\mathfrak{b}_{g_+}, T_{g_+}) = (\mathfrak{b}_{g_+}, T_{g_+})$. Noting that $g_+ < g$, apply Lemma 3.1 and Assumption 3.4 to the pair $(\mathfrak{b}_{g_+}, T_{g_+})$ with the goal of showing that there is an eyeglass move on (S^3, T_g) which, when composed with $h_S h'^{-1}$, is the identity on $\mathfrak{b}_{g_+}$ and remains $h_S h'^{-1}$ on $S^3 - \mathfrak{b}_{g_+}$. Lemma 3.1 says that, under the inductive assumption, this is true for some move $h_{g+} : (\mathfrak{b}_{g_+}, T_{g_+}) \to (\mathfrak{b}_{g_+}, T_{g_+})$ in $\mathcal{E}_-$. By definition, each flip, bubble move, bubble exchange or eyeglass twist used in the construction of h_{g+} is also an eyeglass move on (S^3, T_g). On the other hand, a move in $\mathcal{E}_-$ corresponding to a B_p bubble move in the proof of Lemma 3.1 here corresponds to a bubble move on the bubble $S^3 - \mathfrak{b}_{g_+}$ in (S^3, T_g), and this is also an eyeglass move. So indeed there is an eyeglass move as we seek. Now apply the same argument on the complementary ball $S^3 - \mathfrak{b}_{g_+}$ and deduce that $h_S h'^{-1}$ is an eyeglass move on (S^3, T_g), so h' is eyeglass equivalent to h_S, as desired.

Now consider the reducing sphere S'. We may as well choose labels $g_\pm$ so that $h_S(S') \subset \mathfrak{b}_{g_+}$. Let $\mathfrak{b}_S = h_S^{-1}(\mathfrak{b}_{g_+})$, a genus g_+ bubble for (S^3, T). Apply the first statement again, this time to $h_S|\mathfrak{b}_S : (\mathfrak{b}_S, T_+) \to (\mathfrak{b}_{g_+}, T_{g_+})$, and deduce that $h_S|\mathfrak{b}_{g_+}$ could have been chosen so that $h_S(S') = S_i$ for some $i \le g_+$, that is $h_S(S') \in \{S_i, i = 1, ..., g_+\}$. The same is true, by definition, for the given $h_{S'}$. The argument we have just given that h_S is well-defined up to eyeglass equivalence, repeated now for $h_{S'}$, shows that $h_S \sim h_{S'}$ as required.

The last comment, that under the inductive assumption h_S is well-defined up to eyeglass equivalence, follows simply by taking S' to be a parallel copy of S in the previous argument. □

Our goal is to show that for any genus g Heegaard surface T in S^3, $G(S^3, T) = \mathcal{E}$. Another way of expressing this is that any two orientation preserving homeomorphisms $(S^3, T) \to (S^3, T_g)$ are eyeglass equivalent.

Suppose $F \subset S^3$ is a possibly disconnected closed surface, dividing S^3 into possibly disconnected 3-manifolds M_A and M_B. Suppose further

1. Each component C of M_A has a Heegaard splitting $C = A_C \cup_{T_C} B_C$ in which A_C is a handlebody. In particular, $\partial C = \partial_- B_C$. Here we follow [FS2] and [Sc1] in allowing sphere components of $\partial_- B_C$, so the compression body B_C may be reducible.

2. The symmetric statement is true for each component of M_B.

3. The splitting of each ball component of M_A or M_B has genus ≥ 1. Hence any component that has a genus 0 splitting is a punctured 3-sphere (since it has genus 0) with more than one boundary component (since it is not a ball).

4. The Heegaard splitting (S^3, T) that is obtained by amalgamating all of these Heegaard splittings is of genus g. (See [La, Section 3] for a description of Heegaard surface amalgamation.)

The *defining surface* $F \subset S^3$, together with the Heegaard splittings of M_A and M_B as described above, will be called a *Heegaard split chamber complex* that *supports* the Heegaard splitting (S^3, T). Each component of M_A or M_B is called a *chamber*. Heegaard split chamber complexes are formally defined and explored more extensively in section 5. (See for example Definition 5.9)

Before proceeding we list some Elementary Facts about surfaces and amalgamation that will be useful:

- EF1: A compact surface is non-planar if and only if it contains two simple closed curves that intersect in a single point
- EF2: Consequently, if F is subsurface of a compact surface F' and F is non-planar, then so is F'.
- EF3: If F is a non-planar surface and F' is obtained by removing a finite number of disjoint disks from the interior of F (that is F' is a *punctured* F), then F' is non-planar.
- EF4: Suppose Heegaard splittings (M_1, T_1), (M_2, T_2) of compact orientable 3-manifolds M_1 and M_2 are amalgamated along a closed surface F that is a boundary component of each. Then a punctured copy of F lies in each of T_1, T_2 and the amalgamated Heegaard surface $T \subset M = M_1 \cup_F M_2$. See [La, Figure 12]. In particular, if F is non-planar, so are T_1, T_2 and T.
- EF5: Under the amalgamation just described, a punctured copy of each $T_i, i = 1, 2$ lies in T. In particular, if either T_i is non-planar, so is T. Again see [La, Figure 12].

Lemma 3.6. *Suppose, in a Heegaard split chamber complex in S^3 as described above, S is a sphere component of F. Then, after amalgamation, a slight push-off of S becomes a reducing sphere for the splitting (S^3, T), intersecting T in a single circle that is essential in T.*

Proof. We first show that after amalgamation the part of T lying in each of the ball components of $S^3 - S$ must contain a non-planar surface. There are two cases:

If every chamber within a ball B bounded by S is a punctured 3-sphere, then every component of F within B is a sphere. Pick a sphere (possibly S itself) that is innermost among these components. The ball it bounds contains no other component of F and therefore is a ball chamber. By the third property of Heegaard split chamber complexes described above, the splitting surface for this ball chamber has genus ≥ 1 and so is non-planar. It follows that $T \cap B$ is non-planar, using EF5 above.

On the other hand, if there is a chamber in B that is not a punctured 3-sphere then a boundary component of the chamber is non-planar, so again $T \cap B$ is non-planar, using EF4 above.

We now follow the methodology of [CG]: after amalgamation S becomes a punctured sphere lying in T with some of its boundary components bounding disks in A and others bounding disks in B. Choose a circle $c \subset S$ that divides S into two disks: D_A that contains all disks of $S - T$ that lie in A and D_B that contains all disks of $S - T$ that lie in B. Push int(D_A) slightly into A and int(D_B) slightly into B. The resulting sphere S' intersects T only in the circle c. As we have just shown, each of the two components of $T - S'$ contains a non-planar surface, namely the part of T lying in a ball component of $S^3 - S$. Thus neither component of $T - S'$ is planar so, in particular, neither component is a disk. Hence S' is a reducing sphere for T and $c = S' \cap T$ is essential in T. □

Proposition 3.7. *Suppose, for the Heegaard split chamber complex above, S is an incompressible sphere in a chamber of M_A or M_B. Then there is an orientation preserving homeomorphism $h_S : (S^3, T) \to (S^3, T_g)$ so that $h_S(S) \in \{S_i, i = 1, ..., g-1\}$.*

Moreover, suppose for $\tau \in G(S^3, T)$ *a homeomorphism* $h_{\tau(S)}$ *is similarly defined, for the Heegaard split chamber complex whose defining surface is* $\tau(F)$*. Then, under Assumption 3.4,* $h_{\tau(S)}\tau \sim h_S$*.*

Proof. Let C be the chamber in which S lies, say $C \subset M_A$, so $C = A_C \cup_{T_C} B_C$ with A_C a handlebody.

The main theorem of [Sc1] says that the Heegaard surface T_C in C may be isotoped so that it is aligned with S, that is so that it intersects S in at most one circle.

Claim: When S is aligned with T_C in C and the Heegaard surface $T \subset S$ is created by amalgamation of the splittings of the chambers, S becomes a reducing sphere for the splitting (S^3, T).

There are three cases to consider:

Case 1: T_C cannot be isotoped in C to be disjoint from S.

S is aligned with T_C, and in this case $T_C \cap S$ cannot be empty, so this intersection must be a single circle c. Furthermore c must be essential in T_C: Indeed, if c bounded a disk $D_T \subset T_C$ then the union of D_T and the disk $S \cap A_C$ would be a sphere in the (irreducible) handlebody A_C, and the ball the sphere bounds could be used to isotope D_T through $S \cap A_C$ and so off of S, contradicting the hypothesis of this case.

Since S is separating in S^3, the curve c is separating in T_C. Since c is essential in T_C, neither component of $T_C - c$ is a disk, so both components are non-planar surfaces. By the argument of EF5 above, each side of c in T is then non-planar, so c is essential in T. Under amalgamation, the disk $S \cap A_C$ becomes a disk in A.

It is easy to arrange that, similarly, the disk $S \cap B_C$ remains a disk in B after the amalgamation: isotope the disk $S_B = S \cap B_C$ so that it intersects the defining B−disks for the amalgamation $\mathcal{B} \subset B_C$ only in arcs. Then S_B intersects $B_C - \eta(\mathcal{B})$, which is a collar of ∂C in B_C, only in disks, namely the complement in S_B of the arcs $S_B \cap \mathcal{B}$. Finally, properly isotope the disks $S_B - \eta(\mathcal{B})$ so that they avoid the A-disks of the amalgamation, so S_B ends up entirely in B. Then S is the union of two disks, $S_A \subset A$ and $S_B \subset B$ along their common boundary c, an essential circle in T. Hence S is a reducing sphere for (S^3, T).

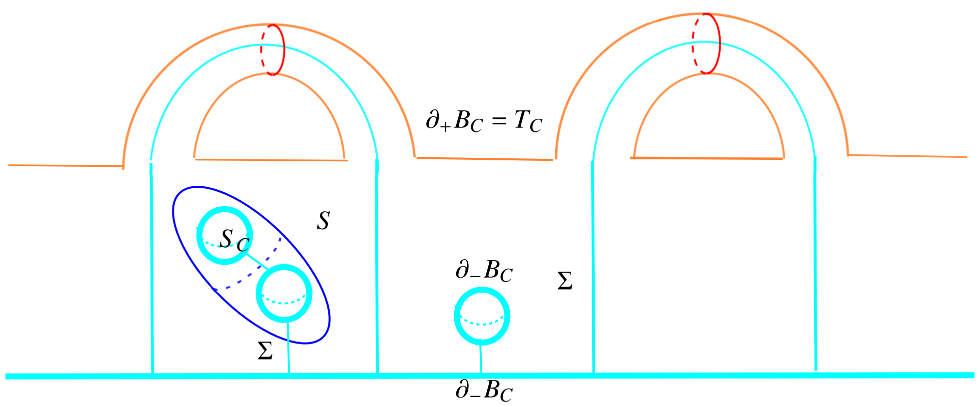

Figure 9: A spine Σ (in aqua) of B_C intersecting S in a single point.

Case 2: T_C can be isotoped in C to be disjoint from S and genus$(T_C) \geq 1$.

Since the chamber $C \subset M_A$, A_C is a handlebody and so is irreducible. Hence if S is disjoint from T_C it must lie entirely in the compression body B_C. This implies that B_C is reducible, so $\partial_- B_C$ contains spheres, some of which lie on the other side of S from $T_C = \partial_+ B_C$ in B_C. Let $S_C \subset \partial_- B_C$ be one of them.

Since S separates S_C from T_C in B_C, any spine of B_C intersects S, and there is some spine Σ of B_C that intersects S in a single point (as pictured in Figure 9). Now isotope T_C to be a neighborhood of that spine, so it intersects S in a single circle c. Amalgamation happens well away from S, so S also ends up intersecting T in c. The circle c divides T_C into a disk, on the side of S containing S_C, and a positive genus surface on the other, since by assumption genus$(T_C) \geq 1$. By Lemma 3.6 the part of T that lies in the ball in S^3 bounded by S_C (not visible in Figure 9) also has genus ≥ 1. Hence each of the surfaces into which S divides T has genus ≥ 1, so S is a reducing sphere for T. Moreover the spine Σ has been chosen so that the disk $S \cap B_C$ is one of the defining disks for the amalgamation, so it automatically persists as a disk in B after amalgamation. Thus S becomes a disk intersecting T in the single essential circle c and so a reducing sphere for (S^3, T).

Case 3: T_C *can be isotoped in* C *to be disjoint from* S *and genus*$(T_C) = 0$.

By assumption (3) on F above, C is a multiply punctured 3-sphere. Since S is incompressible in T_C, each side of S in T_C contains some of the punctures, i. e. some sphere components of $\partial_- C$. Just as in Case 2, applying Lemma 3.6, T_C can be isotoped to intersect S in a single circle and the part of T lying on each side of S must have positive genus, so S is a reducing sphere.

This establishes the Claim; the proof of the first statement then follows from [Wa].

The proof of the second statement follows from the observation that $h_S\tau^{-1}(\tau(S)) \in \{S_i, i = 1, ..., g-1\}$ so for $h_{\tau(S)}$ we could take $h_S\tau^{-1}$. So, under the inductive assumption, Lemma 3.5 says that any other choice of $h_{\tau(S)}$ is eyeglass equivalent to $h_S\tau^{-1}$, or $h_{\tau(S)}\tau \sim h_S$. □

Proposition 3.8. *Suppose* S *and* S' *are not necessarily disjoint spheres, each in possibly different chambers of* M_A *or* M_B *and each is incompressible in the chamber in which it lies. Suppose* $h = h_S : (S^3, T) \to (S^3, T_g)$ *and* $h' = h_{S'} : (S^3, T) \to (S^3, T_g)$ *are homeomorphisms as described in Proposition 3.7, so, in particular,* S *and* S' *are both reducing spheres for* (S^3, T). *Under Assumption 3.4, the homeomorphisms* h *and* h' *are eyeglass equivalent.*

Proof. The proof is an examination of interlocking special cases.

Case 1: S and S' are isotopic rel T.

In this case S and S' both divide T into components $T_\pm$ of genus $g_\pm \geq 1$ where $g_+ + g_- = g$. It follows that $h(S) = S_{g_+}$ or S_{g_-}, say the former. After perhaps composing with the homeomorphism $h_\rho : (S^3, T_g) \to (S^3, T_g)$ defined before Lemma 3.5, we may as well assume that also $h'(S') = S_{g_+}$, so $h'h^{-1}$ leaves S_{g_+} invariant. Apply Lemma 3.5 to two copies of S_{g_+}; the homeomorphism $h'h^{-1}$; and the identity homeomorphism $id_{(S^3,T_g)}$. Deduce that $h'h^{-1} \sim id_{(S^3,T_g)}$. Hence $h \sim h'$.

Case 2: S and S' both lie in the same chamber C of $S^3 - F$ and there is an eyeglass move of T_C in C which carries S to S'.

As discussed in the proof of [FS2, Theorem 5.1], the handle slides that define the eyeglass move in C determine, up to eyeglass twists, parallel handle slides in T after amalgamation. So in this case there is also an eyeglass move of T in S^3 that carries S to a sphere isotopic to S' rel T. The proof of Case 2 then follows from Case 1.

Case 3: S and S' are disjoint.

Apply Lemma 3.5.

Case 4: S and S' both lie in the same chamber C and are isotopic in C.

The main theorem of [FS2] says that there is a sequence of reducing spheres for T_C starting at S and ending at S' so that each successive pair is either isotopic rel T_C, differ by an eyeglass move in (C, T_C), or are disjoint, because they differ by a bubble pass in T_C. The result then follows from Cases 1, 2 and 3.

Case 5: S and S' are isotopic in $S^3 - F$ to disjoint spheres.

If S and S' lie in different chambers of $S^3 - F$ the result follows from Case 3. Suppose S and S' lie in the same chamber C of $S^3 - F$ and can be isotoped *in* C, but not necessarily rel T_C, to be disjoint. According to the main theorem of [Sc1] the Heegaard surface T_C can be aligned with the sphere set $S \cup S'$, so that the S and S' become disjoint reducing spheres for T_C. This new alignment of S and S' may change h and h', but by Case 4 the eyeglass equivalence class of the homeomorphisms will not change. The result then follows from Lemma 3.5.

The general case If S and S' lie in different chambers of $S^3 - F$ the result follows from Case 3. Suppose they lie in the same chamber C. Ignoring T_C for the moment, recall that in classic 3-manifold theory a standard innermost disk argument shows that there is a sequence of incompressible spheres in C beginning with S and ending with S' so that sequential spheres are isotopic in C to disjoint spheres. The result then follows from Case 5. □

The following corollary is then a summary:

Corollary 3.9. *Suppose in a Heegaard split chamber complex (with notation as in conditions (1)-(4) preceding Lemma 3.6), the manifold $S^3 - F$ contains an incompressible sphere, then under Assumption 3.4, F determines a natural eyeglass equivalence class $\tilde{h}_F$ of homeomorphisms $(S^3, T) \to (S^3, T_g)$.*

Moreover, for $\tau \in G(S^3, T)$, $\tilde{h}_{\tau(F)}\tau = \tilde{h}_F$.

Proof. Choose any incompressible sphere S in $S^3 - F$ and let $h_S : (S^3, T) \to (S^3, T_g)$ be a homeomorphism as given by Proposition 3.7. Then Proposition 3.8 shows that, up to eyeglass equivalence, h_S is independent of any choice, including the choice of incompressible sphere S. The last sentence follows from the last sentence of Proposition 3.7. □

To simplify (but slightly abuse) notation, we will often use $h_F : (S^3, T) \to (S^3, T_g)$ to denote any representative of the eyeglass equivalence class $\tilde{h}_F$. With this notation the last sentence of 3.9 would be $h_{\tau(F)}\tau \sim h_F$.

Definition 3.10. *Suppose there are two Heegaard split chamber complexes in S^3 with corresponding associated surfaces $F, F' \subset S^3$ and each Heegaard split chamber complex supports the same Heegaard splitting surface T for S^3. Suppose each of $S^3 - F$ and $S^3 - F'$ contains an incompressible sphere, so (eyeglass equivalence classes of) homeomorphisms $h_F, h_{F'} : (S^3, T) \to (S^3, T_g)$ are defined. If $h_F \sim h_{F'}$ then F and F'* cocertify.

(This definition will be expanded later, see comments following Definition 8.2.)

Lemma 3.11. *Suppose, in Definition 3.10 $F \subset F'$ and a component C' of $S^3 - F'$ contains an incompressible sphere S that is also incompressible in the component $C \supset C'$ of $S^3 - F$ in which it lies. Then F and F' cocertify.*

Proof. Via [Sc1] isotope in C the incompressible sphere to a sphere $S \subset C$ that is aligned with T_C. By the Claim in the proof of Proposition 3.7 S becomes a reducing sphere for the splitting (S^3, T) after amalgamation. Similarly isotope S in C' to a sphere $S' \subset C'$ that is aligned with $T_{C'}$ and again conclude that S' upon amalgamation becomes a reducing sphere for (S^3, T). Then by definition $h_{S'} \in \tilde{h}_{F'}$ and $h_S \in \tilde{h}_F$. Since $C' \subset C$, the isotopies of both S and S' to alignment take place in C so, by Proposition 3.8, $h_{S'}$ also represents $\tilde{h}_F$. Since $\tilde{h}_F$ and $\tilde{h}_{F'}$ have the representative $h_{S'}$ in common, they cocertify. □

Corollary 3.12. *Suppose $F \subset S^3$ is the surface associated with a Heegaard split chamber complex that supports T. Suppose $\tau \in G(S^3, T)$. If the Heegaard split chamber complexes given by F and $\tau(F)$ cocertify, then $\tau \in \mathcal{E}$.*

Proof. Corollary 3.9 shows that $h_F \sim h_{\tau(F)}\tau$. On the other hand, the assumption that F and $\tau(F)$ cocertify implies that $h_{\tau(F)}\tau \sim h_F\tau$. Hence $h_F \sim h_F\tau$ or $id_{(S^3,T)} \sim \tau$, as required. □

3.3 An overview of the argument

Corollary 3.9 suggests a line of attack towards proving $G(S^3, T) = \mathcal{E}$: For each $\tau \in G(S^3, T)$, find a Heegaard split chamber complex in S^3 with associated surface F so that

- the Heegaard split chamber complex supports T
- some chamber contains an incompressible sphere and
- $h_F \sim h_{\tau(F)}$.

There are three conditions required for this example to work that are difficult to realize:

1. The Heegaard splitting of each ball chamber has genus ≥ 1, as required by the definition of a Heegaard split chamber complex. This property is needed to show, as in Lemma 3.6 and Proposition 3.7 above, that an incompressible sphere in a chamber determines a reducing sphere for T.
2. Some chamber contains an incompressible sphere
3. The Heegaard split chamber complexes derived from F and $\tau(F)$ cocertify.

Section 1 gave a brief description of how the **second requirement** is addressed in the general case: given a position of T in S^3, a pair of weakly reducing disks is found for T, with weak reduction defining a chamber complex. Next a particularly useful level sphere (a 'guiding sphere', see Section 9) S is found, and innermost disks in $S - T$ are used to decompose the chamber complex into a typically more complex one. The process is repeated until all circles of intersection are removed. It will be shown that at some stage in this *disk decomposition* process a chamber will contain an incompressible sphere or a sphere that is equally useful for determining an eyeglass equivalence class of homeomorphisms from $(S^3, T) \to (S^3, T_g)$. (See Definition 8.2 of a *certificate*.)

This process raises many technical questions: How to find a guiding sphere whose associated disk decomposition sequence must issue a certificate, Section 10. How to ensure, for a given sphere, that the certificates issued during this process cocertify, Section 8. The most difficult problem is showing that a different choice of such a sphere will result in an equivalent certificate. This takes many steps, occupying Sections 11 through Section 17. The philosophy behind these steps borrows heavily from [FS1] (see next paragraph): the guiding spheres are isotopic in S^3 through appropriate guiding spheres (Section 17) and such an isotopy between them can be broken into small steps, each of which either adds to, subtracts from, or delays a single disk in the associated disk decomposition sequence. We then examine the effect of each such minimal change. The upshot is that, up to cocertification, the certificate issued depends only on the Heegaard split chamber complex itself, and not on the choice of guiding sphere.

To address the **third requirement**, that is, to show that the Heegaard split chamber complexes derived from F and $\tau(F)$ cocertify, we construct a sequence of Heegaard split chamber complexes beginning with F and ending with $\tau(F)$ so that each successive pair cocertify, see Section 18. The heavy lifting for this part of the argument was done in [FS1]. Recall that a Heegaard split chamber complex is derived from a Heegaard splitting (S^3, T) by weak reduction, see Section 5. In [FS1] a series of pairs of weakly reducing disks is found so that, roughly, the resulting Heegaard split chamber complexes are a series beginning with some F and ending with $\tau(F)$, see Section 18. Moreover, successive pairs of weakly reducing disks are related in such a way that Proposition 17.8 guarantees the resulting Heegaard split chamber complexes cocertify. So in the end F and $\tau(F)$ cocertify and the third requirement above is satisfied.

This leaves the **first requirement**, that each chamber in the chamber complex comes with a splitting surface of genus ≥ 1. Establishing this is particularly vexing, and adds great technical complexity to the argument, for reasons we now describe:

It is necessary to eliminate, whenever they arise, any ball chambers that have genus 0 splittings. In the argument below these will be the ball chambers (called *goneballs*) that are absorbed into the surrounding chamber and so their bounding spheres disappear from F. Goneballs can arise in troublesome ways: a chamber that is a handlebody at one stage can be cut up by a disk decomposition to become a collection of balls; if the handlebody had trivial Heegaard splitting, the resulting balls will be goneballs and their boundaries need to be deleted from F. Since chambers may be deleted in this way, what is to prevent F from disappearing entirely, as pieces are cut up in this manner? We will show that unless the chamber complex is of a particularly simple type, called *tiny*, its defining surface will never completely disappear, see Subsection 5.4.

There is an additional but related complication: it is a classical result that a non-trivial Heegaard splitting of a ball or handlebody chamber is stabilized, that is it will contain bubbles (standard genus one summands). It follows from [FS2] that disk decomposition of a Heegaard split chamber gives a Heegaard splitting of the resulting chambers, but it is defined only up to eyeglass moves and passing bubbles through the decomposing disks. As noted above, it will be important to know if a newly created handlebody chamber could have trivial Heegaard splitting, so we want to flag handlebody chambers that, because of how they arise, are known to have non-trivial splittings. This gives rise to the notion of *flagged chamber complexes*: the exact Heegaard splitting of each chamber is typically unknown, but certain handlebody chambers (including ball chambers) are flagged as necessarily having non-trivial splittings. See Section 7.

Following this informal overview, we begin the more formal argument.

4 Chamber complexes and disk decomposition

Suppose M is a compact orientable 3-manifold. Let $F \subset S^3$ be a (typically disconnected) closed surface in int(M) that divides M into two typically disconnected 3-manifolds M_A and M_B, with each component of F incident on one side to M_A and on the other to M_B. Call the collection $\mathbb{C} = (F, M_A, M_B)$ a *chamber complex* with defining surface $F = F(\mathbb{C})$. Each component of $M - F$ is called a *chamber*, with those in M_A called A-chambers and those in M_B called B-chambers.

A Heegaard splitting $M = A \cup_T B$ is a familiar example of a chamber complex in M, with $\mathbb{C} = (T, A, B)$. A Heegaard splitting belongs to a class of chamber complexes, called *tiny*, that, for our purposes, contain too little information to be useful. However, a Heegaard splitting together with a pair of weakly reducing disks (see Section 5 below) does produce a possibly useful chamber complex, through a process called *disk decomposition*, or, more formally, *chamber complex decomposition*, which we now describe.

Phase one of disk decomposition: For $\mathbb{C}$ a chamber complex in M, let $\mathcal{D}$ be a collection of disjoint properly embedded disks, some perhaps inessential, in the chambers of $\mathbb{C}$, with $\partial\mathcal{D}$ disjoint from ∂W, so $\partial\mathcal{D} \subset F$. We will call $\mathcal{D}$ a *disk set* in $\mathbb{C}$. Let $F_{\mathcal{D}} \subset M$ be the surface obtained by doing surgery on $F(\mathbb{C})$ along the disks $\mathcal{D}$, and denote by $\hat{\mathbb{C}}_{\mathcal{D}}$ the associated chamber complex $M - F_{\mathcal{D}}$. (The reason for the circumflex will be given shortly.) In particular, the surgery alters M_A by deleting a bicollar neighborhood of each disk in $\mathcal{D} \cap M_A$ and adding a bicollar neighborhood of each disk in $\mathcal{D} \cap M_B$. The symmetric statement applies to M_B.

Each disk D in $\mathcal{D}$ is the core of a 2-handle used in the surgery as just described. The bicollar neighborhood of ∂D in F will be called the *belt annulus* of the surgery; the two copies of D in $F_{\mathcal{D}}$ that are the result of the surgery are called the *scars* of D. Let C' be a chamber of $\hat{\mathbb{C}}_{\mathcal{D}}$ and $D \in \mathcal{D}$ a disk which leaves a scar or two on $\partial C'$. If the disk D lies outside the new chamber C' then so does the associated belt annulus and we call the scars *external* scars on $\partial C'$. If the disk lies inside C' we call the scars *internal* scars on $\partial C'$. Clearly if D leaves one internal scar on $\partial C'$ it will leave two, since the associated belt annulus lies in C', but the two internal scars may be on different components of $\partial C'$.

One can think of a chamber $\hat{C}$ of $\hat{\mathbb{C}}_{\mathcal{D}}$ as obtained in two stages: it starts as the complement C_- of a bicollar of $\mathcal{D} \cap C$ in a chamber C of $\mathcal{D}$, a result of surgery on $\mathcal{D} \cap C$. Then 2-handles are added along those disks in $\mathcal{D}$ that are incident to ∂C but lie outside C. The chamber $\hat{C}$ is called a *remnant* of C, even though strictly speaking it doesn't lie entirely inside of C.

Phase two of disk decomposition: Among the components of the surgered surface $F_{\mathcal{D}}$ may be spheres that bound balls in M. In the second phase of disk decomposition a collection of these balls is chosen and, for each ball $G \subset M$ in the collection (henceforth called a *goneball*) all components of $F_{\mathcal{D}}$ that lie in G , including the sphere ∂G, are removed from $F_{\mathcal{D}}$. Whether a ball in M bounded by a sphere in $F_{\mathcal{D}}$ will be chosen to be a goneball of the decomposition is a delicate and important part of the theory. (See, for example, Rules 4.4 and 5.6.) After eliminating all components of $F_{\mathcal{D}}$ that lie in goneballs, call the resulting chamber complex $\mathbb{C}_{\mathcal{D}}$ (note: no circumflex).

Any chamber C_D in the chamber complex $\mathbb{C}_{\mathcal{D}}$ is obtained from a chamber $\hat{C}_D$ in $\hat{\mathbb{C}}_{\mathcal{D}}$ by attaching balls, each corresponding to a goneball for the decomposition. We then say that C_D is a remnant of the chamber C in $\mathbb{C}$ if $\hat{C}_D$ is a remnant of C in $\hat{\mathbb{C}}_{\mathcal{D}}$. Notice that a chamber C in $\mathbb{C}$ may have no remnants in $\mathbb{C}_{\mathcal{D}}$. For example, C might be handlebody for which $\mathcal{D}$ contains a complete collection of meridian disks, so the remnants in $\hat{\mathbb{C}}_{\mathcal{D}}$ are all balls. If these all happen to be declared goneballs, then there are no remnants of C in $\mathbb{C}_{\mathcal{D}}$.

For $\mathbb{C}, \mathcal{D}, F, F_{\mathcal{D}}, \mathbb{C}_{\mathcal{D}}$ and $\hat{\mathbb{C}}_{\mathcal{D}}$ as above, suppose F' is a component of the surface $F_{\mathcal{D}}$ such that one of the complementary components W of F' in M is a handlebody. (A ball is considered a genus 0 handlebody). W is not necessarily a chamber of $\hat{\mathbb{C}}_{\mathcal{D}}$, but rather the union of those chambers that lie within it.

Definition 4.1. *The handlebody W is* disky *if each component of $F \cap \text{int}(W)$ is a disk.*

Lemma 4.2. *If W is disky then*

1. *No component of F lies entirely in the interior of W.*
2. *Each component of $F_{\mathcal{D}}$ that lies in the interior of W is a sphere.*
3. *Suppose $D \in \mathcal{D}$ lies in the interior of W. Then the corresponding two scars do not lie on the same component of the surface $F_{\mathcal{D}}$.*

Proof. Take the statements in order:

1) Each component of F is a closed surface, so none can be a disk.

2) Suppose F'' is a component of $F_{\mathcal{D}}$ that lies in the interior of W, and let F''_- be the compact surface obtained from F'' by deleting the interior of all scars of the disk surgery that lie in F''. Then F''_- lies in a component of $F \cap \text{int}(W)$ and, by assumption, this component is a disk. Hence the surface F''_- has no genus, so neither can F''. See Figure 10.

3) Suppose the two scars left by surgery on a disk $D \in (\mathcal{D} \cap \text{int}(W))$ were on the same component F'' of $F_{\mathcal{D}}$. If $F'' = F'$ then the belt annulus for the surgery is a component of $F \cap \text{int}(W)$ that is not a disk, contradicting our assumption that W is disky. Similarly, suppose F'' were in the interior of W, and let α be an arc in F'' running between the two scars. Then the union of the belt annulus and a collar of α would be a punctured torus lying in some component of $F \cap \text{int}(W)$, so that component cannot be a disk, again contradicting our assumption that W is disky. □

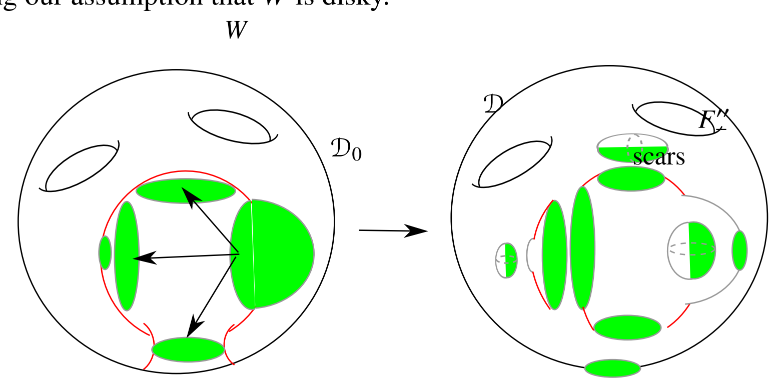

Figure 10: Construction of F''_- in proof of Lemma 4.2 (2)

Lemma 4.3. *Suppose the handlebody W is disky, and F'' is a component of $F_{\mathcal{D}}$ that lies in the interior of W. Then F'' bounds a disky ball in W.*

Proof. By Lemma 4.2 any such component F'' is a sphere and, since the handlebody W is irreducible, it will bound a ball in W. The issue is to show that such a ball is disky. By assumption, $F \cap \mathrm{int}(W)$ consists of a collection E of disks; if $E = \emptyset$ then there are no components of $F_{\mathcal{D}}$ in the interior of W and there is nothing to prove.

Otherwise, since E lies inside W, the annulus in E adjacent to any circle $c \in \partial E$ is a belt annulus for a surgery disk $D \in \mathcal{D}$ that lies in W, with one of its scars on ∂W. Let $\mathcal{D}_0 \subset \mathcal{D}$ be the subset of disks that lie in W; we now induct on $n = |\mathcal{D}_0| > 0$.

Let $D_0 \in \mathcal{D}_0$ be a disk whose boundary is innermost in E among the circles $\partial\mathcal{D}_0$; let E_0 be the subdisk of E that it bounds. Since W is irreducible, the disks D_0 and E_0 together bound a ball B_0 in W. The ball is disjoint from F by Lemma 4.2, so we can think of D_0 as just being parallel to E_0, via the ball B_0. Inductively apply the lemma to $\mathcal{D}' = \mathcal{D}_0 - D_0$, creating a collection S' of spheres, each of which, by inductive assumption, bounds a disky ball in W.

Finally, do surgery on the disk D_0. The result of the surgery is two-fold: The ball B_0 bounded by E_0 and a copy of D_0 becomes a ball chamber that is disky because it is empty. And the sphere S'_0 in S' that contains E_0 is altered, in effect, by isotoping E_0 across B_0 to the other copy of D_0. Depending on which side of S'_0 the disk D_0 lies, this change in S'_0 may add the disk $E_0 \subset F$ to the ball that S'_0 bounds in W. But since E_0 is a disk, it does not change the fact that the ball is disky. See Figure 11. ◻

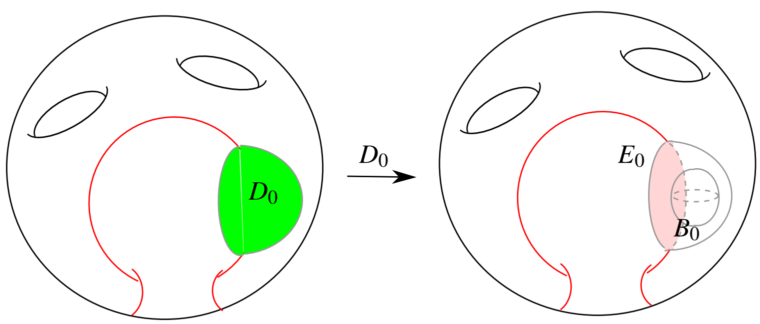

Figure 11: $E_0 \subset F$ in the interior of ball bounded by S'_0. (Figure not incorporating decomposition by $\mathcal{D}' = \mathcal{D}_0 - D_0$.)

The decision process that will be used to decide if sphere components of $F_{\mathcal{D}}$ in a chamber complex decomposition bound goneballs will have certain properties. The ultimate decision process is subtle, but some of those properties can be described already. At this point, view them as rules that disk decompositions follow; later it will be shown that the decision process that will be used satisfies these rules. To that end, declare the rule:

Rule 4.4. *In a chamber complex decomposition, only disky balls can be goneballs.*

Following Lemma 4.3 we then have

Corollary 4.5. *In the second phase of a chamber complex decomposition, any component that is removed from $F_{\mathcal{D}}$ is a sphere that bounds a disky ball.*

Continue with $\mathbb{C}, \mathcal{D}, F, F_{\mathcal{D}}, \hat{\mathbb{C}}_{\mathcal{D}}$ and $\mathbb{C}_{\mathcal{D}}$ be as above. Denote the two-stage operation of chamber complex decomposition just described by

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}.$$

The next rule limits the amount of disturbance caused by adding or removing a single disk from the chosen disk set.

Rule 4.6. *Suppose $\mathcal{D}$ is a disk set in $\mathbb{C}$, $D \in \mathcal{D}$, and $\mathcal{D}_-$ is the disk set $\mathcal{D} - D$. Let S be a sphere component of $F_{\mathcal{D}}$ that does not contain either scar from D. Then S bounds a goneball in $\mathbb{C}_{\mathcal{D}}$ if and only if it bounds a goneball in $\mathbb{C}_{\mathcal{D}_-}$.*

Corollary 4.7. *Suppose $\mathcal{D}$ consists of a single disk. Then a sphere in $F_{\mathcal{D}}$ that does not contain a scar of D does not bound a goneball.*

Proof. Since $\mathcal{D}_- = \emptyset$ such a sphere would necessarily be a component of $F = F_{\mathcal{D}_-}$ and so does not bound a goneball in $\mathbb{C} = \mathbb{C}_{\mathcal{D}_-}$. By Rule 4.6 it does not bound a goneball in $\mathbb{C}_{\mathcal{D}}$. □

An example of a decision process that satisfies both rules is to simply declare each disky ball to be a goneball. For example, it satisfies Rule 4.6 because a sphere $S \subset F_{\mathcal{D}}$ not containing either scar will also be in $F_{\mathcal{D}_-}$ and a ball it bounds is disky in either $\mathbb{C}_{\mathcal{D}}$ or $\mathbb{C}_{\mathcal{D}_-}$ if and only if F intersects the ball only in disks. Unfortunately declaring each disky ball to be a goneball would be too broad for our purposes; too much information would be lost. But the rules so far do suffice to prove an important property:

Proposition 4.8. *Suppose C is a chamber of $\mathbb{C}$ so that every remnant of C in $\mathbb{C}_{\mathcal{D}}$ is a disky handlebody. Then C is a handlebody.*

Proof. We proceed by induction on $|\mathcal{D}|$, noting that if $\mathcal{D} = \emptyset$ there is nothing to prove.

Case 1: There is a disk in $\mathcal{D}$ whose boundary is inessential in ∂C.

In this case, let D be a disk whose boundary is innermost among all such disks, let E be the disk in $\partial C - \mathcal{D}$ bounded by ∂D. Push $\mathrm{int}(E)$ slightly into the chamber of $\mathbb{C}$ on the opposite side of D. That chamber is either C' if D lies in C (the top row of Figure 12) or C itself, if D lies in a chamber C' adjacent to C (the bottom row of Figure 12). Consider the sphere $S = D \cup E$, lying either in a remnant of C' or a remnant of C, and in either case parallel in that remnant to a sphere component S' of $F_{\mathcal{D}}$.

Claim 1: With the proper choice of D for the construction above, S bounds a ball B_S in M containing S'. Moreover, if D lies in C', so S is in a remnant of C, then the subball of B_S bounded by S' is a goneball.

Proof of Claim 1: Suppose first that S lies in a remnant C'_r of C', so D lies inside C, as in the top panel of Figure 12. Then S' is a sphere in the boundary of a chamber C_r of $\hat{\mathbb{C}}_{\mathcal{D}}$ that is a remnant of C. If C_r is a goneball the claim follows. If instead C_r remains as a chamber in $\mathbb{C}_{\mathcal{D}}$, C_r must be a ball, since every remnant of C in $\mathbb{C}_{\mathcal{D}}$ is a handlebody, proving the claim in this case.

Suppose instead that S lies in a remnant C_r of C, so D lies inside C', as in the bottom panel of Figure 12. Then by hypothesis C_r is a handlebody, and so is irreducible. Hence S bounds a ball B_S in C_r. If B_S lies on the side of S containing S' the claim is shown. Suppose B_S lies on the other side of S, so S' bounds a ball in C_r and that ball contains the punctured handlebody $\partial C - E$. Since each remnant of C is disky, it follows that $\partial C - E$ must be a disk, so C is a ball. In this case use instead of D a disk D' in

$\mathcal{D}$ whose boundary is innermost among those in the disk $\partial C - E$ and repeat the construction. The new sphere corresponding to S lies in B_S so it bounds a ball entirely in C_r, a ball that contains the disk E' in $\partial C - E$ bounded by $\partial D'$, i. e. the one that now corresponds to E. This proves Claim 1, using the disk D' instead of D.

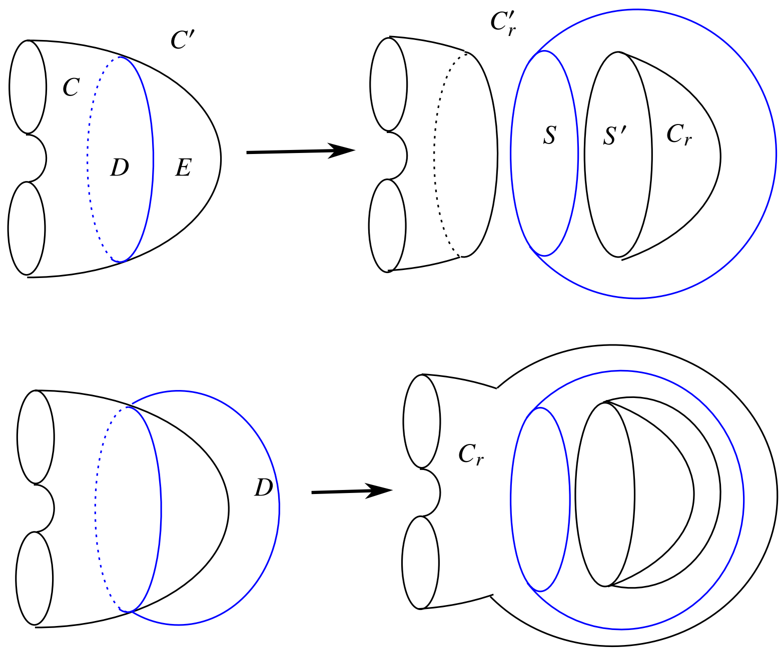

Figure 12: When a disk $D \in \mathcal{D}$ has ∂D inessential in ∂C

Claim 2: The disk D of Claim 1 is inessential in the chamber (either C or C') in which it lies.

Proof of Claim 2: Suppose instead that D is essential in the chamber in which it lies, so there are closed components of F lying between D and E, i. e. in the ball B_S. By Lemma 4.2 the subball of B_S bounded by S' could not then be a disky ball, so by Rule 4.4, S' remains as a sphere component of $F(\mathbb{C}_{\mathcal{D}})$. Claim 1 then implies that D lies in C so S is in a remnant of C'. Then, as in Claim 1, S' is a sphere in the boundary of a chamber of $\hat{\mathbb{C}}_{\mathcal{D}}$ that is a remnant of C. Since S' remains in $F(\mathbb{C}_{\mathcal{D}})$ the ball it bounds in B_S must be a ball remnant B' in $\mathbb{C}_{\mathcal{D}}$ since every remnant of C is a handlebody. But even then there is a contradiction, for every remnant of C is assumed to be disky, and here B' contains closed components of F, contradicting Lemma 4.2, and so proving Claim 2.

Let $\mathcal{D}_- = \mathcal{D} - D$. Observe that since D is inessential by Claim 2, $\mathcal{D}_-$ still satisfies the hypotheses of Proposition 4.8: Indeed, from Case 1 and Rule 4.6, the only difference in the remnants of C if we add D back to $\mathcal{D}_-$ before decomposing is:

- possibly adding a disky ball bounded by S' to the set of remnants of C or
- isotoping E to D, if the ball that S' bounds is a goneball. This is the case when $D \in C'$ and may be the case when $D \in C$.

Neither will affect whether each remnant of C is a disky handlebody. So, by inductive assumption, replacing $\mathcal{D}$ with $\mathcal{D}_-$ leads to the conclusion of Proposition 4.8, namely C is a handlebody. This concludes the proof of Proposition 4.8 for Case 1.

Case 2: Any disk incident to ∂C has essential boundary in ∂C. (In particular, any disk incident to ∂C is essential.)

First consider a remnant C_r of C in $\mathbb{C}_{\mathcal{D}}$, which by hypothesis of the Proposition is a disky handlebody. Since C_r is a remnant of C, each scar on ∂C_r is incident to ∂C, so by the hypothesis of this case its boundary is essential in ∂C. But since C_r is disky, this implies it can have no internal scars, i. e. F is disjoint from $\text{int}(C_r)$, since the belt annulus for an internal scar lies on a component of $\partial C \cap \text{int}(C_r)$ and these are all disks. This in turn implies that $\text{int}(C_r)$ contains no goneballs, so C_r is a remnant of C in $\hat{\mathbb{C}}_{\mathcal{D}}$, before goneballs are absorbed. Thus the remnants of C in $\hat{\mathbb{C}}_{\mathcal{D}}$ consist of a union of disky handlebodies, though among these may be balls that become goneballs in $\mathbb{C}_{\mathcal{D}}$. See left side of Figure 13.

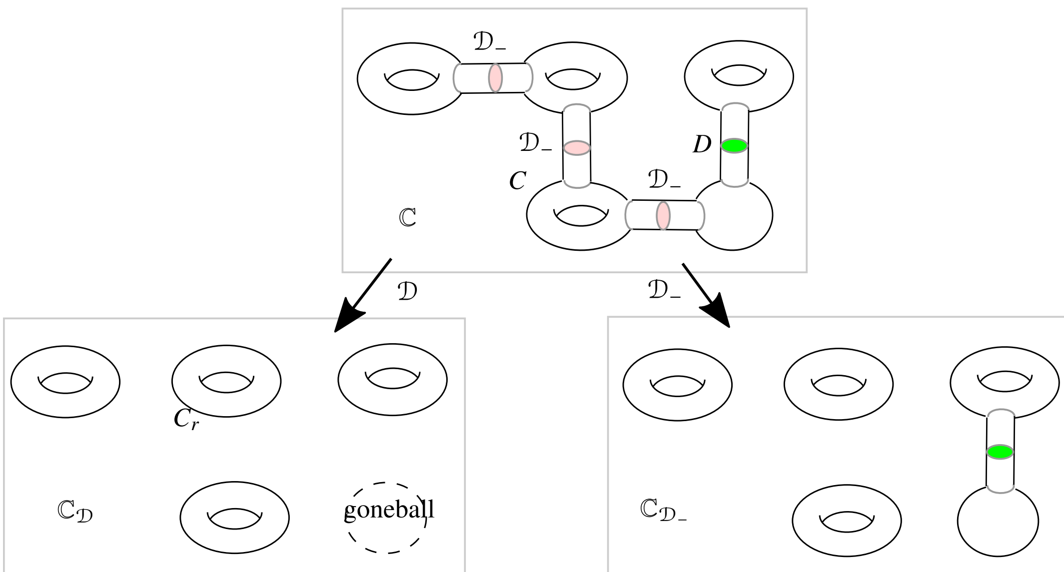

Figure 13: When each disk $D \in \mathcal{D}$ has ∂D essential in ∂C

Let D be any disk in $\mathcal{D}$ and, as in Case 1, let $\mathcal{D}_- = \mathcal{D} - D$. Inductively, it suffices to show that $\mathcal{D}_-$ still satisfies the hypotheses of the Proposition, namely that every remnant of C in $\mathbb{C}_{\mathcal{D}_-}$ is a disky handlebody. If D is not incident to ∂C then removing D from $\mathcal{D}$ has no effect on the remnants of C in $\hat{\mathbb{C}}_{\mathcal{D}_-}$, so the remnants are the same as those in $\hat{\mathbb{C}}_{\mathcal{D}}$. By Rule 4.6, the remnants of C in $\mathbb{C}_{\mathcal{D}_-}$, are then also the same as those in $\mathbb{C}_{\mathcal{D}}$, namely a union of disky handlebodies. Thus in this case $\mathcal{D}_-$ satisfies the hypothesis of Proposition 4.8 as required.

What remains is the case that D is incident to ∂C. See right side of Figure 13. Since there are no internal scars on remnants of C, the two scars in $\hat{\mathbb{C}}_{\mathcal{D}}$ left by D are external scars, so D lies in C. Thus $\hat{\mathbb{C}}_{\mathcal{D}_-}$ is obtained from $\hat{\mathbb{C}}_{\mathcal{D}}$ by simply attaching a 1-handle dual to D to a component, or between two components, of $\hat{\mathbb{C}}_{\mathcal{D}}$. Since each such component is a disky handlebody, the result of adding the 1-handle is also a handlebody, and it is disky because its interior is still disjoint from F. Hence $\hat{\mathbb{C}}_{\mathcal{D}_-}$ is a union of disky handlebodies. $\mathbb{C}_{\mathcal{D}\_}$ is obtained from $\hat{\mathbb{C}}_{\mathcal{D}\_}$ by removing some ball components (the goneballs). Hence $\mathbb{C}_{\mathcal{D}\_}$ is also a union of disky handlebodies, so it satisfies the hypothesis of Proposition 4.8 as required. □

Definition 4.9. *A chamber complex $\mathbb{C}$ in M, with defining surface $F = F(\mathbb{C})$, is* tiny *if either*

- *$F = \emptyset$ or*
- *there is a chamber C of $\mathbb{C}$ so that $M - C$ consists of handlebody chambers. These chambers are called the* designated handlebody chambers*.*

A Heegaard surface in which at least one complementary component is a handlebody is a familiar example of a defining surface of a tiny chamber complex. On the other hand, a weak reduction of a Heegaard surface will yield a chamber complex that is not tiny. That fact (see Proposition 5.18) will be the entryway to our study of Heegaard splittings below. The word "tiny" is used because the tree dual to a chamber complex is particularly small for tiny chamber complexes: at its most complicated, it is a star graph with central vertex corresponding to the chamber C.

Proposition 4.10 (Tinyness pulls back)**.** *Suppose $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$ is a chamber complex decomposition in which the chamber complex $\mathbb{C}_{\mathcal{D}}$ is tiny. Suppose further, if the defining surface of $\mathbb{C}_D$ is not empty, that each of the handlebody chambers named in Definition 4.9 is disky. Then $\mathbb{C}$ is tiny.*

Proof. We first consider the case in which the defining surface of $\mathbb{C}_D$ is not empty. That is, per Definition 4.9, there is a single chamber C' of $\mathbb{C}_{\mathcal{D}}$ so that $M - C'$ consists of a union W' of disky handlebody chambers. Without loss, assume C' is a B-chamber, so each component of W' is an A-chamber adjacent to C' in $\mathbb{C}_{\mathcal{D}}$. See Figure 14.

Since each component of W' is a disky handlebody, any disk $D \in \mathcal{D}$ that lies in $\text{int}(W')$ lies on a disk in F and so has inessential boundary on F; let D be one whose boundary is innermost on F. Surgery on D creates a ball in $\text{int}(W')$, which must be a goneball since W' is irreducible. The effect then of surgery on D followed by declaring the ball gone is merely to isotope $F \cap \text{int}(W')$; it has no effect on either the hypothesis of the proposition or the conclusion. Hence we may inductively assume that $\mathcal{D} \cap \text{int}(W') = \emptyset$, so W' has only external scars.

Let $B_{\mathbb{C}}$ denote the collection of B-chambers in $\mathbb{C}$. C' is a B-chamber in $\mathbb{C}_{\mathcal{D}}$, in fact the only B-chamber in $\mathbb{C}_{\mathcal{D}}$. Recall that it is obtained from $B_{\mathbb{C}}$ in three steps:

- $B_{\mathbb{C}}$ is cut along those disks in $\mathcal{D}$ that lie in $B_{\mathbb{C}}$.
- Two-handles are added along those disks in $\mathcal{D}$ that lie in the A-chambers
- Goneballs are eliminated.

Towards understanding the third step, eliminating goneballs, notice that no goneball can lie in $\text{int}(W')$ since W' has no disks of $\mathcal{D}$ in its interior and, in particular, only external scars. Suppose then that U is a maximal goneball in the B-chamber C'. Since U is maximal and lies in the B-chamber C', the chamber of $\hat{\mathbb{C}}_{\mathcal{D}}$ immediately inside U is an A-chamber. Rule 4.4 says that U is disky. Applying the same argument as just applied to W', we may as well then assume that the only scars on ∂U are external scars, so U itself is an A-chamber in $\hat{\mathbb{C}}_{\mathcal{D}}$. With this simplification, so that the only goneballs of the decomposition are A-chambers, none of the three steps in the process above decreases the number of B-chambers (though the first step may increase it). Since, by hypothesis, there is only one B-chamber in $\mathbb{C}_{\mathcal{D}}$ we conclude that $B_{\mathbb{C}}$ is also a single chamber in $\mathbb{C}$.

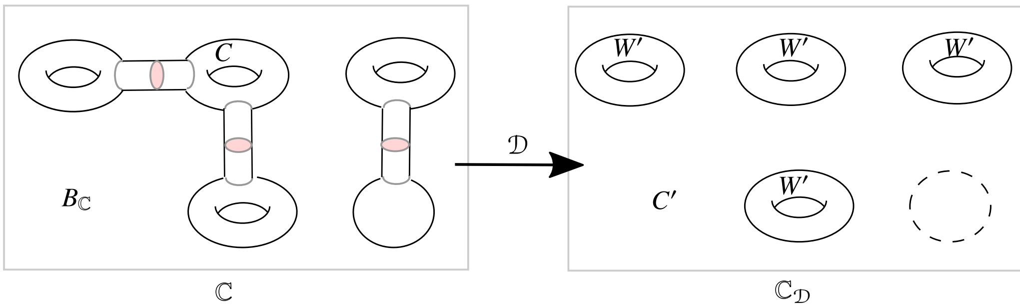

Figure 14: Tininess when the defining surface of $\mathbb{C}_D$ is not empty

Consider any chamber C of $\mathbb{C}$ other than $B_{\mathbb{C}}$. C is then an A-chamber in $\mathbb{C}$ and so has the property that each of its remnants in $\mathbb{C}_{\mathcal{D}}$ is in W', so each is a disky handlebody. It then follows from Proposition 4.8 that C itself is a handlebody. Hence $\mathbb{C}$ consists of a union of handlebodies (the A-chambers) whose complement $B_{\mathbb{C}}$ is a single B-chamber. That is, $\mathbb{C}$ is tiny.

The other possibility from Definition 4.9 is that $\mathbb{C}_{\mathcal{D}}$ is tiny because the defining surface for $\mathbb{C}_{\mathcal{D}}$ is empty. This means that in $\mathbb{C}_{\mathcal{D}}$, M is entirely a B-chamber, say. This implies that either F was empty (completing the argument) or surgery on $\mathcal{D}$ resulted only in spheres bounding goneballs. In that case, the argument is basically the same: Rule 4.4 says that each goneball is disky. Then, examining maximal goneballs as above we can assume without loss that each goneball is an A-chamber with only external scars. This implies that every A-chamber in $\mathbb{C}$ can be recovered from a collection of balls (the maximal goneballs) by attaching 1-handles, dual to the disks that leave external scars. The result is visibly a collection of handlebodies. So again $\mathbb{C}$ is tiny. □

5 Heegaard splittings and chamber complexes

5.1 Review of weak reduction and amalgamation

Chamber complexes are naturally relevant to weakly reducible Heegaard splittings, as we now describe. This first subsection is essentially a review of relevant known results; see, for example, [La, Section 3] to which we refer for relevant notation.

Suppose, in a Heegaard splitting $M = A \cup_T B$ of a compact 3-manifold M, $\mathcal{A}, \mathcal{B}$ are disjoint families of disks, properly embedded in A, B respectively, with $\partial\mathcal{A}, \partial\mathcal{B} \subset T$ and at least one member of each disk family an essential disk. Such a pair of families is called a *weakly reducing* disk family, a notion with a long and important history in the study of Heegaard splittings (cf [CG]). Surger T along $\mathcal{A} \cup \mathcal{B}$ and call the resulting surface $\hat{F}$.

$\hat{F}$ defines a chamber complex $\hat{\mathbb{C}}$ in M as follows. One set of chambers, called the A-chambers, are derived from A in two stages: first bicollars of disks in $\mathcal{A}$ are removed (so these disks lie *outside* the resulting A-chambers) and then bicollar neighborhoods of $\mathcal{B}$ are attached (so these are disks that lie *inside* the resulting A-chambers). The remaining chambers, the B-chambers are derived from B in the symmetric fashion. The result is the perhaps counter-intuitive fact that after the surgery, each A-chamber may contain some disks from $\mathcal{B}$, but none from $\mathcal{A}$, and symmetrically for the B-chambers. In any pair of adjacent chambers, one is an A-chamber and the other is a B-chamber, since the surface between them is (except for some disks) a submanifold of T.

Each A-chamber C (and symmetrically for B-chambers) inherits a natural Heegaard splitting $C = A_C \cup_{T_C} B_C$. See Figure 15. This can be demonstrated by considering the two-stage construction just described, see Figure 15 : In the first stage, (NE to SW in Figure 15) the surface T is compressed in A along $\mathcal{A}$ and a component C' is chosen. Since A was a compression body, so is C'. Let $T_C \subset C'$ be the end of a collar of $\partial_+ C'$ in C', so $T_C \subset \text{int}(C')$ (shown in red in the figure). T_C is a trivial Heegaard surface for C', dividing it into the collar of $\partial_+ C'$ and a copy A_C of C' lying entirely in $\text{int}(C')$. In the second stage of the construction of C (bottom row of Figure 15) 2-handles lying in $\mathcal{B}$ are attached to C' along $\partial_+ C'$, turning the collar of $\partial_+ C'$ into a compression body B_C. Thus T_C remains a Heegaard surface for C. (Note that, as in the outermost surface in the final panel of Figure 15, C may have spherical boundary components, so A_C and B_C would classically be described as punctured compression bodies, as discussed on [Sc1] and [FS2].). The new boundary components of C, that is $\partial C - \partial M = \partial C - \partial_- A$ are, except for the scars of the surgery, subsurfaces of T.

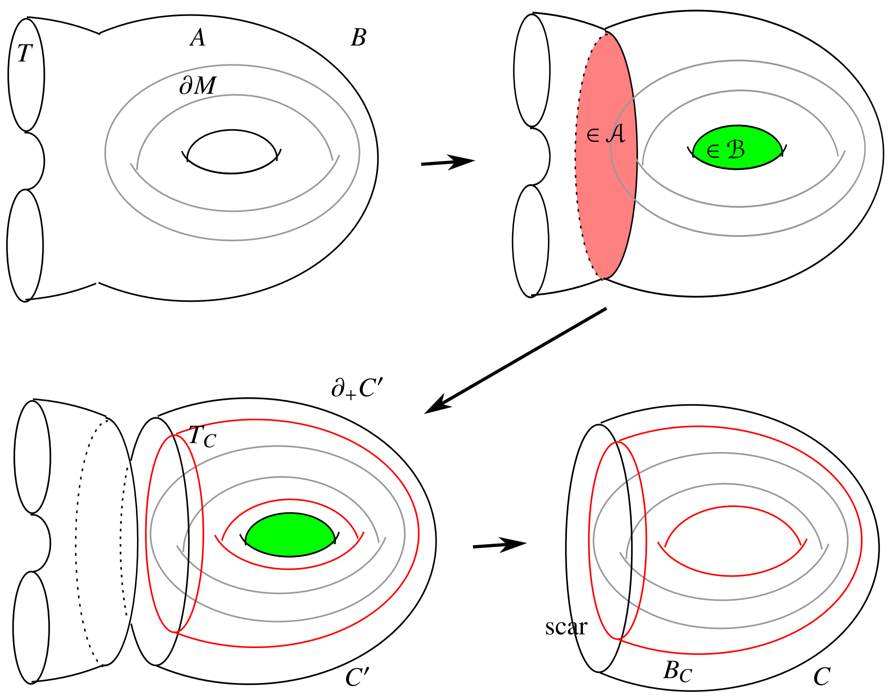

Figure 15: Inherited Heegaard splittings in disk decomposition

Note that in this construction, the result is technically not a Heegaard split chamber complex as defined before Lemma 3.6 because that definition requires that in each A-chamber C, A_C is a handlebody, whereas in the construction above $\partial_- A_C$ may contain components of $\partial_- A$. We will be mostly concerned with a special form of Heegaard splitting in which this is not an issue:

Definition 5.1. *A Heegaard splitting $M = A \cup_T B$ is called a* pure *Heegaard splitting if all the components of ∂M lie in ∂B (or all in ∂A). Put another way, it is pure if either A or B is a handlebody.*

Obviously if ∂M has at most one boundary component then the splitting is pure. Moreover,

Proposition 5.2. *Suppose $\mathbb{C}$ is the chamber complex resulting from weak reduction on a Heegaard splitting of a closed* 3*-manifold. Then the Heegaard splitting of each chamber in $\mathbb{C}$ is pure, so the result of the weak reduction is a Heegaard split chamber complex.*

Proof. Since $M = A \cup_T B$ is closed, both A and B are handlebodies. Suppose C is an A-chamber and consider the construction of its Heegaard structure above. (In Figure 15 take $\partial M = \emptyset$.) In the first stage, A is cut up by $\mathcal{A}$ into handlebodies and a component C' of the result is chosen. Next T_C is defined as the end of a collar of $\partial C'$ in C'. A_C is the complement of the collar, so in particular the surface $\partial C'$ is entirely disjoint from the handlebody A_C. That doesn't change when B_C is created by adding 2-handles to $\partial C'$ to create ∂C. Thus ∂C lies entirely in $\partial_- B_C$, as required. □

There is a natural construction that is inverse to weak reduction: Suppose we are given a chamber complex $\hat{\mathbb{C}}$ with defining surface $\hat{F}$ in which we are able to alternately label the chambers A-chambers and B-chambers, for example when each component of $\hat{F}$ is separating in M. Pick a pure Heegaard splitting $C = A_C \cup_{T_C} B_C$ for each chamber C. These Heegaard splittings induce a Heegaard splitting $M = A \cup_T B$ by amalgamating the Heegaard splittings along $\hat{F}$, and the amalgamated Heegaard splitting is unique up to isotopy, cf [La, Proposition 3.1].

As described in [La], amalgamation in this setting proceeds as follows: for each A-chamber $C = A_C \cup_{T_C} B_C$ choose a complete set of meridian disks $\mathcal{B}_C$ for the compression body B_C. That is, if B_C is cut along $\mathcal{B}_C$ the result is a collar of $\partial_- B_C = \partial C - \partial M$. The disks $\mathcal{B}_C$ define a natural spine γ_C for $B_C \subset C$, namely the union of the surface $\partial C - \partial M$ and a collection of arcs in the interior of C, specifically the arc cocores of the 2-handle neighborhoods of $\mathcal{B}_C$. The arcs are then extended down through the collar using its product structure. Do the dual construction in each B-chamber. Then A can be recovered from $\hat{\mathbb{C}}$ by deleting from each A-chamber C a neighborhood of the graph γ_C and attaching a neighborhood of the graph constructed in each B-chamber. (And, as usual, symmetrically for B.)

An important point for our purposes is that it is possible to choose a *different* spine γ'_C for B_C in each A-chamber C, by originally making a different choice $\mathcal{B}'_C$ of meridian disks. This will result in a different set of arcs and so a different set of of 1-handles but will not change the recovered Heegaard splitting (and dually for B-chambers). As argued more formally in [La] one can get from the arcs of γ_C to those of γ'_C by edge slides, which correspond to handle-slides in B that replace $\mathcal{B}_C$ with $\mathcal{B}'_C$ (and dually for B-chambers).

Two cautionary notes on pure Heegaard splittings:

a). The result of amalgamating pure Heegaard splittings may not be pure. For example, suppose F is a closed surface and $F \times I$ is given a pure Heegaard splitting. Take two copies of this pure Heegaard split

$F \times I$ and attach them along one end of each to get a combined copy of $F \times I$. The Heegaard splitting of $F \times I$ obtained by amalgamating the two splittings will not be pure. See top panel of Figure 16.

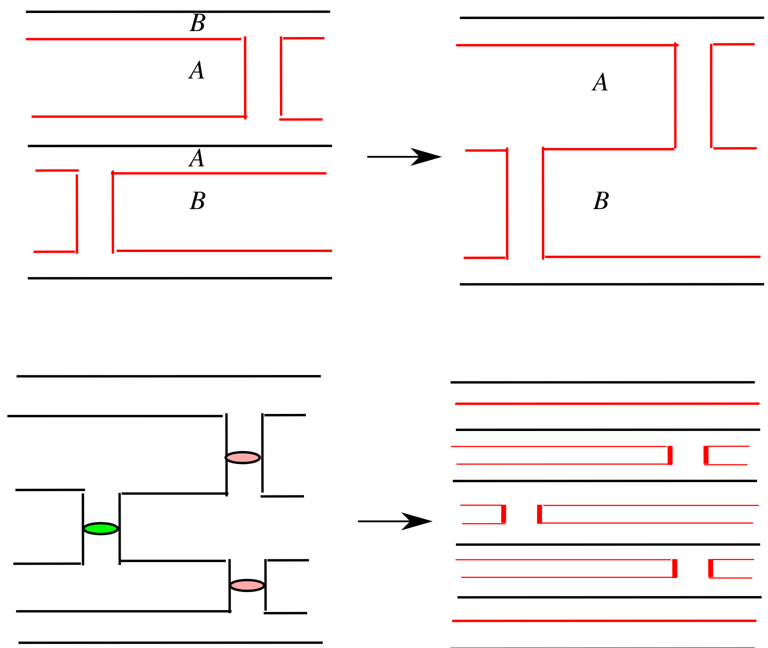

Figure 16: Amalgamation and weak reduction may destroy purity

b) Weak reduction, even of a pure Heegaard splitting, may not result in pure Heegaard splittings in every chamber. For example, in $M = F \times I$ consider the following natural splitting: begin with four horizontal copies of F in the interior of M labeled in order $F_1, ..., F_4$, breaking M into 5 copies of $F \times I$, which we will call collars and which we alternately label A and B, so that both the top and bottom collar are labeled A. In each of the three collars between F_i and $F_{i+1}, i = 1,2,3$ tube F_i to F_{i+1} by a vertical tube. The result is a pure Heegaard splitting $M = A \cup_T B$ of genus $4 \cdot \text{genus}(F)$ in which each collar labeled A becomes part of A, and symmetrically. Now weakly reduce along the meridians of the 3 tubes, two of them disks in A and one of them in B, creating a chamber complex $\mathbb{C}$ with 5 chambers, alternately A-chambers and B-chambers and each a copy of $F \times I$. The middle 3 chambers all have pure splittings, but the chambers incident to $\partial_- M$ are each copies of $F \times I$ split into two copies by a single copy of F, and so are not pure. See bottom panel of Figure 16.

Here are some useful preliminary results. We use the standard convention that the genus of a surface is the sum of the genera over all components.

Lemma 5.3. *Suppose $M = A \cup_T B$ is a pure Heegaard splitting of a 3-manifold M. Then for any collection of components $F \subset \partial M$*

- $\text{genus}(T) \geq \text{genus}(F)$.
- $\text{genus}(T) = \text{genus}(F)$ *only if each component of $\partial M - F$ is a sphere.*

Proof. With no loss assume $\partial M = \partial_- B$. A spine of B consists of the union of $\partial_- B = \partial M$ and a graph γ that intersects $\partial_- B$ exactly in some valence 1 vertices. Since B is connected there has to be at least one such end vertex in each component of $\partial_- B$ and, after edge slides, we can assume there is exactly one in each, and that γ is connected. Recall that for η a regular neighborhood of a graph γ in a 3-manifold, $\chi(\partial\eta) = 2\chi(\gamma)$. T is the boundary of a regular neighborhood of the spine, so it follows immediately from this construction that if $\partial M = \partial_- B$ has n boundary components

$$\begin{aligned} \chi(T) = (\chi(\partial_- B) - n) + (2\chi(\gamma) - n) &= (\chi(\partial_- B) - 2n) + 2\chi(\gamma) \\ &= -2\text{genus}(\partial_- B) + 2\chi(\gamma). \end{aligned}$$

Hence

$$2\text{genus}(T) = 2 - \chi(T) = 2 + 2\text{genus}(\partial_- B) - 2\chi(\gamma) \geq 2\text{genus}(\partial_- B)$$

(with equality if and only if $\chi(\gamma) = 1$, i. e. γ is a tree). Hence

$$\text{genus}(T) \geq \text{genus}(\partial_- B) = \text{genus}(\partial M) = \text{genus}(F) + \text{genus}(\partial M - F)$$

so $\text{genus}(T) \geq \text{genus}(F)$ with equality only if $\text{genus}(\partial M - F) = 0$, that is $\partial M - F$ consists of spheres. □

Lemma 5.4. *Suppose a closed surface F divides a 3-manifold M into two components C_1 and C_2, and each is given a Heegaard splitting $C_i = A_i \cup_{T_i} B_i$. Let T be the Heegaard surface for the splitting of M obtained by amalgamating the two splittings. Then*

$$\chi(T) = \chi(T_1) + \chi(T_2) - \chi(F).$$

In particular, if F is connected, then

$$\text{genus}(T) = \text{genus}(T_1) + \text{genus}(T_2) - \text{genus}(F)$$

Proof. With no loss let A_1 (resp B_2) be the compression body in the splitting by T_1 (resp T_2) that is incident to F. In amalgamation, the construction of the splitting surface T of the amalgamated splitting replaces a p-punctured copy of F in T_1 and a q-punctured copy of F in T_2 with a single $(p+q)$-punctured copy of F. The p punctures correspond to scars parallel to F of the chosen set of meridian disks in A_1, after F_1 is compressed along these disks. A similar and symmetric statement is true for the q punctures. See [La, Figure 12].

Computing Euler characteristics from this description we have

$$\chi(T) = \chi(T_1) + \chi(T_2) - (\chi(F) + p) - (\chi(F) + q) + (\chi(F) + p + q)$$

or, more succinctly

$$\chi(T) = \chi(T_1) + \chi(T_2) - \chi(F).$$

Then, if F is connected,

$$\text{genus}(T) + \text{genus}(F) = \text{genus}(T_1) + \text{genus}(T_2)$$

follows since all surfaces are connected. □

Proposition 5.5. *Suppose M is a 3-manifold with connected boundary in which each closed surface separates. Suppose $\mathbb{C}$ is a chamber complex in M with defining surface F and each chamber in $\mathbb{C}$ has a pure Heegaard splitting. Let $C = A_C \cup_{T_C} \cup B_C$ be the chamber whose boundary contains ∂M. Suppose $M = A \cup_T B$ is the Heegaard splitting of M obtained by amalgamating the Heegaard splittings of all the chambers. Then*

a) $\text{genus}(T_C) \leq \text{genus}(T)$

b) If $\text{genus}(T_C) = \text{genus}(T)$ *then T_C is isotopic to T and*

c) If $\text{genus}(T) = \text{genus}(\partial M)$ *then $M = A \cup_T B$ is a trivially split handlebody, and F consists entirely of spheres bounding balls in M. Moreover, amalgamating the Heegaard splittings in each ball bounded by a component of F gives a trivial splitting of the ball.*

Proof. With no loss, assume $\partial M \subset \partial_- B_C$.

It will be useful to consider the natural graph Γ associated to the chamber complex $\mathbb{C}$, in which each vertex corresponds to a chamber in $\mathbb{C}$ and two such vertices are connected by an edge if they are adjacent, that is there is a component of the defining surface $F = F(\mathbb{C})$ that is incident to both. Since every closed surface in M separates, Γ can contain no cycles, so it is a tree; we can take the vertex v_C corresponding to C as the root of the tree.

Case 1: M is a single chamber, so A is a handlebody and Γ is a single vertex.

In this case T_C is T, so a) and b) are automatic. For c), assume $\text{genus}(T) = \text{genus}(\partial M)$. Then no 2-handles could be attached in the construction of B_C from $\partial_+ B_C = T$, for if any attachment circle were non-separating then $\text{genus}(T) > \text{genus}(\partial M)$ and if any were separating on T then each component of the result would be the Heegaard surface for a separate chamber, contradicting the assumption that M has a single chamber. We conclude that B_C is a collar of ∂M. Since A_C is a handlebody, so then is M.

Case 2: M has two chambers, $C \supset \partial M$ and C'.

Since each component of F separates, $F = C \cap C'$ is connected. By Lemma 5.4

$$\text{genus}(T) = \text{genus}(T_C) + \text{genus}(T_{C'}) - \text{genus}(F).$$

By Lemma 5.3 $\text{genus}(T_{C'}) - \text{genus}(F) \geq 0$, so $\text{genus}(T) \geq \text{genus}(T_C)$, establishing a). Moreover, if $\text{genus}(T) = \text{genus}(T_C)$ then $\text{genus}(T_{C'}) = \text{genus}(F)$ and, by Case 1 applied to C', C' is a trivially split handlebody. But if C' is a trivially split handlebody then amalgamating it with T_C leaves T_C unchanged, so T is isotopic to T_C, establishing b) in this case.

For c) observe that we so far have the inequalities

$$\text{genus}(T) \geq \text{genus}(T_C) \geq \text{genus}(\partial M).$$

Hence if $\text{genus}(T) = \text{genus}(\partial M)$ then both inequalities are equalities. The first equality, by the second claim, shows T_C is isotopic to T. The second equality, by Lemma 5.3 (applied to C and reversing the roles of F and ∂M) says F is a sphere, and we have already established that C' is a trivially split handlebody. Thus C' is a trivially split ball, as claimed. Moreover, Case 1 applied to the amalgamated splitting of M shows M is a trivially split handlebody, completing the proof of c) in this case.

Case 3: Every chamber in M other than C is adjacent to the chamber C. That is, Γ is a star graph: a collection of edges each of which has a single end incident to v_C.

The proof is by induction on the number n of chambers not C. Cases 1 and 2 above apply when $n = 0, 1$, so we can assume $n \geq 2$, that is there are at least 2 chambers other than C. Let C_1, C_2 be two of them, separated from C by surfaces F_1, F_2 respectively. The chamber complex $\mathbb{C}_1$ obtained from $\mathbb{C}$ by amalgamating along F_1 satisfies the hypotheses of Case 3, but with n reduced. So we inductively assume that the Proposition is true for $\mathbb{C}_1$. But further amalgamation of all of $\mathbb{C}_1$ gives Heegaard surface T and does not change ∂M so conclusion c) applies: If $\text{genus}(T) = \text{genus}(\partial M)$ then M is a trivially split handlebody and each $C_i, i \neq 1$ is a trivial split ball. Repeating the argument for the chamber complex $\mathbb{C}_2$ obtained from $\mathbb{C}$ by amalgamating along F_2 shows that $\mathbb{C}_1$ is also a trivially split ball. Thus c) is true in this case.

To prove a) observe that the inductive hypothesis applied to $\mathbb{C}_1$ shows that the Heegaard surface T_1 of the chamber of $\mathbb{C}_1$ containing ∂M has $\text{genus}(T_1) \leq \text{genus}(T)$. But that chamber is obtained from C and C_1 by amalgamation along F_1 so by Case 2) $\text{genus}(T_C) \leq \text{genus}(T_1)$. Hence $\text{genus}(T_C) \leq \text{genus}(T)$ as required to prove a).

Finally, if $\text{genus}(T_C) = \text{genus}(T)$ then we have just shown $\text{genus}(T_1) = \text{genus}(T)$ and $\text{genus}(T_C) = \text{genus}(T_1)$. The first equality shows, by inductive hypothesis, that T_1 and T are isotopic; the second shows, by Case 2), that T_C is isotopic to T_1. Thus T_C and T are isotopic, proving b) in this case.

Case 4: There is an edge in Γ that is not incident to v_C.

There is then such an edge so that deleting the edge from Γ leaves two components: one containing v_C and the other a star graph, as in Case 3. (For example, in a path from v_C to a most distant vertex in Γ choose the second to last edge traversed.). The component F' of F corresponding to that edge then bounds a submanifold $N \subset M$ containing a chamber complex $\mathbb{C}_N$ that satisfies the hypothesis of Case 3. That is, the chamber adjacent to F' in N is adjacent to every other chamber in $\mathbb{C}_N$. Let T_N by the Heegaard surface obtained by amalgamating the chambers of N. The resulting splitting of N is pure, since $F' = \partial N$ is connected.

We induct on the number of chambers in $\mathbb{C}$. Let $\mathbb{C}'$ be the chamber complex in M obtained by amalgamating the chambers of N. The inductive hypothesis applies then to $\mathbb{C}'$ and, since $F' \neq \partial M$, the chamber C is unchanged by the amalgamation within N. Moreover, the result of amalgamating the Heegaard splittings of all the chambers in $\mathbb{C}'$ is the same as amalgamating all those in $\mathbb{C}$, so by inductive hypothesis $\text{genus}(T) \geq \text{genus}(T_C)$ and equality implies that T is isotopic to T_C. This verifies a) and b).

For c), suppose $\text{genus}(T) = \text{genus}(\partial M)$. Again applying the inductive hypothesis to $\mathbb{C}'$ we deduce that $M = A \cup_T B$ is a trivially split handlebody, and all components of F, other than those in the interior of N, are spheres bounding balls in which the amalgamated Heegaard splitting is trivial. In particular, N is a ball trivially split by T_N. Finally, apply Case 3 to $\mathbb{C}_N$: all components of F in the interior of N are also spheres bounding trivially split balls. □

5.2 From weak reduction to chamber complex decomposition

There is a lot of inefficiency in the chamber complex $\hat{\mathbb{C}}$ obtained by surgery on the weakly reducing family $\mathcal{A}, \mathcal{B}$. For example, if two disks in $\mathcal{A}$ are parallel, then surgery as above turns the collar between them into its own A-chamber, one that is simply a ball with trivial Heegaard splitting. This problem is easily addressed by viewing surgery on $\mathcal{A}, \mathcal{B}$ as the first step in a disk-decomposition of the chamber complex $A \cup_T B$ along $\mathcal{A}, \mathcal{B}$, and adopting the following rule:

Rule 5.6. *Following the weak reduction of $A \cup_T B$ along $\mathcal{A}, \mathcal{B}$, let W be a ball in M bounded by a component of $\hat{F}$.*

- *Declare W to be a goneball if and only if the Heegaard splitting of W obtained by amalgamating the Heegaard splittings of all the chambers of $\hat{\mathbb{C}}$ in $\text{int}(W)$ gives a trivial Heegaard splitting of W.*
- *For each goneball G, amalgamate the Heegaard splittings of all the chambers of $\hat{\mathbb{C}}$ in G (including amalgamation along ∂G).*

Denote the resulting chamber complex $\mathbb{C}$ (circumflex removed) and say that it is obtained from the splitting $A \cup_T B$ by *weak reduction* along $\mathcal{A} \cup \mathcal{B}$. In Proposition 5.10 below we show that Rules 4.4 and 5.6 do not conflict, so weak reduction can be viewed as disk decomposition of the chamber complex determined by T. Each chamber in $\mathbb{C}$ is Heegaard split, and furthermore no chamber in $\mathbb{C}$ is a trivially split ball.

When a goneball G is absorbed into the adjacent A-chamber C (or symmetrically), the effect on the spine γ_C of B_C is to replace a collar of ∂G in C by a single vertex whose neighborhood is the ball G. As a result, each A chamber C in $\mathbb{C}$ continues to inherit a natural spine γ_C for B_C, but instead of just a collection of arcs attached to $\partial C - \partial M$ the spine is a graph attached to $\partial C - \partial M$, containing vertices as well as arcs. As before, slides of edges in the graph over other edges and over $\partial C - \partial M$ can be used to replace one spine by another and so a chamber-derived alteration of the disks $\mathcal{B}$.

Proposition 5.7. *Following the weak reduction of $A \cup_T B$ along $\mathcal{A}, \mathcal{B}$, let W be a handlebody in M bounded by a component of $\hat{F}$. Suppose the Heegaard splitting $W = A_W \cup_{T_W} B_W$ obtained by amalgamating the splittings of all the chambers in $\text{int}(W)$ is trivial. Then*

- *$T \cap \text{int}(W)$ consists only of disks and*
- *If W is not a ball after the absorption of the goneballs given by Rule 5.6, then it becomes a single chamber in $\mathbb{C}$. If W is a ball, it becomes a goneball in $\mathbb{C}$.*

Proof. With no loss assume that the chamber in W adjacent to ∂W is an A-chamber, so $\partial W = \partial_- B_W$. Since weak reduction is inverse to amalgamation, it is easy to derive this natural description of the splitting surface T_W: Delete from ∂W the p disks corresponding to internal scars and call the resulting surface ∂W_-. Then attach to ∂W_- the surface $T \cap \text{int}(W)$, which has p boundary components, each corresponding to an internal scar in ∂W. Then push the resulting surface into $\text{int}(W)$ via a collar of ∂W.

The compression body B_W in W has a natural spine: the union of a graph $\gamma \subset \text{int}(W)$ with p end-points attached to $\partial_- B_W = \partial W$ at the centers of the internal scars. We then calculate as in the proof of Lemma 5.3:

$$\chi(T_W) = (\chi(\partial W) - p) + (2\chi(\gamma) - p) = \chi(\partial W) + (2\chi(\gamma) - 2p)$$

By assumption T_W and ∂W are homeomorphic, and this implies $\chi(\gamma) = p$. Let n be the number of components of γ. Since T_W, hence B_W, hence the spine is connected, each component of γ must contain one of the end points of γ on ∂W. This implies $n \leq p$. On the other hand, each component of γ has Euler characteristic ≤ 1. Hence $p = \chi(\gamma) \leq n$. Together, these inequalities that $p = n$ and each component of γ has Euler characteristic = 1. In other words, each component of γ is a tree. This in turn implies that each component of $T \cap \text{int}(W)$ is a disk, verifying the first assertion of the Proposition.

The last assertion c) of Proposition 5.5 further says that each component of $\hat{F}$ lying in the interior of W bounds a goneball, establishing the second claim. □

Corollary 5.8. *The interior of each goneball, as defined in Rule 5.6, intersects T only in disks.*

Proof. Rule 5.6 says that each goneball has trivial Heegaard splitting, so Proposition 5.7 applies. □

The discussion above leads to

Definition 5.9. *A* Heegaard split chamber complex $\mathbb{C}$ *for a compact manifold M is a chamber complex so that:*

- *Each chamber is labeled as either an A-chamber or B-chamber.*
- *Adjacent chambers are labeled A and B*
- *Each chamber has a pure Heegaard splitting*
- *The Heegaard splitting of each ball chamber is non-trivial.*

Note that the last property guarantees that a Heegaard split chamber complex satisfies the troublesome third requirement of the chamber complexes discussed preceding Lemma 3.6.

Proposition 5.10. *Let $M = A \cup_T B$ be a Heegaard split closed manifold and $\mathcal{A}, \mathcal{B}$ be a weakly reducing family of disks for the splitting. Let $\mathbb{C}_T$ denote the chamber complex in M defined by T. That is, the chambers are A and B. Then surgery on $\mathcal{A} \cup \mathcal{B}$ followed by goneball removal, as just described, defines a chamber complex decomposition*

$$\mathbb{C}_T \xrightarrow{\mathcal{A}\cup\mathcal{B}} \mathbb{C},$$

and $\mathbb{C}$ is a Heegaard split chamber complex.

Proof. Surgery on $\mathcal{A} \cup \mathcal{B}$ is the first step in chamber complex decomposition. The next step is goneball absorbtion, where Rule 4.4 requires that any ball that is declared a goneball must be disky. That Rule 5.6 implies this follows immediately from Corollary 5.8. The construction above describes how each chamber is endowed with a pure Heegaard splitting.

It remains to show that the Heegaard splitting of each ball chamber in $\mathbb{C}$ is non-trivial. Suppose S is a sphere in $\hat{F}$ that bounds a ball chamber C in $\mathbb{C}$. Since the ball contains no other chambers, the only components of $\hat{F}$ that lie in the interior of C must be goneballs. Rule 5.6 says that these goneballs are absorbed, so their boundary spheres are removed, by amalgamation of Heegaard splittings. In particular, the resulting Heegaard splitting of C is obtained by amalgamation of all the chambers in int(C). Since C is not itself a goneball, it follows again from Rule 5.6 that this Heegaard splitting is non-trivial, as required. □

5.3 Heegaard split chamber complex decomposition

Suppose $\mathbb{C}$ is a Heegaard split chamber complex in M and $M = A \cup_T B$ is the Heegaard splitting of M obtained by amalgamation of the Heegaard splittings of the chambers of $\mathbb{C}$. Then we say that $\mathbb{C}$ *supports* the splitting $M = A \cup_T B$. So a weak reduction of $A \cup_T B$ as above produces a Heegaard split chamber complex that supports it. (See the proof of Proposition 5.14 below.)

The process of decomposing chamber complexes, as described in Section 4, extends naturally to Heegaard split chamber complexes, as we now describe. Suppose $\mathbb{C}$ is a Heegaard split chamber complex with defining surface F, and $\mathcal{D}$ is a disk set in $\mathbb{C}$ with $\partial\mathcal{D} \subset F$. In each A-chamber $C = A_C \cup_{T_C} B_C$ (and dually for the B-chambers), isotope the splitting surface T_C until it is aligned with the disk set $\mathcal{D}_C = \mathcal{D} \cap C$, as described in [Sc1]. That is, each disk $D \in \mathcal{D}_C$ will either

- intersect T_C in a single circle, and the complementary component $D \cap B_C$ of the circle in D will be a vertical annulus in B_C, or
- lie entirely in B_C, as D_1, D_2 do in Figure 17.

(D cannot lie in A_C since in an A-chamber A_C is disjoint from F.) In either case B_C has a spine Σ (illustrated in Figure 17) that intersects the disks $\mathcal{D}_C$ exactly in $\partial\mathcal{D}_C$. B_C is isotopic to a thin regular neighborhood of the spine Σ, and after the isotopy every disk $D \in \mathcal{D}_C$ intersects T_C in a single circle, with the annulus $D \cap B_C$ vertical in B_C. We say then that $\mathcal{D}$ is *aligned* with $\mathbb{C}$.

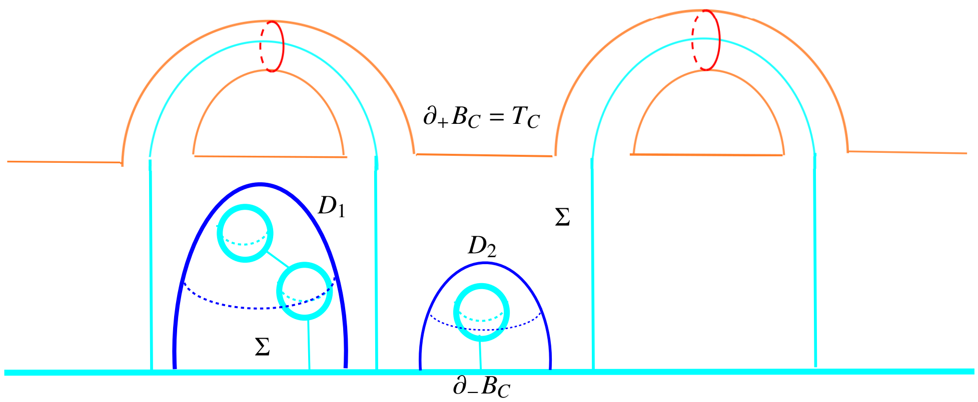

Figure 17: Alignment of disks in $\mathcal{D}_C$ that lie entirely in B_C

The first step in chamber complex decomposition along $\mathcal{D}$ is to surger F along $\mathcal{D}$, making the surface $F_{\mathcal{D}}$. When the Heegaard splittings of the chambers are aligned with $\mathcal{D}$ as above, this has two effects on an A-chamber C:

- Surgery on the B-disks incident to ∂C, lying in adjacent B-chambers, adds 2-handles to C and B_C along $\partial_- B_C$; call the results respectively C_+ and B_{C_+}. The manifold B_{C_+} is still a compression body so, in particular, T_C remains a Heegaard splitting surface for C_+. Note that a B-disk incident to ∂C, lying in an adjacent B-chamber and inessential in that chamber may become part of an essential sphere in C_+, cutting off a punctured ball from C_+

- Further surgery on the disks $\mathcal{D}_C$ creates a surface $T_{\hat{C}_D}$, each of whose components is a Heegaard surface for one of the chambers $\hat{C}_D$ obtained from C_+ by the surgery on $\mathcal{D}_C$. For essential disks, this observation is familiar in Heegaard theory as ∂-reduction of a Heegaard surface (see [Sc1]). Thus $T_{\hat{C}_D}$ is a collection of Heegaard surfaces for the remnants of C in $\hat{\mathbb{C}}_{\mathcal{D}}$.

The result of the operation just described, done simultaneously on all the chambers of $\mathbb{C}$ defines a pure Heegaard splitting of each chamber in $\hat{\mathbb{C}}_{\mathcal{D}}$.

Definition 5.11. *Let $\mathbb{C}_{\mathcal{D}}$ be the Heegaard split chamber complex in M obtained from $\hat{\mathbb{C}}_{\mathcal{D}}$ by the above process, following Rule 5.6 to declare goneballs. Then $\mathbb{C}_{\mathcal{D}}$ is obtained from $\mathbb{C}$ by* Heegaard split chamber complex decomposition *and we write*

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}.$$

This choice of language implies that the process is indeed a chamber complex decomposition, that is that following Rule 5.6 to declare goneballs is consistent with Rules 4.4 and 4.6. For example, we need to show that if G is a goneball as declared by Rule 5.6 then each component of $F \cap G$ is a disk, as was shown for the case $F = T$ in Corollary 5.8. That indeed both rules 4.4 and 4.6 are satisfied will follow from the next two more general propositions, see Corollaries 5.13 and 5.15.

Proposition 5.12. *Let $\mathbb{C}$ be a Heegaard split chamber complex in a 3-manifold M and $F = F(\mathbb{C})$ be its defining surface. Suppose $\mathcal{D}$ is a disk set in $\mathbb{C}$ and*

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is a Heegaard split chamber complex decomposition. Suppose a component of $F_{\mathcal{D}}$ bounds a handlebody W in M. If the Heegaard splitting of W obtained by amalgamating the splittings of all the chambers contained in W is trivial then each component of $F_W = F \cap \operatorname{int}(W)$ is a disk.

Proof. Denote by $M = A \cup_T B$ the Heegaard splitting of M obtained by amalgamating all the splittings of $\mathbb{C}$. Following the argument in Proposition 5.7 we know that $T_W = T \cap \operatorname{int}(W)$ consists of disks.

We now exploit the fact that when Heegaard splittings of adjacent chambers are amalgamated to the defining surface F, F can be viewed as a subsurface of T to which disks are attached. Hence F_W is a subsurface of T_W with possibly disks attached, that is F_W is a planar surface with at most one boundary circle per component. F_W then consists of properly embedded disks and spheres.

An only slightly more complicated argument rules out the existence of spheres. Suppose there were closed components of F_W, necessarily all spheres. In this case, let F_0 be an innermost such sphere, so that it bounds a ball chamber C containing no other components of F_W. Since $\mathbb{C}$ is a Heegaard split chamber complex, C has non-trivial Heegaard splitting, so its Heegaard surface T_C contains a non-separating circle. But when the Heegaard splittings of the chambers are amalgamated, a punctured copy of T_C persists into T, and the non-separating circle can be taken to be disjoint from those punctures. Again this would contradict the fact that T_W consists of disks. We deduce that F_W consists entirely of disks. □

Corollary 5.13. *Let $\mathbb{C}$ be a Heegaard split chamber complex in M, $\mathcal{D}$ be a disk set in $\mathbb{C}$ and $\hat{\mathbb{C}}_{\mathcal{D}}$ be the chamber complex obtained by the process described before Definition 5.11. Then following Rule 5.6 to declare goneballs is consistent with Rule 4.4.*

Proof. Suppose G is a ball in M that is bounded by a sphere in $F_{\mathcal{D}}$. If G is a goneball under Rule 5.6 then the Heegaard splitting obtained by amalgamating all the splittings of chambers in $\hat{\mathbb{C}}_{\mathcal{D}}$ that lie in G is trivial. But by Proposition 5.12 this implies that $F = F(\mathbb{C})$ intersects $\text{int}(G)$ only in disks, that is G is disky. Thus G is also a goneball under Rule 4.4. □

Proposition 5.14. *Suppose $\mathbb{C}$ is a Heegaard split chamber complex for M that supports the Heegaard splitting $M = A \cup_T B$. Suppose $\mathcal{D}$ is an aligned disk set in $\mathbb{C}$ .Let $\hat{\mathbb{C}}_{\mathcal{D}}$ be the chamber complex obtained by the process described before Definition 5.11. Then amalgamating all the induced Heegaard splittings of the chamber complex $\hat{\mathbb{C}}_{\mathcal{D}}$ yields the original Heegaard splitting $M = A \cup_T B$.*

Proof. We are given that amalgamating all the Heegaard splittings of the chambers in $\mathbb{C}$ gives $M = A \cup_T B$. Our assumption is that each disk in $\mathcal{D}$ is aligned with the splitting of the chamber in which the disk lies. Suppose $D \in \mathcal{D}$ lies, without loss, in an A-chamber $C = A_C \cup_{T_C} B_C$ and $C' = A_{C'} \cup_{T_{C'}} B_{C'}$ is the adjacent B-chamber on whose boundary ∂D lies. The process described before Definition 5.11 affects a bicollar neighborhood $D \times [-1,1]$ of D as follows (see Figure 18):

- The ball $\text{int}(D) \times (-1,1)$ is moved from $\text{int}(C)$ to $\text{int}(C')$.
- The annulus $\partial D \times [-1,1]$ is deleted from the spines of $A_{C'}$ and B_C.
- The disks $D \times \{\pm 1\}$ are added to the spines of $A_{C'}$ and B_C.
- The arc $\{0\} \times [-1,1]$ is added to the spine of $A_{C'}$.

On the other hand, amalgamation of the splittings of $\hat{\mathbb{C}}_{\mathcal{D}}$ does exactly the opposite near $D \times [-1,1]$: a tube around $\{0\} \times [-1,1]$, which we can take to be $\partial D \times [-1,1]$, is added to $F_{\mathcal{D}}$ after deleting the disks $D \times \{\pm 1\}$, undoing each of the items above. Extending this observation now to all the disks in $\mathcal{D}$, the result of amalgamating all the Heegaard splittings after the process described before Definition 5.11 is the same as amalgamating before the process, namely $M = A \cup_T B$. □

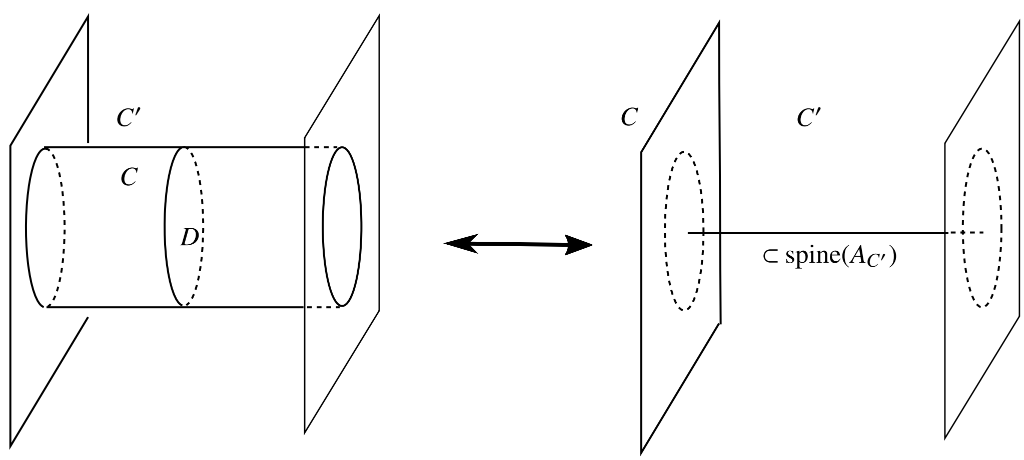

Figure 18: Decomposition near $D \in \mathcal{D}$

Corollary 5.15. *Let $\mathbb{C}$ be a Heegaard split chamber complex in M, $\mathcal{D}$ be a disk set in $\mathbb{C}$ and $\hat{\mathbb{C}}_{\mathcal{D}}$ be the chamber complex obtained by the process described before Definition 5.11. Then following Rule 5.6 to declare goneballs is consistent with Rule 4.6.*

Proof. Suppose G is a ball in M that is bounded by a sphere in $F_{\mathcal{D}}$. Suppose further that the disk $D \in \mathcal{D}$ leaves no scar on ∂G, and define $\mathcal{D}_- = \mathcal{D} - D$. Since D leaves no scar on ∂G, the sphere is also a component of $F_{\mathcal{D}_-}$. Moreover, if we let $\mathcal{D}_\partial \subset \mathcal{D}_-$ be the set of disks in $\mathcal{D}$ that *do* leave scars on ∂G, then ∂G is also a component of $F_{\mathcal{D}_\partial}$ and, as described before Definition 5.11, G inherits a chamber complex structure $\hat{\mathbb{C}}_\partial$, in which each chamber is Heegaard split. Let T_G be the Heegaard splitting of the ball G obtained by amalgamating all these splittings. That is, $\hat{\mathbb{C}}_\partial$ supports T_G.

According to Proposition 5.14 the chamber complexes in G obtained from $\hat{\mathbb{C}}_\partial$ by decomposition along $\mathcal{D} \cap G$ or along $\mathcal{D}_- \cap G$ also support T_G. In particular, under Rule 5.6, G is a goneball in $\mathbb{C}_{\mathcal{D}}$ if and only if T_G is a trivial splitting of G and this is true if and only if G is also a goneball in $\mathbb{C}_{\mathcal{D}_-}$. Thus Rule 4.6 holds. □

Theorem 5.16. *Suppose $\mathbb{C}$ is a Heegaard split chamber complex for M that supports the Heegaard splitting $M = A \cup_T B$ and $\mathcal{D}$ is a disk set in $\mathbb{C}$.After perhaps a proper isotopy of $\mathcal{D}$ within the chambers, not moving T, the chamber complex decompositon*

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is a Heegaard split chamber complex decomposition and the resulting Heegaard split chamber complex $\mathbb{C}_{\mathcal{D}}$ also supports T.

Proof. This would seem to be a straightforward consequence of Proposition 5.14 and the discussion at the beginning of Subsection 5.3. Namely:

1. In each chamber C isotope the splitting surface T_C until it is aligned with $\mathcal{D}$, as is shown to be possible in [Sc1].
2. Mimic the handle slides used in the isotopies within each chamber by handleslides on T itself, as discussed for example in [La].
3. Now that T and $\mathcal{D}$ are aligned, apply Proposition 5.14.

This (ultimately successful) strategy has an obvious conceptual weakness: The process requires isotoping T by various handleslides, so T will typically end up in a different position in M than it began. We will show that the process above can be reframed so that T is fixed throughout and only the disks $\mathcal{D}$ and the chambers $\mathbb{C}$ are allowed to move. We must further ensure that the disks $\mathcal{D}$ stay disjoint, in particular that the boundaries of disks in $\mathcal{D}$ that lie in adjacent chambers don't end up intersecting, after their proper isotopies, in the component of $F = F(\mathbb{C})$ that lies between them. That is the goal of the argument that follows.

Claim: There is an isotopy of $\mathcal{D}$, fixed on $\partial\mathcal{D} \subset F$, so that afterwards $\mathcal{D}$ is aligned with the splitting surface of each chamber.

The proof of the Claim is straightforward and well-known: Let C be a chamber, $\mathcal{D}_C$ be the set of disks in $\mathcal{D}$ that lie in C, and $T_C \subset \text{int}(C)$ be the Heegaard splitting surface of the chamber. Per [Sc1] there is an isotopy of embeddings $\phi_t : T_C \to C$ so that ϕ_0 is the original embedding and $\phi_1(T_C)$ is aligned with $\mathcal{D}_C$. By the isotopy extension theorem, ϕ_t can be extended to an isotopy $\theta_t : C \to C$ in which θ_0 is the identity. Since $T_C \subset \text{int}(C)$, we can also take θ_t to be fixed on ∂C.

Define an isotopy of embeddings $\rho_t : \mathcal{D}_C \to C$ by $\rho_t = (\theta_t)^{-1}|\mathcal{D}_C$. Observe that ρ_0 is the original embedding (since θ_0 is the identity). Moreover, it is easy to check that $\rho_1(\mathcal{D})$ is aligned with T_C. Namely, observe that

$$\theta_1(\rho_1(\mathcal{D}_C) \cap T_C) = \mathcal{D}_C \cap \theta_1(T_C) = \mathcal{D}_C \cap \phi_1(T_C)$$

and the last term consists, by construction, of at most one circle in each component of $\mathcal{D}_C$. Hence $\rho_1(\mathcal{D}_C) \cap T_C$ consists of at most one circle in each component of $\mathcal{D}_C$, that is the disks $\rho_1(\mathcal{D}_C)$ are aligned with T_C. This proves the claim.

We continue our effort to bring $\mathcal{D}$ into alignment with T, not by handleslides that move T (mimicking handleslides within chambers), but by proper isotopies of $\mathcal{D}$. In doing so, we do not use the Claim *per se*, but rather note that the proof of the claim implies that we only need to show how to keep T fixed, during the alignment, in a collar of the boundary of each chamber (i. e. near the defining surface F). Then a way to keep T also fixed outside the collar, that is away from F, is provided by the proof of the Claim. So we focus on how handle slides of the original splitting $M = A \cup_T B$ that mimic handle slides in the splitting of a chamber $C = A_C \cup_{T_C} B_C$ can be replaced *near* F by proper isotopies of $\mathcal{D}$.

Consider a bicollar neighborhood $F_0 \times [-1,1]$ of a component F_0 of F that separates an A-chamber C of $\mathbb{C}$ from a B-chamber C'. Here we take $F_0 \times [0,1] \subset C$ and $F_0 \times [-1,0] \subset C'$. F_0 itself is a subsurface of T whose boundary is capped off by disk scars coming from surgery on a collection of disks $(\mathcal{B}, \partial\mathcal{B}) \subset (B_C, \partial B_C)$ and disk scars coming from surgery on a collection of disks $(\mathcal{A}, \partial\mathcal{A}) \subset (A_{C'}, \partial A_{C'})$. B intersects $F_0 \times [0,1]$ in vertical cylinders, the "legs" of 1-handles in B_C, one for each scar of $\mathcal{B}$ on F_0. We denote the leg correspoinding to a scar b by $b \times [0,1]$. (The symmetric statement is true for $A \cap (F_0 \times [-1,0])$.)

Let $\mathcal{D}_C = \mathcal{D} \cap C$ and $\mathcal{D}_{C'} = \mathcal{D} \cap C'$. Also spanning the collar $F_0 \times [0,1]$ are vertical annuli, each of them a collar neighborhood of the boundary of a disk in $\mathcal{D}_C$. The symmetric statement is true in the collar $F_0 \times [-1,0]$. By general position in the surface F_0 we can take take the legs to be disjoint from the annuli, and, since $\mathcal{D}$ is embedded, we have that $\partial\mathcal{D}_C$ and $\partial\mathcal{D}_{C'}$ are disjoint in F_0.

Consider how a handle slide in B that mimics a handle slide of B_C in C appears in this collar. As it begins, the end of a leg $b_1 \times [0,1]$ is slid via a path $\gamma \subset F_0$ to the end of a leg $b_2 \times [0,1]$ and then up and out of view. At the end of the handle slide the same thing occurs elsewhere in the collar. So it is only this initial move that needs to be understood. By general position we may assume that the path γ is disjoint from all scars. Here are possibilities:

1. the path γ is disjoint from both $\partial\mathcal{D}_C$ and $\partial\mathcal{D}_{C'}$, as in the left side of Figure 19.

2. the path γ is disjoint from $\partial\mathcal{D}_C$ but not $\partial\mathcal{D}_{C'}$.

3. the path γ intersects $\partial\mathcal{D}_C$ and possibly also $\partial\mathcal{D}_{C'}$.

In the first case, where γ is disjoint from $\partial\mathcal{D}$, this first stage of the handle slide can be accomplished, leaving T fixed, by just replacing the disk in $\mathcal{B}$ whose scar is b_2 by the disk b' in B obtained by band-summing b_1 to b_2 along γ. This new disk is disjoint from other scars and, by the assumption of this case, still disjoint from ∂D. See right side of Figure 19.

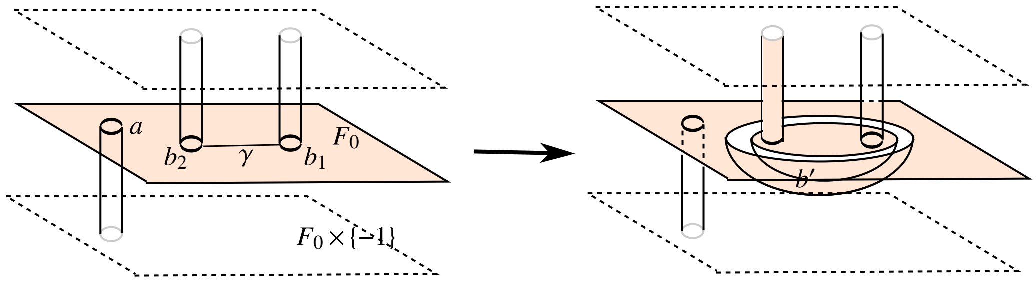

Figure 19: When γ is disjoint from both $\partial\mathcal{D}_C$ and $\partial\mathcal{D}_{C'}$

In the second case, γ may intersect only $\partial\mathcal{D}_{C'}$. These intersection points can be removed by isotoping $\partial\mathcal{D}_{C'}$ along γ towards and then across the disk $b_1 \subset F_0 \subset \partial C'$. See Figure 20. (Another description is that we band sum $\mathcal{D}_{C'}$ at the points it crosses γ to copies of b_1 by bands around subsegments of γ). Such a proper isotopy of $\mathcal{D}_{C'}$ is allowed, and reduces this case to the previous case.

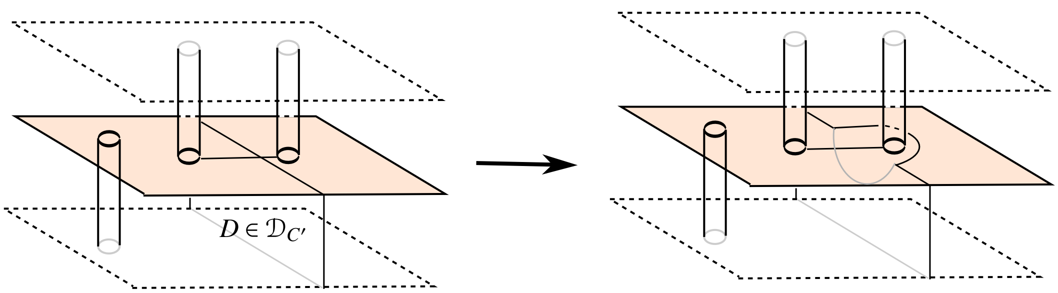

Figure 20: When γ is disjoint from $\partial\mathcal{D}_C$ but not $\partial\mathcal{D}_{C'}$

In the third case, the slide of $b_1 \times [0,1]$ can be broken up into a series of slides along γ, each successive one across a single intersection of γ with $\partial\mathcal{D}_C$. For example, let γ_1 be the segment of γ lying between b_1 and the closest point p of $\gamma \cap \partial\mathcal{D}_C$. Just as in step 2, $\mathcal{D}_{C'}$ can be properly isotoped across b_1 so that it is disjoint from γ_1. The slide of the leg of $b_1 \times [0,1]$ along γ_1 creates a disk intersection of the leg with $\mathcal{D}_C$; as the leg is then straightened from below the disk intersection moves upwards out of view. There is an obvious reinterpretation of this step that keeps the leg (hence T) fixed: instead slide $p \in \partial\mathcal{D}_C$ and neighboring points of $\mathcal{D}_C$ along γ_1 and across the leg $b_1 \times [0,1]$. See Figure 21. Again this creates an intersection disk and again, as the annulus that contains p is made vertical from below, the disk ascends out of view. Note that because we have already cleared all points of $\partial\mathcal{D}_{C'}$ from γ_1, $\partial\mathcal{D}_{C'}$ and $\partial\mathcal{D}_C$ remain disjoint, as required. Continue the slide of the leg along the rest of γ, proceeding as just described across each successive segment of $\gamma - \partial\mathcal{D}_C$, eventually reducing this case to the second case, which we have already considered. □

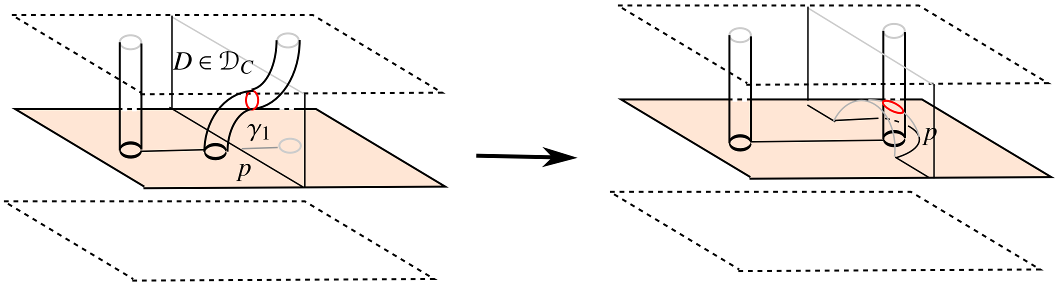

Figure 21: When γ intersects $\partial\mathcal{D}_C$ and possibly also $\partial\mathcal{D}_{C'}$

5.4 Tiny Heegaard split chamber complexes

Proposition 4.10 has an analogue in Heegaard split chamber complex decompositions, as we now describe:

Definition 5.17. *A Heegaard split chamber complex $\mathbb{C}$ is* tiny *if $\mathbb{C}$ is tiny as a chamber complex, and each designated handlebody has a trivial Heegaard splitting.*

Note that in a tiny Heegaard split chamber complex no designated handlebody can be a ball, since, by definition, no chamber in a Heegaard split chamber complex is a trivially split ball. Also, since each designated handlebody has only one boundary component, each must be adjacent to the unique chamber C that is not a designated handlebody. Hence if C is a B-chamber, all the designated handlebodies are A-chambers, and symmetrically.

Proposition 5.18. *Suppose $\mathbb{C}$ is a Heegaard split chamber complex obtained from a Heegaard splitting $M = A \cup_T B$ by weak reduction. Then $\mathbb{C}$ is not tiny.*

Proof. Let $\mathcal{A}, \mathcal{B}$ be the weakly reducing family of disks, and D_A be an essential disk in $\mathcal{A}$. Then after surgery on $\mathcal{A} \cup \mathcal{B}$, D_A lies in a B-chamber C, with ∂D_A on a component of $T \cap \text{int}(C)$. But that component can't be a disk, since ∂D_A is essential on T. Hence C is not a disky handlebody and so, by Proposition 5.7 it is not a trivially split handlebody. To summarize: some B-chamber of $\mathbb{C}$ is not a trivially split handlebody.

The symmetric argument on a disk $D_B \in \mathcal{B}$ shows that there is some A chamber of $\mathbb{C}$ that is not a trivially split handlebody. This immediately implies that the defining surface $F(\mathbb{C}) \neq \emptyset$ and also contradicts the consequence of Definition 5.17 that either all A-chambers or all B-chambers are designated handlebodies, and so must be trivially split. □

Proposition 5.19 (Tinyness pulls back). *Let $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$ be a Heegaard split chamber complex decomposition as given in Definition 5.11, If $\mathbb{C}_{\mathcal{D}}$ is a tiny Heegaard split chamber complex, so is $\mathbb{C}$.*

Proof. Proposition 4.10 and most of its proof apply in this situation: We are given that the designated handlebodies of $\mathbb{C}_{\mathcal{D}}$ have trivial Heegaard splitting so, per Proposition 5.12, each such chamber is disky, and Proposition 4.10 applies. That is, $\mathbb{C}$ is tiny as a chamber complex. What remains to be shown is that the designated handlebodies of $\mathbb{C}$ have trivial Heegaard splittings. The proof of Proposition 4.10 is also valid in this setting; that proof invokes Proposition 4.8 to conclude that all but one chamber of $\mathbb{C}$ is a handlebody, and these become the designated handlebodies of $\mathbb{C}$. Then this strengthened version of Proposition 4.8 implies, in our setting, that each such designated handlebody is trivially split:

Lemma 5.20. *Suppose C is a chamber of $\mathbb{C}$ so that every remnant of C in $\mathbb{C}_{\mathcal{D}}$ is a handlebody with trivial Heegaard splitting. Then C is a handlebody with trivial Heegaard splitting.*

Proof. Proposition 4.8 suffices to conclude that C is a handlebody and the proof of that proposition continues to apply here. According to that proof, all that is required to complete the proof of Lemma 5.20 is this additional claim:

Claim: Suppose C is a handlebody chamber in $\mathbb{C}$ and the disks $\mathcal{D}_C$ in $\mathcal{D}$ that are incident to ∂C all lie within C. Then C is trivially split.

Proof of Claim: Without loss suppose C is an A-chamber and denote by C_r the collection of remnants of C, by hypothesis each a trivially split handlebody A-chamber. Let $F = \partial C$ and let T_C be the chamber's Heegaard surface. Then, by Propositiono 5.5c), $\chi(T_C) \leq \chi(F)$, with equality only if the Heegaard splitting of C is trivial.

Let $\hat{T}_{\mathcal{D}}$ be the union of the Heegaard surfaces for the chambers of $\hat{\mathbb{C}}_{\mathcal{D}}$ that are remnants of C (some of which may be goneballs). When aligned, as described at the beginning of Subsection 5.3, each disk in $\mathcal{D}_C$ intersects F in a single circle (its boundary) and intersects T_C either in a single circle or not at all. In the latter case, the disk lies entirely in B_C and, since $\partial_- B_C$ is incompressible in the compression body B_C, the boundary of the disk is inessential in $\partial_- C$. Hence surgery on such a disk creates a new sphere component of $F_{\mathcal{D}}$ bounding a trivially split ball, so such a surgery adds a sphere component both to $\hat{T}_{\mathcal{D}}$ and to $F_{\mathcal{D}}$. Similarly, surgery on a disk in $\mathcal{D}$ that intersects T in a single circle raises the Euler characteristic of both $\hat{T}_{\mathcal{D}}$ and F by 2. So in the end, surgery on $\mathcal{D}$ raises the Euler characteristic of both $\hat{T}_{\mathcal{D}}$ and F by $2|\mathcal{D}_C|$. (That is, $\chi(F_{\mathcal{D}}) - \chi(F) = 2|\mathcal{D}_C|$.) Consequently, $\chi(\hat{T}_{\mathcal{D}}) \leq \chi(F_{\mathcal{D}})$, again with equality only if C is trivially split.

Consider next what happens when the bounding sphere of a goneball is removed from $F_{\mathcal{D}}$, and $\hat{T}_{\mathcal{D}}$ is altered by amalgamation along that sphere. Since goneballs are exactly those balls that are trivially split, the result is to remove exactly a sphere from both $F_{\mathcal{D}}$ and $\hat{T}_{\mathcal{D}}$, lowering the Euler characteristic of both by 2. Once all spheres bounding goneballs are removed, we then have $\chi(T_{\mathcal{D}}) \leq \chi(\partial C_r)$, again with equality only if $\mathbb{C}$ is trivially split. (Here $T_{\mathcal{D}}$ is the union of all the Heegaard surfaces of the chambers of C_r.) By assumption, each component of $\mathbb{C}_r$ is trivially split. This means that $\chi(T_{\mathcal{D}}) = \chi(\partial C_r)$, so $\chi(T_C) = \chi(F)$ and indeed C is trivially split. This proves the Claim, hence the Lemma and so the Proposition. □

□

6 Sequences of aligned chamber complex decompositions in S^3

There is an important caveat about the structure of a Heegaard split chamber complex decomposition: Typically the Heegaard splitting of a specific chamber of $\mathbb{C}_{\mathcal{D}}$ will not be well-defined. Indeed, even the genus of the chamber may be ambiguous, since, in the process described before Definition 5.11 there is a choice of how to align $\mathcal{D}_C$ with T_C in a chamber C of $\mathbb{C}$. One aspect of this choice (see [FS2]) is that a bubble may be moved across a disk in $\mathcal{D}_C$; if, for example, the disk is separating, this changes the genus of the chambers on each side of the disk. One result of this ambiguity is that $\mathbb{C}_{\mathcal{D}}$ may not be well-defined even as a chamber complex: depending on alignment, a ball chamber in $\hat{\mathbb{C}}_{\mathcal{D}}$ may or may not be trivially split, and therefore may or may not appear as a chamber in $\mathbb{C}_{\mathcal{D}}$.

Soon (see Section 7) this ambiguity will be directly addressed, by describing a *preferred* way of aligning disks, a way that will eventually suffice to certify unambiguously a chamber complex obtained by weak reduction of a Heegaard splitting of S^3. In this section we avoid much of this ambiguity by oversimplifying the full theory. The section is motivational and transitional: We present an application of the theory of disk decompositions developed in Sections 4 and 5, one that extends Proposition 3.8 and so connects to our previous discussions. It also is a model for later arguments we will need in a more complicated setting. We begin with a definition that will only be used in this section, as a way of illustrating and simplifying the full argument that will eventually follow.

Definition 6.1. *In the setting of Definition 5.11, suppose $\mathbb{C}_{\mathcal{D}}$ is a Heegaard split chamber complex obtained from $\hat{\mathbb{C}}_{\mathcal{D}}$ by declaring as goneballs all disky balls. Then*

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is an aligned chamber complex decomposition.

Definition 6.2. *Suppose $\mathbb{C}$ is a Heegaard split chamber complex, that supports a genus g Heegaard splitting $M = A \cup_T B$; C is a chamber in $\mathbb{C}$; and $\mathfrak{b}$ is a genus $g' < g$ bubble in the Heegaard splitting $C = A_C \cup_{T_C} B_C$. Let $\mathbb{C}/\mathfrak{b}$ be the chamber complex, with Heegaard splittings of each chamber, obtained by destabilizing $C = A_C \cup_{T_C} B_C$ along $\mathfrak{b}$, replacing $\mathfrak{b}$ with a neighborhood of a point $*$ in the splitting surface.*

Amalgamating all Heegaard splittings of $\mathbb{C}/\mathfrak{b}$ gives a genus $g - g'$ Heegaard splitting of M obtained by destabilizing T along $\mathfrak{b}$.

Although every chamber in $\mathbb{C}/\mathfrak{b}$ is Heegaard split, $\mathbb{C}/\mathfrak{b}$ may fail to be a Heegaard split chamber complex for a somewhat technical reason: when C is a ball and $\mathfrak{b}$ is a maximal bubble in C, the chamber C in $\mathbb{C}/\mathfrak{b}$ becomes a trivially split ball chamber and this is not allowed in a Heegaard split chamber complex. Nonetheless, any disk set $\mathcal{D}$ in $\mathbb{C}/\mathfrak{b}$ can be aligned with the Heegaard splittings of the chambers, the construction described in Subsection 5.3 carried out, and all disky balls declared goneballs. Denote this aligned chamber complex decomposition

$$\mathbb{C}/\mathfrak{b} \xrightarrow{\mathcal{D}} (\mathbb{C}/\mathfrak{b})_{\mathcal{D}}.$$

Lemma 6.3. *Suppose, in the setting of Definition 6.2, $\mathcal{D}$ is a disk set in $\mathbb{C}$. Then $\mathcal{D}$ can be aligned in $\mathbb{C}$ and in $\mathbb{C}/\mathfrak{b}$ so that $\mathfrak{b}$ remains a bubble in a chamber of $\mathbb{C}_{\mathcal{D}}$ and $\mathbb{C}_{\mathcal{D}}/\mathfrak{b} = (\mathbb{C}/\mathfrak{b})_{\mathcal{D}}$.*

Proof. This is immediate from the definitions: Align the disks $\mathcal{D}$ in $\mathbb{C}/\mathfrak{b}$ so that, by general position, they do not contain the point $*$. This defines also an alignment of $\mathcal{D}$ in $\mathbb{C}$. Following surgery on $\mathcal{D}$ absorb goneballs in both $\hat{\mathbb{C}}$ and $\widehat{\mathbb{C}/\mathfrak{b}}$. They are the same goneballs because in both cases the goneballs are exactly the disky balls. These are defined at the level of chamber complexes, and do not depend on the Heegaard splittings of the chambers. □

Call an alignment as in Lemma 6.3 a $\mathfrak{b}$-*wise alignment*.

Suppose a Heegaard split chamber complex $\mathbb{C}_0$ supports T. For example $\mathbb{C}_0$ could be the Heegaard splitting $A \cup_T B$ itself. Let

$$\vec{\mathbb{C}}: \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

be a sequence of chamber complex decompositions.

Proposition 6.4. *Suppose chambers $C_0 \in \mathbb{C}_0$ and $C_n \in \mathbb{C}_n$ contain, respectively, incompressible spheres S and S'. Then, for iteratively $0 \le i \le n-1$ each disk set $\mathcal{D}_i$ can be aligned in $\mathbb{C}_i$, so that $\mathbb{C}_0$ and the resulting Heegaard split chamber complex structure on $\mathbb{C}_n$ cocertify.*

Proof. We induct on $n \geq 1$.

A standard innermost disk argument shows that there is an incompressible sphere S_0 in the chamber C_0 that is disjoint from the disks $\mathcal{D}_0$. This implies that S_0 will lie in a single chamber C_1 of the chamber complex $\mathbb{C}_1$. Align the disks $\mathcal{D}_0$ as in Theorem 5.16 so that

$$\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1$$

is an aligned disk decompositon. More specifically, in chamber C_0 choose an alignment, per [Sc1], so that the incompressible sphere S_0 is also aligned with the Heegaard splitting surface of the chamber and hence, after amalgamation, also aligned with T, per Proposition 3.7. Following Proposition 3.8 we may as well substitute S_0 for S.

If S_0 is incompressible in C_1, then, according to Propositions 3.7 and Corollary 3.9, $h_{S_0} : (S^3, T) \to (S^3, T_g)$ defines both $h_{F(\mathbb{C}_0)}$ and $h_{F(\mathbb{C}_1)}$. Thus $h_{F(\mathbb{C}_0)} \sim h_{F(\mathbb{C}_1)}$. By inductive assumption $h_{F(\mathbb{C}_1)} \sim h_{F(\mathbb{C}_n)}$, completing the proof in this case.

However, it is possible that the sphere S_0 is compressible in the chamber C_1. (For example, the interior of a ball W that S_0 bounds in C_1 could have contained a single handlebody chamber of $\mathbb{C}_0$, one that disappears in $\mathbb{C}_1$ because it is decomposed by $\mathcal{D}_0$ into a ball that is a goneball.) If S_0 is compressible in C_1 then it bounds a non-trivial bubble $\mathfrak{b}$ in (S^3, T), by [Wa], one that already appears as a bubble in the induced Heegaard splitting of $\mathbb{C}_1$. (The bubble is non-trivial because the presence in int(W) of other chambers means that W is not a trivially split ball, per Proposition 5.12.) In this case, we argue as follows:

For the rest of the decompositions

$$\mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

in the sequence, choose $\mathfrak{b}$-wise alignments. Then the bubble $\mathfrak{b}$ (bounded still by S_0) arrives intact in a single chamber of $\mathbb{C}_n$. In $\mathbb{C}_n/\mathfrak{b}$, S' can be taken to be disjoint from $*$, simply by general position. Then in $\mathbb{C}_n$, $\mathfrak{b}$ is disjoint from S'. Then Lemma 3.5 shows that h_{S_0} and $h_{S'}$ are eyeglass equivalent. Hence $h_{F(\mathbb{C}_0)} \sim h_{S_0} \sim h_{S'} \sim h_{F(\mathbb{C}_n)}$ as required. □

Having shown there exists a choice of alignment for which $\mathbb{C}_0$ and $\mathbb{C}_n$ cocertify, we next show that any choice of alignments will do. The critical point is that we understand how different alignments of disk-sphere sets in a Heegaard split 3-manifold are related: By [FS2] they differ by a sequence of eyeglass moves and passing bubbles through the disk-sphere set. We begin with a lemma and proposition that are not specific to Heegaard splittings of S^3.

Lemma 6.5. *Suppose $\mathbb{C}_0$ is a Heegaard split chamber complex in M that supports the splitting $M = A \cup_T B$, and*

$$\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1$$

is a chamber complex decomposition. Suppose $E \subset \mathbb{C}_1$ is a disk-sphere set.

Let $\mathcal{D}_0^x$ and $\mathcal{D}_0^y$ be possibly different alignments of the disk set $\mathcal{D}_0$ in $\mathbb{C}_0$, with resulting Heegaard split chamber complexes $\mathbb{C}_1^x$ and $\mathbb{C}_1^y$ respectively, and similarly let E^x, E^y be alignments of the disk-sphere set E in $\mathbb{C}_1^x$ and $\mathbb{C}_1^y$ respectively. Then a sequence of bubble passes and eyeglass moves will change the alignment $\mathcal{D}_0^x$ to $\mathcal{D}_0^y$ and E^x to E^y.

Proof. By [FS2] there is a series of bubble passes and eyeglass moves that change the alignment $\mathcal{D}_0^x$ to $\mathcal{D}_0^y$. The proof of the Lemma is by induction on p, the number of bubble passes required. If $p = 0$ then, after an eyeglass move, we can take $\mathcal{D}_0^x = \mathcal{D}_0^y$, so $\mathbb{C}_1^x = \mathbb{C}_1^y$ and then apply [FS2] to the disk sets E^x and E^y in $\mathbb{C}_1^x = \mathbb{C}_1^y$.

So suppose $p \geq 1$ and assume the Lemma is true whenever the number of bubble passes needed to change the alignment $\mathcal{D}_0^x$ to $\mathcal{D}_0^y$ is less than p. Let $\mathfrak{b}$ be the first ball that is passed, through a disk $D \in \mathcal{D}_0^x$, in changing the alignment $\mathcal{D}_0^x$ to $\mathcal{D}_0^y$. Say $\mathfrak{b}$ is passed from a chamber $C \in \hat{\mathbb{C}}_1^x$ to a chamber $C' \in \hat{\mathbb{C}}_1^x$, with possibly $C = C'$. (Enthusiasts will note that if $C = C'$ then the bubble pass is also an eyeglass move, by Lemma 2.5, so p can be reduced and we are done.)

Consider the subset E^C of the aligned disk set E^x that lies in C. As aligned, E^C may well intersect the interior of $\mathfrak{b}$. However there is *some* alignment $E^{\mathfrak{b}}$ of E^C in C that is disjoint from the bubble $\mathfrak{b}$. This follows from general position and [Sc1] applied to E^C in $\hat{\mathbb{C}}_1/\mathfrak{b}$. Then by [FS2] the alignment $E^{\mathfrak{b}}$ can be obtained from E^C by eyeglass moves and bubble passes through E in C. Thus we may as well assume that $E^C = E^{\mathfrak{b}}$.

By further bubble passes of $\mathfrak{b}$ through E^C we can move $\mathfrak{b}$ adjacent to the disk $D \in \mathcal{D}_0^x$ and complete the pass of $\mathfrak{b}$ through D. Then only $p-1$ further bubble passes are needed to move $\mathcal{D}_0^x$ to $\mathcal{D}_0^y$, completing the inductive step. □

This argument generalizes, though the statement and argument become more complex. Let $\mathbb{C}_0$ be a Heegaard split chamber complex in M that supports the Heegaard splitting $M = A \cup_T B$. Suppose

$$\vec{\mathbb{C}}: \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a sequence of chamber complex decompositions and E is a disk-sphere set in $\mathbb{C}_n$.

Suppose that the disks $\mathcal{D}_i$ are aligned at each successive stage of the decomposition, that

$$\vec{\mathbb{C}}^x: \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0^x} \mathbb{C}_1^x \xrightarrow{\mathcal{D}_1^x} \mathbb{C}_2^x \xrightarrow{\mathcal{D}_2^x} \dots \xrightarrow{\mathcal{D}_{n-1}^x} \mathbb{C}_n^x$$

is the resulting aligned chamber complex decomposition sequence, and that E is aligned with the resulting Heegaard splitting of $\mathbb{C}_n^x$. Suppose further that for some $0 \leq i \leq n$ there is a bubble $\mathfrak{b}$ in one of the chambers C of $\mathbb{C}_i$ and the bubble $\mathfrak{b}$ remains disjoint from each set of disks $\mathcal{D}_j, j \geq i$ and from $E \subset \mathbb{C}_n$. That is, the alignment of the disks in the remainder of the sequence is $\mathfrak{b}$-wise. Then say that $\mathfrak{b}$ is an *intact bubble* for the sequence of aligned chamber complex decompositions.

Suppose next that the alignment of some disk set $\mathcal{D}_i$ is altered simply by a bubble pass of $\mathfrak{b}$ to another chamber C' of $\mathbb{C}_{i+1}$, and each successive alignment of the $\mathcal{D}_j$, $j \geq i+1$ and the alignment of E all are $\mathfrak{b}$-wise, so that $\mathfrak{b}$ remains an intact bubble for the resulting aligned chamber complex decomposition

$$\vec{\mathbb{C}}^y : \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0^y} \mathbb{C}_1^y \xrightarrow{\mathcal{D}_1^y} \mathbb{C}_2^y \xrightarrow{\mathcal{D}_2^y} \dots \xrightarrow{\mathcal{D}_{n-1}^y} \mathbb{C}_n^y$$

and alignment of E. (By construction $\mathbb{C}_k^x = \mathbb{C}_k^y$ and $\mathcal{D}_{k-1}^x = \mathcal{D}_{k-1}^y$ for $1 \leq k \leq i$.) Denote the two alignments of E by E^x and E^y respectively.

Definition 6.6. *In the construction above, the aligned chamber complex decomposition sequence $\vec{\mathbb{C}}^y$ and the alignment E^y of E are said to be obtained from $\vec{\mathbb{C}}^x$ and E^x by an* ancestral bubble pass. *This includes the degenerate case, in which $\vec{\mathbb{C}}^x = \vec{\mathbb{C}}^y$ but the alignments E^x, E^y differ by a bubble pass in $\mathbb{C}_n$.*

Proposition 6.7. *Suppose aligned chamber complex decomposition sequences $\vec{\mathbb{C}}^x$ and $\vec{\mathbb{C}}^y$ and alignments E^x, E^y are chosen for the chamber complex decomposition sequence $\vec{\mathbb{C}}$ and the disk-sphere set E in $\mathbb{C}_n$. Then there is a sequence of ancestral bubble passes and eyeglass moves that changes the pair $(\vec{\mathbb{C}}^x, E^x)$ to $(\vec{\mathbb{C}}^y, E^y)$.*

Proof. The proof begins much as the proof of Lemma 6.5. By [FS2] there is a series of bubble passes and eyeglass moves that change the alignment $\mathcal{D}_0^x$ to $\mathcal{D}_0^y$. The proof is by induction on the pair (n, p), lexicographically ordered, where p is the minimal number of bubble passes in such a series. Lemma 6.5 covers the case $n = 1$, so assume $n \geq 2$. In case $p = 0$, we can take $\mathcal{D}_0^x = \mathcal{D}_0^y$, so $\mathbb{C}_1^x = \mathbb{C}_1^y$ and then apply the inductive assumption to the shorter sequences

$$\mathbb{C}_1^x \xrightarrow{\mathcal{D}_1^x} \mathbb{C}_2^x \xrightarrow{\mathcal{D}_2^x} \dots \xrightarrow{\mathcal{D}_{n-1}^x} \mathbb{C}_n^x \supset E^x$$

and

$$\mathbb{C}_1^y \xrightarrow{\mathcal{D}_1^y} \mathbb{C}_2^y \xrightarrow{\mathcal{D}_2^y} \dots \xrightarrow{\mathcal{D}_{n-1}^y} \mathbb{C}_n^y \supset E^y.$$

So we may suppose $p \geq 1$ and, as in the proof of Lemma 6.5, let $\mathfrak{b}$ be the first bubble that is passed, through a disk $D \in \mathcal{D}_0^x$, from a chamber $C \in \hat{\mathbb{C}}_1^x$ to a chamber $C' \in \hat{\mathbb{C}}_1^x$. Let $\mathcal{D}_0^w$ be the disk set $\mathcal{D}_0$ as realigned by this bubble pass. Let

$$\mathbb{C}_1^x \xrightarrow{\mathcal{D}_1^z} \mathbb{C}_2^z \xrightarrow{\mathcal{D}_2^z} \dots \xrightarrow{\mathcal{D}_{n-1}^z} \mathbb{C}_n^z \supset E^z$$

and

$$\mathbb{C}_1^w \xrightarrow{\mathcal{D}_1^w} \mathbb{C}_2^w \xrightarrow{\mathcal{D}_2^w} \dots \xrightarrow{\mathcal{D}_{n-1}^w} \mathbb{C}_n^w \supset E^w$$

be $\mathfrak{b}$-*wise* alignments of the aligned chamber complex sequences, with $\mathfrak{b} \subset C$ in $\mathbb{C}_1^x$ (as is given) and $\mathfrak{b} \subset C'$ in $\mathbb{C}_1^w$. (Here the significant difference between the given $\mathcal{D}_i^x$ and the new $\mathcal{D}_i^z$, $i \geq 1$, is that each $\mathcal{D}_i^z$ is $\mathfrak{b}$-wise aligned.) By definition, the two sequences differ by an ancestral bubble pass and by inductive assumption (on n) the former sequence differs from

$$\mathbb{C}_1^x \xrightarrow{\mathcal{D}_1^x} \mathbb{C}_2^x \xrightarrow{\mathcal{D}_2^x} \dots \xrightarrow{\mathcal{D}_{n-1}^x} \mathbb{C}_n^x \supset E^x$$

by a sequence of ancestral bubble passes and eyeglass moves. Extending to the left we then have sequences

$$\vec{\mathbb{C}}^x : \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0^x} \mathbb{C}_1^x \xrightarrow{\mathcal{D}_1^x} \mathbb{C}_2^x \xrightarrow{\mathcal{D}_2^x} \ldots \xrightarrow{\mathcal{D}_{n-1}^x} \mathbb{C}_n^x \supset E^x,$$

$$\vec{\mathbb{C}}^z : \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0^x} \mathbb{C}_1^z \xrightarrow{\mathcal{D}_1^z} \mathbb{C}_2^z \xrightarrow{\mathcal{D}_2^z} \ldots \xrightarrow{\mathcal{D}_{n-1}^z} \mathbb{C}_n^z \supset E^z,$$

$$\vec{\mathbb{C}}^w : \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0^w} \mathbb{C}_1^w \xrightarrow{\mathcal{D}_1^w} \mathbb{C}_2^w \xrightarrow{\mathcal{D}_2^w} \ldots \xrightarrow{\mathcal{D}_{n-1}^w} \mathbb{C}_n^w \supset E^w$$

$$\vec{\mathbb{C}}^y : \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0^y} \mathbb{C}_1^y \xrightarrow{\mathcal{D}_1^y} \mathbb{C}_2^y \xrightarrow{\mathcal{D}_2^y} \ldots \xrightarrow{\mathcal{D}_{n-1}^y} \mathbb{C}_n^y \supset E^y$$

and we have shown that any pair of the first three ($\vec{\mathbb{C}}^x, \vec{\mathbb{C}}^z$, and $\vec{\mathbb{C}}^w$) differ by a sequence of ancestral bubble passes and eyeglass moves. Now notice that, by construction, the two aligned chamber complex decompositions $\mathbb{C}_0 \xrightarrow{\mathcal{D}_0^w} \mathbb{C}_1^w$ and $\mathbb{C}_0 \xrightarrow{\mathcal{D}_0^y} \mathbb{C}_1^y$ differ by a sequence of eyeglass moves and at most $p-1$ bubble passes. Hence, by inductive assumption (on p), the last two sequences (hence also the first and the last) differ by a sequence of ancestral bubble passes and eyeglass moves, as required. □

We now turn to S^3. Throughout the remaining part of this section, make the inductive Assumption 3.4. That is, the Goeritz group of S^3 is the eyeglass group on splittings of genus $< g$.

Suppose a Heegaard split chamber complex $\mathbb{C}_0$ supports the genus g Heegaard splitting $S^3 = A \cup_T B$ and

$$\vec{\mathbb{C}} : \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a sequence of chamber complex decompositions. Suppose further that $\mathbb{C}_n$ contains an incompressible sphere S.

Lemma 6.8. *Suppose*

$$\vec{\mathbb{C}}^x : \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0^x} \mathbb{C}_1^x \xrightarrow{\mathcal{D}_1^x} \mathbb{C}_2^x \xrightarrow{\mathcal{D}_2^x} \ldots \xrightarrow{\mathcal{D}_{n-1}^x} \mathbb{C}_n^x \supset S^x$$

and

$$\vec{\mathbb{C}}^y : \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0^y} \mathbb{C}_1^y \xrightarrow{\mathcal{D}_1^y} \mathbb{C}_2^y \xrightarrow{\mathcal{D}_2^y} \ldots \xrightarrow{\mathcal{D}_{n-1}^y} \mathbb{C}_n^y \supset S^y$$

are sequences of aligned chamber complex decompositions resulting from possibly different choices of alignment at each stage. Suppose further that the sequences differ by an ancestral bubble pass, of the bubble $\mathfrak{b}$. (In particular S is aligned with $\mathbb{C}_n^x$ and $\mathbb{C}_n^y$ and is disjoint from $\mathfrak{b}$.) Denote S as so aligned S^x and S^y respectively. Then $\mathbb{C}_n^x$ and $\mathbb{C}_n^y$ cocertify.

Proof. We sketch the proof; more detail appears in the earlier proof of Proposition 3.8.

Let $S_{\mathfrak{b}}$ be the sphere $\partial\mathfrak{b}$, defining, after amalgamation, a reducing sphere for the original Heegaard splitting $A \cup_T B$. By Lemma 3.5 $h_{S^x} \sim h_{S_{\mathfrak{b}}} \sim h_{S^y} : (S^3, T) \to (S^3, T_g)$ as required. □

Now return to the setting of Proposition 6.4: Suppose $S^3 = A \cup_T B$ is a genus g Heegaard splitting of S^3, a Heegaard split chamber complex $\mathbb{C}_0$ supports T and

$$\vec{\mathbb{C}} : \quad : \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a sequence of chamber complex decompositions.

Proposition 6.9. *Suppose chambers $C_0 \in \mathbb{C}_0$ and $C_n \in \mathbb{C}_n$ each contain incompressible spheres. Then, for* any *alignment of each disk set $\mathcal{D}_i$ in $\mathbb{C}_i$, the Heegaard split chamber complexes $\mathbb{C}_0$ and $\mathbb{C}_n$ cocertify.*

Proof. By Proposition 6.4 there is some alignment $\vec{\mathbb{C}}^x$ for the sequence so that $\mathbb{C}_0$ and $\mathbb{C}_n^x$ cocertify. By Proposition 6.7 any other alignment $\vec{\mathbb{C}}^y$ can be obtained from $\vec{\mathbb{C}}^x$ by a sequence of ancestral bubble passes and eyeglass moves. Lemma 6.8 shows that then $\mathbb{C}_n^x$ and $\mathbb{C}_n^y$ cocertify. Hence $\mathbb{C}_0$ and $\mathbb{C}_n^y$ cocertify. □

Corollary 6.10. *Suppose (S^3, T) is a genus g Heegaard splitting of S^3 and*

$$\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a sequence of aligned chamber complex decompositions of Heegaard split chamber complexes supporting T. If S and S' are each incompressible spheres in possibly different chambers of the sequence, then those Heegaard split chamber complexes cocertify. □

Unfortunately, declaring every disky ball to be a goneball, as is done throughout this Section, erases too much information. A critical example is this: consider a sequence of two aligned chamber complex decompositions of Heegaard split chamber complexes

$$\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2$$

and suppose that a chamber C_1 of $\mathbb{C}_1$ is a handlebody that is not disky in the first decomposition. We know from Proposition 5.12 that the Heegaard splitting of C_1 is non-trivial. Now suppose that the disks of $\mathcal{D}_1$ that are incident to the handlebody C_1 are a complete set of meridians for C_1. The result is a ball chamber of $\mathbb{C}_2$ that has non-trivial Heegaard splitting. As shown in the proof of Proposition 3.7, this is (under the inductive Assumption 3.4) enough information to determine an eyeglass equivalence class $(S^3, T) \to (S^3, T_g)$. But if we take all disky balls to be goneballs, this ball would be absorbed because it is disky under the decomposition $\mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2$. Thus we lose the information it contains.

Motivated by this example, we will use a more complicated rule to declare that, under certain conditions, a disky ball should *not* be declared a goneball. The rule will focus on how handlebodies are treated in a disk decomposition. The critical step is to be more restrictive in how we allow disks to be aligned.

7 Preferred alignment and flagged chamber complexes

Return now to the general case, which we briefly review: $M = A \cup_T B$ is a Heegaard splitting of a compact 3-manifold; $\mathbb{C}$ is a Heegaard split chamber complex in M that supports T; $\mathcal{D}$ is an aligned disk set in $\mathbb{C}$; $\hat{\mathbb{C}}_{\mathcal{D}}$ is the chamber complex obtained by surgery on $\mathcal{D}$; and $\mathbb{C}_{\mathcal{D}}$ is the Heegaard split chamber complex obtained by using Rule 5.6 to declare goneballs. Thus

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is a Heegaard split chamber complex decomposition.

Because the disks are aligned, each chamber inherits a Heegaard splitting, as described before Definition 5.11. We consider 3-balls to be (genus 0) handlebodies and introduce the terminology:

Definition 7.1. *A Heegaard split handlebody chamber in $\mathbb{C}$ is* empty *if the Heegaard splitting is trivial. If the Heegaard splitting is non-trivial, the handlebody chamber is called* occupied*. A* flagged *chamber complex is a Heegaard split chamber complex in which each handlebody chamber is labelled either empty or occupied.*

Two flagged chamber complexes are the same as flagged chamber complexes if they are the same as chamber complexes and the flagging of each handlebody chamber is the same. The Heegaard splittings of any given chamber are not necessarily isotopic, nor even of the same genus.

Definition 7.2. *The alignment of a disk set $\mathcal{D}$ in $\mathbb{C}$ is a* preferred alignment *if, in each chamber C of $\mathbb{C}$, it has these properties:*

1. *If any remnant of C is* not *a disky handlebody, then each disky handlebody remnant is empty.*
2. *If every remnant of C is a disky handlebody, then, per Proposition 4.8, C is a handlebody. In this case:*
   (a) *If C is empty, so is every remnant*
   (b) *If C is occupied, then exactly one remnant of C is occupied.*

Corollary 7.3. *Suppose $\mathcal{D}$ is a disk set that is in preferred alignment in a flagged chamber complex $\mathbb{C}$. Suppose $\mathbb{C}$ has no occupied handlebody chambers. Then, under the decomposition*

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

each handlebody chamber in $\mathbb{C}_{\mathcal{D}}$ is empty if and only if it is disky.

Proof. If a handlebody chamber in $\mathbb{C}_{\mathcal{D}}$ is not disky, then it is occupied per Proposition 5.12, so the interest is in the other direction. Suppose C' is a disky handlebody chamber in $\mathbb{C}_{\mathcal{D}}$, a remnant of a chamber C in $\mathbb{C}$. By hypothesis, C is not an occupied handlebody so, per Definition 7.2(2a), if every remnant of C is a disky handlebody then every remnant, including C', is empty. On the other hand, if not every remnant of C is a disky handlebody then per 7.2(1) C' is empty. Thus in every case C' is empty, as required. □

We will first show that any disk set has a preferred alignment. A critical fact from Heegaard theory is that any Heegaard splitting of a handlebody is standard, that is it is a stabilization of the trivial splitting. (See, for example, the discussion at the end of Section 6 or, in more detail, the proof of Proposition 8.1 below.) In particular, if $H = A_H \cup_{T_H} B_H$ is a Heegaard split handlebody with non-trivial splitting then there is a bubble $\mathfrak{b}$ for the splitting so that if the bubble is destabilized (for example by replacing $\mathfrak{b}$ with a ball containing just a properly embedded equatorial disk), the splitting becomes trivial. That is, the splitting surface becomes parallel to ∂H. Such a bubble will be called a *maximal bubble* in H.

Lemma 7.4. *Suppose*

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is a Heegaard split chamber complex decomposition and $H = A_H \cup_{T_H} B_H$ is an occupied disky handlebody remnant of a chamber $C \in \mathbb{C}$, with $C = A_C \cup_{T_C} B_C$. Then there is an isotopy of T_H in H so that afterwards

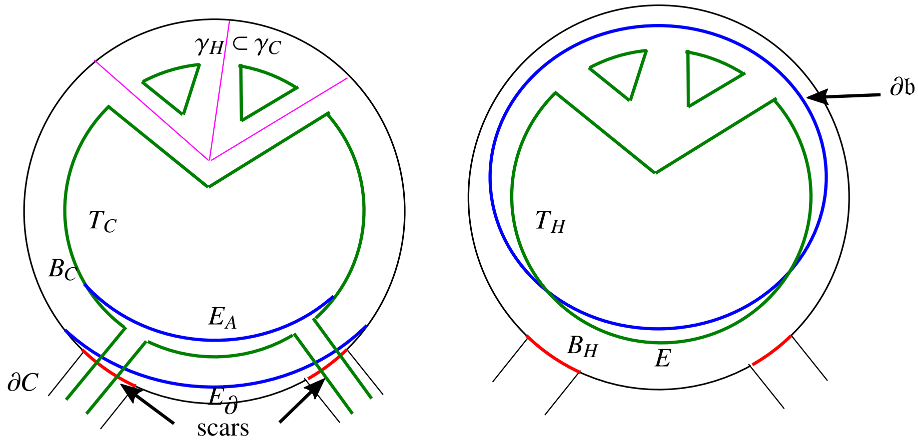

Figure 22: Case 1 of Lemma 7.4

- *There is a disk $E \subset T_H$ so that $T_H - E \subset T_C$ and*
- *There is a maximal bubble $\mathfrak{b}$, disjoint from E, in the splitting of H that is also a bubble in the splitting of C.*

Proof. With no loss assume C and hence H are A-chambers. We briefly set terminology and review the results of Heegaard split chamber complex decomposition from Subsection 5.3. Let Σ_C be a spine for B_C consisting of the union of ∂C and a graph $\gamma_C \subset C$ incident to ∂C only in its ends. We may as well focus on those disks in $\mathcal{D}$ that are incident to C and retreat to denoting these as $\mathcal{D}$. Since $\mathcal{D}$ is aligned, each disk $D \in \mathcal{D}$ intersects the spine Σ_C only in ∂D; the interior of D lies either in C or in a chamber adjacent to C. In particular, $\mathcal{D}$ is disjoint from the graph γ_C. Except for scars left on ∂H by $\mathcal{D}$, $\partial H \subset \partial C$. If $D \in \mathcal{D}$ leaves an external scar on ∂H then D lies in C; if it leaves an internal scar then $\mathrm{int}(D)$ lies in an adjacent chamber. There is a spine Σ_H for B_H which consists of the union of ∂H with two graphs in H (whose union we denote γ_H): the graph $\gamma_C \cap H$ and a graph $\gamma_I \subset H$ determined by the internal scars in ∂H and goneballs in the interior of H. Each edge in γ_I is dual to a disk in $\mathcal{D}$ whose interior lies outside C. More specifically, each end of an edge of γ_I lies either on an internal scar in ∂H or on a vertex in $\gamma_I \cap \mathrm{int}(H)$ that corresponds to a goneball. Because the compression body B_H is connected, each component of γ_H has at least one end on ∂H.

We consider increasingly complicated cases:

Case 1: H is a ball and has no internal scars or goneballs. That is $\gamma_I = \emptyset$.

Connect the (external) scars of $\mathcal{D}$ on ∂H by a collection of arcs in ∂H which, by general position, can be taken to be disjoint from the ends of γ_H in ∂H. Viewing these arcs as edges and the scars as vertices of a graph in ∂H, choose the arcs so that the resulting graph is a tree. Let $E_\partial \subset \partial H$ be a (disk) regular neighborhood in ∂H of the union of the scars and arcs. Since that union is disjoint from the ends of γ_H we can take E_∂ also to be disjoint from the ends of γ_H.

The compression body B_C is a regular neighborhood of the spine Σ_C. Since $\mathcal{D}$ is aligned and so disjoint from γ_C, $\mathcal{D}$ intersects B_C only in a collar of ∂B_C. Hence B_C intersects each scar $\mathfrak{s}$ on ∂H in a collar of $\partial\mathfrak{s}$ in ∂H. A regular neighborhood of Σ_H, to which we can isotope B_H, can then be constructed from B_C in an obvious way: take the union of $B_C \cap H$ and a collar in H of each scar on ∂H. See Figure 22. With this construction, T_H lies in T_C except for the collection of disks in T_H that are parallel to the scars. If we then take E to be the disk in T_H parallel to E_∂, we have $T_H - E \subset T_C$, verifying the first assertion of the lemma in this case.

Since E is disjoint from γ_H, the interior of E can be pushed, rel ∂E, into A_H. Call the result E_A. Since H is a ball, the complement of E_∂ in ∂H is also a disk; let E_B be a copy of the disk $\partial H - E_\partial$ pushed up into the compression body B_H until its boundary lies on $\partial E = \partial E_A$ in T_H. The union of E_A and E_B is a sphere bounding a ball in H that contains all of $T_H - E$. That ball is then a maximal bubble for H that is also a bubble for the splitting of C, verifying the lemma in this case.

Case 2: H is a ball that may contain internal scars and goneballs.

Suppose $\mathfrak{s}$ is an internal scar on ∂H, a scar left by a disk in $\mathcal{D}$ that lies in an adjacent chamber C'. See Figure 23. Since by assumption H is a *disky* remnant, $\partial\mathfrak{s}$ bounds a disk component $E_\mathfrak{s}$ of $\partial C \cap \text{int}(H)$. Since H is irreducible, the disks $\mathfrak{s}$ and $E_\mathfrak{s}$ are isotopic rel ∂ in H, and all ball components of $\hat{\mathbb{C}}_\mathcal{D}$ must be goneballs and therefore trivially split. In particular the Heegaard splitting surface of the adjacent chamber C' does not contribute any genus to T_H.

Let $\partial H_\mathfrak{s}$ be the copy of ∂H obtained by replacing each interior scar $\mathfrak{s}$ with the corresponding disk $E_\mathfrak{s} \subset \partial C$. Now apply the argument of Case 1 to the ball in H bounded by $\partial H_\mathfrak{s}$ instead of ∂H, with the arcs in ∂H between external scars in that construction chosen to avoid also internal scars. This completes the proof in this case. (Note that with this construction, T_H is disjoint from the ball in H between the scar $\mathfrak{s}$ and the disk $E_\mathfrak{s}$.)

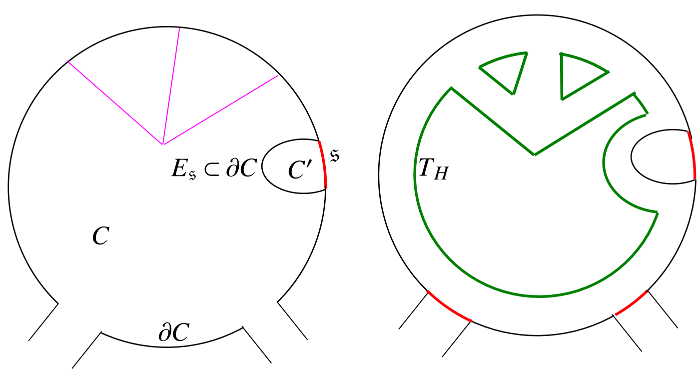

Figure 23: Case 2 of Lemma 7.4

Case 3: The general case.

Define the disk $E \subset T_H$ parallel to a disk $E_\partial \subset \partial H$ as in Case 1, with arcs chosen in its construction to also be disjoint from internals scars. Essentially the same argument as in Cases 1 and 2 shows that we can ensure that $T_H - E \subset T_C$. So we proceed to the second assertion of the Lemma: there is a maximal bubble in H that is disjoint from E and is also a bubble for C.

Let $\mathcal{D}_+$ be a minimal complete collection of meridian disks in H. That is, $\mathcal{D}_+$ is a properly embedded collection of disks in H so that the closed complement of a collar of $\mathcal{D}_+$ in H is a single ball W. By general position, we may take $\partial\mathcal{D}_+ \subset \partial H$ to be disjoint from E and all internals scars, and per [Sc1] we can take $\mathcal{D}_+$ to be aligned with the Heegaard surface T_H. This means in particular that the graph $\gamma_H \subset \Sigma_H$ can be taken to be disjoint from the meridian disks $\mathcal{D}_+$, so it lies entirely in the ball W.

Let $E_{\partial+}$ be the disk in ∂W obtained by band summing E_∂ to each copy of $\mathcal{D}_+$ in ∂W. (For each disk $D \in \mathcal{D}_+$ there are two copies $D_\pm$ in ∂W.) See Figure 24. The complement in the sphere ∂W of $E_{\partial+}$ is a disk that lies entirely in ∂H. In parallel fashion expand the disk E_A (per Case 1 a copy of E with interior pushed into A) to a disk E_{A+} properly embedded in A by band summing E_A to each disk $D_\pm \cap A, D \in \mathcal{D}_+$. Since $\partial W - E_+$ is a disk in ∂H, it follows as in Case 1 that $\partial E_{A+} \subset T_H$ bounds a disk in B_H. The union of the two disks is then a sphere parallel to ∂W and the ball $\mathfrak{b}$ it bounds in W is disjoint from E and contains all but the ends of γ_H. So $\mathfrak{b}$ is a maximal bubble, as required. ☐

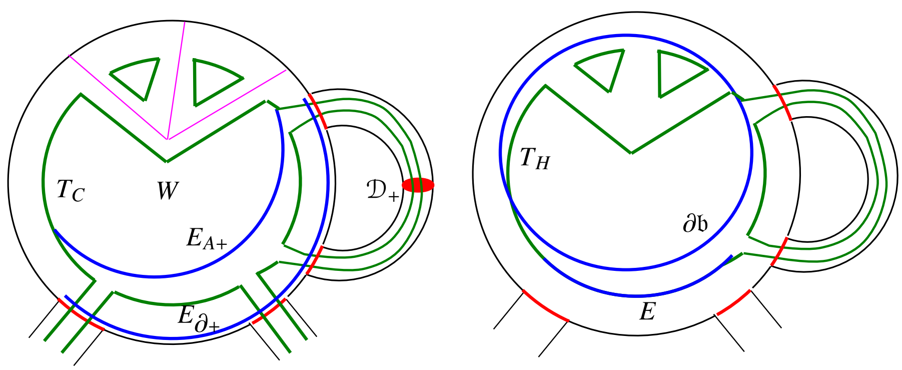

Figure 24: Case 3 of Lemma 7.4

Proposition 7.5. *Suppose $\mathbb{C}$ is a Heegaard split chamber complex in $M = A \cup_T B$ that supports T, and $\mathcal{D}$ is a disk set in $\mathbb{C}$. Then $\mathcal{D}$ has a preferred alignment.*

Proof. As discussed in Subsection 5.3, $\mathcal{D}$ can be aligned. The strategy will be to begin with any alignment and alter it to a preferred alignment.

We first describe how to realign $\mathcal{D}$ so as to achieve Definition 7.2(1). Suppose some remnant $C' \in \mathbb{C}_\mathcal{D}$ of a chamber $C \in \mathbb{C}$ is not a disky handlebody, and let $H = \{H_1, ..., H_k\}$ be the collection of occupied disky handlebody remnants of C. If $H = \emptyset$ there is nothing to prove. Otherwise, let $\gamma = \{\gamma_1, ..., \gamma_k\}$ be a collection of arcs in C with the following properties:

- Each γ_i is a path in the Heegaard surface T_C that is transverse to $\mathcal{D}$ and runs from a point in $\mathrm{int}(C')$ to a point in $\mathrm{int}(H_i)$.
- Among all such families of arcs, $|\gamma \cap \mathcal{D}| \geq k$ is minimized.

Let $D \in \mathcal{D}$ be the disk containing the closest intersection point on γ_1 to its endpoint in H_1. Let C'' be the remnant that is on the other side of D. Since $|\gamma \cap \mathcal{D}|$ is minimized, γ_1 is otherwise disjoint from D and $C'' \neq H_1$. H_1 is assumed to be occupied; let $\mathfrak{b}$ be a maximal bubble for the splitting $H_1 = A_1 \cup_{T_1} B_1$ satisfying the conclusion of Lemma 7.4. That is, $\mathfrak{b}$ is also a bubble for the Heegaard splitting $C = A_C \cup_{T_C} B_C$.

Take the endpoint of γ_1 in H_1 to be a point in the circle $c_{\mathfrak{b}} = \partial\mathfrak{b} \cap T_C$, so the subpath $\gamma_H = \gamma_1 \cap H_1$ is an arc in T_C from D to the sphere $\partial\mathfrak{b}$. Tube sum D to $\partial\mathfrak{b}$ along the path γ_H, creating a new disk D'. Observe that D' intersects T_C in a single circle, namely the band sum along γ_H of the circles $D \cap T_C$ and $\partial\mathfrak{b} \cap T_C$. Thus D' is an aligned disk, and is isotopic to D in C because $\mathfrak{b}$ is a ball. Moreover, the bubble $\mathfrak{b}$ lies on the same side of D' as C'', not H. Thus D' is a realignment of D that makes H an empty handlebody. If C'' is not a handlebody, or is a handlebody that was already occupied, then k is reduced by 1. If C'' was an empty handlebody then after the realignment C'' becomes an occupied handlebody, leaving k unchanged. But the path γ_1 now has its endpoint in C'' and no longer intersects the realigned D', so $|\gamma_1 \cap \mathcal{D}|$ (with D realigned) is reduced by 1. Continue the realignment of disks until $k = |\gamma \cap \mathcal{D}| = 0$ as required.

To establish Definition 7.2(2a) note that if any remnant of C is a disky occupied handlebody then Lemma 7.4 shows C contains a bubble and so cannot have trivial splitting; it is occupied. To establish 7.2(2b), use the same argument as for 7.2(1) above, but make an arbitrary choice of any remnant for C'. □

Definition 7.6. *Suppose $\mathbb{C}$ is a flagged chamber complex, $\mathcal{D}$ is a disk set in $\mathbb{C}$ with a preferred alignment, and $\hat{\mathbb{C}}_{\mathcal{D}}$ is the chamber complex obtained by surgery on $\mathcal{D}$. Let $\mathbb{C}_{\mathcal{D}}$ be the flagged chamber complex obtained by using Rule 5.6 to declare goneballs. Then the Heegaard split chamber complex decomposition*

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is called a flagged *chamber complex decomposition.*

For much of our argument, simply keeping track of the flagging of the chamber complexes (that is, whether a handlebody chamber is empty or occupied) will suffice.

Proposition 7.7. *Suppose, for $\mathcal{D}$ a disk set in $\mathbb{C}$,*

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is a flagged chamber complex decomposition.

1. *Any ball chamber in $\mathbb{C}_{\mathcal{D}}$ is occupied. (That is, $\mathbb{C}_{\mathcal{D}}$ is a Heegaard split chamber complex.)*
2. *If any remnant of a chamber C in $\mathbb{C}$ is an occupied disky handlebody then C is an occupied handlebody and every remnant of C in $\mathbb{C}_{\mathcal{D}}$ is a disky handlebody.*
3. *Suppose $\mathcal{D}$ is given a different preferred alignment and $\mathbb{C}'_{\mathcal{D}}$ is the resulting flagged chamber complex structure. If $\mathbb{C}'_{\mathcal{D}}$ and $\mathbb{C}_{\mathcal{D}}$ differ as flagged chamber complexes, then $\mathbb{C}$ contains an occupied handlebody chamber whose remnants in $\mathbb{C}_{\mathcal{D}}$ and $\mathbb{C}'_{\mathcal{D}}$ are all disky handlebodies.*

Proof. (1) If a ball chamber were not occupied, it would be empty so, by Definition 7.1, it would have trivial Heegaard splitting. But then, by Rule 5.6, it would have been a goneball.

(2) The proof is essentially the same as that of Corollary 7.3: If a remnant of C is an occupied disky handlebody then from the contrapositive of Definition 7.2(1) every remnant of C is a disky handlebody. Then the contrapositive of Definition 7.2(2a) says that C must be an occupied handlebody.

(3) In a flagged chamber complex decomposition, the flagging of the chambers after surgery, as described in Definition 7.2, depends only on the disks $\mathcal{D}$ and not on how they are aligned with the Heegaard surfaces in the chamber, except in satisfying Definition 7.2(2b): If C is an occupied handlebody chamber of $\mathbb{C}$ and each remnant of C is a disky handlebody, then a choice of alignment is made to ensure that exactly one remnant (of possibly several) is occupied. Hence if $\mathbb{C}'_{\mathcal{D}}$ and $\mathbb{C}_{\mathcal{D}}$ differ, so $\hat{\mathbb{C}}_{\mathcal{D}}$ and $\hat{\mathbb{C}}'_{\mathcal{D}}$ differ, it must be because of a different such choice. That is, in $\hat{\mathbb{C}}_{\mathcal{D}}$ one remnant is an occupied handlebody, and in $\hat{\mathbb{C}}'_{\mathcal{D}}$ a different remnant is. Thus, per statement (2), C is the required occupied handlebody chamber of $\mathbb{C}$. □

Suppose, as described in the proof of Proposition 7.7(3), two flagged chamber complex decompositions $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$ and $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}'_{\mathcal{D}}$ differ only because, for some occupied handlebody chambers in $\mathbb{C}$, different alignments of the disks $\mathcal{D}$ in the chambers result in different remnants being occupied.

Definition 7.8. *The decompositions $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$ and $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}'_{\mathcal{D}}$ are* sibling *decompositions, and the occupied handlebody chambers in $\mathbb{C}$ are called* parent *chambers in the sibling decompositions.*

Proposition 7.5. is an existence statement; we now move towards a parallel uniqueness statement.

Lemma 7.9. *Suppose $H = A \cup_T B$ is a Heegaard split handlebody. Suppose $\mathfrak{b}^p$ and $\mathfrak{b}^q$ are not necessarily disjoint bubbles for T of genus $p \le q$ respectively. Then*

- *There is a genus $(q-p)$ bubble $\mathfrak{b}'$ for T that is disjoint from $\mathfrak{b}^p$, and a homeomorphism $h : (H,T) \to (H,T)$ so that h is the identity on ∂H, and $\mathfrak{b}^q$ is the tube sum of $h(\mathfrak{b}^p)$ and $h(\mathfrak{b}')$.*
- *Under Assumption 3.4, if* $\text{genus}(T) - \text{genus}(H) \le g-1$ *then we may take h to be an eyeglass move.*

Proof. Pick an aligned disk set $\mathcal{D}$ in H so that $H - \eta(\mathcal{D})$ is a single ball. This implies $|\mathcal{D}| = \text{genus}(H)$ and each disk is non-separating. Then surgery on $\mathcal{D}$ gives a genus $g' = \text{genus}(T) - \text{genus}(H)$ Heegaard splitting of the ball which, by [Wa], is unique. More explicitly, echoing the notation surrounding Figure 7, there is a nested sequence of g' bubbles $\mathfrak{b}_1 \subset ... \subset \mathfrak{b}_{g'}$ for T so that each sphere $S_i = \partial \mathfrak{b}_i$ intersects T in a single circle c_i, and $T/\mathfrak{b}_{g'}$ is parallel to ∂H.

More specifically, align $\mathcal{D}$ first with $T/\mathfrak{b}^q$ so that in the associated alignment of $\mathcal{D}$ with T, $\mathfrak{b}^q$ is disjoint from $\mathcal{D}$. Then with no loss of generality we may assume that $\mathfrak{b}^q = \mathfrak{b}_q$. For the first statement above, it then suffices to find a homeomorphism $h(H,T) \to (H,T)$ so that $h(\mathfrak{b}^p) = \mathfrak{b}_p$. Denote this alignment of $\mathcal{D}$ by $\mathcal{D}^q$.

Similarly, let $\mathcal{D}^p$ be an alignment of $\mathcal{D}$ found by first aligning with $T/\mathfrak{b}^p$ and then observing that the associated alignment of $\mathcal{D}^p$ with T is disjoint from $\mathfrak{b}^p$. According to [FS2] there is a sequence of bubble passes and eyeglass moves that will take $\mathcal{D}^p$ to $\mathcal{D}^q$ and, since each disk in $\mathcal{D}$ is non-separating disk, it follows from Lemma 2.5 that even the bubble passes are eyeglass moves. (To pass a bubble through a disk $D \in \mathcal{D}$ choose the arc β in the proof to be disjoint from all other disks in $\mathcal{D}$.) Thus there is an eyeglass move that takes $\mathcal{D}^p$ to $\mathcal{D}^q$. Put another way, there is an eyeglass move $h_1 : (H,T) \to (H,T)$ so that $h_1(\mathfrak{b}^p)$ lies in the ball $H - \eta(\mathcal{D}^q)$. By [Wa] there is then a homeomorphism $h_2 : (H,T) \to (H,T)$ so that $h_2(h_1(\mathfrak{b}^p)) = \mathfrak{b}_p$. This completes the proof of the first statement, with $h = h_2 h_1$.

To prove the second statement, just note that T induces a genus g' splitting on $H - \eta(\mathcal{D}^q)$, so if $g' \le g-1$ then we may take h_2 to be an eyeglass move, so $h_2 h_1$ is also an eyeglass move. □

Throughout the remainder of this section $\mathbb{C}$ is a flagged chamber complex in M supporting a genus g Heegaard splitting, with $\mathcal{D}$ a disk set in $\mathbb{C}$. We further continue with Assumption 3.4. Let $\mathcal{D}_\pm$ denote two preferred alignments of $\mathcal{D}$ that result in the same flagged chamber complex decompositions $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}_\pm}$. That is, to repeat from Definition 7.1, $\mathbb{C}_{\mathcal{D}_\pm}$ are identical chamber complexes with identical flagging (handlebody chambers either empty or occupied), though the underlying Heegaard splittings of each chamber may differ.

Proposition 7.10. *There is a sequence of preferred alignments of $\mathcal{D}$ in $\mathbb{C}$*

$$\mathcal{D}_- = \mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3 ..., \mathcal{D}_n = \mathcal{D}_+$$

so that

a) *for each $1 \leq i \leq n$, $\mathcal{D}_i$ is obtained from $\mathcal{D}_{i-1}$ by eyeglass moves and a single bubble pass, possibly through multiple disks of $\mathcal{D}_{i-1}$.*

b) *The result $\mathbb{C}_{\mathcal{D}_i}$ of each flagged chamber complex decomposition*

$$\mathbb{C} \xrightarrow{\mathcal{D}_i} \mathbb{C}_{\mathcal{D}_i}$$

is the same flagged chamber complex as $\mathcal{D}_\pm$.

Proof. The central theorem of [FS2] is that there is such a sequence of alignments, but without the condition that each $\mathcal{D}_i$ is in *preferred* alignment and that each resulting Heegaard split chamber complex $\mathbb{C}_{\mathcal{D}_i}$ has the same flagging. We begin with such a sequence and alter it to achieve these conditions.

Claim: It suffices to find a sequence of (not necessarily preferred) alignments satisfying a) and b) so that each disky handlebody chamber of $\mathbb{C}_{\mathcal{D}}$ that is empty in $\mathbb{C}_{\mathcal{D}_-}$ is also empty in each $\mathbb{C}_{\mathcal{D}_i}$.

Proof of claim: We will *assume we have found such a sequence of alignments* and show how the proof of the proposition follows.

First observe that by Proposition 5.12 any handlebody chamber of $\mathbb{C}_{\mathcal{D}}$ that is not disky is occupied regardless of alignments, so the only difference in possible flaggings of the $\mathbb{C}_{\mathcal{D}_i}$ is in the flagging of disky handlebodies. Since $\mathbb{C}_{\mathcal{D}_-}$ has a preferred alignment, it satisfies Definition 7.2(1), namely that if any remnant of C is not a disky handlebody, then each disky handlebody remnant is empty. But by the assumption, each disky handlebody remnant is then empty in $\mathbb{C}_{\mathcal{D}_i}$, so each $\mathbb{C}_{\mathcal{D}_i}$ also satisfies Definition 7.2(1).

Similarly, if each remnant of a chamber C is a disky handlebody (so C is a handlebody, per Proposition 4.8), then Definition 7.2(2a) says that if C is empty, every remnant is empty in $\mathbb{C}_{\mathcal{D}_-}$ and so, by assumption, also in each $\mathbb{C}_{\mathcal{D}_i}$. Thus each $\mathbb{C}_{\mathcal{D}_i}$ satisfies Definition 7.2(2a). On the other hand, if C is occupied, Definition 7.2(2b) says that exactly one remnant, say $C' \in \mathbb{C}_{\mathcal{D}_-}$ is occupied, so the others are all empty. By assumption, all remnants of C in each $\mathbb{C}_{\mathcal{D}_i}$, except possibly C', are then empty. But since C is occupied it has non-trivial splitting, that is the genus of its Heegaard splitting surface is greater than that of its boundary. So decomposition cannot leave only trivially split handlebodies. Thus some remnant of C must be empty in each $\mathbb{C}_{\mathcal{D}_i}$, and C' is the only possibility. Thus exactly one remnant of C (namely C') in each

$\mathbb{C}_{\mathcal{D}_i}$ is occupied, and so each $\mathbb{C}_{\mathcal{D}_i}$ also satisfies Definition 7.2(2b), with the same flagging as $\mathbb{C}_{\mathcal{D}_-}$. Thus each $\mathcal{D}_i$ is in preferred alignment, as required.

Having established the claim, we proceed to show that in fact there is a sequence of alignments satisfying a) and b) so that each disky handlebody chamber of $\mathbb{C}_{\mathcal{D}}$ that is empty in $\mathbb{C}_{\mathcal{D}_-}$ is also empty in each $\mathbb{C}_{\mathcal{D}_i}$. This follows from a min-max argument that we now describe.

Let H be the collection of disky handlebodies in $\mathbb{C}_{\mathcal{D}}$ that are designated as empty in $\mathbb{C}_{\mathcal{D}_-}$ and so by assumption, also in $\mathbb{C}_{\mathcal{D}_+}$. Since these are flagged chamber complexes, no chamber in H is a ball.

Consider any sequence of (not necessarily preferred) alignments, given by [FS2], that begins with $\mathcal{D}_-$ and ends with $\mathcal{D}_+$ so that each alignment in the sequence differs from the previous alignment by eyeglass moves before and after a single bubble pass, possibly through multiple disks. For each $0 \le i \le n$, let $\sigma_i \ge 0$ be the difference between the sum of all the genera of the Heegaard surfaces for H in $\mathbb{C}_{\mathcal{D}_i}$ and the sum of the genera of ∂H. If we can find a sequence so that each $\sigma_i = 0$, then each Heegaard splitting would be trivial, so each chamber in H would continue to be empty in each $\mathbb{C}_{\mathcal{D}_i}$. Following the claim above, this would conclude the proof of the lemma.

In search of a sequence of alignments for which $\sigma_i = 0$, let

$$\mathcal{D}_- = \mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3 ..., \mathcal{D}_n = \mathcal{D}_+$$

be one that minimizes, in lexicographic order, the pair (m, x), where m is the maximum of σ_i throughout the sequence, and x is the number of integers i for which $\sigma_i = m$. The aim is to show that $m = 0$. Towards a proof by contradiction, suppose $m \ge 1$ and let i be the smallest integer for which $\sigma_i = m$. We know that $1 \le i \le n-1$ since $\sigma_0 = \sigma_n = 0$.

So we have

1. $\sigma_{i-1} < m$
2. $\sigma_i = m$
3. $\sigma_{i+1} \le m$.

The first statement implies that the realignment of $\mathcal{D}_{i-1}$ to $\mathcal{D}_i$ moves a bubble from a remnant $R_1 \notin H$ to a remnant $R_2 \in H$. The third statement means that in the realignment of $\mathcal{D}_i$ to $\mathcal{D}_{i+1}$, which moves a bubble from remnant R_3 to remnant R_4, we cannot have both $R_3 \notin H$ and $R_4 \in H$.

Suppose that $R_2 \ne R_3$. In this case, the bubble in R_3 could be passed to R_4 before the bubble passes from R_1 to R_2. This does not affect σ_{i-1} or σ_{i+1} but may change σ_i, say to σ'_i. Since we cannot have both $R_3 \notin H$ and $R_4 \in H$, $\sigma'_i \le \sigma_{i-1}$. (1) then implies $\sigma'_i < m$. Thus we have reduced x (or m if always $j > i \implies \sigma_j < m$).

Next suppose $R_2 = R_3$, and just call it R. We have already seen that this remnant must be in H, and so a disky handlebody. Since bubbles can only be passed between remnants of the same chamber, we deduce that R_1, R, R_4 are all remnants of the same chamber of $\mathbb{C}$. In particular, it is possible to directly pass a bubble from R_1 to R_4. Let g_1, g_4 be respectively the genera of the bubble $\mathfrak{b}_1$ that is moved into R from R_1 and the bubble $\mathfrak{b}_4$ that is moved from R into R_4. There are two cases to consider:

Case 1: $g_1 \le g_4$

Let $D_1 \subset \partial R$ be the disk through which $\mathfrak{b}_1$ is passed into R from R_1 as it lies in $\mathcal{D}_{i-1}$. We proceed as though D_1 is the only disk through which $\mathfrak{b}_1$ is passed, so it is passed from one remnant of C to an adjacent one; if it is passed through multiple chambers the argument only needs to be altered by taking multiple parallel copies of D_1, corresponding to how the thin tube from R to $\mathfrak{b}_1$ passes through other disks in $\mathcal{D}_{i-1}$ in the bubble pass. Similarly let $D_4 \subset \partial R$ be the disk (or parallel copies of the disk) in $\mathcal{D}_{i-1}$ through which $\mathfrak{b}_4$ is passed into R_4. (Possibly $R_1 = R_4$ and $D_1 = D_4$.)

Let $(D_s, \partial D_s) \subset (R_1, \partial D_1)$ be a disk parallel to D_1 obtained by tube-summing $\mathfrak{b}_1$ to D_1. Similarly, let H_+ be the handlebody obtained from R by the same tube-summing, replacing D_1 with D_s. Then H_+ contains both $\mathfrak{b}_1$ and $\mathfrak{b}_4$. According to Lemma 7.9 there is an eyeglass move $h : H_+ \to H_+$ and a bubble $\mathfrak{b}'$ of genus $g_4 - g_1$ so that $\mathfrak{b}_4$ is the tube sum of $\mathfrak{b}'$ and $h(\mathfrak{b}_1)$. Replace by an eyeglass move $D_1 \in \mathcal{D}_{i-1}$ with $D'_1 = h(D_1)$. See the highly schematic Figure 25.

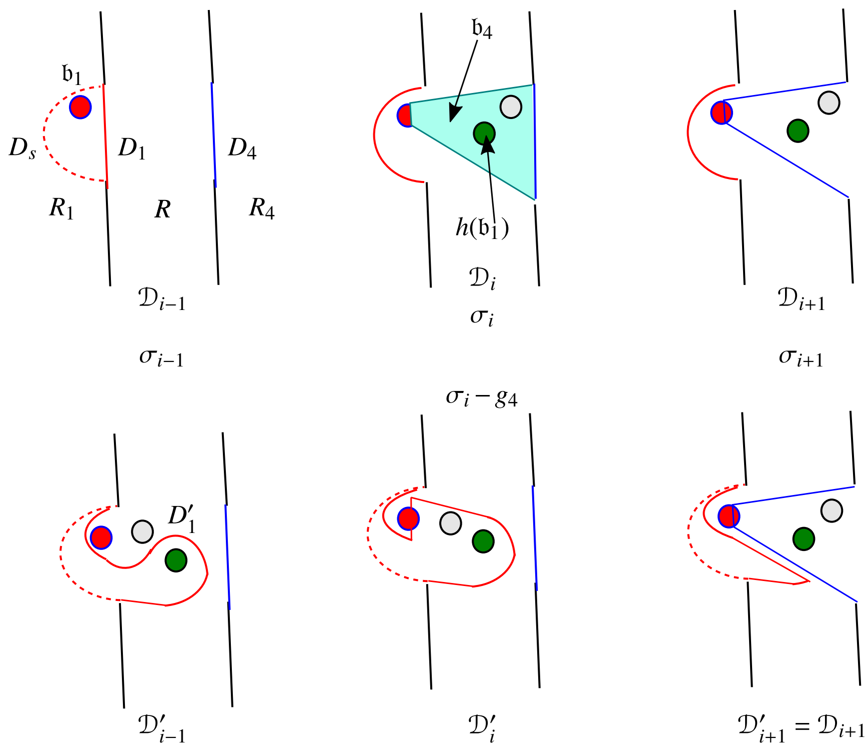

Figure 25: Case 1 when $R_2 = R_3 = R$

Do the transition from the alignment of $\mathcal{D}_{i-1}$ to that of $\mathcal{D}_{i+1}$ with these bubble passes instead of the given two: Tube sum D'_1 to $\mathfrak{b}'$, after which $\mathfrak{b}_4$ lies entirely in R_1. This changes σ to $\sigma_i - g_4$, since $R_1 \notin H$. Then pass the bubble $\mathfrak{b}_4$ directly from R_1 into R_4. This realignment leaves D_4 aligned so that R_4 contains $\mathfrak{b}_4$. The upshot of this replacement of D_1 with D'_1 in $\mathcal{D}_{i-1}$ is that the first occurrence of $\sigma_i = m$ in the sequence is replaced with $m - g_4$. Thus x is reduced by one (or m is reduced, if always $j > i \implies \sigma_j < m$).

Case 2: $g_4 \leq g_1$

The argument is much the same, using H_+ as defined above, with the following alteration: According to Lemma 7.9 there is a homeomorphism $h : H_+ \to H_+$ and a bubble $\mathfrak{b}'$ of genus $g_1 - g_4$ so that $h(\mathfrak{b}_1)$ is the tube sum of $\mathfrak{b}'$ and $\mathfrak{b}_4$.

Do the transition from the alignment of $\mathcal{D}_i$ to that of $\mathcal{D}_{i+1}$ with these bubble passes instead of the given two: First replace D_1 with $h(D_1)$. After this realignment $\mathfrak{b}_4 \subset h(\mathfrak{b}_1)$ lies entirely in R_1. Next move $\mathfrak{b}_4$ directly to R_4. This move does not raise $\sigma = \sigma_{i-1}$ and may lower it by g_4 if $R_4 \notin H$. Next move $\mathfrak{b}'$ into R, returning the alignment to $\mathcal{D}_{i+1}$. This transition replaces the first occurrence of $\sigma_i = m$ in the sequence with a term no larger than σ_{i-1}. Thus again x is reduced by one (or m is reduced, if always $j > i \implies \sigma_j < m$).

In any case we have contradicted the assumption that (m, x) is minimal. We conclude that $m = 0$, as required. □

Definition 7.11. *A flagged chamber complex is tiny if it is a tiny chamber complex, as in Definition 4.9, and, in the case that $F(\mathbb{C}) \neq \emptyset$, each of the designated handlebodies is empty.*

In particular, if $F(\mathbb{C})$ is a single component dividing M into two handlebodies, so either of them can play the role of designated handlebody, the flagged chamber complex is tiny unless *both* handlebodies are occupied.

Proposition 7.12 (Tinyness pulls back)**.** *Suppose*

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is a flagged chamber complex decomposition. If $\mathbb{C}_{\mathcal{D}}$ is tiny, so was $\mathbb{C}$.

Proof. This is essentially a restatement of Proposition 5.19. □

Corollary 7.13. *Suppose*

$$\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a sequence of flagged chamber complex decompositions. Then if $\mathbb{C}_0$ is not tiny, neither is $\mathbb{C}_n$. In particular, $F(\mathbb{C}_n) \neq \emptyset$.

8 Occupied handlebodies and certification

In this section we demonstrate that the principal results from Section 6 continue to apply in the context of flagged chamber complexes, and with a broader version of certification. Continue with Assumption 3.4, that $G(S^3, T') = \mathcal{E}$ whenever $\text{genus}(T') \leq g - 1$.

Proposition 8.1. *Let $\mathbb{C}$ be a flagged chamber complex for S^3 that supports the genus g Heegaard splitting (S^3, T). Suppose $C = A_C \cup_{T_C} B_C$ is an occupied handlebody chamber of $\mathbb{C}$.*

1. *There is a reducing sphere S for the Heegaard surface T_C and a corresponding homeomorphism $h_C : (S^3; T) \to (S^3; T_g)$ so that $h_C(S) \in \{S_i, i = 1, ..., g-1\}$. Moreover, the eyeglass equivalence class of h_C is independent of S.*

2. *If C' is another occupied handlebody chamber of $\mathbb{C}$ then h_C and $h_{C'}$ are eyeglass equivalent.*

3. *If S is an incompressible sphere in a chamber of $\mathbb{C}$ then h_C is eyeglass equivalent to the homeomorphism h_S given in Proposition 3.7.*

4. *Suppose $\mathcal{D}$ is a disk set in $\mathbb{C}$ in preferred alignment, and $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$ is the corresponding flagged chamber complex decomposition. Suppose all remnants of C are disky handlebodies and C' is the chamber in $\mathbb{C}_{\mathcal{D}}$ that, per Definition 7.2, is the unique remnant of C that is occupied. Then h_C and $h_{C'}$ are eyeglass equivalent.*

5. *For any $\tau \in G(S^3, T)$ and $h_{\tau(C)}$ similarly defined, $h_{\tau(C)}\tau \sim h_C$.*

Proof. We prove each statement in turn:

1. Since C is an occupied handlebody, T_C is a non-trivial Heegaard splitting. That is, $b = \text{genus}(T_C) - \text{genus}(\partial C) \geq 1$. Following Lemma 7.9 there is a genus b bubble $\mathfrak{b}$ in T_C and, so long as $b < g$, $\mathfrak{b}$ is unique up to eyeglass moves.

   We have
   $$g = \text{genus}(T) \geq \text{genus}(T_C) = b + \text{genus}(\partial C)$$
   so indeed $b < g$ unless C is a ball and T_C is genus g. But this is impossible, for the complement of C in S^3 would then be a trivially split ball, and so a goneball, so ∂C could not be a component of F. We conclude that $\mathfrak{b}$ is unique up to eyeglass moves and take $\partial\mathfrak{b}$ for the required reducing sphere.

2. This follows from Lemma 3.5.

3. This again follows from Lemma 3.5, since the handlebody chamber C cannot be the same as the chamber with incompressible sphere.

4. For the non-trivial Heegaard splitting of handlebody C', let $\mathfrak{b}'$ be a bubble whose boundary defines $h_{C'}$, as we have just described. According to Lemma 7.4, $\mathfrak{b}'$ can be placed so that it is also a bubble for T_C. According to Lemma 7.9 an eyeglass move will put $\mathfrak{b}'$ inside the bubble $\mathfrak{b}$ whose boundary defines h_C. The result then again follows from Lemma 3.5.

5. This follows from the second statement in Proposition 3.7.

□

Motivated by Proposition 8.1(1-3), we have

Definition 8.2. *A flagged chamber complex $\mathbb{C}$ in S^3* certifies *if some chamber contains an incompressible sphere or some chamber is an occupied handlebody. An incompressible sphere in such a chamber, or the boundary of a bubble in an occupied handlebody, is called* a certificate issued by $\mathbb{C}$.

Corollary 8.3. *Let $\mathbb{C}$ be a flagged chamber complex for S^3 that supports the genus g Heegaard splitting (S^3, T). If $\mathbb{C}$ certifies then all homeomorphisms $(S^3, T) \to (S^3, T_g)$ determined by certificates in $\mathbb{C}$ are eyeglass equivalent.*

Let $h_{\mathbb{C}} : (S^3, T) \to (S^3, T_g)$ denote any homeomorphism given by such a certificate; by Corollary 8.3 $h_{\mathbb{C}}$ is well defined up to eyeglass equivalence. If $\mathbb{C}'$ is another flagged chamber complex that certifies and the homeomorphisms $h_{\mathbb{C}}$ and $h_{\mathbb{C}'}$ are eyeglass equivalent, then we say that $\mathbb{C}$ and $\mathbb{C}'$ *cocertify* and write $\mathbb{C} \sim \mathbb{C}'$.

Corollary 8.4. *If* $\mathbb{C}$ *certifies and* $\tau \in G(S^3, T)$ *then* $h_{\tau(\mathbb{C})}\tau \sim h_{\mathbb{C}}$.

Proof. This follows from 8.1(5). □

Suppose $S^3 = A \cup_T B$ is a genus g Heegaard splitting of S^3, the Heegaard split chamber complex $\mathbb{C}_0$ supports T, and we are given a sequence of flagged chamber complex decompositions

$$\vec{\mathbb{C}}: \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

Suppose further that both $\mathbb{C}_0$ and $\mathbb{C}_n$ certify.

In analogy to Proposition 6.4 we have.

Proposition 8.5. *For iteratively* $0 \le i \le n$ *there is a preferred alignment of each* $\mathcal{D}_i$ *so that* $\mathbb{C}_0$ *and* $\mathbb{C}_n$ *cocertify.*

Proof. The proof is by induction on n.

Initial Case: $n = 1$

Following Proposition 7.5, take any preferred alignment of $\mathcal{D}_0$ in C_0.

One possibility is that in the flagged chamber complex decomposition $\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1$ there is an occupied handlebody chamber in $\mathbb{C}_0$ for which every remnant in $\mathbb{C}_1$ is a disky handlebody. Then the result follows from Proposition 8.1(4). So assume that this is not the case. Then by Proposition 7.7(2) a handlebody in $\mathbb{C}_1$ is empty if and only if it is disky. This implies that different choices of preferred alignment of $\mathcal{D}_0$ have no effect on the flagging of $\mathbb{C}_1$. In particular, every disky ball in $\hat{\mathbb{C}}_1$ is a goneball.

Consider a chamber $C = A_C \cup_{T_C} B_C$ in $\mathbb{C}_0$ that certifies. There is a (reducing sphere) certificate $S \subset C$ for T_C that is either incompressible in C or, if C is an occupied handlebody, bounds a bubble $\mathfrak{b}$ for T_C. In either case there is a realignment of $\mathcal{D}_0$ in $\mathbb{C}$ that is disjoint from S, in the former case by the proof of Proposition 6.4 and in the latter case by considering the Heegaard splitting $\mathbb{C}_0/\mathfrak{b}$. We first need to show that this realignment of $\mathcal{D}_0 \cap C$ can be done in a way that is still a preferred alignment and then show that the homeomorphisms h_S and $h_{S'}$ determined by S and some certificate S' for $\mathbb{C}_1$ are eyeglass equivalent.

Consider the remnant C' of C in $\hat{\mathbb{C}}_1$ that contains S.

Subcase 1: C' is not a disky handlebody. (In particular, C' is not a goneball, and so C' remains a chamber in $\mathbb{C}_1$.)

For example, this is always the case if S is incompressible in C and so is either incompressible in C' or bounds a ball containing components of F.

The disk set $\mathcal{D}_0 \cap C$ can be given a preferred alignment by emptying any remnant of C that is a disky handlebody by passing bubbles to C', see Definition 7.2. If S is incompressible in C' then it is a certificate for $\mathbb{C}_1$ and we are done. If there is a certificate S' for $\mathbb{C}_1$ in a chamber of $\mathbb{C}_1$ other than C', then S and S'

are disjoint, so h_S and $h_{S'}$ are eyeglass equivalent by Lemma 3.5. This implies $h_{\mathbb{C}_1} \sim h_{S'} \sim h_S \sim h_{\mathbb{C}_0}$ and again we are done.

So we are reduced to the case where C' is the only certifying chamber in $\mathbb{C}_1$ and S bounds a bubble $\mathfrak{b}$ in C'. If C' contains an incompressible sphere, then aligning it with the Heegaard surface for C' in $\mathbb{C}_1/\mathfrak{b}$ gives a certificate S' for $\mathbb{C}_1$ that is disjoint from $\mathfrak{b}$ and we are done as above. If C' does not contain an incompressible sphere then, since it contains a certificate for $\mathbb{C}_1$, it must be an occupied handlebody in which S' bounds a bubble. By Lemma 7.9 the spheres S and S' can be made disjoint by an eyegelass move. By Lemma 3.5 this again implies $h_{\mathbb{C}_1} \sim h_{S'} \sim h_S \sim h_{\mathbb{C}_0}$. This concludes the argument in this subcase.

Subcase 2: C' is a disky handlebody.

In this case, to obtain a preferred alignment of $\mathcal{D}_0$ in $\mathbb{C}$ we may need to empty C', see Definition 7.2. Recall that this is done by finding a maximal bubble in C', one that is also a bubble for T_C, and tube summing its boundary to a possible series of disks in $\mathcal{D}_0$. Lemma 7.9 shows that such a maximal bubble can be found that contains $\mathfrak{b}$ within it, so $\mathfrak{b}$ ends up in a remnant of C that is not a disky handlebody. Then Subcase 1 applies. This concludes the argument for Subcase 2 and hence the Initial Case $n = 1$.

The inductive step, $n \geq 2$:

Inductively suppose the result is known for any sequence of length $\leq n-1$. Then we may as well assume that $\mathbb{C}_i$ does not certify for $1 \leq i \leq n-1$, else we could break the sequence into two sequences of smaller length. In particular, in none of the decompositions does an occupied handlebody have all remnants disky handlebodies, for when that occurs in a flagged chamber complex decomposition both the occupied handlebody and one of the remnants certify, see Definitions 7.6 and 7.2. So, just as in the case $n = 1$, the alignment of each $\mathcal{D}_i$ is preferred if and only if it has this property: each handlebody remnant is empty if and only if it is disky. As a result, different choices of preferred alignment throughout the decomposition sequence have no effect on the flagging of the flagged chamber complexes.

Since $\mathbb{C}_1$ does not certify, no chamber in $\mathbb{C}_1$ contains an incompressible sphere. Also, no occupied handlebody in $\mathbb{C}_0$ has only handlebody remnants in $\mathbb{C}_1$, since at least one such remnant would certify. There is a certificate for $\mathbb{C}_0$ in some chamber C of $\mathbb{C}_0$. As in the case $n = 1$ we can find such a certificate S and realign $\mathcal{D}_0 \cap C$ so that the sphere S is disjoint from $\mathcal{D}_0$. Then, regardless of whether S is an incompressible sphere or bounds a bubble in C, S must compress in the chamber of $\mathbb{C}_1$ that contains it, so it bounds a bubble $\mathfrak{b}$ in that chamber.

Following further the case $n = 1$, the disks $\mathcal{D}_0 \cap C$ can be further realligned so that all disky handlebody remnants of C are emptied, by bubble passes to a remnant C' of C that is not a disky handlebody, and this can be done in a way that leaves the bubble $\mathfrak{b}$ intact. The result is that the first decomposition again has a preferred alignment, but now with $\mathfrak{b}$ lying intact in the remnant $C' \in \mathbb{C}_1$. Finally note that C' cannot be a handlebody at all: by construction it is not a disky handlebody and if it were not disky it would be occupied and so it would certify.

Consider the flagged chamber complex $\mathbb{C}_1/\mathfrak{b}$. Since $\mathfrak{b}$ does not lie in a handlebody, there is no change in flagging. Via Proposition 7.5 there is a preferred alignment for the disks $\mathcal{D}_1$ so that

$$\mathbb{C}_1/\mathfrak{b} \xrightarrow{\mathcal{D}_1} \mathbb{C}_2/\mathfrak{b}$$

is a flagged chamber complex decomposition. Since C' is not a handlebody, not every remnant of C' in $\mathbb{C}_2$ is a disky handlebody and, unless $n = 2$, $\mathbb{C}_2$ does not certify, so in fact not every remnant is a handlebody at all. By isotoping $*$ in $\mathbb{C}_1/\mathfrak{b}$ (or, equivalently, passing the bubble $\mathfrak{b}$ in $\mathbb{C}_1$) we can ensure that $*$ lies in a remnant of C' in $\mathbb{C}_2/\mathfrak{b}$ that is not a handlebody. Continue in this manner until $*$ reaches $\mathbb{C}_n/\mathfrak{b}$ and lies in a chamber C'' that is not a disky handlebody (but may be an occupied handlebody). The proof now concludes as it did for of $n = 1$, Subcase 1. □

Just as Proposition 8.5 for flagged decompositions parallels Proposition 6.4, we now work towards finding a parallel of Proposition 6.9 for flagged decompositions. The argument is mostly just a reprise of the previous argument. We begin with an analogue of Lemma 6.5.

Lemma 8.6. *Suppose $\mathbb{C}_0$ is a flagged chamber complex in S^3 that supports the splitting $S^3 = A \cup_T B$, and*

$$\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1$$

is a flagged chamber complex decomposition.

Let $\mathcal{D}_0^x$ and $\mathcal{D}_0^y$ be possibly different preferred alignments of the disk set $\mathcal{D}_0$ in $\mathbb{C}_0$, with resulting flagged chamber complexes $\mathbb{C}_1^x$ and $\mathbb{C}_1^y$ respectively. Then if $\mathbb{C}_0$ and $\mathbb{C}_1^x$ cocertify, so do $\mathbb{C}_0$ and $\mathbb{C}_1^y$.

Proof. If any occupied handlebody chamber C in $\mathbb{C}_0$ is decomposed entirely into disky handlebody chambers in $\mathbb{C}_1$, then from Proposition 8.1(4) $\mathbb{C}_0$ and $\mathbb{C}_1$ cocertify in any preferred alignment and we are done. So we will assume there is no such decomposition of a chamber. Then by Proposition 7.7(2) a handlebody in $\mathbb{C}_1$ is empty if and only if it is disky. This implies that $\mathbb{C}_1^x$ and $\mathbb{C}_1^y$ have the same flagging.

In the case that $\mathbb{C}_1^x$ certifies because a chamber contains an incompressible sphere, the proof is the same as that of Lemma 6.5, with this augmentation: Following Proposition 7.10, the various bubble passes that are needed can be done in such a way that at every stage the disk alignments are preferred and give the same flagging.

In the case that $\mathbb{C}_1^x$ certifies because a chamber is an occupied handlebody, the proof is similar. From [FS2] there is a sequence of bubble passes and eyeglass moves that change the alignment $\mathcal{D}_0^x$ to $\mathcal{D}_0^y$. From Proposition 7.10 this can be done so that in the sequence the alignments are always preferred and the flagging never changes. So it suffices to consider the case of a single bubble pass. With no loss of generality then assume that $\mathbb{C}_1^x$ and $\mathbb{C}_1^y$ differ by a single bubble pass. If the bubble pass does not involve an occupied handlebody, there is nothing to prove. So assume that $\mathbb{C}_1^y$ is obtained from $\mathbb{C}_1^x$ by passing a bubble $\mathfrak{b}$ out of an occupied handlebody chamber C' of $\mathbb{C}_1$.

Since the flagging remains the same after the bubble passes out of C', $\mathfrak{b}$ is not a maximal bubble for C' in $\mathbb{C}_1^x$. Following Lemma 7.9 there is a bubble $\mathfrak{b}^y$ in C' that is disjoint from $\mathfrak{b}$ and the tube sum of the two bubbles $\mathfrak{b}^x$ is a maximal bubble for C' in $\mathbb{C}_1^x$. In particular, $\mathfrak{b}^y$ is a maximal bubble for C' in $\mathbb{C}_1^y$. Since the spheres $S^x = \partial\mathfrak{b}^x$ and $S^y = \partial\mathfrak{b}^y$ are disjoint, it follows from Lemma 3.5 that the homeomorphisms $h_{S^x}, h_{S^y} : (S^3, T) \to (S^3, T_g)$ are eyeglass equivalent, so $\mathbb{C}_1^x$ and $\mathbb{C}_1^y$ cocertify, as required. □

Continue with the hypotheses of Proposition 8.5:

Proposition 8.7. *No matter which preferred alignment occurs at each step in $\vec{\mathbb{C}}$, the flagged chamber complexes $\mathbb{C}_0$ and $\mathbb{C}_n$ cocertify.*

Proof. The case $n = 1$ is Lemma 8.6, so assume $n \geq 2$. We can further inductively assume that none of the $\mathbb{C}_i, 1 \leq i \leq n-1$ certify, so each handlebody chamber is empty. In this case, the argument proceeds essentially as in Proposition 6.9: Proposition 6.7 and Lemma 6.8 remain true for flagged chamber complex decompositions, using Proposition 7.10 to ensure that throughout the argument all decomposing disks are in preferred alignment. Then the proof of Proposition 6.9 suffices if $\mathbb{C}_n$ contains an incompressible sphere. If instead $\mathbb{C}_n$ certifies because it contains an occupied handlebody, the proof concludes as does the proof of Lemma 8.6. ◻

Following Proposition 8.7 it is possible to make the following definition:

Definition 8.8. *Suppose (S^3, T) is a genus g Heegaard splitting of S^3, and*

$$\vec{\mathbb{C}}: \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a sequence of flagged chamber complex decompositions supporting T. If any chamber complex $\mathbb{C}_i$ certifies we say that the sequence $\vec{\mathbb{C}}$ certifies and let $h_{\vec{\mathbb{C}}} : (S^3, T) \to (S^3, T_g)$ be (the eyeglass equivalence class of) any homeomorphism given by a certifying flagged chamber complex in the sequence.

Two flagged chamber complex sequences $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ cocertify (written $\vec{\mathbb{C}} \sim \vec{\mathbb{C}}'$) if both sequences certify and $h_{\vec{\mathbb{C}}} \sim h_{\vec{\mathbb{C}}'}$.

For example, suppose $\tau \in G(S^3, T)$ and $\vec{\mathbb{C}}$ is a flagged chamber complex decomposition sequence supporting T as above. There is a natural way to define a similar flagged chamber complex decomposition sequence $\tau(\vec{\mathbb{C}})$, namely

$$\tau(\vec{\mathbb{C}}): \quad \tau(\mathbb{C}_0) \xrightarrow{\tau(\mathcal{D}_0)} \tau(\mathbb{C}_1) \xrightarrow{\tau(\mathcal{D}_1)} \tau(\mathbb{C}_2) \xrightarrow{\tau(\mathcal{D}_2)} \dots \xrightarrow{\tau(\mathcal{D}_{n-1})} \tau(\mathbb{C}_n)$$

Corollary 8.9. *If $\vec{\mathbb{C}}$ certifies and $\tau \in G(S^3, T)$ then $h_{\tau(\vec{\mathbb{C}})}\tau \sim h_{\vec{\mathbb{C}}}$.*

Proof. This follows immediately from Corollary 8.4. ◻

9 Guiding spheres

Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 with defining surface $F = F(\mathbb{C})$, and $S \subset S^3$ is an embedded sphere transverse to F. In this and following sections we intend to associate to each such sphere S a sequence $\overrightarrow{(\mathbb{C}, S)}$ of flagged chamber complex decompositions so that:

1. For $\mathbb{C}$ not tiny and $S_t, 0 \leq t \leq 1$ a sweep-out of S^3 by spheres, there is a non-empty interval $(a, b) \subset I$ so that, for each $t \in (a, b)$, $\vec{\mathbb{C}}_{S_t}$ certifies.

2. For any $t, t' \in (a, b)$ the sequences $\vec{\mathbb{C}}_{S_t}$ and $\vec{\mathbb{C}}_{S_{t'}}$ cocertify.

3. If, at any point $\mathbb{C}_i \xrightarrow{\mathcal{D}_i} \mathbb{C}_{i+1}$ in the sequence $\overrightarrow{(\mathbb{C}, S)}$, a disk E, disjoint from S, is added to $\mathcal{D}_i$, the resulting sequence of flagged chamber complex decompositions and the sequence $\vec{\mathbb{C}}_S$ cocertify.

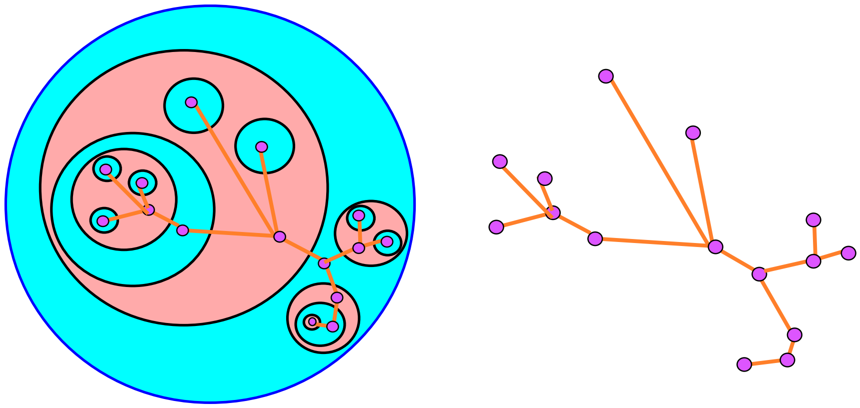

Figure 26: $S \cap F$ and its associated tree

Let $F = F(\mathbb{C})$ be the defining surface of $\mathbb{C}$ and, in the standard way, associate to S and F a tree Y, in which each vertex corresponds to a component of $S - F$ and each edge corresponds to a (circle) component of $S \cap F$. An edge associated to a circle $c \subset S - F$ connects the two vertices that correspond to the components of $S - F$ that are incident to c. Then the leaves of Y correspond to the disk components of $S - F$, and the outermost edges in Y correspond to innermost circles in S of $F \cap S$. See Figure 26.

Definition 9.1. *For any edge $e \in Y$, with end vertices $v_\pm$, $Y - e$ consists of two trees $Y_\pm$, with $v_\pm \in Y_\pm$. Let*

$$m_\pm = max\{d(v_\pm, v_\ell) | v_\ell \text{ a leaf in } Y \text{ lying in } Y_\pm\}$$

and $\rho_Y(e) = min\{m_+, m_-\}$. See Figure 27.

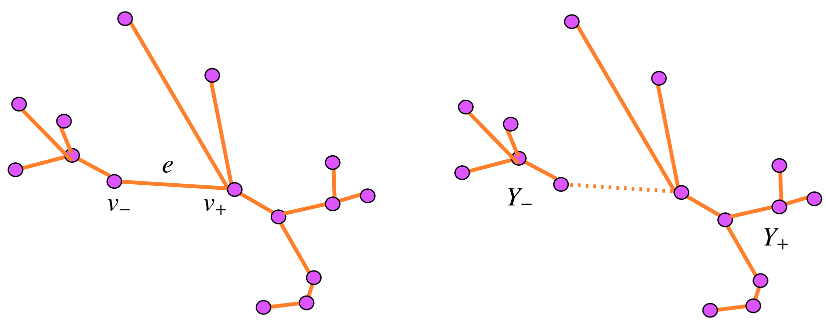

Figure 27: Edge e and its trees Y_{pm}; $m_+ = 4, m = m_- = 2$

For example, for e an outermost edge, one of its incident vertices, v_- say, is a leaf, so $Y_- = v_-$ and $\rho_Y(e) = m_- = 0$.

Lemma 9.2. *Suppose e is an edge in a tree Y.*

1. *Suppose, with no loss of generality, that $\rho_Y(e) = m_- \le m_+$ and e' is an edge in Y_-. Then $\rho_Y(e') < \rho_Y(e)$.*

2. *Let Y' be the tree obtained from Y by trimming: that is, removing all leaves and outermost edges from Y. For $e \in Y'$, $\rho_{Y'}(e) = \rho_Y(e) - 1$.*

3. $0 \leq \rho_Y(e) \leq \lfloor \frac{\text{diam}(Y)-1}{2} \rfloor$

Proof. We prove the items in turn:

1. Let Y'_- be the component of $Y - e'$ that is contained in Y_- and v'_- the end of e' in Y'_-. For any leaf v'_ℓ of Y'_- that lies in Y, the path from v_- to v'_ℓ in Y_- passes through e', so

$$\rho_Y(e) = m_- \geq d(v_-, v'_\ell) \geq d(v'_-, v'_\ell) + 1.$$

Since this is true for each leaf v'_ℓ of Y'_-,

$$\rho_Y(e) \geq max\{d(v'_-, v'_\ell) | v'_\ell \text{ a leaf in } Y'_- \text{ lying in } Y\} + 1 \geq \rho_Y(e') + 1$$

2. Note that any leaf v'_ℓ in Y'_+ that lies in Y' is necessarily incident to some outermost edges of Y, so $d(v_+, v'_\ell) \leq m_+ - 1$. This implies that

$$m'_+ = max\{d(v_+, v'_\ell) | v'_\ell \text{ a leaf in } Y'_+ \text{ lying in } Y'\} \leq m_+ - 1.$$

On the other hand, if v_ℓ is the most distant leaf from v_+ in Y_+ that lies in Y then the outermost edge incident to v_ℓ is incident to a leaf v'_ℓ in Y'_+, so $m'_+ \geq m_+ - 1$. Together these imply $m'_+ = m_+ - 1$. Similarly $m'_- = m_- - 1$. Hence

$$\rho_{Y'}(e) = min\{m'_+, m'_-\} = min\{m_+, m_-\} - 1 = \rho_Y(e) - 1,$$

as required.

3. Repeat the process of trimming $\lfloor \frac{\text{diam}(Y)-1}{2} \rfloor$ times. Each step reduces the diameter of the tree by 2 and, by conclusion (2), reduces the maximum of ρ on any remaining edge by 1. Thus it suffices to verify the case in which diam$(Y) = 1$ or 2, i. e. Y is a single edge or a star graph, depending on the parity of diam(Y). In both these cases, $\rho(e) = 0$ as required.

□

Partition the collection of circles $S \cap F$ as follows: For $0 \leq i \leq \lfloor \frac{\text{diam}(Y)-1}{2} \rfloor$ let $\mathfrak{c}_i$ be the set of all circles for which the corresponding edge e in Y has $\rho_Y(e) = i$.

Lemma 9.3. *Suppose $c \in \mathfrak{c}_i$. Then c bounds a disk D in S such that each circle $c' \subset \text{int}(D) \cap F$ lies in some $\mathfrak{c}_j$, $j < i$.*

Proof. Let Y be the tree corresponding to $S \cap F$ and apply Lemma 9.2(1) to the edge in Y corresponding to c. The trees $Y_\pm$ are trees describing the circles $S \cap F$ lying in the complementary disks $D_\pm$ of c in S, with $v_\pm$ corresponding to the components of $D_\pm - F$ that are adjacent to c. With no loss of generality suppose D_- corresponds to Y_- with $i = \rho_Y(e) = m_-$. Then Lemma 9.2(1) says that any component c' of $\text{int}(D_-) \cap F$ has corresponding edge e' in Y' with $\rho_Y(e') < \rho_Y(e) = i$. Setting $j = \rho_Y(e')$ we have $c' \in \mathfrak{c}_j$. □

Each circle in $\mathfrak{c}_0$ corresponds to a leaf in Y, so each $c \in \mathfrak{c}_0$ bounds a disk in $S - F$. Let $\mathcal{D}_0$ be the collection of disks, viewed as a disk set in $\mathbb{C}$. Choose a preferred alignment for $\mathcal{D}_0$ and consider the flagged chamber complex decomposition $\mathbb{C} \xrightarrow{\mathcal{D}_0} \mathbb{C}_1$. Before the decomposition each circle in $\mathfrak{c}_1$ bounded a disk in S that intersected F only in components of $\mathfrak{c}_0$, by Lemma 9.3. So after the disk decomposition along $\mathcal{D}_0$, which surgers away all components of $\mathfrak{c}_0$, each circle in $\mathfrak{c}_1$ bounds a disk in S that is disjoint from the defining surface $F_1 = F(\mathbb{C}_1)$. More correctly, this is true of each *remaining* circle in $\mathfrak{c}_1$, for some circles in $\mathfrak{c}_1$ may lie on the boundary of goneballs of the decomposition by $\mathcal{D}_0$, and so not still appear in $F_1 \cap S$. In any case, the collection of disks in $S - F_1$ bounded by (remaining) circles in $\mathfrak{c}_1$ will be denoted $\mathcal{D}_1$ and a choice of preferred alignment defines a flagged chamber complex decomposition $\mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2$. Continue in this manner through $\mathbb{C}_{n-1} \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$, where $n - 1 = \lfloor \frac{\text{diam}(Y)-1}{2} \rfloor$ or $n = \lfloor \frac{\text{diam}(Y)+1}{2} \rfloor$. In the end, $F_n = F(\mathbb{C}_n)$ is then disjoint from S.

There are potentially two sources of ambiguity in this construction:

One ambiguity is this: in the last decomposition in the sequence, $\mathbb{C}_{n-1} \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$, it's possible that the defining surface $F_{n-1} = F(\mathbb{C}_{n-1})$ intersects S in a single circle, so $|\mathfrak{c}_{n-1}| = 1$. This circle divides S into two disks, one lying in the incident A-chamber of $\mathbb{C}_{n-1}$ and the other lying in the incident B-chamber. The description above would allow either of these disks to be used as the disk set $\mathcal{D}_{n-1}$ in the last decomposition. Will it make a difference which one we use? This issue will be addressed later - see Lemma 10.4.

The second potential source of ambiguity is that *prima facie* the construction above depends at each stage on the choice of preferred alignment of the disks $\mathcal{D}_i$. So the process is more accurately described as defining a *family* of decomposition sequences, which may differ at several points in the choice of preferred alignment. We will now show that if any sequence in the family certifies, then they all cocertify.

Definition 9.4. *Given $\mathbb{C}$ a flagged chamber complex in S^3, S a sphere transverse to $F(\mathbb{C})$, and Y the tree associated to the circles $F \cap S$ in S, a sequence of flagged chamber complex decompositions*

$$\vec{\mathbb{C}}: \quad \mathbb{C} = \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \dots \xrightarrow{\mathcal{D}_{k-1}} \mathbb{C}_k$$

as constructed above is said to be guided by *the sphere S. If $k = n = \lfloor \frac{\text{diam}(Y)+1}{2} \rfloor$ the sequence is* complete*.*

The collection of all complete sequences will be denoted $\overrightarrow{(\mathbb{C}, S)}$.

Continue with Assumption 3.4, that $G(S^3, T') = \mathcal{E}$ whenever genus$(T') \leq g - 1$.

Proposition 9.5. *If $\overrightarrow{(\mathbb{C}, S)}$ contains more than one complete sequence of flagged chamber complex decompositions then each sequence in $\overrightarrow{(\mathbb{C}, S)}$ certifies and all sequences cocertify.*

Proof. Suppose

$$\vec{\mathbb{C}}: \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

and

$$\vec{\mathbb{C}}': \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}'_1 \xrightarrow{\mathcal{D}'_1} \dots \xrightarrow{\mathcal{D}'_{n-1}} \mathbb{C}'_n$$

are any two sequences in $\overrightarrow{(\mathbb{C}, S)}$. and $i \geq 1$ is the smallest value for which $\mathbb{C}_i \neq \mathbb{C}'_i$. Then $\mathbb{C}_{i-1} = \mathbb{C}_{i-1}$ and $\mathcal{D}_{i-1} = \mathcal{D}'_{i-1}$ are the same as disk sets but are given different preferred alignments. In particular, by Proposition 7.7(3), the decompositions $\mathbb{C}_{i-1} \xrightarrow{\mathcal{D}_{i-1}} \mathbb{C}_i$ and $\mathbb{C}_{i-1} \xrightarrow{\mathcal{D}'_{i-1}} \mathbb{C}'_i$ contain sibling decompositions. That is, there is at least one parent chamber C in $\mathbb{C}_{i-1}$ so that C is an occupied handlebody and $\mathcal{D}_{i-1}$ is given a different preferred alignment in C for the two decompositions. Then C certifies for both $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ □

If $\vec{\mathbb{C}} \in \overrightarrow{(\mathbb{C}, S)}$ certifies, denote a homeomorphism $h_{\vec{\mathbb{C}}} : (S^3, T) \to (S^3, T_g)$ (see Definition 8.8) by $h_{(\mathbb{C}, S)}$. By Proposition 9.5 $h_{(\mathbb{C}, S)}$ is well-defined up to eyeglass equivalence; that is, it does not depend on the choice of decomposition sequence $\vec{\mathbb{C}} \in \overrightarrow{(\mathbb{C}, S)}$.

Corollary 9.6. *If $\vec{\mathbb{C}} \in \overrightarrow{(\mathbb{C}, S)}$ certifies and $\tau \in G(S^3, T)$ then $h_{(\tau(\mathbb{C}), \tau(S))}\tau \sim h_{(\mathbb{C}, S)}$.*

Proof. This follows immediately from Corollary 8.9. □

10 Balanced and almost balanced spheres

Definition 10.1. *Suppose $S \subset S^3$ is a sphere transverse to $F = F(\mathbb{C})$ in a flagged chamber complex $\mathbb{C} \subset S^3$, with X, Y the complementary components of S.*

Then S is balanced *for $\mathbb{C}$ if the compact surfaces $F \cap X$ and $F \cap Y$ are either both planar (planar balanced) or both non-planar (non-planar balanced).*

S is almost balanced *if* $\text{genus}(F \cap X) = 0$ *and* $\text{genus}(F \cap Y) = 1$ *or vice versa.*

Proposition 10.2. . *Suppose*

$$\vec{\mathbb{C}} : \quad \mathbb{C} = \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a complete flagged chamber complex decomposition sequence guided by $S \subset S^3$. That is, $\vec{\mathbb{C}} \in \overrightarrow{(\mathbb{C}, S)}$. Then

- *If S is planar (resp non-planar) balanced in any chamber complex in the sequence, then it is in every chamber complex in the sequence. In this case, call S planar (resp non-planar) balanced for the sequence.*
- *If S is balanced for the sequence and $\mathbb{C}$ is not tiny, then $\mathbb{C}_n$ certifies, so $\vec{\mathbb{C}}$ certifies.*

Proof. For the first statement, note that the effect on $F \cap X$, say, of decomposing along $\mathcal{D}_i$ is 2-fold:

1. Cap off some of boundary components of $F \cap X$ with disks.
2. Delete some spheres from F (the boundaries of the goneballs).

Neither of these steps affects the genus of $F \cap X$ or $F \cap Y$.

For the second statement, note that since $\mathbb{C}$ is not tiny, $\mathbb{C}_n$ is not empty (Corollary 7.13). Let $F = F(\mathbb{C}_n) \neq \emptyset$ be the defining surface for $\mathbb{C}_n$. When the sequence is non-planar balanced, there are (non-planar) closed components of F in both X and Y. This implies that S is a reducing sphere for the chamber of C_n in which it lies, so S itself is a certificate. If the sequence is planar balanced, then each component of F is a closed planar surface, i. e. a sphere. Since there are no empty balls in a flagged chamber complex, an innermost sphere in F bounds an occupied ball, which certifies (see Definition 8.2). □

Similarly we have:

Proposition 10.3. *Suppose*

$$\vec{\mathbb{C}}: \quad \mathbb{C} = \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a complete flagged chamber complex decomposition sequence guided by $S \subset S^3$. That is, $\vec{\mathbb{C}} \in \overrightarrow{(\mathbb{C}, S)}$. Then

- *If S is almost balanced in any chamber complex in the sequence, then it is in every chamber complex in the sequence. In this case, call S almost balanced for the sequence.*
- *If S is almost balanced for the sequence and $\mathbb{C}$ is not tiny, then $\mathbb{C}_n$ certifies, so $\vec{\mathbb{C}}$ certifies.*

Proof. The proof of the first statement is unchanged.

For the proof of the second, observe that since the sequence is complete, $F_n = F(\mathbb{C}_n)$ is disjoint from S, so both $F_n \cap X$ and $F_n \cap Y$ are closed surfaces. Since F_n is almost balanced, $F_n \cap X$ has genus 0 and $F_n \cap Y$ has genus 1, or vice versa. In any case F_n is the union of a torus and some spheres. If there are spheres, at least one must bound an occupied ball and so certifies. If F is just a torus it bounds a solid torus Y in S^3, or perhaps complementary solid tori $Y_\pm$. Since $\mathbb{C}$ is not tiny, $\mathbb{C}_n$ is not tiny (Corollary 7.13), so Y cannot be empty nor, in the second case, can both solid tori $Y_\pm$ be empty. Hence one of these is an occupied solid torus and so is a certificate for $\mathbb{C}_n$. □

We now briefly return to what was described, preceding Definition 9.4, as the first source of ambiguity in the construction of a sequence of flagged chamber complex decompositions guided by S. Suppose $\mathbb{C}$ is a flagged chamber complex supporting the genus g splitting $S^3 = A \cup_T B$ and S is a sphere in S^3 intersecting $\mathbb{C}$ in a single circle c and dividing S^3 into two 3-balls, X and Y. The circle c divides S into two disks D_A lying in an A-chamber C_A of $\mathbb{C}$ and D_B lying in a B-chamber C_B.

Lemma 10.4. *Let $\mathbb{C}_A$ and $\mathbb{C}_B$ be the flagged chamber complexes that result from the decompositions $\mathbb{C} \xrightarrow{D_A} \mathbb{C}_A$ and $\mathbb{C} \xrightarrow{D_B} \mathbb{C}_B$ respectively. Suppose S is balanced or almost balanced for $\mathbb{C}$ so both $\mathbb{C}_A$ and $\mathbb{C}_B$ certify. Then $\mathbb{C}_A$ and $\mathbb{C}_B$ cocertify.*

Proof. Let c be the circle described before the statement of the lemma.

Case 1: c is essential in $F = F(\mathbb{C})$.

Then c divides F into two non-planar components and F is non-planar balanced. We show that our rule for certifying a non-planar balanced chamber complex cocertifies $\mathbb{C}_A$ and $\mathbb{C}_B$:

As usual, align $\mathcal{D}_A$ in $\mathbb{C}_A$ and $\mathcal{D}_B$ in $\mathbb{C}_B$, via an appropriate choice of spine for the compression bodies $B_{\mathbb{C}_A}$ in $\mathbb{C}_A$ and $A_{\mathbb{C}_B}$ in $\mathbb{C}_B$. Since c is essential in F it is essential in the Heegaard surface T of S^3 that the chamber complex supports. Hence S is a reducing sphere for T. In particular there is a homeomorphism $h : (S^3, T) \to (S^3, T_g)$ as described before Definition 3.3 in which $h(S) = S_i$ for some $1 \leq i \leq g-1$. But this coincides with the definition of the homeomorphism $h_S : (S^3, T) \to (S^3, T_g)$ in Proposition 3.7, and that is the homeomorphism which certifies S in both $\mathbb{C}_A$ and $\mathbb{C}_B$. So $\mathbb{C}_A$ and $\mathbb{C}_B$ cocertify in this case.

Case 2: c bounds a disk $D_F \subset F \cap X$ but does not bound a disk in $F \cap Y$ (or vice versa).

Let S_A be the sphere $D_A \cup D_F$, a component of $F(\hat{\mathbb{C}}_A)$ and, symmetrically, S_B be the sphere $D_B \cup D_F$, a component of $F(\hat{\mathbb{C}}_B)$. One possibility is that S_A is essential in $\mathbb{C}_A \cap X$, perhaps because the ball $B_{X \cap A}$ in X bounded by S_A contains other components of $F(\mathbb{C}_A)$ or because it is occupied, and so not a goneball in $\hat{\mathbb{C}}_A$. Then S_A, pushed slightly out of $B_{X \cap A}$, is also essential in $\mathbb{C}_A$ because it separates $B_{X \cap A}$ from the (non-spherical) component of $F(\mathbb{C}_A)$ that contains the circle c. Thus it certifies for $\mathbb{C}_A$. See Figure 28.

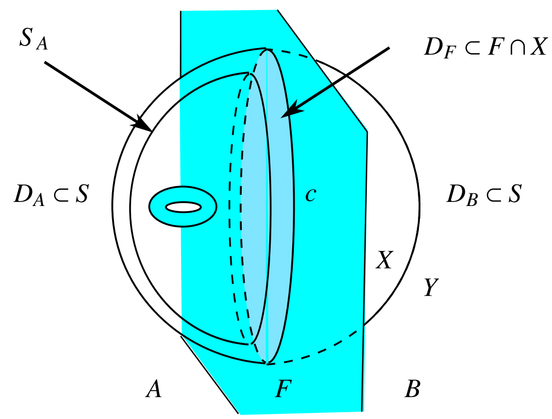

Figure 28: Case 2: c bounds a disk in $F \cap X$ but not in $F \cap Y$

If S_B similarly certifies for $\mathbb{C}_B$ then S_A and S_B define disjoint reducing spheres for T and the result follows from Lemma 3.5. On the other hand, if S_B does not certify for $\mathbb{C}_B$ then $B_{X \cap B}$ is a goneball. In this case, decomposing by D_B has no effect, that is $\mathbb{C}_B$ and $\mathbb{C}$ are isotopic, so $\mathbb{C}$ also certifies. By Proposition 8.7, $\mathbb{C}_A \sim \mathbb{C} \sim \mathbb{C}_B$ as required.

Finally, if S_A is inessential in $\mathbb{C}_A \cap X$ and S_B is inessential in $\mathbb{C}_B \cap X$, so $B_{X \cap B}$ and $B_{X \cap A}$ are goneballs, then $\mathbb{C}_A$ and $\mathbb{C}_B$ are isotopic to $\mathbb{C}$, so $\mathbb{C}_A$ and $\mathbb{C}_B$ again cocertify.

Case 3: c lies on a sphere component S_F of F, dividing it into disks $D_X \subset X$ and $D_Y \subset Y$.

Since S_F does not bound a goneball in $\mathbb{C}$, it either divides S^3 into two occupied balls or is parallel to an incompressible sphere in an adjoining chamber of $\mathbb{C}$. Thus S_F is a certificate for $\mathbb{C}$. Then, per Proposition 8.7 , $\mathbb{C}_A \sim \mathbb{C} \sim \mathbb{C}_B$ as required. □

This dispenses with the "first ambiguity" preceding Definition 9.4.

Suppose $\mathbb{C} \subset S^3$ is a flagged chamber complex, with $F = F(\mathbb{C})$ in general position with respect to the height projection $p : S^3 \to [-1, 1]$. Recall that $S^3 - \{poles\}$ is swept out by the family of level spheres

S_s, where $S_s = p^{-1}(s), s \in (-1,1)$. We will consider these level spheres as potential guiding spheres for decomposition of $\mathbb{C}$.

Definition 10.5. *Let $F = F(\mathbb{C})$, $\mathfrak{g} = \text{genus}(F)$.*

For each $s \in (-1,1)$ that is a regular value of $p|F$, let $F_{a(bove)}(s)$ be the part of F lying above S_s, that is $F_a(s) = F \cap p^{-1}([s,1])$. Similarly, let $F_{b(elow)}(s)$ be the part of F lying below S_s, that is $F_b(s) = F \cap p^{-1}([-1,s])$. Finally, let $g_a(s) = \text{genus}(F_a(s))$ and $g_b(s) = \text{genus}(F_b(s))$.

Lemma 10.6. *The functions g_a, g_b have these properties:*

1. *As s ascends from -1 to 1, g_b ascends from 0 to $\mathfrak{g}$ and g_a descends from $\mathfrak{g}$ to 0.*
2. *As s ascends through a critical point of $p|F : F \to [-1,1]$, $g_b(s)$ may increase by 1 but it will not decrease. Symmetrically, $g_a(s)$ may decrease by 1 but it will not increase.*
3. *As S_s ascends through a critical point of $p|F$, at most one of $g_a(s), g_b(s)$ will change.*

Proof. The first statement follows from the definition.

For the second and third statements, observe that passing through a critical point on F of index 0 (a minimum) or 2 (a maximum) simply adds or subtracts a disk from F_a and F_b, which has no effect on genus. So the interest is in index 1 critical points, that is saddle points of tangency.

Suppose, with little loss in generality, that as s ascends through a saddle tangency, two circles in $F \cap S_s$ fuse into one. See Figure 29. (If instead one circle is divided in two, just turn the following argument upside down.) Thus a band is added to F_b with its ends on two different boundary circles. If the two circles lie on different components of F_b, the genus of F_b does not change. If the two lie on the same component, g_b ascends by 1. This proves the second statement. For the third statement, consider F_a: the effect on F_a is to cut out a band between two boundary components, effectively replacing a pair of pants in F_a with an annulus. This has no effect on g_a. This proves the third statement. □

Corollary 10.7. *Let*

$$s_b = \sup\{s \in [0,1] | g_b(s) = 0\} \qquad s_a = \inf\{s \in [0,1] | g_a(s) = 0\}.$$

Then

1. *$s_a \neq s_b$.*
2. *If $s_a < s_b$ then for all $s_a < s < s_b$, S_s is planar balanced.*
3. *If $s_b < s_a$ then for all $s_b < s < s_a$, S_s is non-planar balanced.*
4. *If $s_a < s_b$ then for s just below s_a and s just above s_b, S_s is almost balanced.*
5. *If $s_b < s_a$ and $p|F$ has no index 1 critical points for $s \in (s_b, s_a)$ then for s just below s_b and s just above s_a, S_s is almost balanced.*

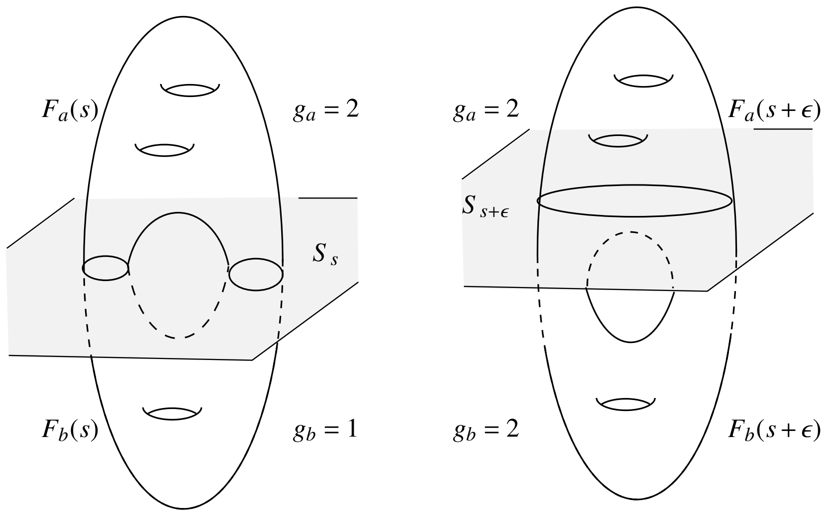

Figure 29: s ascends through a saddle and two circles fuse

Proof. All but the last two statements follow directly from the definitions and Lemma 10.6; the last two require little more:

It follows from Lemma 10.6(2) that for s just above s_b, $g_b(s) = 1$ and, if $s_a < s_b$, $s \in \{s|g_s(a) = 0\}$ so also $g_a(s) = 0$. So S_s is almost balanced. A symmetric argument shows that it is also almost balanced for s just below s_a. This proves (4).

Suppose $s_b < s_a$. Then for s just above s_b, $g_b(s) = 1$ and $g_a(s) \geq 1$; similarly for s just below s_a, $g_a(s) = 1$ and $g_b(s) \geq 1$. Since $p|F$ has no index one critical points in (s_b, s_a), neither g_a or g_b changes in that interval. Hence for any $s \in (s_b, s_a)$, $g_a(s) = 1 = g_b(s)$. It follows from Lemma 10.6(3) that for s just above s_a, $g_a(s) = 0$ and $g_b(s) = 1$ so S_s is almost balanced. The symmetric argument shows that for s just below s_b, S_s is almost balanced. This proves (5). □

Suppose that the chamber complex $\mathbb{C}$ is not tiny. Then for any s between s_a and s_b it follows from Corollary 10.7 and Propositions 10.2 and 9.5 that the homeomorphism

$$h_s = h_{(\mathbb{C},S_s)} : (S^3, T) \to (S^3, T_g)$$

is well-defined up to eyeglass equivalence. What remains unclear is whether the eyeglass equivalence class of h_s depends on s. The central issue is whether the eyeglass equivalence class can change as s passes through a critical point of $p|F$. We will show (Theorem 16.2) that in fact the eyeglass equivalence class does not change. The proof is technically complex; it occupies the next five sections.

11 Deflation and bullseyes

Suppose $\mathbb{C} \subset S^3$ is a flagged chamber complex supporting the genus g Heegaard splitting $S^3 = A \cup_T B$. We continue under the inductive Assumption 3.4.

Let H be an occupied handlebody chamber of $\mathbb{C}$ that is not a ball. In the following description we assume that H is an A-chamber, but everything can also be taken symmetrically, see Definition 11.1. Let C be the B-chamber adjacent to H and C_A be an A-chamber adjacent to C so that $C_A \neq H$. Since the associated Heegaard splitting of $C = A_C \cup_{T_C} B_C$ is pure, there is a spine for the compression body A_C in which an edge e is incident to ∂H on one end and ∂C_A on the other.

Let $(D_e, \partial D_e) \subset (A_C, T_C)$ be the disk dual to e in A_C, that is a meridian of the tube in A_C that corresponds to a regular neighborhood of the edge e, so the tube runs from H to C_A. Alter D_e by tube-summing with a maximal bubble $\mathfrak{b}$ in H and let $\mathbb{C}_d$ be the resulting flagged chamber complex. (More precisely, after amalgamation of $\mathbb{C}$ to the original splitting, D_e becomes a disk in A and $\mathfrak{b}$ a bubble for T; it is these that are tube-summed to create the disk D_e' that replaces D_e in $\mathbb{C}_d$.) In effect, the bubble $\mathfrak{b}$ has been moved from H to the chamber C_A. See Figure 30. Since $\mathfrak{b}$ is maximal in H, H changes from an *occupied* handlebody in $\mathbb{C}$ to an *empty* handlebody in $\mathbb{C}_d$. It is natural to call this process a *deflation* of H. That is, the flagged chamber complex $\mathbb{C}_d$ is obtained from $\mathbb{C}$ by deflation (of H to C_A) . Note that the process is well-defined up to eyeglass equivalence by a choice of H and C_A, since a bubble move is an eyeglass move.

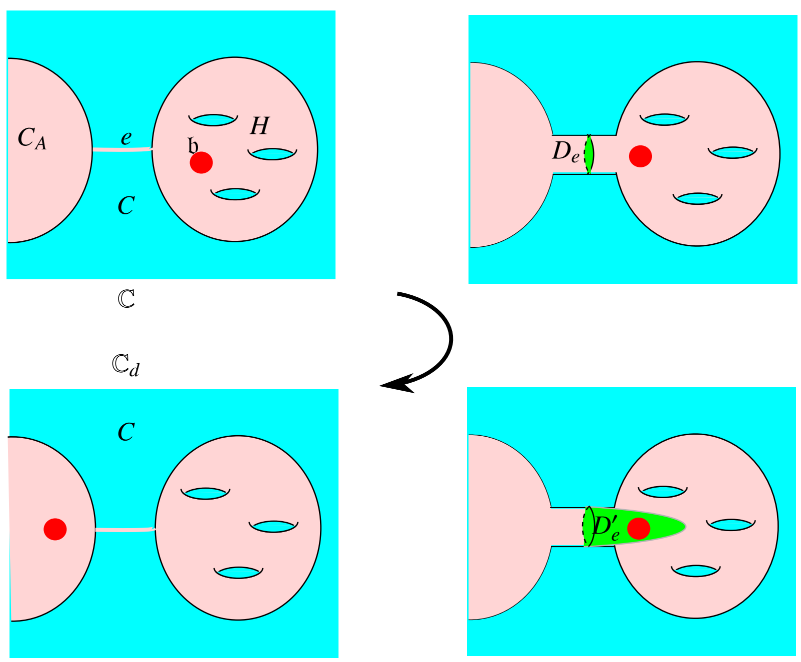

Figure 30: Simple deflation

More generally, the process can be applied to a collection of edges in a spine of A_C, each of which is incident to ∂H and an A-chamber adjacent to C, but not necessarily the same A-chamber for all the edges. That is, redefine the meridian A-disks dual in A_C to the collection of edges by tube-summing them with disjoint non-trivial bubbles in H, bubbles which together constitute a maximal bubble in H. Generalizing further, bubbles from H may be sent to further A-chambers $\{C_{A_i}\}$ throughout $\mathbb{C}$, not just A-chambers adjacent to C, by further tube-summing to A-disks in B-chambers. See Figure 31. The only requirement is that in the resulting flagged chamber complex $\mathbb{C}_d$, the chamber H is empty. In this general form we have:

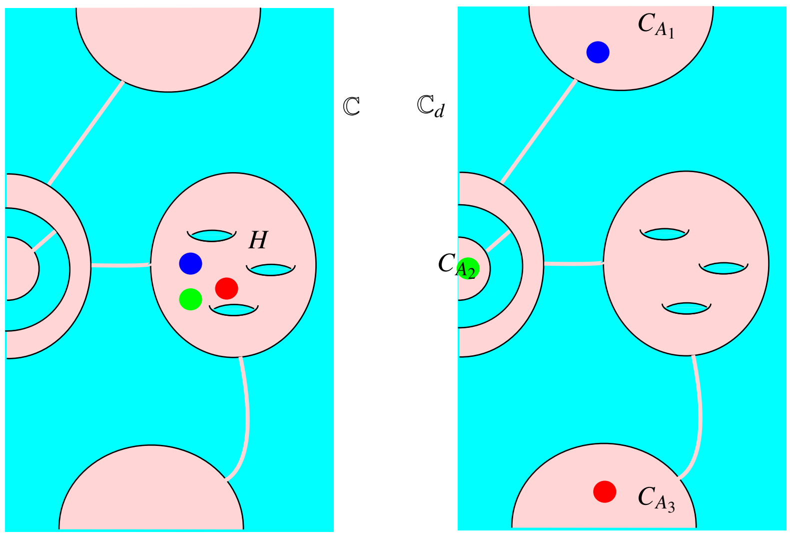

Figure 31: General deflation

Definition 11.1. *The flagged chamber complex* $\mathbb{C}_d$ *is obtained from* $\mathbb{C}$ *by* handlebody deflation *(of* H *to the collection of chambers* $\{C_{A_i}\}$*). Deflation of an occupied handlebody* B*-chamber of* $\mathbb{C}$ *that is not a ball is defined symmetrically.*

As is true for C_A a single chamber, the change of flagged chamber complex from $\mathbb{C}$ to $\mathbb{C}_d$ is well-defined up to eyeglass equivalence by a choice of H, the collection $\{C_{A_i}\}$, and the genus of the bubbles moved into each. The effect of deflation on the flagging of $\mathbb{C}$ is to change H to an empty handlebody chamber in $\mathbb{C}_d$ (hence the term deflation) and, for those chambers in $\{C_{A_i}\}$ that are empty handlebodies, to change the flagging to occupied.

Proposition 11.2. *Suppose the flagged chamber complex* $\mathbb{C}_d$ *is obtained from* $\mathbb{C}$ *by deflating an occupied non-ball handlebody* H*. If* $\mathbb{C}_d$ *certifies then* $\mathbb{C}_d$ *and* $\mathbb{C}$ *cocertify.*

Proof. The occupied handlebody H is a certificate for $\mathbb{C}$ so $\mathbb{C}$ certifies. With no loss assume, as above, that H is an A-chamber deflated into the set of A-chambers $\{C_{A_i}\}$. Suppose $\mathbb{C}_d$ certifies and let C_d be a certifying chamber of $\mathbb{C}_d$. Since H is empty in $\mathbb{C}_d$, $C_d \neq H$. The only other chambers whose flagging is affected by the deflation of H are $\{C_{A_i}\}$ so, unless $C_d \in \{C_{A_i}\}$, the chamber C_d also certifies for $\mathbb{C}$, as required. So suppose $C_d \in \{C_{A_i}\}$, i. e. after the bubble pass one or more of the chambers in $\{C_{A_i}\}$, call it C_a, certifies for $\mathbb{C}_d$.

The only change in C_a from $\mathbb{C}$ to $\mathbb{C}_d$ is that a bubble $\mathfrak{b}$ from H has been passed into it. Since the chamber C_a certifies in $\mathbb{C}_d$, it is either an occupied handlebody or it contains an incompressible sphere. If $\mathbb{C}_a$ is an occupied handlebody then, per Proposition 8.1(1), $\partial\mathfrak{b} \subset \mathbb{C}_a$ is a certificate for $\mathbb{C}_d$ as well as $\mathbb{C}$ completing the proof. Suppose, on the other hand, that the chamber C_a contains an incompressible sphere

S. By [Sc1] S can be aligned with the Heegaard surface T_a for C_a in $\mathbb{C}$, so S becomes a certificate for $\mathbb{C}$. In constructing $\mathbb{C}_d$ pass the bubble $\mathfrak{b}$ into the chamber C_a so that $\mathfrak{b}$ remains disjoint from the aligned S. Then the sphere S certifies for both $\mathbb{C}$ and $\mathbb{C}_d$, as required. □

The same construction can be done when H is an occupied ball, but an important difference is that once the occupied ball is deflated, so H has a genus 0 splitting, H must be declared a goneball, else $\mathbb{C}_d$ is not a flagged chamber complex. That is, the sphere ∂H is removed from the defining surface F and H is absorbed into its neighboring chamber (the chamber C in the construction above). This difference requires a change in strategy to arrive at the equivalent of Proposition 11.2.

A set of disjoint spheres in a ball B^3 is *nested* if only one complementary component is a ball.

Definition 11.3. *Suppose $F_{\mathfrak{b}}$ is a nested set of spheres in a 3-ball B^3, with ∂B^3 the outermost sphere. Call the collar between any pair of spheres a* shell. *A* bullseye *is the flagged chamber complex in B^3 whose defining surface is $F_{\mathfrak{b}}$, each shell has a genus 0 (pure) Heegaard splitting, and the ball $B_{\mathfrak{b}}$ bounded by the innermost sphere is occupied with maximal bubble $\mathfrak{b}$.*

A bullseye in a chamber complex $\mathbb{C}$ is maximal *if it is not contained in any larger bullseye.*

We now extend the notion of deflation to include bullseyes, not just occupied non-ball handlebodies. Suppose $F_{\mathfrak{b}}$ is a bullseye in a chamber complex $\mathbb{C}$ and, with no loss of generality, the chamber adjacent to $F_{\mathfrak{b}}$ and outside it is a B-chamber C. Let $\{C_{A_i}\}$ (resp. $\{C_{B_i}\}$) be a collection of A-chambers (resp. B-chambers) in $\mathbb{C}$, other than those contained in $F_{\mathfrak{b}}$. Use the same process as described above for deflating a handlebody to deflate $\mathfrak{b}$ either into the chambers $\{C_{A_i}\}$, if the ball chamber of $F_{\mathfrak{b}}$ is an A-chamber, or into $\{C_{B_i}\}$, if the ball chamber in $F_{\mathfrak{b}}$ is a B-chamber. (In the former case, the number of nested spheres in $F_{\mathfrak{b}}$ is odd; in the latter it is even.) After the deflation, each sphere in $F_{\mathfrak{b}}$ bounds a goneball, so the spheres in $F_{\mathfrak{b}}$ disappear in the resulting flagged chamber complex.

Definition 11.4. *The flagged chamber complex $\mathbb{C}_d$ is obtained from $\mathbb{C}$ by* bullseye deflation *(of $F_{\mathfrak{b}}$ to $\{C_{A_i}\}$ or $\{C_{B_i}\}$) or, informally, deleting a bullseye. Similarly, we informally say that $\mathbb{C}$ is obtained from $\mathbb{C}_d$ by inserting a bullseye into C.*

Proposition 11.5. *Suppose the flagged chamber complex $\mathbb{C}_d$ is obtained from $\mathbb{C}$ by deflating a bullseye $F_{\mathfrak{b}}$. If $\mathbb{C}_d$ certifies then $\mathbb{C}_d$ and $\mathbb{C}$ cocertify.*

Proof. The proof proceeds as for Proposition 11.2, with one difference. In the case of bullseye deflation, the chamber C changes, because all the spheres in $F_{\mathfrak{b}}$ disappear. That is, the bullseye becomes a nested set of goneballs that are absorbed into C to become a chamber we denote C_d. See Figure 32. The proof then proceeds much as in Proposition 11.2 except, unlike the situation in Proposition 11.2, it is possible that C turns into the lone certifying chamber C_d for $\mathbb{C}_d$. We examine this possibility:

If C_d certifies because it contains an incompressible sphere, let $S \subset C_d$ be such an incompressible sphere. By general position (with a point in the center of the ball chamber of $F_{\mathfrak{b}}$) S can be isotoped to be disjoint from $F_{\mathfrak{b}}$ and so lie in C. Align S with the Heegaard splitting of C. Then it is aligned also with the Heegaard splitting of C_d regardless of whether $C \in \{C_{B_i}\}$. The sphere S then certifies for both $\mathbb{C}$ and $\mathbb{C}_d$ as required.

Finally, suppose C_d certifies because it is an occupied handlebody. There are two possibilities: $C \in \{C_{B_i}\}$ and $C \notin \{C_{B_i}\}$. If $C \in \{C_{B_i}\}$ then a sub-bubble $\mathfrak{b}'$ of $\mathfrak{b}$ becomes a bubble in C_d. $\partial\mathfrak{b}'$ is disjoint

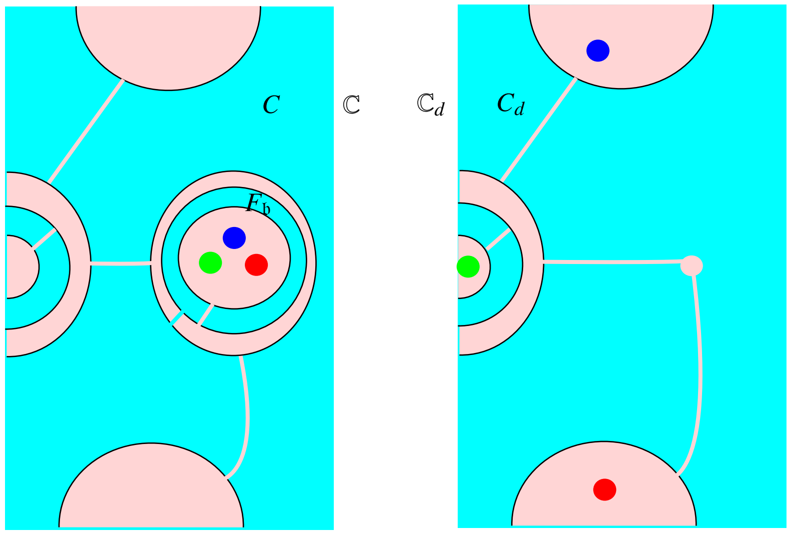

Figure 32: A nested set of goneballs absorbed into C

from $\partial\mathfrak{b}$, which certifies for $\mathbb{C}$, so by Lemma 3.5 $\partial\mathfrak{b}'$ also certifies for $\mathbb{C}$. After deflation, the sphere $\partial\mathfrak{b}' \subset C_d$, may be taken to be disjoint from the boundary of the maximal bubble of C_d, by Lemma 7.9, so $\partial\mathfrak{b}'$ also certifies for $\mathbb{C}_d$. Hence $\mathbb{C}$ and $\mathbb{C}_d$ cocertify, as required.

On the other hand, if $C \notin \{C_{B_i}\}$, so the entire bubble $\mathfrak{b}$ is dispersed away from C_d and still C_d is occupied, the maximal bubble $\mathfrak{b}_d$ of C_d must have come from the chamber C in $\mathbb{C}$. Since, in $\mathbb{C}$, $\mathfrak{b}_d$ lies in C, it is disjoint in C from $\mathfrak{b}$. Hence, by Lemma 3.5, $\partial\mathfrak{b}_d$ and $\partial\mathfrak{b}$ cocertify, as required. □

Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 and $\mathbb{C}_d$ is a flagged chamber complex obtained by (non-ball) handlebody deflation. Suppose $\mathcal{D}$ is a disk set in $\mathbb{C}$ in preferred alignment and

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is the resulting flagged chamber complex decomposition. Similarly construct the flagged chamber complex decomposition

$$\mathbb{C}_d \xrightarrow{\mathcal{D}} (\mathbb{C}_d)_{\mathcal{D}}$$

Proposition 11.6. *If $\mathbb{C}_d$ does not certify, then either $(\mathbb{C}_d)_{\mathcal{D}} = \mathbb{C}_{\mathcal{D}}$ or $(\mathbb{C}_d)_{\mathcal{D}}$ is obtained from $\mathbb{C}_{\mathcal{D}}$ by handlebody deflation or by bullseye deflation.*

Proof. With no loss of generality assume the deflated non-ball handlebody H is an A-chamber and the chambers C and $\{C_{A_i}\}$ are as described preceding Definition 11.1. The chambers $\{C_{A_i}\}$ cannot contain a handlebody because when H is deflated into $\{C_{A_i}\}$ the result would be an occupied handlebody in $\mathbb{C}_d$, contradicting the assumption that $\mathbb{C}_d$ does not certify. So the only difference between $\mathbb{C}$ and $\mathbb{C}_d$ as flagged chamber complexes is that H is occupied in the former but is empty in the latter. Thus the difference between $\mathbb{C}_{\mathcal{D}}$ and $(\mathbb{C}_d)_{\mathcal{D}}$, if any, is in the flagging of the remnants of H, perhaps including the disappearance of remnants as goneballs. So we examine these remnants:

Unless each remnant of H is a disky handlebody, the flagging of the remnants in both $\mathbb{C}_{\mathcal{D}}$ and $(\mathbb{C}_d)_{\mathcal{D}}$ is determined by Definition 7.2: Each handlebody remnant of H is empty if and only if it is disky. Hence in this case $\mathbb{C}_{\mathcal{D}} = (\mathbb{C}_d)_{\mathcal{D}}$ as flagged chamber complexes. On the other hand, if each remnant of H is a disky handlebody then, per Definition 7.2, in $\mathbb{C}_{\mathcal{D}}$ there is exactly one handlebody remnant H' (possibly a ball) that is occupied, while in $(\mathbb{C}_d)_{\mathcal{D}}$ each remnant is empty. We examine this situation more closely:

Since $\{C_{A_i}\}$ contains no handlebodies, some remnant in $\mathbb{C}_{\mathcal{D}}$ of each C_{A_i} is not a disky handlebody, by Proposition 4.8. Let R be such a remnant of some C_{A_i}. Consider the flagged chamber complex obtained from $\mathbb{C}_{\mathcal{D}}$ by deflating H' into R. Since R is not a disky handlebody it is either not a handlebody or it is an occupied handlebody; in either case moving a bubble into R will not change the flagging of R. Thus this deflation changes $\mathbb{C}_{\mathcal{D}}$ to the same flagged chamber complex as $(\mathbb{C}_d)_{\mathcal{D}}$, as required.

If H' is non-ball handlebody then the deflation is a handlebody deflation. If H' is a ball then its deflation turns it into a goneball, so $(\mathbb{C}_d)_{\mathcal{D}}$ is a bullseye deflation of $\mathbb{C}_{\mathcal{D}}$, where the bullseye is centered on the ball H'. (In fact, it is not hard to see that the bullseye consists only of H'.) □

There is a similar result for bullseye deflation, but it is a bit more difficult to state and prove. Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 and $\mathbb{C}_d$ is the flagged chamber complex obtained by deflating a bullseye $F_{\mathfrak{b}}$ in $\mathbb{C}$. Suppose $\mathcal{D}$ is a disk set in $\mathbb{C}$ in preferred alignment and

$$\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$$

is the resulting flagged chamber complex decomposition. Similarly construct the flagged chamber complex decomposition

$$\mathbb{C}_d \xrightarrow{\mathcal{D}'} (\mathbb{C}_d)_{\mathcal{D}'},$$

where $\mathcal{D}' \subset \mathcal{D}$ is the subcollection of disks not incident to any sphere in $F_{\mathfrak{b}}$. (Since the spheres in $F_{\mathfrak{b}}$ disappear in $\mathbb{C}_d$ it would not make sense to include disks incident to them among the decomposing disk set in $\mathbb{C}_d$.)

Proposition 11.7. *Continue with the notation just given. If $\mathbb{C}_d$ does not certify, then either $(\mathbb{C}_d)_{\mathcal{D}'} = \mathbb{C}_{\mathcal{D}}$ or $(\mathbb{C}_d)_{\mathcal{D}'}$ is obtained from $\mathbb{C}_{\mathcal{D}}$ by bullseye deflation.*

Proof. As usual, without loss of generality, assume the chamber in $\mathbb{C}$ adjacent to $F_{\mathfrak{b}}$ is a B-chamber C which, after deflation and absorption of the nested set of goneballs into C, becomes a chamber we denote $C_d \in \mathbb{C}_d$. Let $\mathcal{D}_{\mathfrak{b}} = \mathcal{D} - \mathcal{D}'$, the subcollection of disks that are incident to spheres in $F_{\mathfrak{b}}$. The boundary ∂D of any disk $D \in \mathcal{D}_{\mathfrak{b}}$ divides the sphere $S \subset F_{\mathfrak{b}}$ in which it lies into two disks $E_\pm$. If D lies in the occupied ball or a shell of $F_{\mathfrak{b}}$ then it is parallel to one or both of $E_\pm$ and so is inessential in C. If $(D, \partial D) \subset (C, \partial F_{\mathfrak{b}})$ and neither sphere $D \cup E_\pm$ bounds a ball in C, then the two spheres (which become parallel spheres in C_d, where $F_{\mathfrak{b}}$ is gone) would be essential in C_d, and so would certify for $\mathbb{C}_d$, contradicting our assumption. We conclude that each disk in $\mathcal{D}_{\mathfrak{b}}$ is inessential in the chamber of $\mathbb{C}$ in which it lies.

Claim: The remnants of $F_{\mathfrak{b}}$ in $\mathbb{C}_{\mathcal{D}}$ consist of a bullseye $F'_{\mathfrak{b}}$ that intersects C only in a collar of $\partial F_{\mathfrak{b}}$. The occupied ball component of $F'_{\mathfrak{b}}$ still has $\mathfrak{b}$ as a maximal bubble.

The proof of the claim is straightforward but a bit tedious - here is a sketch of the major steps: Since each disk of $\mathcal{D}$ that is incident to $\partial F_{\mathfrak{b}}$ and lies in C is inessential, there is a collar of $\partial F_{\mathfrak{b}}$ that contains them all. Then each remnant of $F_{\mathfrak{b}}$ in $\hat{\mathbb{C}}_{\mathcal{D}}$ lies in a ball $F^+_{\mathfrak{b}}$, the union of $F_{\mathfrak{b}}$ and the collar, and is bounded by spheres, so it is a punctured ball. Among the remnants of each shell in $F_\mathfrak{b}$ is one containing an incompressible sphere, so that remnant is not a handlebody; similarly the adjacent remnant of the chamber C is not a handlebody. Hence, per Definition 7.2, each ball remnant in $\mathbb{C}_\mathcal{D}$ of each shell and of C is a goneball. This implies that each remnant of the ball $B_\mathfrak{b}$ in $\mathbb{C}_\mathcal{D}$ is a ball and, per Definition 7.2, exactly one is occupied, and so must contain all of $\mathfrak{b}$.

Each component of $F_\mathcal{D}$ that lies in $F_\mathfrak{b}^+$ is a sphere; the ball it bounds in $F_\mathfrak{b}^+$ is a goneball in $\mathbb{C}_\mathcal{D}$ unless it contains $\mathfrak{b}$. Hence the components of $F(\mathbb{C}_\mathcal{D})$ that lie in $F_\mathfrak{b}^+$ consist of a collection of nested spheres centered on a ball containing $\mathfrak{b}$; in other words, they define a bullseye $F'_\mathfrak{b}$, as claimed. See Figure 33.

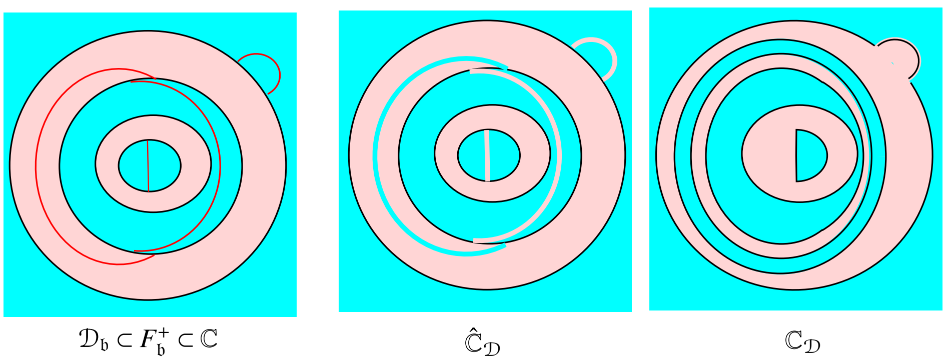

Figure 33: From bullseye to bullseye

Since each ball remnant of C in $F'_\mathfrak{b}$ is a goneball, the Heegaard surface T_C for C (as well as the splittings of other chambers in $\mathbb{C}$) is unaffected by decomposition along $\mathcal{D}_\mathfrak{b}$. So the original preferred alignment of $\mathcal{D}$ in $\mathbb{C}$ remains an alignment of $\mathcal{D}'$ in $\mathbb{C}$. That alignment is easily made into a preferred alignment of $\mathcal{D}'$ in $\mathbb{C}_d$ as follows: $\mathcal{D}'$ is already in preferred alignment in chambers not C_d nor among the chambers $\{C_{A_i}\}$ (say) into which $\mathfrak{b}$ is deflated, since these are the only chambers changed by the deflation. Whichever chamber, in $\{C_{A_i}\}$ or C_d, that $\mathfrak{b}$ is deflated into cannot be a handlebody, since $\mathbb{C}_d$ does not certify, so there is some remnant of that chamber that is not a disky handlebody; choose the alignment of $\mathcal{D}'$ in that chamber so that $\mathfrak{b}$ appears in a remnant that is not a disky handlebody. This does not change the flagging and thereby ensures that $\mathcal{D}'$ is in preferred alignment with $\mathbb{C}_d$.

Now consider the effect of $\mathcal{D}$ surgery *away* from the bullseye $F_\mathfrak{b}$, that is surgery via $\mathcal{D}'$ on the surface $F - F_\mathfrak{b}$. The resulting surface $\hat{F}$ can be viewed in two ways: Since $F - F_\mathfrak{b}$ is also the defining surface for $\mathbb{C}_d$, $\hat{F}$ can be described as the first (surgery) step in the construction of the decomposition $\mathbb{C}_d \xrightarrow{\mathcal{D}'} (\mathbb{C}_d)_{\mathcal{D}'}$. And $\hat{F}$ is also that part of the surface $F_\mathcal{D}$ that lies outside the new bullseye $F'_\mathfrak{b}$.

It remains only to understand the effect of the second step in disk decomposition - eliminating goneballs after surgery. We have already determined that eliminating goneballs from surgery on $F_\mathfrak{b}$ results in the bullseye $F'_\mathfrak{b} \subset \mathbb{C}_\mathcal{D}$. So we need only consider sphere components of $\hat{F}$. Any such sphere bounds a ball (in fact two balls) in S^3. If such a ball is a goneball in $\mathbb{C}_\mathcal{D}$ then that ball is unaffected by a deflation of $F'_\mathfrak{b}$ (since $F'_\mathfrak{b}$ is necessarily outside the goneball) so it remains a goneball in $(\mathbb{C}_d)_{\mathcal{D}'}$.

Consider the symmetric question: if a sphere component of $\hat{F}$ bounds a goneball in $(\mathbb{C}_d)_{\mathcal{D}'}$, does it bound a goneball in $\mathbb{C}_\mathcal{D}$? The goneball in $(\mathbb{C}_d)_{\mathcal{D}'}$ may or may not contain the deflated $F'_\mathfrak{b}$. If the goneball does not contain $F'_\mathfrak{b}$ then the goneball lies also in $\mathbb{C}_\mathcal{D}$ and so the sphere does not appear in either $\mathbb{C}_\mathcal{D}$ or $(\mathbb{C}_d)_{\mathcal{D}'}$. Suppose, on the other hand, that there is a sphere component of $\hat{F}$ that bounds a goneball in

$(\mathbb{C}_d)_{\mathcal{D}'}$ and that goneball does contain the deflated $F'_\mathfrak{b}$. Let S be an innermost one. Since the ball it bounds contains $F'_\mathfrak{b}$ it is not a goneball in $\mathbb{C}_\mathcal{D}$, so S persists as a sphere in the defining surface of $\mathbb{C}_\mathcal{D}$. On the other hand, observe this: since a goneball must have genus 0 splitting, the region between S and $\partial F'_\mathfrak{b}$ in $\mathbb{C}_\mathcal{D}$, a collar of $\partial F'_\mathfrak{b}$, must have genus 0 splitting. Hence the region is a genus 0 shell. We can then regard S as yet another nested sphere component in $F'_\mathfrak{b}$ and continue the argument. In the end, the only components of $\hat{F}$ that appear in $\mathbb{C}_\mathcal{D}$ but not in $(\mathbb{C}_d)_{\mathcal{D}'}$ are spheres that just add shells to $F'_\mathfrak{b}$. In particular, they are all eliminated by a bullseye deflation, in this case of of a bullseye that properly contains $F'_\mathfrak{b}$. So $(\mathbb{C}_d)_{\mathcal{D}'}$ is obtained from $\mathbb{C}_\mathcal{D}$ by bullseye deflation, as required. □

Digressing momentarily, Proposition 11.7 gives an easy and early example of a *guiding set of disks*. These will be further discussed in Section 13.

Definition 11.8. *Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 and $\overline{\mathcal{D}}$ is a finite set of disjoint disks in S^3 transverse to $F = F(\mathbb{C})$ so that $F \cap \overline{\mathcal{D}} = \partial\mathcal{D}$, for some $\mathcal{D} \subset \overline{\mathcal{D}}$. Then $\overline{\mathcal{D}}$ is a* guiding set of disks *for the flagged chamber complex decomposition $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_\mathcal{D}$. Slightly abusing notation we can then write $\mathbb{C} \xrightarrow{\overline{\mathcal{D}}} \mathbb{C}_\mathcal{D}$.*

For example, in Proposition 11.7, $\mathcal{D}$ is a disk set for the decomposition $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_\mathcal{D}$ and $\mathcal{D}$ is also a *guiding disk set* for the decomposition $\mathbb{C}_d \xrightarrow{\mathcal{D}'} (\mathbb{C}_d)_\mathcal{D}$. So, with no loss of meaning, we could also write the latter $\mathbb{C}_d \xrightarrow{\mathcal{D}} (\mathbb{C}_d)_\mathcal{D}$.

In general we will write $\mathbb{C}' \dashrightarrow \mathbb{C}$ to denote that $\mathbb{C}'$ is obtained from $\mathbb{C}$ by either (non-ball) handlebody deflation or bullseye deflation. (The direction of the arrow is meant to indicate that the defining surface $F(\mathbb{C}')$ embeds in $F(\mathbb{C})$.) Thus, in both Propositions 11.6 and 11.7 above, $\mathbb{C}_d \dashrightarrow \mathbb{C}$. Applying Definition 11.8 these two propositions, together with Propositions 11.2 and 11.5, can then be summarized as follows:

Corollary 11.9. *Suppose $\mathbb{C}, \mathbb{C}'$ are flagged chamber complexes for which $\mathbb{C}' \dashrightarrow \mathbb{C}$ and suppose $\mathcal{D}$ is a disk set in $\mathbb{C}$.*

If $\mathbb{C}'$ certifies, then $\mathbb{C}'$ and $\mathbb{C}$ cocertify.

If $\mathbb{C}'$ does not certify, then the following square of flagged chamber complexes commutes:

$$\begin{array}{ccc} \mathbb{C}' & \xrightarrow{\mathcal{D}} & \mathbb{C}'_\mathcal{D} \\ \Big\downarrow & & \Big\downarrow \\ \mathbb{C} & \xrightarrow{\mathcal{D}} & \mathbb{C}_\mathcal{D} \end{array}$$

Broaden now to sequences of decompositions:

Definition 11.10. . *Suppose*

$$\vec{\mathbb{C}}: \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

and

$$\vec{\mathbb{C}}': \quad \mathbb{C}'_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}'_1 \xrightarrow{\mathcal{D}_1} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}'_n$$

are flagged chamber complex decomposition sequences so that for each $i \geq 0$, $\mathbb{C}'_i \dashrightarrow \mathbb{C}_i$. Then the decomposition sequence $\vec{\mathbb{C}}'$ is a deflation *of the decomposition sequence $\vec{\mathbb{C}}$.*

Suppose

$$\vec{\mathbb{C}}: \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a flagged chamber complex decomposition sequence and $\mathbb{C}'_0$ is a flagged chamber complex such that $\mathbb{C}'_0 \dashrightarrow \mathbb{C}_0$. Iteratively define a flagged chamber complex decomposition sequence

$$\vec{\mathbb{C}}'^{m(ax)}: \quad \mathbb{C}'_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}'_1 \xrightarrow{\mathcal{D}_1} \dots \xrightarrow{\mathcal{D}_{m-1}} \mathbb{C}'_m$$

by the following process: If $\mathbb{C}'_i \dashrightarrow \mathbb{C}_i$ take $\mathbb{C}'_{i+1}$ to be the result of flagged chamber complex decomposition of $\mathbb{C}'_i$ by (the guiding set of disks) $\mathcal{D}_i$. Continue until either $m = n$ or $\mathbb{C}'_{m+1}$ is not a deflation of $\mathbb{C}_{m+1}$. The sequence $\vec{\mathbb{C}}'^m$ is a deflation of the sequence

$$\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \dots \xrightarrow{\mathcal{D}_{m-1}} \mathbb{C}_m$$

so the following definition is natural:

Definition 11.11. *$\vec{\mathbb{C}}'^m$ is the maximal deflationary subsequence of $\vec{\mathbb{C}}'$.*

Corollary 11.12. *Suppose*

$$\vec{\mathbb{C}}: \quad \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is a flagged chamber complex decomposition sequence supporting $S^3 = A \cup_T B$ and $\mathbb{C}'_0$ is a flagged chamber complex such that $\mathbb{C}'_0 \dashrightarrow \mathbb{C}_0$. Let $\vec{\mathbb{C}}'^m$ be the maximal deflationary subsequence of $\vec{\mathbb{C}}'$.

- *If $\vec{\mathbb{C}}'^m$ does not certify then $m = n$.*
- *If $\vec{\mathbb{C}}'^m$ does certify then $\vec{\mathbb{C}}'$ and $\vec{\mathbb{C}}$ cocertify.*

Proof. Iteratively apply Corollary 11.9. □

12 Reordering disks

12.1 Parallel vs sequential flagged chamber complex decompositions

Suppose $\mathbb{C}$ is a flagged chamber complex supporting $S^3 = A \cup_T B$, with $F = F(\mathbb{C})$ its defining surface. Suppose $\mathcal{D}_x, \mathcal{D}_y$ are disjoint disk sets in $\mathbb{C}$. In this section we compare the results of flagged chamber complex decomposition using these disk sets in series vs in parallel. That is, we compare the two flagged chamber complexes that appear in the bottom row of this diagram:

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{\mathcal{D}_x} & \mathbb{C}_x \\ {\scriptstyle \mathcal{D}_x \cup \mathcal{D}_y} \downarrow & & \downarrow {\scriptstyle \mathcal{D}_y} \\ \mathbb{C}_{x+y} & & \mathbb{C}_{xy} \end{array}$$

In general the answer is complex: for starters, notice that some disks in $\mathcal{D}_y$ may be incident to the boundary of goneballs in $\mathbb{C}_x$, so the disks $\mathcal{D}_y$ become not a disk set for the decomposition $\mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$ but a guiding disk set. In light of such difficulties, eventually we will restrict our interest to the case in which $\mathcal{D}_x$ or $\mathcal{D}_y$ consists of a single disk. See Subsection 12.2.

We first identify circumstances in which $\mathbb{C}_{x+y} = \mathbb{C}_{xy}$. To that end:

Definition 12.1. *Suppose H is a handlebody in S^3 such that $\partial H \subset F_{\mathcal{D}_x \cup \mathcal{D}_y}$, the surface obtained from F by surgery on $\mathcal{D}_x \cup \mathcal{D}_y$. We say that H is* coherent *for the decompositions by $\mathcal{D}_x \cup \mathcal{D}_y$ in the diagram above when*

- *H is a chamber in $\mathbb{C}_{x+y}$ if and only if it is a chamber in $\mathbb{C}_{xy}$ and*
- *if H is a chamber in both $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$ then it has the same flagging.*

Here is an easy example:

Lemma 12.2. *Suppose there is a component of $F_{\mathcal{D}_x \cup \mathcal{D}_y}$ that lies in* int(H) *and is not a sphere. Then H is not a chamber in either $\mathbb{C}_{x+y}$ or $\mathbb{C}_{xy}$. Hence H is coherent.*

Proof. Let F_{x+y} be a component of $F_{\mathcal{D}_x \cup \mathcal{D}_y}$ that lies in int(H) and is not a sphere. Since F_{x+y} is not a sphere it is not contained in a goneball of $\mathbb{C}_{x+y}$ (cf Lemma 4.2(2)) so it is a component of the defining surface $F(\mathbb{C}_{x+y})$ that lies inside H. Hence H is not a handlebody chamber in $\mathbb{C}_{x+y}$.

The complement of the scars in F_{x+y} is a non-planar connected surface contained in a closed component $F_0 \subset F$. Similarly the complement of just the scars of $\mathcal{D}_y$ in F_{x+y} is contained in a non-planar closed surface F_x of $F_{\mathcal{D}_x}$. Since both F_0 and F_x contain a non-planar subsurface they are not spheres. Since F_x is not a sphere it is not contained in a goneball of $\mathbb{C}_x$, again by Lemma 4.2(2). In particular it is a component of the defining surface $F(\mathbb{C}_x)$. F_{x+y} is a component of the surface that results from surgery by $\mathcal{D}_y$ on F_x. Since it is not a sphere it also is not contained in a goneball of $\mathbb{C}_{xy}$. Hence F_{x+y} is a closed component of the defining surface for $\mathbb{C}_{xy}$ lying in the interior of H. Thus H is not a handlebody chamber in $\mathbb{C}_{xy}$. □

Returning to the definition of coherent, note that H could fail to be a chamber in $\mathbb{C}_{x+y}$ for two reasons: H contains a component of the defining surface $F(\mathbb{C}_{x+y})$ in its interior, as in Lemma 12.2, or H could be a goneball. Similarly, H could fail to be a chamber in $\mathbb{C}_{xy}$ because it has a component of the defining surface $F(\mathbb{C}_{xy})$ in its interior, or because H is a goneball under the sequence of decompositions

$$\mathbb{C} \xrightarrow{\mathcal{D}_x} \mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}.$$

Suppose the latter. H could become a goneball under the second decomposition, so the complement ∂H_- of the scars of $\mathcal{D}_y$ in ∂H is part of a surface in $F(\mathbb{C}_x)$. On the other hand, ∂H_- may not be part of a surface in $F(\mathbb{C}_x)$ if ∂H_- is part of a sphere $G \subset F = F(\mathbb{C})$ that bounds a goneball in $\mathbb{C}_x$. In the last case, if H also has no other components of $\mathbb{C}_{xy}$ in its interior, it is natural to call H a goneball of the decomposition sequence $\mathbb{C} \xrightarrow{\mathcal{D}_x} \mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$. As we will see, such a ball H need not be a goneball in $\mathbb{C}_{x+y}$.

The following lemma explains why coherence is useful:

Lemma 12.3. *Suppose every handlebody in S^3 whose boundary is in $F_{\mathcal{D}_x \cup \mathcal{D}_y}$ is coherent. Then $\mathbb{C}_{x+y} = \mathbb{C}_{xy}$ as flagged chamber complexes.*

Proof. Let F_0 be a component of $F_{\mathcal{D}_x \cup \mathcal{D}_y}$. If F_0 is not a sphere, then, as in the proof of Lemma 12.2, F_0 is a component of both defining surfaces $F(\mathbb{C}_{x+y})$ and $F(\mathbb{C}_{xy})$. If F_0 is a sphere, then it divides S^3 into two 3-balls, $H_\pm$. By assumption both $H_\pm$ are coherent, so each is a goneball in $\mathbb{C}_{x+y}$ if and only if it is a goneball in $\mathbb{C}_{xy}$. In particular, if either $H_\pm$ is a goneball in one of $\mathbb{C}_{x+y}$ or $\mathbb{C}_{xy}$ then it is a goneball in both and F_0 is not in either $F(\mathbb{C}_{x+y})$ and $F(\mathbb{C}_{xy})$. On the other hand, neither of $H_\pm$ is a goneball in $\mathbb{C}_{x+y}$ if and only if neither is a goneball in $\mathbb{C}_{xy}$ and, in this case, F_0 does lie in both $F(\mathbb{C}_{x+y})$ and $F(\mathbb{C}_{xy})$. Thus, ignoring flagging, $\mathbb{C}_{x+y} = \mathbb{C}_{xy}$ as chamber complexes.

Now consider flagging of a handlebody chamber H in one of $\mathbb{C}_{x+y}$ or $\mathbb{C}_{xy}$ and so in both. Suppose H is a handlebody chamber in one of $\mathbb{C}_{x+y}$ or $\mathbb{C}_{xy}$. Since H is coherent, it is then a handlebody chamber in both $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$, and the flagging of H is the same. Thus $\mathbb{C}_{x+y} = \mathbb{C}_{xy}$ as flagged chamber complexes. □

Here is a helpful example of incoherence in a handlebody: Let C be a chamber in $\mathbb{C}$ that a disk D, properly embedded in C, divides into two components: U, not a handlebody, and V, the complement in a handlebody H of a regular neighborhood η of a properly embedded arc $\alpha \subset H$. Let E be a meridian disk in η, so E is properly embedded in η and meets α in a single point. Note that $F \cap \mathrm{int}(H)$ is the annulus $\partial\eta$. See Figure 34.

Since $F \cap \mathrm{int}(H)$ is an annulus and not a collection of disks, H is occupied after the decomposition $\mathbb{C} \xrightarrow{D \cup E} \mathbb{C}_{D+E}$ and after the sequence of decompositions

$$\mathbb{C} \xrightarrow{D} \mathbb{C}_D \xrightarrow{E} \mathbb{C}_{DE}.$$

This means that H is a coherent handlebody for $\mathbb{C}_{D+E}$ and $\mathbb{C}_{DE}$. On the other hand, after the decomposition $\mathbb{C} \xrightarrow{E} \mathbb{C}_E$ the disk D divides $\mathbb{C}_E$ into U and the handlebody H. For the next decomposition $\mathbb{C}_E \xrightarrow{D} \mathbb{C}_{ED}$, Definition 7.2 requires that E be aligned so that H is empty in $\mathbb{C}_{ED}$. Thus H is not a coherent handlebody for $\mathbb{C}_{D+E}$ and $\mathbb{C}_{ED}$.

Replacing D in this example of incoherence by a disk family consisting of D and a disk D' parallel to E has no effect on the above argument: For $\mathcal{D} = D \cup D'$, H is coherent for $\mathbb{C}_{\mathcal{D}+E}$ and $\mathbb{C}_{\mathcal{D}E}$ but incoherent for $\mathbb{C}_{\mathcal{D}+E}$ and $\mathbb{C}_{E\mathcal{D}}$. This leads to the following observation, using analogous notation:

Lemma 12.4. *Let $\mathcal{D} \cup E$ be a disk family in a flagged chamber complex $\mathbb{C}$ in S^3. Suppose the disk E is parallel to a disk $D \in \mathcal{D}$ within a chamber C of $\mathbb{C}$. Then*

1. *Unless the chamber C_D of $\mathbb{C}_{\mathcal{D}}$ containing E is an occupied handlebody, $\mathbb{C}_{\mathcal{D}} = \mathbb{C}_{\mathcal{D}E}$. If C_D is an occupied handlebody then for one of the two resulting siblings in $\mathbb{C}_{\mathcal{D}E}$, $\mathbb{C}_{\mathcal{D}} = \mathbb{C}_{\mathcal{D}E}$.*
2. *Unless the chamber C is an occupied handlebody with only disky handlebody remnants in $\mathbb{C}_{\mathcal{D}}$, the chamber complex $\mathbb{C}_{\mathcal{D}} = \mathbb{C}_{\mathcal{D}+E}$. The same is almost true in the remaining case: If C is an occupied handlebody with only disky handlebody remnants in $\mathbb{C}_{\mathcal{D}}$ then C has sibling remnants in $\mathbb{C}_{\mathcal{D}+E}$ and deleting a single such sibling gives the remnants of C in $\mathbb{C}_{\mathcal{D}}$.*
3. *either $\mathbb{C}_{\mathcal{D}+E} = \mathbb{C}_{\mathcal{D}E}$ or C_D is an occupied handlebody chamber in $\mathbb{C}_{\mathcal{D}}$ and, in one or both of two resulting siblings in $\mathbb{C}_{\mathcal{D}E}$, $\mathbb{C}_{\mathcal{D}+E} = \mathbb{C}_{\mathcal{D}E}$.*

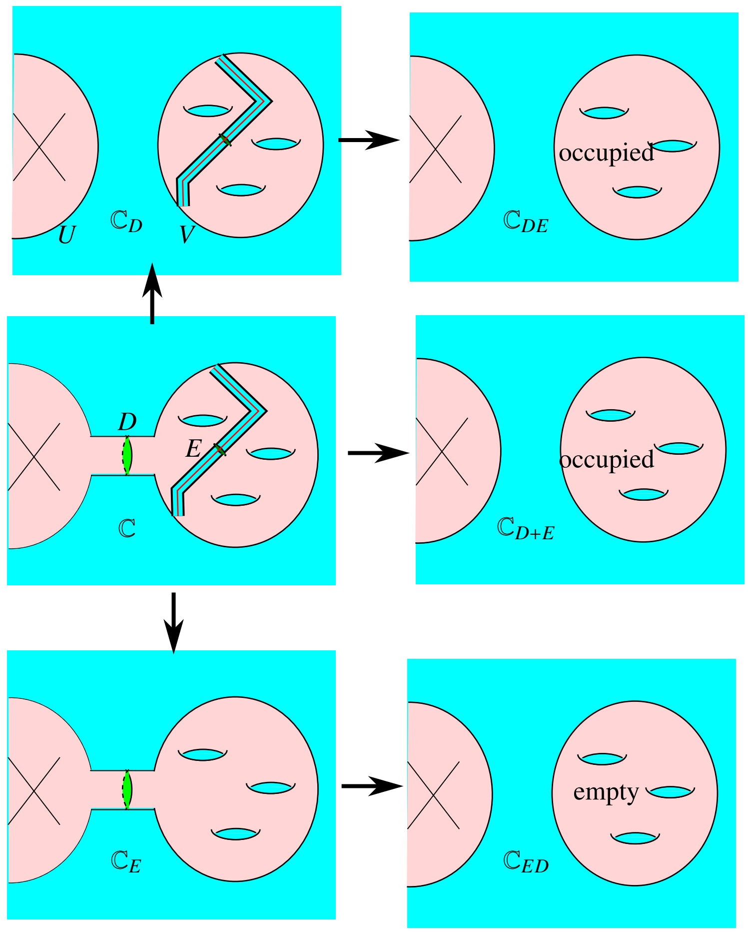

Figure 34: $\mathbb{C}_{DE} = \mathbb{C}_{D+E} \neq \mathbb{C}_{ED}$

Proof. Since D and E are parallel, after surgery on $\mathcal{D} \cup E$ there is a sphere G in $F_{\mathcal{D}\cup E}$ bounding a disky ball B_G with one scar of each of the disks D and E on G. Moreover the other remnant $C_- = C_D - B_G$ of C_D in $\mathbb{C}_{\mathcal{D}E}$ is homeomorphic to C_D, so it is a handlebody if and only if C_D is. If C_D is not a handlebody, so also C_- is not a handlebody, or if C_D is an empty handlebody, then per Definition 7.2 the ball B_G is empty and so a goneball in $\mathbb{C}_{\mathcal{D}E}$. Then $\mathbb{C}_{\mathcal{D}} = \mathbb{C}_{\mathcal{D}E}$ (with C_- replacing C_D). If C_D is occupied then the same statement is true for one of the two siblings of C_D. This proves 1).

The remnants of C in $\mathbb{C}_{\mathcal{D}+E}$ are the same as in $\mathbb{C}_{\mathcal{D}}$, with the addition of the disky ball B_G. If B_G is a goneball, we are done as in the proof of 1). The only way in which B_G could not be a goneball, per Definition 7.2, is if C is an occupied handlebody with only disky remnants. In the latter case, only the sibling in which B_G is occupied does not also occur among the sibling remnants of C in $\mathbb{C}_{\mathcal{D}}$. This proves 2).

The proof of 3) follows almost immediately. Following 1), we may as well assume that $\mathbb{C}_{\mathcal{D}} = \mathbb{C}_{\mathcal{D}E}$. The result follows from 2), unless C is an occupied handlebody with only disky handlebody remnants in $\mathbb{C}_{\mathcal{D}}$. In this case, C_- is also disky, so C also has only disky handlebody remnants in $\mathbb{C}_{\mathcal{D}+E}$. Then per Definition 7.2 the remnants of C in both $\mathbb{C}_{\mathcal{D}+E}$ and $\mathbb{C}_{\mathcal{D}E}$ have the same description of siblings, one corresponding to exactly one of the remnants of $\mathbb{C}_{\mathcal{D}+E}$ being occupied, with remnant B_G a goneball except when it is the occupied remnant. □

Note: Switching the order of $\mathcal{D}$ and E makes a difference. The example preceding Lemma 12.4 shows that there may be a handlebody in S^3 whose boundary is in $F_{\mathcal{D}\cup E}$ and which is not even coherent for $\mathbb{C}_{\mathcal{D}+E}$ and $\mathbb{C}_{E\mathcal{D}}$.

Return now to the discussion of decomposition by $\mathcal{D}_x$ and $\mathcal{D}_y$.

Definition 12.5. *Suppose H is a handlebody in S^3 so that $\partial H \subset F_{\mathcal{D}_x \cup \mathcal{D}_y}$. Then a disk $D \in \mathcal{D}_x \cup \mathcal{D}_y$* is incident to H *if $D \subset \text{int}(H)$ or D leaves a scar on ∂H. The handlebody H is an x*-handlebody *(resp. y-handlebody) if H is incident only to disks in $\mathcal{D}_x$ (resp. $\mathcal{D}_y$). An x*-handlebody *(resp. y-handlebody) H is an x*-handlebody *(resp. y-handlebody)* chamber *in $\mathbb{C}_{xy}$ if it is a chamber there. That is, ∂H is in the defining surface $F(\mathbb{C}_{xy})$ and no other chamber lies in* int(H).

Lemma 12.6. *If H is an x-handlebody chamber in $\mathbb{C}_{xy}$ then H is also a chamber of $\mathbb{C}_x$.*

Proof. Let C be the chamber of $\mathbb{C}_x$ which has H as a remnant. If any disks in $\mathcal{D}_y$ are incident to C then each remnant of C in $\hat{\mathbb{C}}_{xy}$ has a scar from $\mathcal{D}_y$, so the same is true after getting rid of goneballs. Since H has no scars from $\mathcal{D}_y$, no disks from $\mathcal{D}_y$ are incident to C, and C is unaffected by the decomposition by $\mathcal{D}_y$. That is, $C = H$ as required. □

Suppose $\mathbb{C}$ contains no occupied handlebody chambers. Then by Corollary 7.3 a handlebody chamber in $\mathbb{C}_x$ or a handlebody chamber in $\mathbb{C}_{x+y}$ is empty if and only if it is disky.

Lemma 12.7. *Suppose $\mathbb{C}$ contains no occupied handlebody chambers. Then:*

- *Each x-handlebody chamber of $\mathbb{C}_{xy}$ is empty if and only if it is a disky chamber under the decomposition $\mathbb{C} \xrightarrow{\mathcal{D}_x} \mathbb{C}_x$.*
- *If $\mathbb{C}_x$ contains no occupied handlebody chambers then each y-handlebody chamber of $\mathbb{C}_{xy}$ is empty if and only if it is disky under the decomposition $\mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$.*
- *If $\mathbb{C}_x$ does contain occupied handlebody chambers, there is a preferred alignment of $\mathcal{D}_y$ in $\mathbb{C}_x$ so that each y-handlebody chamber of $\mathbb{C}_{xy}$ is empty if and only if it is disky under the decomposition $\mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$. Any other preferred alignment of $\mathcal{D}_y$ gives a sibling decomposition $\mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$.*

Proof. The first statement follows from Corollary 7.3 and Lemma 12.6. The second statement follows from Corollary 7.3.

The third statement is more difficult to prove. Observe first that by the argument in Corollary 7.3, any handlebody remnant of a chamber in $\mathbb{C}_x$ that is not an occupied handlebody is empty if and only if it is disky. Moreover the argument extends, via Definition 7.2, even to occupied handlebody chambers in $\mathbb{C}_x$ so long as at least one remnant in $\mathbb{C}_{xy}$ is not a disky handlebody. So we need only describe how to align those disks in $\mathcal{D}_y$ that are incident to each chamber C_x in $\mathbb{C}_x$ of the following type: C_x is an occupied handlebody and every remnant of C_x in $\mathbb{C}_{xy}$ is a disky handlebody.

This is how it is done: Since C_x is occupied, by Corollary 7.3 C_x is not disky. So there is some component F_x of $F \cap \text{int}(C_x)$ that is not a disk. F_x must be incident to some disk in $\mathcal{D}_x$ else it would be a subsurface of $F(\mathbb{C}_x)$ and not lie in the interior of C_x.

The subcollection of disks in $\mathcal{D}_y$ that are incident to ∂C_x is disjoint from F_x, since $\text{int}(\mathcal{D}_y)$ is disjoint from F. Hence the surface F_x will lie in the interior of some remnant C_{xy} of C_x in $\mathbb{C}_{xy}$, so that remnant too is incident to $\mathcal{D}_x$. In particular, C_{xy} is not a y-handlebody. Following Definition 7.2, choose a preferred alignment where every (disky) handlebody remnant of C_x other than C_{xy} is empty. Then in particular each y-handlebody remnant is both disky and empty, as required. $\square$

Lemma 12.8. *Suppose $\mathbb{C}$ does not certify and H is a handlebody in S^3 so that $\partial H \subset F_{\mathcal{D}_x \cup \mathcal{D}_y}$. If H is an x-handlebody it is coherent. If H is a y-handlebody then it is coherent in one of the sibling decompositions of $\mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$.*

Proof. If appropriate (i. e. if $\mathbb{C}_x$ contains occupied handlebody chambers) choose the alignment for $\mathcal{D}_y$ in $\mathbb{C}_x$ given in Lemma 12.7 to ensure that each y-handlebody chamber of $\mathbb{C}_{xy}$ is empty if and only if it is disky under the decomposition $\mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$. Also from Lemma 12.7 we have that each x-handlebody chamber of $\mathbb{C}_{xy}$ is empty if and only if it is disky under the decomposition $\mathbb{C} \xrightarrow{\mathcal{D}_x} \mathbb{C}_x$.

Case 1: x-handlebodies

The proof will be by contradiction. Suppose there is an x-handlebody that is not coherent and let H be an innermost one that is not coherent. Suppose first that there are no components of $F_{\mathcal{D}_x \cup \mathcal{D}_y}$ contained in the interior of H. Then $\mathbb{C} \xrightarrow{\mathcal{D}_x} \mathbb{C}_x$ is disky (for $F \cap \text{int}(H)$ has no components at all) and hence empty (possibly a goneball) in $\mathbb{C}_{xy}$. Similarly $\mathbb{C} \xrightarrow{\mathcal{D}_x \cup \mathcal{D}_y} \mathbb{C}_{x+y}$ is disky and hence H is empty in $\mathbb{C}_{x+y}$. Since H is empty in both $\mathbb{C}_{xy}$ and $\mathbb{C}_{x+y}$ it is coherent, a contradiction.

We deduce then that there must be components of $F_{\mathcal{D}_x \cup \mathcal{D}_y}$ in the interior of H. Suppose a subsurface of $F_{\mathcal{D}_x \cup \mathcal{D}_y}$ in the interior of H becomes in $\mathbb{C}_x$ a component F_x of $F(\mathbb{C}_x)$. Then H is not a chamber in $\mathbb{C}_x$ (since it contains a component of $F(\mathbb{C}_x)$ in its interior) and, since H is incident to no disks in $\mathcal{D}_y$, H remains not a chamber in $\mathbb{C}_{xy}$. Since no disk in $\mathcal{D}_y$ is incident to F_x, F_x is also a component of $\mathbb{C}_{x+y}$ unless it is a goneball there. But if it were a goneball there, and not in $\mathbb{C}_{xy}$ it would be a further in handlebody that is not coherent, contradicting our choice of H. We conclude that F_x remains as a surface in $\mathbb{C}_{x+y}$ so H is not a chamber there either. But this implies H is coherent, a contradiction.

We deduce then that $F_{\mathcal{D}_x \cup \mathcal{D}_y} \cap \text{int}(H)$ must become a collection of spheres bounding goneballs in $\mathbb{C}_x$, hence in $\mathbb{C}_{xy}$. Then, by our consistency hypothesis, the spheres bound goneballs also in $\mathbb{C}_{x+y}$. Hence in each of $\mathbb{C}_{xy}$ and $\mathbb{C}_{x+y}$, H is either a chamber or is itself a goneball. Whether H is occupied or empty (so perhaps a goneball) is determined in $\mathbb{C}_{x+y}$ by whether the components of $F \cap \text{int}(H)$ consist entirely of disks, by Corollary 7.3. By the first statement of Lemma 12.7 the same is true in $\mathbb{C}_{xy}$. Hence H is coherent a final contradiction.

Case 2: y-handlebodies

The proof will be by contradiction, similar to Case 1. Suppose there is a y-handlebody that is not coherent and let H be an innermost one that is not coherent. Suppose there is a component F_0 of $F_{\mathcal{D}_x \cup \mathcal{D}_y}$

that lies in int(H) and remains in $F(\mathbb{C}_{xy})$, so H is not a chamber in $\mathbb{C}_{xy}$. By Lemma 12.2 F_0 is a sphere, so it bounds a ball in H. The ball it bounds is not a goneball in $\mathbb{C}_{xy}$ so, by the consistency hypothesis, F_0 does not bound a goneball in $\mathbb{C}_{x+y}$. This implies that H is not a chamber in either $\mathbb{C}_{xy}$ or $\mathbb{C}_{x+y}$, contradicting our assumption that H is not coherent.

We deduce then that int(H) is disjoint from $F(\mathbb{C}_{xy})$ so H is a chamber in $\mathbb{C}_{xy}$. A symmetric argument shows that H is also a chamber in $\mathbb{C}_{x+y}$. Suppose $F \cap \text{int}(H)$ has a component F_0 that is not a disk, so H is an occupied handlebody in $\mathbb{C}_{x+y}$.

Subcase 2a: $F_0 \subset \text{int}(H)$ is a closed surface.

Since H is a y-handlebody, F_0 is unaffected by surgery on $\mathcal{D}_x$ so it persists as a component of $F_{\mathcal{D}_x}$. Since F_0 does not bound a goneball in $\mathbb{C}$, it does not bound one in $\mathbb{C}_s$, so F_0 persists as a component of $F(\mathbb{C}_x)$. In particular, H is not disky under the decomposition $\mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$. This implies that H is an occupied handlebody in $\mathbb{C}_{xy}$, contradicting the hypothesis that H is not coherent.

Subcase 2b: $\partial F_0 \neq \emptyset$.

Let $\partial_- H = \partial H \cap F$, that is ∂H with all scars removed. Note that since F_0 is not a disk it is either non-planar or has more than one boundary component; in either case $F_0 \cup \partial_- H$ is non-planar. In particular, the subsurface of $F_{\mathcal{D}_x}$ that contains it cannot be a sphere, so $F_0 \cup \partial_- H$ persists as a subsurface of $F(\mathbb{C}_x)$. In particular, F_0 persists, and, just as in Subcase 2a this implies that H is an occupied handlebody in $\mathbb{C}_{xy}$, contradicting the hypothesis that H is not coherent.

We are thus reduced to the case that $F \cap \text{int}(H)$ consists entirely of disks, so H is disky in $\mathbb{C}_{x+y}$. Since we are assuming $\mathbb{C}$ does not certify, and therefore has no occupied handlebody chambers, this implies that H is empty (and so a goneball if H is a ball) in $\mathbb{C}_{x+y}$. It's possible that in $\mathbb{C}_x$ the surface $\partial_- H$ is part of a sphere $G \subset F$ that bounds a goneball in $\mathbb{C}_x$. This would imply that ∂H is planar, so H is a ball, and would make H a goneball of the decomposition sequence $\mathbb{C} \xrightarrow{\mathcal{D}_x} \mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$, as described before Lemma 12.3.

On the other hand, if $F \cap \partial H$ does remain in $F(\mathbb{C}_{D_x})$, what we can conclude is that H is disky in the decomposition $\mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$ and so, per our choice of preferred alignment of $\mathcal{D}_y$, H is empty. (In particular, it is again a goneball if H is a ball.) In either case, H is then coherent for $\mathbb{C}_{xy}$ and $\mathbb{C}_{x+y}$, a final contradiction. □

12.2 When a disk set is a singleton

We take the discussion above one step further:

Definition 12.9. *Suppose H is a handlebody in S^3 so that $\partial H \subset F_{\mathcal{D}_x \cup \mathcal{D}_y}$. H is an xe-handlebody (resp. ey-handlebody) if H is incident only to disks in $\mathcal{D}_x$ (resp. $\mathcal{D}_y$) except for a single disk E(xeptional) which is also incident to H but lies in $\mathcal{D}_y$ (resp. $\mathcal{D}_x$).*

We proceed to two lemmas that are parallel in spirit and whose proofs will use the same figures. To keep their use disentangled, in Figures 36, 37 and 39 the subscripts $xe, x+e$ will refer to the case in which $\mathcal{D}_y = E$ and the subscripts $ey, e+y$ will refer to the case in which $\mathcal{D}_x = E$.

Lemma 12.10. *Suppose $\mathbb{C}$ does not certify, H is an ey-handlebody and, if $\mathbb{C}_x$ contains an occupied handlebody, $\mathcal{D}_y$ is given the preferred alignment in $\mathbb{C}_x$ of Lemma 12.7. Then either H is coherent or*

- *E lies in* $\mathrm{int}(H)$ *and*
- *either ∂E is non-separating in $F - \partial\mathcal{D}_y$ or ∂E is non-separating in F and E is parallel to a disk in $\mathcal{D}_y$ and*
- *H lies in a chamber*[1] *of $\mathbb{C}_{xy}$ and H is either*
  - *empty (and so a goneball if H is a ball), the deflation of a handlebody chamber in $\mathbb{C}_{x+y}$ or*
  - *the deflation of a single bullseye of $\mathbb{C}_{x+y}$ in H. (Possibly H is part of the bullseye.)*

Proof. As usual, let $F_{E\cup\mathcal{D}_y}$ be the surface obtained from $F = F(\mathbb{C})$ by surgery on $E \cup \mathcal{D}_y$. Following Lemma 12.2 we will assume that each component of $F_{E\cup\mathcal{D}_y}$ that lies in $\mathrm{int}(H)$ is a sphere and so bounds a ball in H. Let $\mathfrak{S}$ be this set of spheres

Special case: Each ball in H bounded by a sphere in $\mathfrak{S}$ is coherent.

Claim: In this special case, H lies entirely in a chamber of $\mathbb{C}_{x+y}$ if and only if it lies entirely in a chamber of $\mathbb{C}_{xy}$.

Proof of claim: For both $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$, H is contained in a chamber if and only if each ball in H bounded by a sphere in $\mathfrak{S}$ is a goneball, for this determines whether there is any chamber contained in $\mathrm{int}(H)$. But in this special case, each ball bounded by a sphere in $\mathfrak{S}$ is coherent, so the ball it bounds is a goneball in $\mathbb{C}_{x+y}$ if and only if it is a goneball in $\mathbb{C}_{xy}$. This proves the claim.

Following the claim, but still in the special case, we will assume that H lies entirely in a chamber (possibly it is itself an entire chamber) of both $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$ and need to determine under what circumstances H is not coherent. That is,

- If H is a ball, when is it a goneball in one of $\mathbb{C}_{x+y}$ or $\mathbb{C}_{xy}$ but not the other?
- If H is not a ball, so it is itself a chamber, when is it empty in one of $\mathbb{C}_{x+y}$ or $\mathbb{C}_{xy}$ but not the other?

We examine possible subcases. For all but the last, H will turn out to be coherent:

Subcase 1a: There is a component F_0 of $F \cap \mathrm{int}(H)$ that is not a disk and not incident to E.

Since F_0 is not a disk, H is not a disky handlebody in the decomposition $\mathbb{C} \to \mathbb{C}_{x+y}$. Since F_0 is not incident to E, F_0 is not affected by the decomposition $\mathbb{C} \xrightarrow{E} \mathbb{C}_x$. This implies that H is also not a disky handlebody in the decomposition $\mathbb{C}_x \to \mathbb{C}_{xy}$. H is then an occupied handlebody in both $\mathbb{C}_{x+y}$ or $\mathbb{C}_{xy}$, so H is coherent.

Subcase 1b: A component F_e of $F \cap \mathrm{int}(H)$ is incident to E and surgery on E (in the decomposition $\mathbb{C} \xrightarrow{E} \mathbb{C}_x$) does not turn F_e into the union of disks and spheres.

Essentially the same argument applies: F_e can't be a disk, so H is occupied in $\mathbb{C}_{x+y}$; some non-disk component of the surgered F_e remains as a component of $F(\mathbb{C}_x)$ so H is occupied in $\mathbb{C}_{xy}$.

Subcase 2: E does not lie in $\mathrm{int}(H)$, so it leaves either one or two external scars on ∂H.

Following Subcase 1a, we can assume every component of $F \cap \mathrm{int}(H)$ is a disk, hence H is disky and so, by Corollary 7.3 H is an empty chamber in $\mathbb{C}_{x+y}$. H is also a disky handlebody remnant of the

[1]The phrase 'H lies in a chamber' means that either H is itself a chamber, or H is a goneball in a chamber.

decomposition $\mathbb{C}_x \to \mathbb{C}_{xy}$ but this does not immediately imply that H is empty in $\mathbb{C}_{xy}$: per Definition 7.2 H might be the remnant of an occupied handlebody C_x in $\mathbb{C}_x$ all of whose other remnants are also disky handlebodies. We explore this possibility further:

Suppose this were the case, and C is the chamber in $\mathbb{C}$ of which C_x is the occupied handlebody remnant. If the disk E lies outside C then the scars of E in ∂C_x would be internal scars. By the hypothesis of this subcase we may therefore assume that E lies in the interior of C. The one or two remnants of C in $\mathbb{C}_x$ are then subsets of C, so F is disjoint from these remnants and the remnant C_x in particular is disky. Per Definition 7.2 the only way that $\mathbb{C}_x$ can be a disky remnant of C and still be occupied is if C is an occupied handlebody with all remnants disky. But a hypothesis of the Lemma is that $\mathbb{C}$ does not certify, so C is not an occupied handlebody. From the contradiction we conclude that H is empty in $\mathbb{C}_{xy}$ and so H is coherent.

Following Subcase 2, we henceforth assume that $E \subset \text{int}(H)$.

Subcase 3: $F \cap \text{int}(H)$ consists of disks, so H is either a goneball or an empty chamber in $\mathbb{C}_{x+y}$.

Let F_e be the component of $F \cap \text{int}(H)$ on which ∂E lies. Since H is irreducible, E is parallel in H to a subdisk F_0 of F_e. Then the decomposition $\mathbb{C} \xrightarrow{E} \mathbb{C}_x$ has essentially no effect on $\mathbb{C}$: after removing the goneball bounded by $E \cup F_0$, all that has changed is a proper isotopy of F_e that replaces F_0 with E (and the removal of any disks in $\mathcal{D}_y$ that were incident to F_0, but these only give rise to goneballs in $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$). Then the decomposition $\mathbb{C}_x \to \mathbb{C}_{xy}$ also leaves H either a goneball or an empty chamber in $\mathbb{C}_{xy}$, so H is coherent.

Subcase 4: All that remains: $F \cap \text{int}(H)$ has a single non-disk component F_e, F_e is incident to E, and surgery on E turns F_e into the union of disks and spheres.

F_e could be a torus bounding an empty solid torus, a once-punctured torus, or an annulus, and in each case E is a meridian disk. (See the left column in Figures 36 and 37.) It follows that ∂E is non-separating in F and, unless $\mathcal{D}_y$ also contains a disk whose boundary is parallel to ∂E, ∂E is also non-separating in $F - \partial\mathcal{D}_y$. If $\mathcal{D}_y$ does contain a disk D whose boundary is parallel to ∂E then D and E are parallel, since otherwise between them would lie a closed component of $F \cap \text{int}(H)$, contradicting the hypothesis of this subcase. So, in the end, we are in the situation described in the statement of the Lemma: ∂E **is non-separating in** $F - \partial D_y$ **or** ∂E **is non-separating in** F **and** E **is parallel to a disk in** $\mathcal{D}_y$. In this case H is an occupied handlebody in $\mathbb{C}_{x+y}$, since F_e is not a disk, but may be empty in $\mathbb{C}_{xy}$, since $\mathbb{C}_x \xrightarrow{\mathcal{D}_y} \mathbb{C}_{xy}$ is disky. Since H could be incoherent, we explore this situation further:

Consider the chamber C_x in $\mathbb{C}_x$ which has H as a remnant. If every other remnant of C_x is also a disky handlebody, then C_x is a handlebody and it must be occupied because F_e is not a disk. Since $E \subset \text{int}(H)$, each other remnant of C_x is a disky y-handlebody, so per Lemma 12.7, each is empty. This implies that H must be occupied in $\mathbb{C}_{xy}$, and H is coherent. Finally, if any remnant of C_x in $\mathbb{C}_{xy}$ is not a disky handlebody then Definition 7.2 says that a preferred alignment of $\mathcal{D}_y$ leaves H deflated to such a remnant, an outcome allowed in Lemma 12.10. This concludes the proof in the special case.

The general case:

Following the special case, the remaining possibility is that there is a sphere in $\mathfrak{S}$ bounding a ball in H that is not coherent. Let G be an innermost such sphere. By the special case, E lies in the ball B_G that G bounds in H and B_G is occupied in $\mathbb{C}_{x+y}$ but deflates into another chamber in $\mathbb{C}_{xy}$. Since B_G is occupied in $\mathbb{C}_{x+y}$, it is a chamber, so H does not lie inside a chamber in $\mathbb{C}_{x+y}$. If any sphere in $\mathfrak{S}$ does not bound a goneball in $\mathbb{C}_{xy}$ then H is also not a chamber in $\mathbb{C}_{xy}$ and so is coherent. See Figure 35.

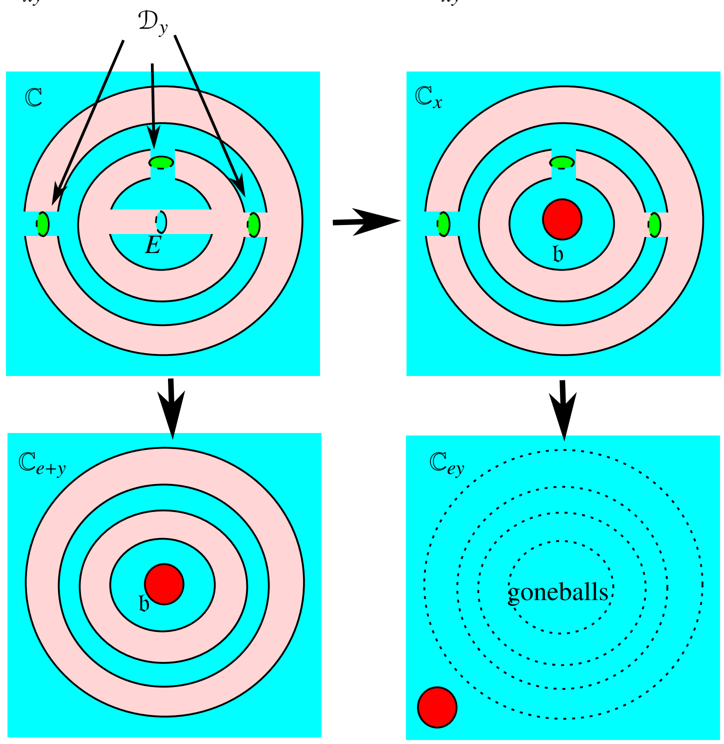

Figure 35: Bullseye deflation $\mathbb{C}_{ey} \dashrightarrow \mathbb{C}_{e+y}$

We are left with the case that every sphere in $\mathfrak{S}$ bounds a goneball in $\mathbb{C}_{xy}$. Any such sphere not containing G in the interior of the ball it bounds will be a y-handlebody, since $E \subset B_G$, so it follows from Lemma 12.8 that the ball will also be a goneball in $\mathbb{C}_{x+y}$. So all spheres in $\mathfrak{S}$ that don't bound goneballs in $\mathbb{C}_{x+y}$ are nested around G. Let G' be an outermost sphere in this nested collection of spheres $\mathfrak{S}_e \subset \mathfrak{S}$ bounding goneballs in $\mathbb{C}_{xy}$ but not in $\mathbb{C}_{x+y}$, or ∂H itself if H is a ball. Since G' bounds a goneball in $\mathbb{C}_{xy}$, the ball has trivial Heegaard splitting. It follows that if $\mathfrak{S}_e$ has more than one sphere, the collar between any two has trivial splitting. Thus $\mathfrak{S}_e$ is a bullseye in $\mathbb{C}_{x+y}$, the final possibility allowed by Lemma 12.10. □

The next lemma is more complicated to state, mostly because it has to accomodate the possibility that even if $\mathcal{D}_y = E$, there may be two xe-handlebody chambers in $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$. Namely if E divides a handlebody chamber $C_x \subset \mathbb{C}_x$ into two handlebodies, then they each are xe-handlebody chambers, since E leaves an external scar on both of them. According to Definition 7.2 if C_x is an occupied handlebody, there are two preferred alignments of E: H is empty and H' is occupied, or vice versa. The different alignments produce sibling flaggings of $\mathbb{C}_{xy}$. One sibling is obtained from the other by deflation of H' into H (which is a bullseye deflation if H' is then a goneball) or vice versa. In this case it is natural to call replacing the flagging of the chambers H and H' in $\mathbb{C}_{x+y}$ with either of the flaggings in $\mathbb{C}_{xy}$ a *sibling deflation*.

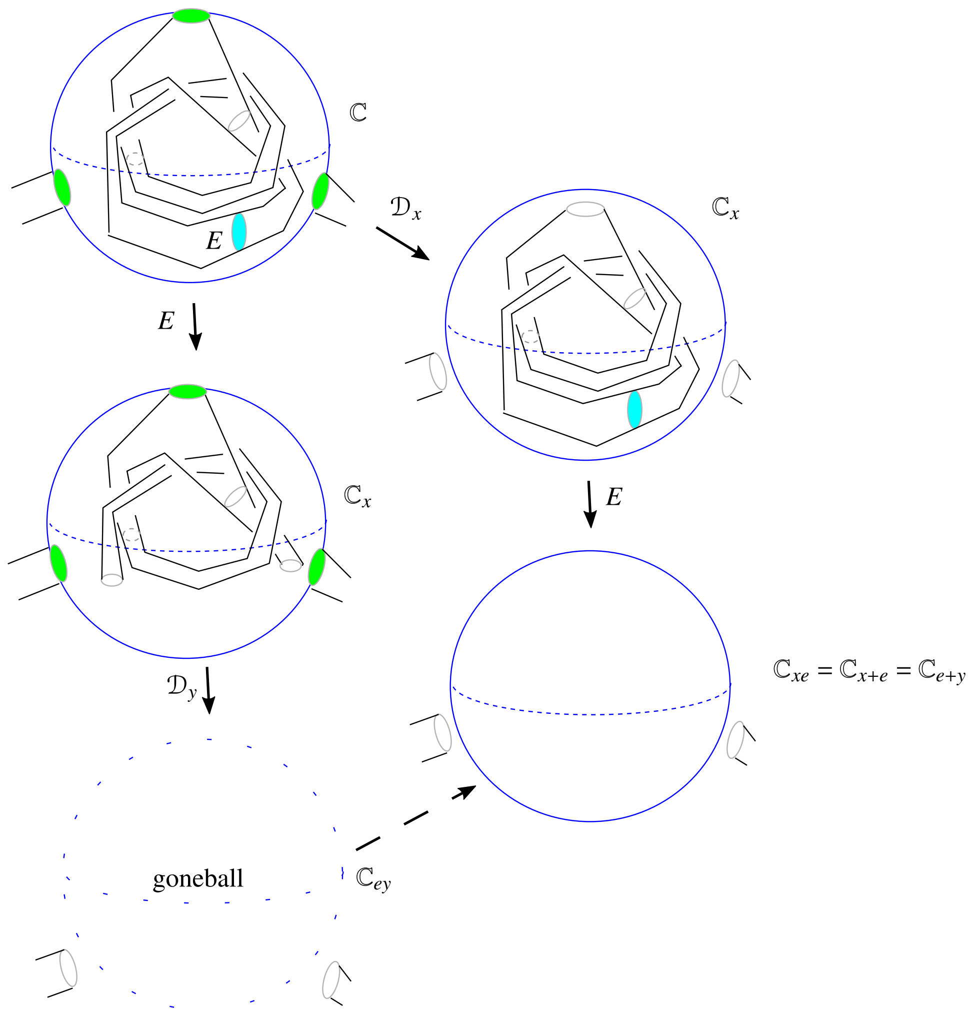

Figure 36: H a ball, F_e a punctured torus

Lemma 12.11. *Suppose $\mathbb{C}$ does not certify and H is an xe-handlebody. Then either*

- *H is coherent, or*
- *$\mathbb{C}_x$ has an occupied handlebody chamber containing E and, in one of the two sibling deflations, H is coherent, or*
- *H is a chamber in $\mathbb{C}_{xy}$, either the deflation of an occupied chamber in $\mathbb{C}_{x+y}$ or the result of the deflation of a single bullseye of $\mathbb{C}_{x+y}$ in H.*

Furthermore, in the third instance, when neither of the first two outcomes applies,

- *E leaves exactly one scar on ∂H,*

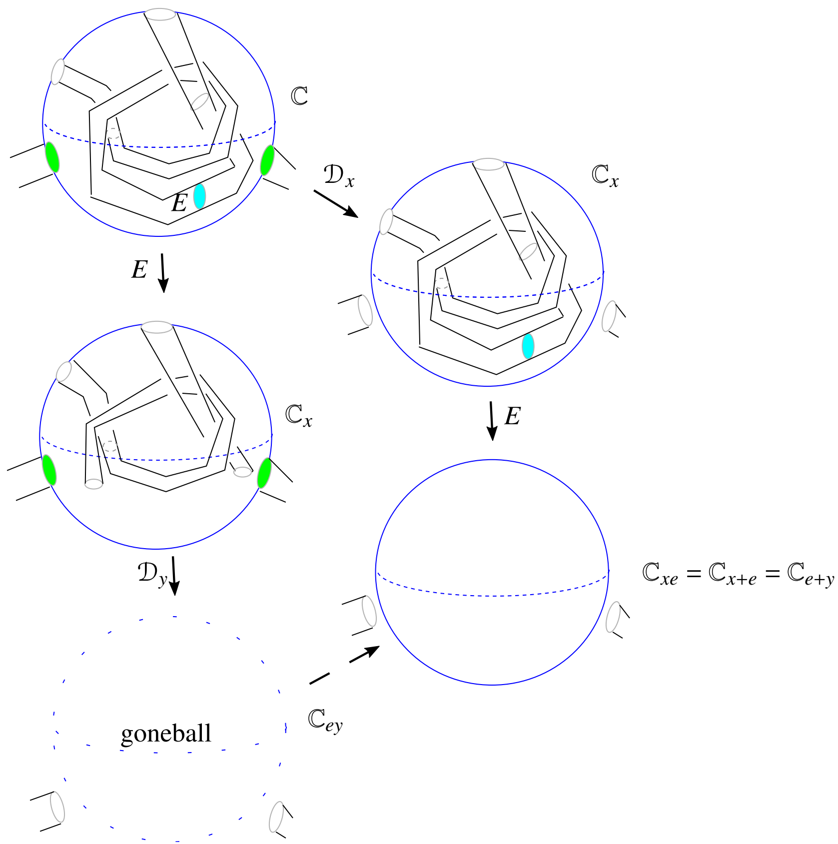

Figure 37: H a ball, F_e an annulus

- *∂E is separating in a component of $F - \partial\mathcal{D}_x$, and*
- *E is not parallel to any disk in $\mathcal{D}_x$.*

Proof. If E is parallel to a disk in $\mathcal{D}_x$, Lemma 12.4 says one of the first two outcomes occurs, and we are done. So henceforth we assume that E is not parallel to a disk in $\mathcal{D}_x$.

Here are some preliminary observations:

- If H (or H' in the case in which sibling deflation may arise) is coherent we are done. So we assume H (and H', if relevant) is not coherent.
- Following Lemma 12.2, unless every component of $F_{\mathcal{D}_x \cup E}$ lying in $\text{int}(H)$ is a sphere, H is not a chamber in either $\mathbb{C}_{xy}$ nor $\mathbb{C}_{x+y}$ and so is coherent. So we henceforth assume that every component of $F_{\mathcal{D}_x \cup E}$ lying in $\text{int}(H)$ is a sphere. Let $\mathfrak{S}$ denote the collection of all such spheres.
- Sibling deflation may arise when E is a separating disk in an occupied handlebody chamber of $\mathbb{C}_x$. If that chamber has H as one of the two remnants, each ball bounded by a sphere in $\mathfrak{S}$ is an x-handlebody and so, per Lemma 12.8 the sphere is coherent.

On the last point, sibling deflation, there is more to say:

The Sibling Claim: Suppose that E is a separating disk in an occupied handlebody chamber C_x of $\mathbb{C}_x$. In at least one of the two sibling deflations of C_x in $\mathbb{C}_{xy}$, either

- H is coherent, or
- H is the deflation of an occupied chamber in $\mathbb{C}_{x+y}$ or
- the only remnant of C_x in the interior of H is a coherent ball.

.

Proof of the Sibling Claim: Since $\mathbb{C}$ has no occupied handlebodies, the handlebody $\mathbb{C}_x$ is occupied if and only if $F \cap \text{int}(C_x)$ contains a non-disk component, by Corollary 7.3. Hence one or both of the handlebody chamber remnants $C_x - E$ are occupied in $\mathbb{C}_{x+y}$. It follows that either chamber (but not necessarily both) can be made coherent by a choice of sibling deflation in $\mathbb{C}_{xy}$. This proves the Sibling Claim if either of these chambers is H. Another possibility is that both handlebody chambers lie in the interior of H, and so are necessarily balls. We have seen then that at least one ball is occupied in $\mathbb{C}_{x+y}$ and exactly one in $\mathbb{C}_{xy}$. Thus H is not a chamber in either $\mathbb{C}_{x+y}$ or $\mathbb{C}_{xy}$ so H is coherent.

The final possibility is that one of the handlebody chamber remnants is a ball B_H in the interior of H and the other lies outside H (in fact it is then the complement $S^3 - H$ since its boundary is connected). B_H is either empty or occupied in $\mathbb{C}_{x+y}$. By choosing the right sibling deflation we can match that status for B_H in $\mathbb{C}_{xy}$. This makes B_H coherent and so proves the Sibling Claim.

It is possible that a sphere in $\mathfrak{S}$ bounds a ball that is not coherent - that is, it is a goneball in one of $\mathbb{C}_{xy}$ or $\mathbb{C}_{x+y}$ but not the other (regardless of the choice of sibling deflation, when E lies in an occupied handlebody of $\mathbb{C}_x$). If this occurs, let B_G be an innermost such ball, B_G can't be an x-handlebody, by Lemma 12.8 so E is incident to B_G. By the Sibling Claim, we may assume that if E lies in an occupied handlebody in $\mathbb{C}_x$, B_G is not one of its remnants in $\mathbb{C}_{xy}$.

In view of these possibilities, in the discussion below we will use the term *target manifold* to refer either to B_G, if there is a B_G as above, or to H itself if there is no B_G. In either case, any handlebody contained in the interior of the target manifold is a coherent ball.

Let $F_x = F(\mathbb{C}_x)$ denote the defining surface of $\mathbb{C}_x$. We consider possible scars that E might leave on the boundary of the target manifold in $F_{\mathcal{D}_x \cup E}$. Only in the last case, when there is exactly one external scar on the boundary of the target manifold, does the possibility of sibling deflations arise:

Claim 1: There is a scar of E on the boundary of the target manifold (either $G = \partial B_G$ or ∂H.)

The argument is essentially the same for either target manifold, so we assume here the target manifold is B_G. If there is no scar on G then G is among the components of $F_{\mathcal{D}_x}$, the surface obtained by $\mathcal{D}_x$ surgery on F. Notice that B_G must be a chamber in both $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$, for if there are occupied balls in the interior of one, they would be occupied in the other (because B_G is innermost among the non-coherent), making B_G not a chamber and therefore coherent, contradicting hypothesis.

If $F \cap \text{int}(B_G)$ consists of disks, then B_G is disky and hence a goneball in both $\mathbb{C}_{x+y}$ and $\mathbb{C}_x$, by Corollary 7.3. Hence B_G is also a goneball in $\mathbb{C}_{xy}$, again contradicting that B_G is not coherent. On the other hand, if $F \cap \text{int}(B_G)$ is not all disks. Then B_G is an occupied ball in both $\mathbb{C}_{x+y}$ and $\mathbb{C}_x$. The latter implies that B_G is also an occupied ball in $\mathbb{C}_{xy}$, since E is not incident to B_G; B_G is its own and only remnant in $\mathbb{C}_{xy}$. This would make B_G coherent, a final contradiction that proves Claim 1.

Claim 2: E does not leave two external scars on the target manifold.

Again the argument for target manifold B_G applies also to H, so we can suppose that B_G is the target manifold and consider what happens if E leaves two external scars on G. Then the chamber C_x in $\mathbb{C}_x$ of which B_G is a remnant is obtained from G by attaching a 1-handle dual to E, and so is a handlebody in $\mathbb{C}_x$. B_G is the sole remnant of C_x under the decomposition $\mathbb{C}_x \to \mathbb{C}_{xy}$ and its interior is disjoint from F_x, so B_G is occupied in $\mathbb{C}_{xy}$ if and only if C_x is occupied in $\mathbb{C}_x$. Since $\mathbb{C}$ has no occupied handlebodies, the handlebody $\mathbb{C}_x$ is occupied if and only if $F \cap \mathrm{int}(C_x)$ contains a non-disk component, by Corollary 7.3. But $F \cap \mathrm{int}(C_x)$ contains a non-disk component if and only if $F \cap \mathrm{int}(B_G)$ contains a non-disk component and the latter is equivalent to B_G being occupied in $\mathbb{C}_{x+y}$. Hence B_G is occupied in $\mathbb{C}_{xy}$ if and only if it is occupied in $\mathbb{C}_{x+y}$. Hence B_G is coherent, a contradiction that proves the claim.

Claim 3: E does not leave two internal scars on the target manifold.

The case when the target manifold is B_G is representative. If E leaves two internal scars on G then the belt annulus is a non-disk component of both $F \cap \mathrm{int}(B_G)$ and $F_x \cap \mathrm{int}(B_G)$, so B_G is occupied in both $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$. Hence B_G is coherent, a contradiction that proves the claim.

Claim 4: E does not leave one internal scar on the the target manifold

This is the most difficult claim to verify, for the argument involves some unusual computation. To avoid the danger of over-simplifying we will assume that the more complicated H is the target manifold; the argument when B_G is the target manifold is then just a special case. Suppose that E leaves one internal scar on ∂H. Then the other scar must be on a component of $F_{\mathcal{D}_x \cup E}$ that lies in the interior of H and so is a sphere $S \in \mathfrak{S}$. Put another way, in $F_{\mathcal{D}_x}$ the belt annulus of E connects the scar of E on ∂H to the disk complement in S of the scar of E on S. So the scar of E on ∂H is parallel in H to a disk component F_0 of $F_x \cap \mathrm{int}(H)$ (the union of the belt annulus of E and a subdisk of S). Since E is the only disk in the decomposition $\mathbb{C}_x \to \mathbb{C}_{xy}$, the disk F_0 is the only component of $F_x \cap \mathrm{int}(H)$.

By assumption the ball B_S that S bounds in H is coherent: unless it is a goneball in both, it persists in both $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$, so H would not be a chamber in either and therefore is coherent, a contradiction. So we henceforth assume that B_S is a goneball in both $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$. The former means that $F \cap \mathrm{int}(B_S)$ consists of disks, and the latter means that the chamber C_x of $\mathbb{C}_x$ of which H is a remnant is just the complement in H of the ball (parallel to B_S) between the scar of E in ∂H and F_0. That is, H is parallel to C_x via this ball. In particular, H is the only remnant of C_x so, via Definition 7.2, H is occupied in $\mathbb{C}_{xy}$ if and only if C_x is occupied in $\mathbb{C}_x$.

We wish now to examine $F \cap \mathrm{int}(H)$. F_0 is a disk containing scars of $\mathcal{D}_x$; the complement of the scars is a subset of F. In particular, if $F \cap \mathrm{int}(H)$ contains any closed components they would be disjoint from F_0. A closed component of F can't lie in B_S, since B_S is empty in $\mathbb{C}_{x+y}$, so it would have to lie in the chamber C_x of $\mathbb{C}_x$, making C_x, hence H, occupied in $\mathbb{C}_{xy}$ as well as $\mathbb{C}_{x+y}$. Then H would be coherent, a contradiction. So we need only consider the case in which $F \cap \mathrm{int}(H)$ contains no closed components.

Here is a useful observation about compact orientable surfaces with no closed components: For any closed connected orientable surface Q, $\chi(Q) \leq 2$ with equality if and only if Q is a sphere. It follows that for any compact connected surface Q, $\chi(Q) + |\partial Q| \leq 2$, with equality if and only if Q is planar, and so, when $\partial Q \neq \emptyset$, $\chi(Q) - |\partial Q| \leq 0$ with equality if and only if Q is planar and has a single boundary component, i. e. is a disk. More generally, this is then true for any compact orientable surface, so long as there are no closed (in fact no sphere) components. We have thus verified this claim:

Claim: Suppose Q is a compact orientable surface with no closed components. Then $\chi(Q)-|\partial Q|\leq 0$ with equality if and only if Q consists entirely of disks.

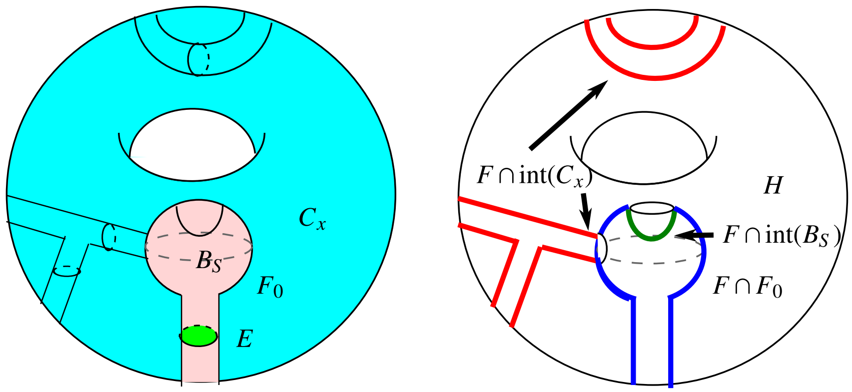

Figure 38: $F\cap\mathrm{int}(H)=(F\cap\mathrm{int}(C_x))\cup(F\cap F_0)\cup(F\cap\mathrm{int}(B_S))$

To apply the Claim, begin with the observation

$$F\cap\mathrm{int}(H)=(F\cap\mathrm{int}(C_x))\cup(F\cap F_0)\cup(F\cap\mathrm{int}(B_S)). \tag{1}$$

See Figure 38. For the purpose of this computation, include all boundary circles on each surface, so each surface is compact. We then have

$$\chi(F\cap\mathrm{int}(H))-\chi(F\cap\mathrm{int}(C_x))=\chi(F\cap F_0)+\chi(F\cap\mathrm{int}(B_S)). \tag{2}$$

Note that $F\cap F_0$ is the disk F_0 with all the scars of $\mathcal{D}_x$ that lie in F_0 removed. Hence $\chi(F\cap F_0)=1-|\text{scars on } F_0|$. These scars on F_0 are of two types: those that come from components of F lying in $\mathrm{int}(B_S)$ (all of them disks, since B_S is a goneball) and those coming from components of F lying in $\mathrm{int}(C_x)$. Each disk in $F\cap\mathrm{int}(B_S)$ contributes 1 to $\chi(F\cap\mathrm{int}(B_S))$ and -1 to $\chi(F\cap F_0)$ so

$$\chi(F\cap F_0)+\chi(F\cap\mathrm{int}(B_S))=1-|\text{scars on } F_0 \text{ left by } F\cap\mathrm{int}(C_x)|. \tag{3}$$

Another way to characterize the number of scars left on F_0 by $F\cap\mathrm{int}(C_x)$ is that it is the difference of the number of scars left on ∂C_x and those left on ∂H. Furthermore, the total number of scars left on ∂H is one more than the number left by ∂C_x, since it includes ∂E. We then obtain:

$$1-|\text{scars on } F_0 \text{ left by } F\cap\mathrm{int}(C_x)|=|\partial(F\cap\mathrm{int}(H))|-|\partial(F\cap\mathrm{int}(C_x))|. \tag{4}$$

Combining (2), (3) and (4) we have

$$\chi(F\cap\mathrm{int}(H))-\chi(F\cap\mathrm{int}(C_x))=|\partial(F\cap\mathrm{int}(H))|-|\partial(F\cap\mathrm{int}(C_x))| \tag{5}$$

which implies

$$\chi(F\cap\mathrm{int}(H))-|\partial(F\cap\mathrm{int}(H))|=\chi(F\cap\mathrm{int}(C_x))-|\partial(F\cap\mathrm{int}(C_x))|. \tag{6}$$

Now apply the Claim: The left side of (6) is zero if and only if H is disky and hence empty; the right side of (6) is zero if and only if C_x is disky and hence empty. Thus H is empty in $\mathbb{C}_{x+y}$ if and only if it is empty in $\mathbb{C}_{xy}$. That is, H is coherent, a contradiction that proves Claim 4.

Claim 5: If the target manifold contains exactly one external scar of E then H satisfies one of the three outcomes listed in Lemma 12.11.

The argument for this claim depends on whether H or B_G is the target manifold.

Case 1: The target manifold is B_G.

Let C_x be the chamber in $\mathbb{C}_x$ of which B_G is a remnant. Since E leaves exactly one external scar on G, E separates C_x into B_G and another remnant R_x, so B_G is disky under the decomposition $\mathbb{C}_x \to \mathbb{C}_{xy}$.

Subcase 1a: C_x is a handlebody.

We first show that C_x cannot be an empty handlebody. Since $\mathbb{C}$ contains no occupied handlebodies, C_x can only be empty if $F \cap \mathrm{int}(C_x)$ consists entirely of disks, by Corollary 7.3. Since $\mathrm{int}(B_G) \subset \mathrm{int}(C_x)$ this implies that $F \cap \mathrm{int}(B_G)$ consists entirely of disks, so B_G is empty in $\mathbb{C}_{x+y}$. Since C_x is empty in $\mathbb{C}_x$, Definition 7.2 says B_G is also empty in $\mathbb{C}_{xy}$. This contradicts the requirement that B_G is not coherent.

We deduce that C_x is an occupied handlebody. In this case, apply the Sibling Claim: the first two outcomes of that claim imply the first two outcomes of the Lemma; the third only arises when the target manifold is H, and then it contradicts Claim 4, completing the proof of the Lemma in this subcase.

Subcase 1b: C_x is not a handlebody.

In this case R_x cannot be a handlebody and in particular cannot be a ball, and so it can't be a chamber in $\mathrm{int}(H)$. Then $R_x \subset S^3 - H$. Moreover, Definition 7.2 says that in this case the preferred alignment of E leaves B_G empty in $\mathbb{C}_{x+y}$ and so a goneball. Our assumption is that B_G is not coherent, so B_G must be occupied in $\mathbb{C}_{x+y}$. The difference then between H in $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$ is that the latter is obtained from the former by deflation of B_G. This can be viewed as a bullseye deflation, the third outcome allowed. See the right side of Figure 39.

Case 2: The target manifold is H.

The proof is highly analogous to Case 1, so we merely sketch: As in Case 1, the component C_x of $\mathbb{C}_x$ of which H is a remnant is divided by E into two chambers, H and R_x. If C_x is an empty handlebody then, just as in Case 1, H is empty in both $\mathbb{C}_{x+y}$ and $\mathbb{C}_{xy}$ and so is coherent, completing the proof. If $\mathbb{C}_x$ is an occupied handlebody then the Sibling Claim applies; the first two outcomes of that Claim imply the first two claims of the Lemma, and the third outcome contradicts Claim 4. Finally, if C_x (so also R_x) is not a handlebody, then the preferred alignment given by Definition 7.2 leaves H empty in $\mathbb{C}_{xy}$. Thus if H is empty in $\mathbb{C}_{x+y}$ it is coherent; if it is occupied then $\mathbb{C}_{xy}$ is obtained from $\mathbb{C}_{x+y}$ by deflation into R_x, implying the third outcome of the Lemma.

Having established the claims, we now prove the Lemma: Suppose a ball $B_G \subset \mathrm{int}(H)$ presents itself, as described before Claim 1. Claim 1 says that E leaves scars on G, claims 2 and 3 show that E leaves exactly one scar on G, and Claim 4 says that it cannot be an internal scar. Thus E must leave one external scar on G, in which case Claim 5 shows that one of the three outcomes of Lemma 12.11 occurs.

On the other hand, if no B_G satisfies the criteria described before Claim 1 then Claims 1, 2, 3 and 4 show that, unless ∂H contains a single external scar of E, H is coherent. Claim 5 shows that when ∂H does contain a single external scar of E, one of the three outcomes of Lemma 12.11 occurs.

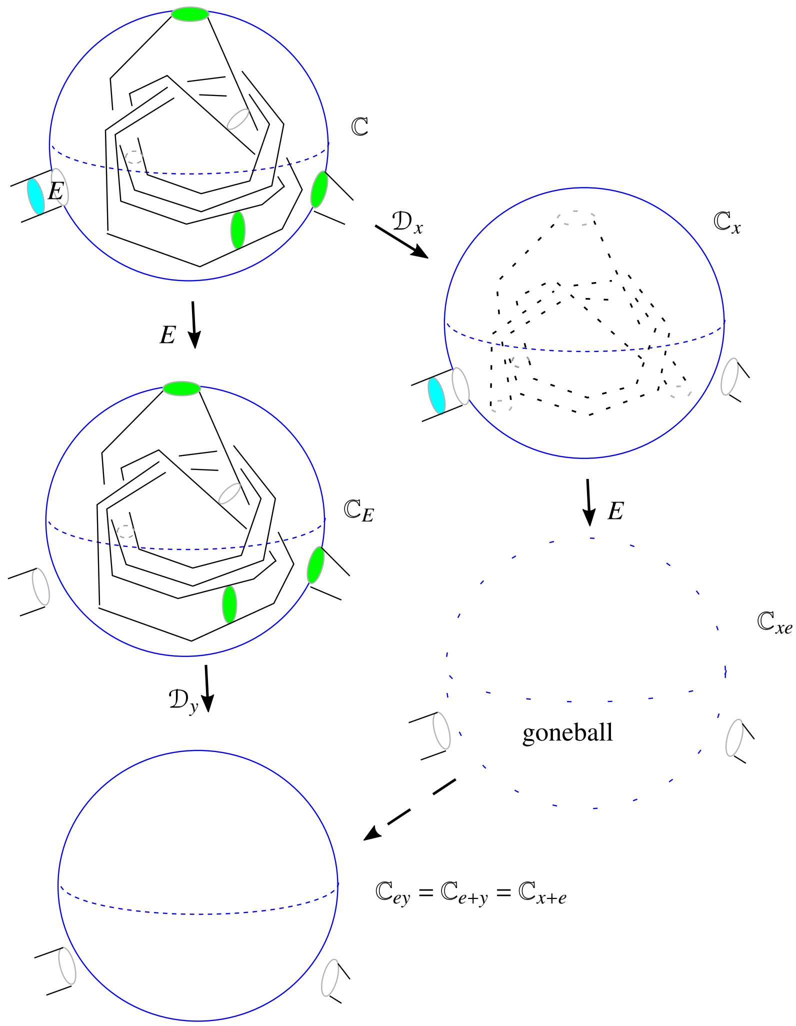

Figure 39: E leaves one external scar

We now examine when neither of the first two outcomes occurs, to verify the final three claims of Lemma 12.11. We have already concluded that E is not parallel to any disk in $\mathcal{D}_x$. The proof so far shows also that the target manifold must have a single external scar, and then Claim 5 shows that in this case ∂H contains exactly one scar of E, an internal scar when the target manifold is B_G and an external scar when the target manifold is H. All that remains is to show that ∂E is separating in $F - \partial\mathcal{D}_x$:

The definition of F_x implies that the complement of the scars of $\mathcal{D}_x$ in F_x is a collection of components of $F - \partial\mathcal{D}_x$. Not removing the scars of the disks in $\mathcal{D}_x$, that is, adding disks back, does not affect the number of components in the surface, so to show that ∂E is separating in a component of $F - \partial D_x$ it suffices to show that ∂E is separating in a component of F_x. But that follows immediately from the fact that E leaves only one scar on ∂H: ∂H is a component of $F_x - \partial E$ that is distinct from the component on which the other scar of E lies. □

Proposition 12.12. *Suppose $\mathbb{C}$ is a flagged chamber complex supporting $S^3 = A \cup_T B$, $\mathbb{C}$ does not certify and $\mathcal{D}_x, \mathcal{D}_y$ are disjoint disk sets in $\mathbb{C}$. If either $|\mathcal{D}_x|$ or $|\mathcal{D}_y| = 1$, then, after an appropriate choice of sibling if $\mathbb{C}_x$ contains occupied handlebody chambers, either $\mathcal{D}_{xy} = \mathcal{D}_{x+y}$ or $\mathcal{D}_{xy} \dashrightarrow \mathcal{D}_{x+y}$.*

In more detail: Suppose $\mathcal{D} \cup E$ is a disk set in $\mathbb{C}$, where E is a single disk. Let

- *$\mathbb{C}_{DE}$ be the flagged chamber complex given by the composition $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}} \xrightarrow{E} \mathbb{C}_{DE}$*
- *$\mathbb{C}_{ED}$ be the flagged chamber complex given by the composition $\mathbb{C} \xrightarrow{E} \mathbb{C}_E \xrightarrow{\mathcal{D}} \mathbb{C}_{ED}$ and $\mathcal{D}$ the preferred alignment described in Lemma 12.8*
- *$\mathbb{C}_{D+E}$ be the flagged chamber complex given by the decomposition $\mathbb{C} \xrightarrow{\mathcal{D} \cup E} \mathbb{C}_{D+E}$.*

(Note that here D in the subscripts refers not to a single disk but to the family $\mathcal{D}$ of disks.)

Then either

- *$\mathbb{C}_{DE} = \mathbb{C}_{D+E} = \mathbb{C}_{ED}$ or*
- *$\mathbb{C}_{DE} = \mathbb{C}_{D+E} \dashleftarrow \mathbb{C}_{ED}$ or*
- *$\mathbb{C}_{DE} \dashrightarrow \mathbb{C}_{D+E} = \mathbb{C}_{ED}$ or*
- *E lies in an occupied handlebody in $\mathbb{C}_{\mathcal{D}}$ and its sibling $\mathbb{C}'_{DE}$ satisfies $\mathbb{C}'_{DE} = \mathbb{C}_{D+E} = \mathbb{C}_{ED}$*

If $\mathbb{C}_{D+E} \neq \mathbb{C}_{ED}$ then either ∂E is non-separating in $F - \partial\mathcal{D}$ or ∂E is non-separating in F and E is parallel to a disk in $\mathcal{D}$.

Proof. The proof breaks naturally into two cases:

Case 1: $\mathbb{C}_{ED} \neq \mathbb{C}_{D+E}$.

By Lemma 12.3 there is a handlebody H in S^3 whose boundary is in $F_{\mathcal{D}\cup E}$ and which is not coherent in $\mathbb{C}_{ED}$ and $\mathbb{C}_{D+E}$. Then applying Lemma 12.10 to H, either ∂E is non-separating in $F - \partial\mathcal{D}$ or ∂E is non-separating in F and E is parallel to a disk in $\mathcal{D}$. Then by Lemma 12.11 every handlebody in S^3 whose boundary is in $F_{\mathcal{D}\cup E}$ is coherent for $\mathbb{C}_{DE}$ and $\mathbb{C}_{D+E}$. So, by Lemma 12.3, $\mathbb{C}_{DE} = \mathbb{C}_{D+E}$. Thus what is left to show in this case is that $\mathbb{C}_{ED} \dashrightarrow \mathbb{C}_{D+E}$.

There cannot be two disjoint handlebodies in S^3 whose boundaries are in $F_{\mathcal{D}\cup E}$ and which are not coherent for $\mathbb{C}_{ED}$ and $\mathbb{C}_{D+E}$ since Lemma 12.10 says that the disk E lies in the interior of each such handlebody. Then there is a maximal such handlebody - that is, one that contains any other - so without loss take H to be the maximal one. By Lemma 12.10 H is contained in a chamber of $\mathbb{C}_{ED}$ and is obtained from a chamber of $\mathbb{C}_{D+E}$ by deflation of a chamber or of a bullseye in $\mathbb{C}_{D+E}$. All chambers of $\mathbb{C}_{D+E}$ outside of H are the same in $\mathbb{C}_{D+E}$ and $\mathbb{C}_{ED}$ by the same proof as that of Lemma 12.3. Thus $\mathbb{C}_{ED} \dashrightarrow \mathbb{C}_{D+E}$ as required.

Case 2: $\mathbb{C}_{DE} \neq \mathbb{C}_{D+E}$

The proof is analogous to that of Case 1, but a bit more complicated. By Lemma 12.3 there is a handlebody H in S^3 whose boundary is in $F_{\mathcal{D}\cup E}$ and which is not coherent in $\mathbb{C}_{DE}$ and $\mathbb{C}_{D+E}$. Then according to Lemma 12.11 E is not parallel to a disk in $\mathcal{D}$ and ∂E is separating in $F - \partial\mathcal{D}$. Then by Lemma 12.10 every handlebody in S^3 whose boundary is in $F_{\mathcal{D}\cup E}$ is coherent for $\mathbb{C}_{ED}$ and $\mathbb{C}_{D+E}$ so by Lemma 12.3 $\mathbb{C}_{ED} = \mathbb{C}_{D+E}$. Thus what is left to show in this case is that $\mathbb{C}_{DE} \dashrightarrow \mathbb{C}_{D+E}$.

According to Lemma 12.11 E leaves exactly one scar on ∂H. If it is an internal scar, then per Claim 5 in the proof of Lemma 12.11, E lies in the interior of a chamber C_x of $\mathbb{C}_x$ dividing C_x into a ball in int(H) and a chamber R_x adjacent to and outside of H, as in Figure 38. Unless R_x is also a handlebody, the choice of preferred alignment of E has no effect on the flagging of chambers outside of H. In particular, there can be no handlebody chamber incoherent for $\mathbb{C}_{DE}$ and $\mathbb{C}_{D+E}$ that is disjoint from H. In this case, the argument concludes as for Case 1, using Lemma 12.11 instead of Lemma 12.10. On the other hand, if R_x is a handlebody (and so, as noted in Lemma 12.11 Claim 5, Subcase 1a, the only chamber of $\mathbb{C}_{x+y}$ that lies outside of H) then the choice of alignment of E in C_x does affect the flagging of R_x. The argument in this case is much like the argument in the case that E leaves an external scar on H, so we turn to that case, noting that the argument applies here as well.

If E leaves an external scar on H then it is possible that there is a disjoint handlebody H' in S^3 just like H: $\partial H'$ is in $F_{\mathcal{D}\cup E}$, H' is not coherent for $\mathbb{C}_{DE}$ and $\mathbb{C}_{D+E}$, and E also leaves an external scar on $\partial H'$. Then, just as is true for H, H' lies in a handlebody chamber in $\mathbb{C}_{DE}$. Thus there is a handlebody chamber $C_x \in \mathbb{C}_{\mathcal{D}}$ containing E, with E dividing C_x into H and H'.

We now reprise part of the argument of Claim 4a in Lemma 12.10: C_x must be an occupied handlebody, for if C_x is empty then H and H' would both be empty in $\mathbb{C}_{D+E}$ and $\mathbb{C}_{DE}$, contradicting the assumption that H is not coherent. Then, per Definition 7.2, C_x is the parent of two siblings in $\mathbb{C}_{DE}$: one in which H is occupied and H' is empty, the other in which H' is occupied and H is empty.

Since C_x is occupied, $F \cap \text{int}(C_x)$ has non-disk components by Corollary 7.3. There are three cases to consider, depending on where these non-disk components lie:

- If some non-disk components lie in H and some in H', then H and H' are both occupied in $\mathbb{C}_{D+E}$. Since H is not coherent, it must then be empty in $\mathbb{C}_{DE}$. In this case the sibling for C_{DE} is the one in which H is empty and H' is occupied, that is the one that results from an alignment for E which deflates H into H', leaving H' occupied in $\mathbb{C}_{DE}$ as well. Thus $\mathbb{C}_{DE} \dashrightarrow \mathbb{C}_{D+E}$, an allowed outcome.

- If each component of $F \cap \text{int}(H)$ is a disk, then H is empty in $\mathbb{C}_{D+E}$. Since H is not coherent it is then occupied in $\mathbb{C}_{DE}$. But in this case the sibling C'_{DE} of C_{DE} is one in which H is empty and H' is occupied. Then in this sibling $\mathbb{C}'_{DE} = \mathbb{C}_{D+E}$, an allowed outcome.

- If each component of $F \cap \text{int}(H')$ is a disk then H' is empty and H is occupied in $\mathbb{C}_{D+E}$. Since H is not coherent, it is empty in $\mathbb{C}_{DE}$. In this case the sibling C'_{DE} of C_{DE} has H occupied and H' empty, just as is true for $\mathbb{C}_{D+E}$. Thus $\mathbb{C}'_{DE} = \mathbb{C}_{D+E}$, an allowed outcome.

□

Corollary 12.13. *Suppose $\mathbb{C}$ is a flagged chamber complex supporting $S^3 = A \cup_T B$, $\mathbb{C}$ does not certify and $\mathcal{D} \cup E$ is a disk set in $\mathbb{C}$, where E is a single disk. In the notation of Proposition 12.12, one of these is true (for some choice of sibling, if siblings exist):*

- $\mathbb{C}_{DE} = \mathbb{C}_{ED}$,
- $\mathbb{C}_{DE} \dashrightarrow \mathbb{C}_{ED}$ *or*
- $\mathbb{C}_{DE} \dashleftarrow \mathbb{C}_{ED}$

We denote the three possibilities above by $\mathbb{C}_{DE} \leftrightarrow \mathbb{C}_{ED}$. Then one way of expressing the conclusion of Corollary 12.13 is to say that this diagram commutes:

$$\begin{array}{ccc} \mathbb{C}_0 & \xrightarrow{\mathcal{D}} & \mathbb{C}_{\mathcal{D}} \\ & & \downarrow E \\ E \downarrow & & \mathbb{C}_{DE} \\ & & \updownarrow \\ \mathbb{C}_E & \xrightarrow{\mathcal{D}} & \mathbb{C}_{ED} \end{array}$$

13 Guiding disk sets and disk addition

Recall from Section 11 the notion of guiding disks (Definition 11.8) for flagged chamber decompositions. The example given there involved decomposition of two chamber complexes that are related by bullseye deflation. Here is another example: Suppose $\mathcal{D} \cup E$ is a disk set in a flagged chamber complex $\mathbb{C}$. Let $\mathbb{C}'$ be the result of the flagged chamber complex decomposition $\mathbb{C} \xrightarrow{E} \mathbb{C}'$ and let $\mathcal{D}'$ be the set of disks in $\mathcal{D}$ whose boundaries lie on $\mathbb{C}'$. The decomposing set $\mathcal{D}'$ is contained in $\mathcal{D}$, but may not be all of $\mathcal{D}$, because it does include those disks in $\mathcal{D}$ that lie on goneballs of the decomposition $\mathbb{C} \xrightarrow{E} \mathbb{C}'$. Nevertheless $\mathcal{D}$ remains as a guiding set of disks for the decomposition $\mathbb{C}' \xrightarrow{\mathcal{D}'} \mathbb{C}'_{\mathcal{D}'}$.

We have already encountered another example of guiding disk sets, in the context of guiding spheres in Section 9. Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 with defining surface $F = F(\mathbb{C})$, and $S \subset S^3$ is an embedded sphere transverse to F. Consider a sequence $\vec{\mathbb{C}}$ of flagged chamber complex decompositions guided by S, as given in Definition 9.4 and described after the proof of Lemma 9.3: The circles $F \cap S$ are partitioned into sets $\mathfrak{c}_i, 0 \leq i \leq n$ and to each $\mathfrak{c}_i$ a collection of disjoint disks $\overline{\mathcal{D}}_i \subset S$ is assigned. The associated sequence $\vec{\mathbb{C}}$ of flagged chamber complex decompositions guided by S has, as each disk set $\mathcal{D}_i$ those disks in $\overline{\mathcal{D}}_i$ that remain incident to $F(\mathbb{C}_i)$ after all goneballs in earlier decompositions have been removed. Thus each set of disks $\overline{\mathcal{D}}_i$ is a guiding set of disks for the decomposition $\mathbb{C}_i \xrightarrow{\mathcal{D}_i} \mathbb{C}_{i+1}$ not the disk set itself. Consistent with the notation of Definition 11.8 we could then write the sequence in Definition 9.4 as

$$\mathbb{C} = \mathbb{C}_0 \xrightarrow{\overline{\mathcal{D}}_0} \mathbb{C}_1 \xrightarrow{\overline{\mathcal{D}}_1} \mathbb{C}_2 \xrightarrow{\overline{\mathcal{D}}_2} \dots \xrightarrow{\overline{\mathcal{D}}_{k-1}} \mathbb{C}_k$$

Joining this example with the previous one, suppose E is a properly embedded disk in $\mathbb{C}$ that is disjoint from the guiding sphere S and, as above, let $\mathbb{C}'$ be the result of the flagged chamber complex decomposition $\mathbb{C} \xrightarrow{E} \mathbb{C}'$. Then it is natural to construct a sequence $\vec{\mathbb{C}}'$ of flagged chamber complex decompositions

$$\vec{\mathbb{C}}' : \quad \mathbb{C}' = \mathbb{C}'_0 \xrightarrow{\mathcal{D}'_0} \mathbb{C}'_1 \xrightarrow{\mathcal{D}'_1} \mathbb{C}'_2 \xrightarrow{\mathcal{D}'_2} \dots \xrightarrow{\mathcal{D}'_{n-1}} \mathbb{C}'_n \tag{13.1}$$

by setting each $\mathcal{D}'_i$ to be the collection of disks in $\mathcal{D}_i$ (and so in $\overline{\mathcal{D}}_i$) whose boundaries lie on $\mathbb{C}'_i$. Then, just as for $\vec{\mathbb{C}}$, each decomposition $\mathbb{C}'_i \xrightarrow{\mathcal{D}'_i} \mathbb{C}'_{i+1}$ has guiding disk set $\overline{\mathcal{D}}_i$. So the sequence could also be written

$$\vec{\mathbb{C}}' : \quad \mathbb{C}' = \mathbb{C}'_0 \xrightarrow{\overline{\mathcal{D}}_0} \mathbb{C}'_1 \xrightarrow{\overline{\mathcal{D}}_1} \mathbb{C}'_2 \xrightarrow{\overline{\mathcal{D}}_2} \dots \xrightarrow{\overline{\mathcal{D}}_{n-1}} \mathbb{C}'_n.$$

One way to describe the relation between the sequences above is that the sequence $\vec{\mathbb{C}}'$ is what the sequence $\vec{\mathbb{C}}$ would be if we had inserted decomposition with the disk E before beginning the sequence. Generalizing, we could decompose by E at a later stage of the decomposition sequence. That is, for each $0 \le k \le n$, define a flagged chamber complex decomposition sequence

$$\vec{\mathbb{C}}^k_E : \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \dots \xrightarrow{\mathcal{D}_{k-1}} \mathbb{C}_k \xrightarrow{E} \mathbb{C}^k_k \xrightarrow{\mathcal{D}^k_k} \mathbb{C}^k_{k+1} \xrightarrow{\mathcal{D}^k_{k+1}} \dots \xrightarrow{\mathcal{D}^k_{n-1}} \mathbb{C}^k_n \tag{13.2}$$

where the disks $\mathcal{D}^k_i \subset \overline{\mathcal{D}}_i, k \le i \le n$ are those disks whose boundaries lie on $F(\mathbb{C}^k_i)$. In particular, $\vec{\mathbb{C}}'$ of Sequence 13.1 is $\vec{\mathbb{C}}^0_E$ with the first decomposition $\mathbb{C} \xrightarrow{E} \mathbb{C}'$ dropped. As usual, a simplified way of writing each sequence is

$$\vec{\mathbb{C}}^k_E : \mathbb{C}_0 \xrightarrow{\overline{\mathcal{D}}_0} \mathbb{C}_1 \xrightarrow{\overline{\mathcal{D}}_1} \dots \xrightarrow{\overline{\mathcal{D}}_{k-1}} \mathbb{C}_k \xrightarrow{E} \mathbb{C}^k_k \xrightarrow{\overline{\mathcal{D}}_k} \mathbb{C}^k_{k+1} \xrightarrow{\overline{\mathcal{D}}_{k+1}} \dots \xrightarrow{\overline{\mathcal{D}}_{n-1}} \mathbb{C}^k_n$$

Lemma 13.1. *Suppose $\mathbb{C}$ and $\mathbb{C}'$ are flagged chamber complexes and their defining surfaces have the property that $F(\mathbb{C}') \subset F(\mathbb{C})$. Suppose $\mathbb{C} \xrightarrow{\mathcal{D}} \mathbb{C}_{\mathcal{D}}$ and $\mathbb{C}' \xrightarrow{\mathcal{D}'} \mathbb{C}'_{\mathcal{D}'}$ have a guiding disk set $\overline{\mathcal{D}}$ in common. Then $\mathcal{D}$ is a guiding disk set for $\mathcal{D}'$.*

Proof. Each $D' \in \mathcal{D}'$ is a disk in $\overline{\mathcal{D}}$ whose boundary lies on $F(\mathbb{C}')$. Since $F(\mathbb{C}') \subset F(\mathbb{C})$, $\partial D'$ also lies on $F(\mathbb{C})$, so $D' \in \mathcal{D}$. Thus D' is a disk in $\mathcal{D}$ whose boundary lies in $F(\mathbb{C}')$. Conversely, suppose $D \in \mathcal{D}$ has its boundary on $F(\mathbb{C}')$. Since $D \in \mathcal{D}$, $D \in \overline{\mathcal{D}}$ and, since $\overline{\mathcal{D}}$ is a guiding disk set for $\mathcal{D}'$ and $\partial D \subset F(\mathbb{C}')$, $D \in \mathcal{D}'$. Thus $\mathcal{D}'$ consists exactly of those disks in $\mathcal{D}$ whose boundaries lie on $F(\mathbb{C}')$, as required. □

Proposition 13.2. *Let $\mathbb{C}$ be a flagged chamber complex that supports $S^3 = A \cup_T B$ and*

$$\vec{\mathbb{C}} : \mathbb{C} = \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

be a flagged chamber complex decomposition sequence. Suppose

- *there is a guiding disk set $\overline{\mathcal{D}}_i$ for each decomposition $\mathbb{C}_i \xrightarrow{\mathcal{D}_i} \mathbb{C}_{i+1}$.*
- *E is a proper disk in a chamber of $\mathbb{C}$, and E is disjoint from all disks in all $\overline{\mathcal{D}}_i$*
- *$\vec{\mathbb{C}}'$ is the sequence defined in Formula 13.1 above, so each $\overline{\mathcal{D}}_i$ is a guiding disk set for $\mathbb{C}'_i \xrightarrow{\mathcal{D}'_i} \mathbb{C}'_{i+1}$*
- *for each $0 \le k \le n$ and decomposition sequence $\vec{\mathbb{C}}^k_E$ defined in 13.2 above each $\overline{\mathcal{D}}_i$ is also a guiding disk set for the corresponding decomposition in $\vec{\mathbb{C}}^k_E$.*

If $\mathbb{C}_n$, $\mathbb{C}'_n$ and all $\mathbb{C}^k_n, 0 \le k \le n$ certify, then the sequences $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ cocertify.

Proof. Case 1: $\mathbb{C}$ certifies.

Then the sequence $\vec{\mathbb{C}}$ and the sequence

$$\vec{\mathbb{C}}^0_E : \mathbb{C} \xrightarrow{E} \mathbb{C}' = \mathbb{C}'_0 \xrightarrow{\mathcal{D}'_0} \mathbb{C}'_1 \xrightarrow{\mathcal{D}'_1} \dots \xrightarrow{\mathcal{D}'_{n-1}} \mathbb{C}'_n$$

cocertify because the chamber complex $\mathbb{C}$ certifies for both. (The same argument shows that all sequences $\vec{\mathbb{C}}^k_E$ cocertify.) Similarly the sequences $\vec{\mathbb{C}}^0_E$ and $\vec{\mathbb{C}}'$ cocertify because, by hypothesis, $\mathbb{C}'_n$ certifies and, as noted above, the latter sequence is contained in the former sequence (by dropping $\mathbb{C} \xrightarrow{E} \mathbb{C}'$). Thus we have $\vec{\mathbb{C}} \sim \vec{\mathbb{C}}^0_E \sim \vec{\mathbb{C}}'$ as required.

Case 2: $\mathbb{C}$ does not certify, but $\mathbb{C}_1$ does.

Since $\mathbb{C}$ does not certify then per Corollary 12.13, with some choice of siblings if there are siblings, the following diagram commutes:

$$\begin{array}{ccccccccc}
\mathbb{C} & \xrightarrow{\mathcal{D}_0} & \mathbb{C}_1 & \xrightarrow{\mathcal{D}_1} & \dots & \xrightarrow{\mathcal{D}_{n-1}} & \mathbb{C}_n & = \vec{\mathbb{C}} \\
\Big\downarrow E & & \Big\downarrow E & & & & & \\
 & & \mathbb{C}^1_1 & \xrightarrow{\mathcal{D}^1_1} & \dots & \xrightarrow{\mathcal{D}^1_{n-1}} & \mathbb{C}^1_n & = \vec{\mathbb{C}}^1_E \\
 & & \updownarrow & & & & & \\
\mathbb{C}' & \xrightarrow{\mathcal{D}'_0} & \mathbb{C}'_1 & \xrightarrow{\mathcal{D}'_1} & \dots & \xrightarrow{\mathcal{D}'_{n-1}} & \mathbb{C}'_n & = \vec{\mathbb{C}}'
\end{array}$$

We henceforth ignore the issue of choosing siblings, since different choices of sibling give cocertifying sequences, via the certifying parent chamber. (See the proof of Proposition 9.5.)

We would like to apply Corollary 11.12 to the sequences $\vec{\mathbb{C}}'$ and $\vec{\mathbb{C}}^1_E$. The Corollary requires that if $\mathbb{C}'_1$, say, deflates to $\mathbb{C}^1_1$ (i. e. $\mathbb{C}^1_1 \dashrightarrow \mathbb{C}'_1$) that the disks in $\mathcal{D}^1_1$ are a subset of the disks in $\mathcal{D}'_1$, namely those that have their boundaries on the defining surface for $\mathbb{C}^1_1$. In our language, $\mathcal{D}'_1$ needs to be a guiding disk set for $\mathcal{D}^1_1$. Moreover, it is required that the same true for each $\mathcal{D}'_i$ and $\mathcal{D}^1_i$ throughout the maximal deflationary sequence. But since, by hypothesis, $\overline{\mathcal{D}}_i$ is a guiding set of disks for both $\mathcal{D}'_i$ and $\mathcal{D}^1_i$, this is true, via Lemma 13.1.

So we can apply Corollary 11.12: Since $\vec{\mathbb{C}}^1_E$ certifies by the hypothesis that $\mathbb{C}^1_n$ certifies, $\vec{\mathbb{C}}^1_E$ and $\vec{\mathbb{C}}'$ cocertify. A symmetric argument leads to the same conclusion in the case that $\mathbb{C}^1_1$ deflates to $\mathbb{C}'_1$ or, most simply, when $\mathbb{C}^1_1 = \mathbb{C}'_1$. Since, in this case, we are assuming that $\mathbb{C}_1$ certifies, it certifies for the sequences that contain it, namely $\vec{C}$, and $\vec{\mathbb{C}}^1_E$. So we have $\vec{\mathbb{C}} \sim \vec{\mathbb{C}}^1_E \sim \vec{\mathbb{C}}'$, as required.

We are left with the possibility that $\mathbb{C}_1$ does not certify. The argument just concluded suggests an inductive approach, which we now pursue:

Inductive Step: Suppose the sequence $\vec{\mathbb{C}}'$ and, for $0 \le k \le \ell$, all $\vec{\mathbb{C}}^k_E$ cocertify. (The proof of Cases 1 and 2 show that this is true for $\ell = 1$.) Then either $\vec{\mathbb{C}} \sim \vec{\mathbb{C}}'$ and we are done, or the same statement is true with ℓ replaced by $\ell + 1$.

If $\mathbb{C}_\ell$ certifies, then it certifies for sequences that contain it, including $\vec{\mathbb{C}}$, and $\vec{\mathbb{C}}^\ell_E$. Thus $\vec{\mathbb{C}} \sim \vec{\mathbb{C}}^\ell_E$. By inductive hypothesis, $\vec{\mathbb{C}}^\ell_E \sim \vec{\mathbb{C}}'$. Hence $\vec{\mathbb{C}} \sim \vec{\mathbb{C}}'$, as required.

If $\mathbb{C}_\ell$ does not certify, Corollary 12.13 says the following diagram commutes:

$$\begin{array}{ccccccccccccc}
\mathbb{C} & \xrightarrow{\mathcal{D}_0} & \mathbb{C}_1 & \xrightarrow{\mathcal{D}_1} & \ldots & \xrightarrow{\mathcal{D}_{\ell-1}} & \mathbb{C}_\ell & \xrightarrow{\mathcal{D}_\ell} & \mathbb{C}_{\ell+1} & \xrightarrow{\mathcal{D}_{\ell+1}} & \ldots & & = \vec{\mathbb{C}} \\
 & & & & & & \Big\downarrow E & & \Big\downarrow E & & & & \\
 & & & & & & & & \mathbb{C}^{\ell+1}_{\ell+1} & \xrightarrow{\mathcal{D}^{\ell+1}_{\ell+1}} & \ldots & & = \vec{\mathbb{C}}^{\ell+1}_E \\
 & & & & & & & & \Updownarrow & & & & \\
 & & & & & & \mathbb{C}^\ell_\ell & \xrightarrow{\mathcal{D}^\ell_\ell} & \mathbb{C}^\ell_{\ell+1} & \xrightarrow{\mathcal{D}^\ell_{\ell+1}} & \ldots & & = \vec{\mathbb{C}}^\ell_E
\end{array}$$

Applying Corollary 11.12 much as in Case 2, $\vec{\mathbb{C}}^\ell_E$ and $\vec{\mathbb{C}}^{\ell+1}_E$ cocertify. Hence $\vec{C}'$ and, for $0 \le k \le \ell+1$, all $\vec{\mathbb{C}}^k_E$ cocertify, completing the inductive step.

Continue to iterate the inductive step, showing that $\vec{\mathbb{C}}'$ and each $\vec{\mathbb{C}}^k_E$ cocertify, until a k for which $\mathbb{C}_k$ certifies is reached. We know that we will eventually reach such a k, since $\vec{\mathbb{C}}$ certifies. At that point, $\vec{\mathbb{C}}^k_E$ cocertifies with both $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$, so $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ cocertify, as required. □

Theorem 13.3. *Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 that is not tiny, and S is a balanced or almost balanced sphere transverse to $F(\mathbb{C})$. Let E be a disk disjoint from S properly embedded and essential in a chamber of $\mathbb{C}$ that is not an empty torus. Let $\mathbb{C}'$ be the flagged chamber complex obtained from $\mathbb{C}$ by decomposition along E. Suppose also that S is balanced or almost balanced for $\mathbb{C}'$. Then any sequence in $\overrightarrow{(\mathbb{C}, S)}$ and any sequence in $\overrightarrow{(\mathbb{C}', S)}$ cocertify.*

Note: the requirements that E be essential in its chamber and that C not be an empty torus will be removed later, see Corollary 15.11.

Proof. Let F, F' denote the defining surfaces $F(\mathbb{C}), F(\mathbb{C}')$ respectively. The circles $F \cap S$ and $F' \cap S$ determine decomposition sequences $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ as described in Section 9. Since E is essential and not the meridian of an empty torus, surgery on E creates no new goneballs, so in fact $F \cap S = F' \cap S$. Thus the guiding disk sets $\overline{\mathcal{D}}_i$ for the two decomposition sequences $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ are the same. Import the notation for $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ from the paragraph that precedes Proposition 13.2 to these decomposition sequences $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$.

We are given that the chamber complexes $\mathbb{C}$ and $\mathbb{C}'$, related by the decomposition $\mathbb{C} \xrightarrow{E} \mathbb{C}'$, are each balanced or almost balanced. As noted in the proof of Proposition 10.2 during the decomposition sequence the genus of the defining surface above S and below S does not change. Surgery on E may or may not lower by one the genus of the defining surface on the side of S in which E lies, but we are given that S is balanced or almost balanced for $\mathbb{C}'$ as well as for $\mathbb{C}$. It follows that for any $0 \le k \le m$ the chamber complex $\mathbb{C}^k_k$ defined by the decomposition $\mathbb{C}_k \xrightarrow{E} \mathbb{C}^k_k$ in $\vec{\mathbb{C}}^k_E$ is also balanced or almost balanced.

Then Propositions 10.2 and 10.3 imply that the last term in each of the sequences $\vec{\mathbb{C}}, \vec{\mathbb{C}}$ and each $\vec{\mathbb{C}}^k_E$ certifies. Proposition 13.2 then says that $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ cocertify, as claimed. □

14 Delaying a disk

In this section we consider what happens in a sequence of two flagged chamber complex decompositions if, when this is possible, a decomposition disk in the first decomposing disk set is transferred to the second decomposing disk set.

Suppose $\mathbb{C}_0$ is a flagged chamber complex supporting $S^3 = A \cup_T B$, and

$$\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2$$

is a sequence of flagged chamber complex decompositions. Suppose E is a properly embedded disk in a chamber of $\mathbb{C}_0$ and E is disjoint from both $\mathcal{D}_0$ and $\mathcal{D}_1$. There are a variety of decomposition sequences that can be defined in this situation. Here is the notation we will use for them:

- $\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{E} \mathbb{C}_{0E}$
- $\mathbb{C}_1 \xrightarrow{E} \mathbb{C}_E \xrightarrow{\mathcal{D}'_1} \mathbb{C}_{E1}$
- $\mathbb{C}_0 \xrightarrow{\mathcal{D}_0 \cup E} \mathbb{C}_{0+E}$
- $\mathbb{C}_1 \xrightarrow{E \cup \mathcal{D}_1} \mathbb{C}_{E+1}$

In the second of these sequences $\mathcal{D}'_1$ has the usual meaning: those disks in $\mathcal{D}_1$ whose boundaries lie on the defining surface of $\mathbb{C}_E$ (and so do not lie on the boundary of goneballs after decomposition by E). Note that if ∂E lies on the boundary of a goneball in $\hat{\mathbb{C}}_1$, then it does not lie on the defining surface for $\mathbb{C}_1$, so decomposing on a disk set containing E is the same as decomposing on the disk set without E. So in this situation $\mathbb{C}_{E1} = \mathbb{C}_{E+1} = \mathbb{C}_2$.

Lemma 14.1. *Suppose neither $\mathbb{C}_0$ nor $\mathbb{C}_1$ certify. Then, up to a choice of siblings, if siblings exist, either*

1. $\mathbb{C}_{0E} = \mathbb{C}_{0+E}$ *and* $\mathbb{C}_{E1} = \mathbb{C}_{E+1}$ *or*
2. $\mathbb{C}_{0E} \dashrightarrow \mathbb{C}_{0+E}$ *and* $\mathbb{C}_{E1} = \mathbb{C}_{E+1}$ *or*
3. $\mathbb{C}_{0E} = \mathbb{C}_{0+E}$ *and* $\mathbb{C}_{E1} \dashrightarrow \mathbb{C}_{E+1}$.

Proof. Let F_0, F_1 be the defining surfaces for respectively $\mathbb{C}_0, \mathbb{C}_1$. Suppose first that ∂E is separating in a component of $F_0 - \partial\mathcal{D}_0$. Then it is separating in a component of F_1, because F_1 is obtained from $F_0 - \partial\mathcal{D}_1$ simply by adding (scar) disks to $\partial\mathcal{D}_1$ and perhaps deleting some spheres. Then ∂E (if it is not on the boundary of a goneball) is also separating in a component of $F_1 - \partial D_1$. It follows from the last sentence of Proposition 12.12 that $\mathbb{C}_{E1} = \mathbb{C}_{E+1}$ and, further applying Proposition 12.12, that outcome (1) or (2) of Lemma 14.1 occurs.

On the other hand, if ∂E is non-separating in $F_0 - \partial\mathcal{D}_0$ then by Lemma 12.11 $\mathbb{C}_{0E} = \mathbb{C}_{0+E}$. Then by Proposition 12.12, outcome (1) or (3) occurs. □

We now put this in the context of decomposition sequences. Suppose $\overline{\mathcal{D}}_1, ..., \overline{\mathcal{D}}_n$ are a sequence of guiding disk sets in S^3 so that $\overline{\mathcal{D}}_1$ determines the decomposition $\mathbb{C}_1 \xrightarrow{E \cup \mathcal{D}_1} \mathbb{C}_{E+1}$ above. Then the guiding disk sets $\overline{\mathcal{D}}_i$ determine extensions of the three sequences that we label as follows:

$$\vec{\mathbb{C}}^{0+E} : \mathbb{C}_0 \xrightarrow{\mathcal{D}_0 \cup E} \mathbb{C}_{0+E} \xrightarrow{\overline{\mathcal{D}}_1} \mathbb{C}_2^{0+E} \xrightarrow{\overline{\mathcal{D}}_2} \mathbb{C}_3^{0+E} \xrightarrow{\overline{\mathcal{D}}_3} ... \xrightarrow{\overline{\mathcal{D}}_{n-1}} \mathbb{C}_n^{0+E}$$

$$\vec{\mathbb{C}}^{E1} : \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{E} \mathbb{C}_{0E} \xrightarrow{\overline{\mathcal{D}}_1} \mathbb{C}_2^{0E} \xrightarrow{\overline{\mathcal{D}}_2} \mathbb{C}_3^{0E} \xrightarrow{\overline{\mathcal{D}}_3} ... \xrightarrow{\overline{\mathcal{D}}_{n-1}} \mathbb{C}_n^{0E}$$

$$\vec{\mathbb{C}}^{E+1} : \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{E \cup \overline{\mathcal{D}}_1} \mathbb{C}_{E+1} \xrightarrow{\overline{\mathcal{D}}_2} \mathbb{C}_3^{E+1} \xrightarrow{\overline{\mathcal{D}}_3} ... \xrightarrow{\overline{\mathcal{D}}_{n-1}} \mathbb{C}_n^{E+1}$$

Lemma 14.2. *Suppose all three decomposition sequences certify. Then all three decomposition sequences cocertify.*

Proof. If $\mathbb{C}_0$ certifies, this is obvious: $\mathbb{C}_0$ is a term in all three sequences, so if it certifies then all three sequences cocertify.

If $\mathbb{C}_1$ certifies, then $\vec{\mathbb{C}}^{E+1}$ and $\vec{\mathbb{C}}^{E1}$ cocertify for the same reason: $\mathbb{C}_1$ is in each sequence. If further $\mathbb{C}_0$ does not certify, we can apply Proposition 12.12 to conclude that either $\mathbb{C}_{0E} = \mathbb{C}_{0+E}$ or $\mathbb{C}_{0E} \dashrightarrow \mathbb{C}_{0+E}$. In the former case the sequences $\vec{\mathbb{C}}^{E1}$ and $\vec{\mathbb{C}}^{0+E}$ also cocertify because, starting with $\mathbb{C}_{0E} = \mathbb{C}_{0+E}$, they are the same sequence and before that in $\vec{\mathbb{C}}^{0+E}$ is only $\mathbb{C}_0$, which does not certify. In the latter case, when $\mathbb{C}_{0E} \dashrightarrow \mathbb{C}_{0+E}$, they cocertify by Corollary 11.12.

Suppose neither $\mathbb{C}_0$ nor $\mathbb{C}_1$ certify. Then following Lemma 14.1 the other decompositions can be arranged towards the left in exactly one of the following three diagrams, which differ only in which of two possible deflations (shown as vertical dashed arrows) are in fact equalities:

$$\begin{array}{ccccccccccl}
\mathbb{C}_0 & \xrightarrow{\mathcal{D}_0} & \mathbb{C}_1 & \xrightarrow{\overline{\mathcal{D}}_1} & \mathbb{C}_{E+1} & \xrightarrow{\overline{\mathcal{D}}_2} & \mathbb{C}_3^{E+1} & \xrightarrow{\overline{\mathcal{D}}_3} & ... & & = \vec{\mathbb{C}}^{E+1} \\
 & \searrow & \downarrow E & & \| & & & & & & \\
 & \mathcal{D}_0 \cup E & \mathbb{C}_{0E} & \xrightarrow{\overline{\mathcal{D}}_1} & \mathbb{C}_{E1} & \xrightarrow{\overline{\mathcal{D}}_2} & ... & & & & = \vec{\mathbb{C}}^{E1} \\
 & & \| & \searrow & & & & & & & \\
 & & \mathbb{C}_{0+E} & \xrightarrow{\overline{\mathcal{D}}_1} & \mathbb{C}_1^{0+E} & \xrightarrow{\overline{\mathcal{D}}_2} & ... & & & & = \vec{\mathbb{C}}^{0+E}
\end{array} \qquad (14.1.1)$$

$$\begin{array}{ccccccccccl}
\mathbb{C}_0 & \xrightarrow{\mathcal{D}_0} & \mathbb{C}_1 & \xrightarrow{\overline{\mathcal{D}}_1} & \mathbb{C}_{E+1} & \xrightarrow{\overline{\mathcal{D}}_2} & \mathbb{C}_3^{E+1} & \xrightarrow{\overline{\mathcal{D}}_3} & ... & & = \vec{\mathbb{C}}^{E+1} \\
 & \searrow & \downarrow E & & \| & & & & & & \\
 & \mathcal{D}_0 \cup E & \mathbb{C}_{0E} & \xrightarrow{\overline{\mathcal{D}}_1} & \mathbb{C}_{E1} & \xrightarrow{\overline{\mathcal{D}}_2} & ... & & & & = \vec{\mathbb{C}}^{E1} \\
 & & \downarrow & \searrow & & & & & & & \\
 & & \mathbb{C}_{0+E} & \xrightarrow{\overline{\mathcal{D}}_1} & \mathbb{C}_1^{0+E} & \xrightarrow{\overline{\mathcal{D}}_2} & ... & & & & = \vec{\mathbb{C}}^{0+E}
\end{array} \qquad (14.1.1)$$

$$\begin{array}{ccccccccc}
\mathbb{C}_0 & \xrightarrow{\mathcal{D}_0} & \mathbb{C}_1 & \xrightarrow{\overline{\mathcal{D}}_1} & \mathbb{C}_{E+1} & \xrightarrow{\overline{\mathcal{D}}_2} & \mathbb{C}_3^{E+1} \xrightarrow{\overline{\mathcal{D}}_3} \ldots & = \vec{\mathbb{C}}^{E+1} \\
 & & \downarrow E & & \uparrow & & & \\
 & \mathcal{D}_0 \cup E \searrow & \mathbb{C}_{0E} & \xrightarrow{\overline{\mathcal{D}}_1} & \mathbb{C}_{E1} & \xrightarrow{\overline{\mathcal{D}}_2} & \ldots & = \vec{\mathbb{C}}^{E1} \\
 & & \| & & & & & \\
 & & \mathbb{C}_{0+E} & \xrightarrow{\overline{\mathcal{D}}_1} & \mathbb{C}_1^{0+E} & \xrightarrow{\overline{\mathcal{D}}_2} & \ldots & = \vec{\mathbb{C}}^{0+E}
\end{array} \tag{14.1.1}$$

In the first diagram it is obvious that all three sequences cocertify: because of the equalities, all three sequences are in fact the same sequence after the terms $\mathbb{C}_{E+1}, \mathbb{C}_{E1}$ and $\mathbb{C}_1^{0+E}$. Similarly the sequences $\vec{\mathbb{C}}^{E+1}$ and $\vec{\mathbb{C}}^{E1}$ become the same sequence after those terms in the second diagram and so cocertify; the same is true (even one step earlier) for $\vec{\mathbb{C}}^{E1}$ and $\vec{\mathbb{C}}^{0+E}$ in the third diagram.

Finally, by Corollary 11.12, both $\vec{\mathbb{C}}^{E1}$ and $\vec{\mathbb{C}}^{0+E}$ cocertify in the second diagram and $\vec{\mathbb{C}}^{E+1}$ and $\vec{\mathbb{C}}^{E1}$ cocertify in the third. □

15 Ghost circles and timing of disks

Let $\mathbb{C}$ be a flagged chamber complex in S^3 supporting $A \cup_T B$, and $S \subset S^3$ be a sphere transverse to $\mathbb{C}$. It was shown in Section 9 how S defines a flagged chamber complex decomposition sequence for $\mathbb{C}$. We briefly recall the process, filling in further description, assisted by augmented notation: For F the defining surface for $\mathbb{C}$, the innermost in S circles $\mathfrak{c}_0$ of $F \cap S$ bound disks in S which we regard as a disk set $\mathcal{D}_0$ for $\mathbb{C}$. The set of circles $\bar{\mathfrak{c}}_1$ in $F \cap S$ (denoted simply $\mathfrak{c}_1$ in Section 9) consists of those circles which bound disks in S whose interiors may contain circles in $\mathfrak{c}_0$ but otherwise are disjoint from F. Call this set of disks $\overline{\mathcal{D}}_1$. Proceeding in this manner partition the entire collection of circles $F \cap S$ into sets $\bar{\mathfrak{c}}_0 = \mathfrak{c}_0, \bar{\mathfrak{c}}_1 ..., \bar{\mathfrak{c}}_n$, where $n = \lfloor \frac{\text{diam}(Y)+1}{2} \rfloor$, so that each circle in $\bar{\mathfrak{c}}_k$ bounds a disk D in S, and any circle of $F \cap S$ that lies in the interior of D lies in some $\bar{\mathfrak{c}}_i, i < k$. Denote the collection of such disks in S bounded by the circles in $\bar{\mathfrak{c}}_k$ by $\overline{\mathcal{D}}_k$.

The decomposition sequence

$$\vec{\mathbb{C}}: \quad \mathbb{C} = \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \ldots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

is then iteratively defined, the first step being the decomposition $\mathbb{C} = \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1$. After this decomposition two important things happen to the sets of circles $\bar{\mathfrak{c}}_i, 1 \leq i \leq n$:

- The circles $\mathfrak{c}_0$ have been surgered away, so the interior of each disk in $\overline{\mathcal{D}}_1$ is disjoint from the defining surface F_1 for $\mathbb{C}_1$.
- Some of the circles in $\bar{\mathfrak{c}}_1$ may lie on the goneballs of the decomposition, and so no longer lie in $S \cap F_1$.

Then let $\mathcal{D}_1 \subset \overline{\mathcal{D}}_1$ be the subset of disks that are left, i. e. the subset of disks whose boundaries lie on F_1.

Since, after the decomposition by $\mathcal{D}_0$, the interiors of the disks $\mathcal{D}_1$ are disjoint from F_1, they are a disk set in $\mathbb{C}_1$. Use these disks to define the decomposition $\mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2$. Continue in this manner to get the entire decomposition sequence, noting that each $\mathcal{D}_i$ in the construction consists of those disks in $\overline{\mathcal{D}}_i$ whose boundaries lie on the defining surface $F_i = F(\mathbb{C}_i)$, and not on the boundaries of goneballs from earlier decompositions in the sequence. In other words:

Lemma 15.1. *For each $1 \leq i \leq n-1$, the disk set $\overline{\mathcal{D}}_i$ is a guiding disk set for the decomposition $\mathbb{C}_i \xrightarrow{\mathcal{D}_i} \mathbb{C}_{i+1}$.* □

Definition 15.2. *For each $1 \leq i \leq n-1$, the circles $\bar{\mathfrak{c}}_i - \mathfrak{c}_i$ (that is $\partial\overline{\mathcal{D}}_i - \partial\mathcal{D}_i$) are called the* ghost circles *of such a decomposition sequence in $\overrightarrow{(\mathbb{C}, S)}$. The circles $\mathfrak{c}_i = F_i \cap S$, that is the circles that bound the disks $\mathcal{D}_i \subset S$, will be called the F-circles of the decomposition sequence.*

The disk set $\mathcal{D}_i$ may well contain ghost circles in its interior; these indicate where $\mathcal{D}_i$ intersects spheres bounding goneballs in earlier decompositions, and so can be ignored. Nevertheless, the ghost circles do play a role in the construction of the decomposition sequence, namely they are part of what determines, at the beginning of the decomposition sequence, before they are ghost circles, where in the sequence of decompositions a particular sub-disk of S appears.

Following up on this idea, it is reasonable to consider how introducing ghost circles in S at the beginning of the decomposition process (when the sets of circles $\mathfrak{c}_i$ for the decomposition sequence are being defined) might affect the decomposition sequence. For example, if a ghost circle c' were to be added, at the beginning of the decomposition sequence, parallel to a circle $c \in \mathfrak{c}_{i-1}$ and just inside the disk $D \in \mathcal{D}_{i-1}$ that c bounds, one effect would be to redefine D as a disk in $\mathcal{D}_i$, because c is no longer i-th innermost in $F \cap S$ but $(i+1)$-st innermost in the collection of circles $\mathfrak{c} \cup c'$. There could well be a similar delay to the use of other disks in the following $\mathcal{D}_k, i < k < n$ in a manner that may be a bit difficult to describe - see Figure 40. Our next goal is to understand to some extent how adding or removing ghost circles (and thereby changing the location of the decomposing disks in the sequence of disk decompositions) affects the certification of a decomposition sequence. The news is good: eventually we will show that adding or removing ghost circles may alter the decomposition sequence, as we have seen, but the new decomposition sequence cocertifies with the old.

To this end we generalize the discussion of Section 9 to spheres in S transverse to the defining surface F of a flagged chamber complex $\mathbb{C}$ in S^3 and consider a finite collection of circles $\bar{\mathfrak{c}}$ in S containing all of the circles $\mathfrak{c} = F \cap S$ (which we have distinguished as *F-circles*) but augmented by other possible circles $\bar{\mathfrak{c}} - F$ in S called ghost circles. The collection of circles $\bar{\mathfrak{c}}$ determines a sequence of flagged chamber complex decompositions guided by disk sets $\overline{\mathcal{D}}_i$, where each circle in $\bar{\mathfrak{c}}$ is the boundary of a disk in some $\overline{\mathcal{D}}_i$, as described above and in Section 9. (This setting naturally arises in the middle of a flagged chamber complex decomposition sequence, as we have seen.)

As in Section 9, let Y be the tree associated to the circles $\bar{\mathfrak{c}} \subset S$, and, as in Definition 9.4, let

$$\vec{\mathbb{C}}: \quad \mathbb{C} = \mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \mathbb{C}_2 \xrightarrow{\mathcal{D}_2} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

be the complete sequence of flagged chamber complex decompositions guided by the sphere S, where $n = \lfloor \frac{\text{diam}(Y)+1}{2} \rfloor$. It's perfectly possible that for some values of i, $\bar{\mathfrak{c}}_i$ consists entirely of ghost circles, so

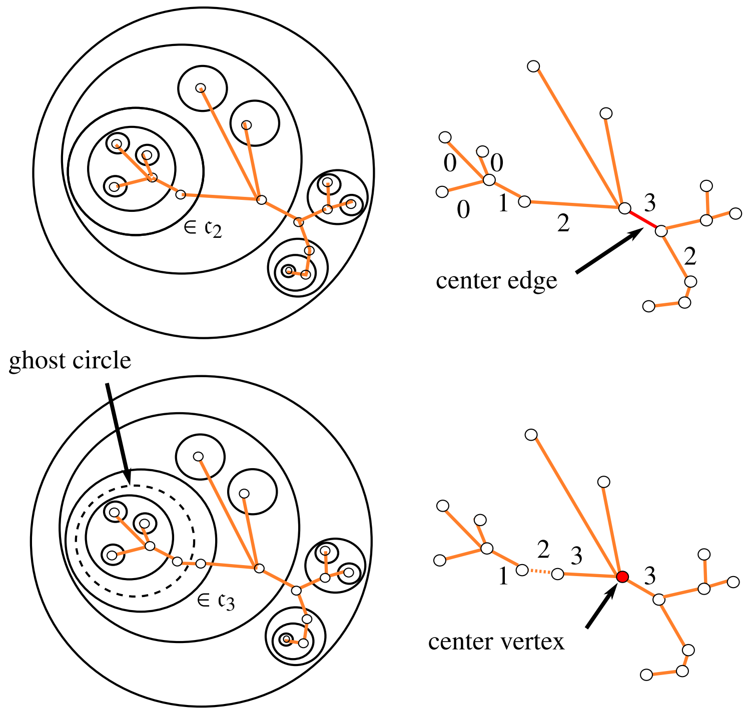

Figure 40: Adding a ghost circle, $i = 3$

none of the disks $\overline{\mathcal{D}}_i$ have their boundaries on F and so $\mathcal{D}_i = \emptyset$. In this case the decomposition $\mathbb{C}_i \xrightarrow{\mathcal{D}_i} \mathbb{C}_{i+1}$ is just the identity: $\mathbb{C}_i = \mathbb{C}_{i+1}$.

Recalling Definition 9.1, each edge e in Y (corresponding to a circle $c \in \bar{\mathfrak{c}}$) is assigned a value

$$0 \leq \rho_Y(e) \leq \lfloor \frac{\text{diam}(Y) - 1}{2} \rfloor.$$

For each vertex $v \in Y$ we can similarly assign $\rho_Y(v) = \max\{\rho_Y(e) \mid e \text{ an edge incident to } v\}$.

Here are some elementary observations. For e an edge in Y, let $m^e_\pm$ be the numbers defined for e in Definition 9.1.

Lemma 15.3. *There is an edge f in Y such that $m^f_+ = m^f_-$ if and only if* diam(Y) *is odd. In this case $\rho_Y(f) = \frac{\text{diam}(Y)-1}{2}$ and f is unique: it is the mid-edge of each path in Y whose length is* diam(Y). *Call f the* center edge *of the tree Y. See Figure 40.*

Proof. Let e be any edge in Y, with endpoints v_+ and v_- belonging to the trees Y_+ and Y_- respectively in $Y - e$, as discussed in Definition 9.1. If there is a leaf v_ℓ of Y that belongs to Y_+ and is a distance $m > \frac{\text{diam}(Y)-1}{2}$ from v_+, then v_ℓ is a distance $m + 1 > \frac{\text{diam}(Y)+1}{2}$ from v_-. It follows from the definition of diameter that no leaf of Y that belongs to Y_- can be a distance more than

$$\text{diam}(Y) - m - 1 = \frac{\text{diam}(Y) - 1}{2} + \frac{\text{diam}(Y) + 1}{2} - m - 1 < \frac{\text{diam}(Y) - 1}{2}$$

from v_-. Hence $m_- \leq \frac{\text{diam}(Y)-1}{2} < m \leq m_+$ and e is not as f is described in the Lemma.

Thus an edge f as described in the Lemma has the property that any leaf of Y in $Y_\pm$ has distance at most $\frac{\text{diam}(Y)-1}{2}$ from respectively the vertices $v_\pm$. This implies that any path in $Y_\pm$ can have length at most $\text{diam}(Y)-1$, so in particular no path γ in Y of length $\text{diam}(Y)$ can lie entirely in either of $Y_\pm$. This implies that f must be in γ and in fact be the mid-edge of γ, defining f uniquely and establishing that $\rho_Y(f) = m^f_\pm = \frac{\text{diam}(Y)-1}{2}$.

Conversely, suppose $\text{diam}(Y)$ is odd and f is the mid-edge of a path γ in Y of length $\text{diam}(Y)$. If any leaf in Y lying in Y_+ (resp Y_-) is a distance greater than $\frac{\text{diam}(Y)-1}{2}$ from v_+ (resp v_-) then conjoining a path from that leaf to v_+ with the part of $\gamma - v_+$ containing v_- would give a path of length greater than $\text{diam}(Y)$, a contradiction. Hence both $m^f_\pm = \frac{\text{diam}(Y)-1}{2}$ as required. □

Quite similarly we have:

Lemma 15.4. *There is a vertex w in Y that is incident to more than one edge e with $\rho_Y(e) = \rho_Y(w)$ if and only if* $\text{diam}(Y)$ *is even. In this case w is unique and $\rho_Y(w) = \frac{\text{diam}(Y)}{2} - 1$. Call w the* center vertex *of the tree Y. See Figure 40.*

Proof. Consider any vertex v in Y and suppose there is a path γ in Y from v to a leaf of Y and that the length of γ is $m > \frac{\text{diam}(Y)}{2}$. Let e be the edge of γ that is adjacent to v and observe that, by definition of diameter, any path from v to a leaf of Y that is longer than $\text{diam}(Y) - m < m$ must pass through e. Definition 9.1 then implies that $\rho_Y(e)$ is the length of the longest path in Y *not passing through e* from v to a leaf of Y, and any edge e' incident to v other than e has $\rho_Y(e') \leq \rho_Y(e) - 1$. In particular, e is the only edge incident to v with $\rho_Y(e) = \rho_Y(v)$, so e is not an edge as described in the lemma.

It follows that if w is as described in the lemma, then the length of any path from w to a leaf in Y has length $\leq \frac{\text{diam}(Y)}{2}$. So, by definition of diameter, any path in Y whose length is $\text{diam}(Y)$ must pass through w and have its length bisected by w. Thus $\text{diam}(Y)$ is even and the distance from w to a most distant leaf will be $\frac{\text{diam}(Y)}{2}$. In particular, any edge e incident to w will find the vertex at its other end a distance of at most $\frac{\text{diam}(Y)}{2} - 1$ from a leaf. Thus $\rho_Y(e) \leq \frac{\text{diam}(Y)}{2} - 1$. Since this is true for all edges incident to w, $\rho_Y(w) \leq \frac{\text{diam}(Y)}{2} - 1$. Furthermore, since any path in Y whose length is $\text{diam}(Y)$ must pass through w, there are at least two edges $e_\pm$ incident to w lying on such a path, and then $\rho_Y(e_\pm) = \frac{\text{diam}(Y)}{2} - 1$. It follows that for both edges $\rho_Y(e_\pm) = \frac{\text{diam}(Y)}{2} - 1 = \rho_Y(w)$ as required.

Conversely, if $\text{diam}(Y)$ is even, γ is a path in Y of length $\text{diam}(Y)$, and w is the vertex that bisects γ, the same argument shows that for e either of the two edges in γ incident to w, $\rho_Y(e) = \rho_Y(w)$. □

Lemma 15.5. *Suppose $P \subset S - \bar{\mathfrak{c}}$ is the planar surface corresponding to a vertex $v \in Y$. If $P \subset D \in \overline{\mathcal{D}}_k$, then $\rho_Y(v) \leq k$.*

Proof. For $D \in \overline{\mathcal{D}}_k$, $\partial D \in \bar{\mathfrak{c}}_k$ and from Lemma 9.2 each circle in $\bar{\mathfrak{c}} \cap \text{int}(D)$ lies in some $\bar{\mathfrak{c}}_j$, $j < k$. Hence if $P \subset D$, each circle $c \in \partial P$ lies in some $\bar{\mathfrak{c}}_j$, $j \leq k$. Corresponding to this, we have that each edge e in Y that is incident to v has $\rho_Y(e) \leq k$. Hence by definition $\rho_Y(v) \leq k$ as required. □

Lemma 15.6. *Suppose a ghost circle is added to S in a planar surface $P \subset S - \bar{\mathfrak{c}}$ corresponding to a vertex $v \in Y$ and*

$$\vec{\mathbb{C}}' : \quad \mathbb{C} = \mathbb{C}_0 \xrightarrow{\mathcal{D}'_0} \mathbb{C}'_1 \xrightarrow{\mathcal{D}'_1} \mathbb{C}'_2 \xrightarrow{\mathcal{D}'_2} \ldots$$

is the resulting sequence of flagged chamber complex decompositions. Suppose $\rho_Y(v) \geq 1$. *Then for each* $1 \leq i \leq \rho_Y(v)$, $\mathcal{D}'_{i-1} = \mathcal{D}_{i-1}$ *and* $\mathbb{C}'_i = \mathbb{C}_i$.

Proof. If $i \leq \rho_Y(v)$ then $\rho_Y(v) \nleq i-1$. It follows from Lemma 15.5 that any disk $D \in \overline{\mathcal{D}}_{i-1}$ does not contain P and so is unaffected by the addition of the ghost circle. Thus, for each $1 \leq i \leq \rho_Y(v)$, $\overline{\mathcal{D}}'_{i-1} = \overline{\mathcal{D}}_{i-1}$. Now argue iteratively: Since $C'_0 = C_0$ and $\overline{\mathcal{D}}'_0 = \overline{\mathcal{D}}_0$ it follows that $\mathcal{D}'_0 = \mathcal{D}_0$. Hence $\mathbb{C}'_1 = \mathbb{C}_1$, so if $\overline{\mathcal{D}}'_1 = \overline{\mathcal{D}}_1$ then $\mathcal{D}'_1 = \mathcal{D}_1$. The argument continues iteratively until $i = \rho_Y(v)$: For each $1 \leq i \leq \rho_Y(v)$, $\mathbb{C}'_{i-1} = \mathbb{C}_{i-1}$ and $\overline{\mathcal{D}}'_{i-1} = \overline{\mathcal{D}}_{i-1}$ imply $\mathcal{D}'_{i-1} = \mathcal{D}_{i-1}$, so $\mathbb{C}'_i = \mathbb{C}_i$, as required. □

Corollary 15.7. *Suppose* diam(Y) *is even and a ghost circle is added to* S *in the planar surface* $P \subset S - \bar{\mathfrak{c}}$ *corresponding to the center vertex* $v \in Y$. *Then* $\vec{\mathbb{C}}' = \vec{\mathbb{C}}$.

Proof. Since diam(Y) is even, the complete sequence $\vec{\mathbb{C}}$ has last term $\mathbb{C}_n$ with $n = \text{diam}(Y)/2 = \rho_Y(v)$. If $\rho_Y(v) = 0$ then Y is a star graph, each circle in $\bar{\mathfrak{c}}$ bounds a unique disk in $\overline{\mathcal{D}}_0$ and both $\vec{\mathbb{C}}'$ and $\vec{\mathbb{C}}$ are simply the sequence $\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1$. Indeed, after this decomposition the only circle remaining in S is the added ghost circle, so $S \cap F(\mathbb{C}_1) = \emptyset$ and there is no further decomposition in the sequence. If, on the other hand, $\rho_Y(v) \geq 1$ then by Lemma 15.6, for all $0 \leq i \leq n$, $\mathbb{C}'_i = \mathbb{C}_i$ and after these decompositions, the only circle that might remain is again the added ghost circle. □

Corollary 15.8. *Suppose* diam(Y) *is odd and a ghost circle is added to* S *in a planar surface* $P \subset S - \bar{\mathfrak{c}}$ *corresponding to a vertex at the end of the center edge* $e \in Y$. *Then if* $\vec{\mathbb{C}}$ *certifies,* $\vec{\mathbb{C}}'$ *and* $\vec{\mathbb{C}}$ *cocertify.*

Proof. One possibility is that diam$(Y) = 1$, that is Y is the single edge e. In this case $\bar{\mathfrak{c}}$ is a single circle c (corresponding to e) dividing S into two disks D_A and D_B, as in Lemma 10.4. Say the ghost circle is added to $D_A = P$. If c is also a ghost circle, then F is disjoint from S and there is no further decomposition. In particular, $\vec{\mathbb{C}}' = \vec{\mathbb{C}}$. If, on the other hand, c is an F-circle, so $c = F \cap S$ then it is shown in Lemma 10.4 that if $\vec{\mathbb{C}}$ certifies, the choice of whether to decompose $\mathbb{C}$ along D_A or D_B makes no difference: the two results cocertify. Adding a ghost circle to D_A has the sole effect of declaring that in $\mathbb{C}'$ the decomposition will be along D_B (since D_A is no longer a disk component of $S - \bar{\mathfrak{c}}$). Hence $\vec{\mathbb{C}}'$ and $\vec{\mathbb{C}}$ cocertify.

If, on the other hand, diam$(Y) \geq 3$ then, per Lemma 15.3 $\rho_Y(v) = \rho_Y(e) = \frac{\text{diam}(Y)-1}{2} \geq 1$. Then according to Lemma 15.6 the decomposition sequences $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ are identical through $\mathbb{C}'_{\frac{\text{diam}(Y)-1}{2}} = \mathbb{C}_{\frac{\text{diam}(Y)-1}{2}}$. But for $\mathbb{C}_{\frac{\text{diam}(Y)-1}{2}}$, $\bar{\mathfrak{c}}$ has shrunk to a single circle, and Y is a single edge. Thus we can apply the argument just given for this case and conclude that $\vec{\mathbb{C}}' \sim \vec{\mathbb{C}}$ as required. □

Proposition 15.9. *Suppose* $\mathbb{C}$ *is a flagged chamber complex in* S^3 *that is not tiny. Let* $S \subset S^3$ *be a sphere, transverse to the defining surface* $F = F(\mathbb{C})$, *that is either balanced or almost balanced. Let* $\bar{\mathfrak{c}}$ *be a collection of circles in* S *containing the set* $\mathfrak{c} = F \cap S$ *and let* $\bar{\mathfrak{c}}'$ *be the union of* $\bar{\mathfrak{c}}$ *and a disjoint (ghost) circle* c' *in* S. *Let* $\vec{\mathbb{C}}$ *and* $\vec{\mathbb{C}}'$ *be the sequences of flagged chamber complex decompositions determined by* $\bar{\mathfrak{c}}$ *and* $\bar{\mathfrak{c}}'$ *respectively. Then* $\vec{\mathbb{C}}$ *and* $\vec{\mathbb{C}}'$ *cocertify.*

Proof. Let Y be the tree associated to $\bar{\mathfrak{c}}$, $P \subset S - \bar{\mathfrak{c}}$ be the planar surface component to which the ghost circle c' is added, and v be the vertex in Y corresponding to P. The proof is by induction on the distance between v and the center vertex or edge of Y. Corollaries 15.7 and 15.8 show that the proposition is true when that distance is 0.

Following Lemma 15.3, as used in the proof of Corollary 15.7, the decomposition sequences are identical until the tree has been shrunk to the point that v is a leaf. So we may as well assume that v is a leaf of Y, so P is an innermost disk of $S - \overline{\mathfrak{c}}$. Let $\overline{\mathfrak{c}}''$ be the collection of circles $\overline{\mathfrak{c}} \cup c''$, where c'' is a circle parallel to ∂P but just *outside* P, that is in the component P'' of $S - \overline{\mathfrak{c}}$ that is adjacent to P. See Figure 41. Then the vertex v'' corresponding to P'' is closer to the center edge or vertex of Y so, by inductive assumption, the sequence of flagged chamber complex decompositions $\vec{\mathbb{C}}''$ determined by $\overline{\mathfrak{c}}''$ and $\vec{\mathbb{C}}$ cocertify. So all that remains is to show that $\vec{\mathbb{C}}''$ and $\vec{\mathbb{C}}'$ cocertify.

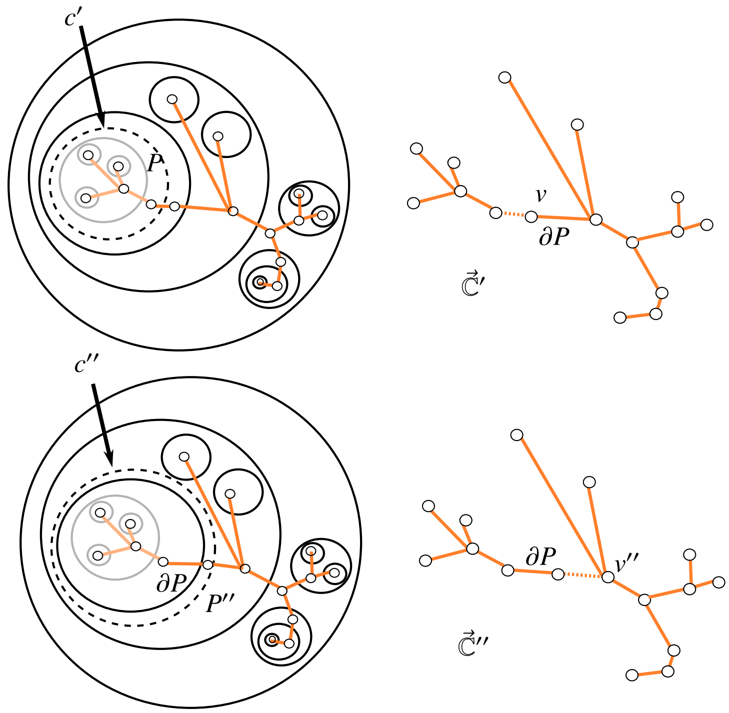

Figure 41: Moving a ghost circle, so $\partial P \in \mathfrak{c}'_3 \to \partial P \in \mathfrak{c}''_2$

If ∂P is a ghost circle, then $\overline{\mathfrak{c}}'' = \overline{\mathfrak{c}}'$ so in fact $\vec{\mathbb{C}}'' = \vec{\mathbb{C}}'$ and we are done. So assume that ∂P is an F-circle in $\overline{\mathfrak{c}}$ and note that the only difference between $\overline{\mathfrak{c}}''$ and $\overline{\mathfrak{c}}'$ is the side on which a ghost circle is added. This switch in sides has no effect on the tree and so no effect on the disks in the decomposition sequence $\vec{\mathbb{C}}'$ except this: In $\vec{\mathbb{C}}''$ the disk P is in $\mathcal{D}_0$, since P is an innermost disk, whereas in $\mathbb{C}'$ the disk P is in $\mathcal{D}_1$ because the disk bounded by ∂P contains the single added ghost circle c'.

Apply Lemma 14.2 to this situation, where here the disk P plays the role of E in that lemma, $\vec{\mathbb{C}}''$ here corresponds to $\vec{\mathbb{C}}^{0+E}$ there and $\vec{\mathbb{C}}'$ here corresponds to $\vec{\mathbb{C}}^{E+1}$ there. Since S is balanced or almost balanced, both sequences $\vec{\mathbb{C}}''$ and $\vec{\mathbb{C}}'$ certify, as does the sequence $\vec{\mathbb{C}}^{P1}$, defined for P as $\vec{\mathbb{C}}^{E1}$ is defined for E in the preamble to Lemma 14.2. The conclusion of Lemma 14.2 then says that $\vec{\mathbb{C}}''$ and $\vec{\mathbb{C}}'$ cocertify, so $\vec{\mathbb{C}} \sim \vec{\mathbb{C}}'' \sim \vec{\mathbb{C}}'$, as required. □

Corollary 15.10. *Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 that is not tiny and $S \subset S^3$ is a sphere transverse to $F = F(\mathbb{C})$ which is either balanced or almost balanced for $\mathbb{C}$. Let $\overline{\mathfrak{c}}$ and $\overline{\mathfrak{c}}'$ be collections* *of circles in S, each containing the set $\mathfrak{c} = F \cap S$. Let $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ be the sequences of flagged chamber complex decompositions determined by $\bar{\mathfrak{c}}$ and $\bar{\mathfrak{c}}'$ respectively. Then $\vec{\mathbb{C}}$ and $\vec{\mathbb{C}}'$ cocertify.*

Proof. It suffices to prove the Corollary in the case that $\bar{\mathfrak{c}} = \mathfrak{c}$. In this case $\bar{\mathfrak{c}}'$ is obtained from $\bar{\mathfrak{c}}$ by adding some number m of ghost circles. But this case follows by applying Proposition 15.9 m times. □

Corollary 15.11. *Theorem 13.3 remains true even when the chamber C containing E is an empty torus, or when E is inessential.*

Proof. Referring to the proof of Theorem 13.3, the requirement that E be essential and C not be an empty torus is to ensure that the original decomposition $C = C_0 \xrightarrow{E} C'$ creates no goneballs. The boundaries of such goneballs would intersect $F = F(\mathbb{C})$ in a family of circles $\bar{\mathfrak{c}}_G$ in $F \cap S$ that continue to determine which circles in $F \cap S$ bound disks in each $\mathcal{D}_i$ during the subsequent decomposition

$$\vec{\mathbb{C}}' : \quad \mathbb{C}' = \mathbb{C}'_0 \xrightarrow{\mathcal{D}'_0} \mathbb{C}'_1 \xrightarrow{\mathcal{D}'_1} \mathbb{C}'_2 \xrightarrow{\mathcal{D}'_2} \dots \xrightarrow{\mathcal{D}'_{n-1}} \mathbb{C}'_n \tag{13.1}$$

used in the proof of Theorem 13.3.

In contrast, it is the collection of circles $F(\mathbb{C}') \cap S$ that determines a decomposition sequence in $\overrightarrow{(\mathbb{C}', S)}$ and the goneballs of the decomposition $C = C_0 \xrightarrow{E} C'$ do not appear in the definition of $F' = F(\mathbb{C}')$. In other words, $F \cap S = (F' \cap S) \cup \bar{\mathfrak{c}}_G$, so $\bar{\mathfrak{c}}_G$ acts as a collection of ghost circles in the decomposition sequence $\vec{\mathbb{C}}'$. Corollary 15.10 says that the addition of such ghost circles makes no difference in certification. That is, we may as well assume that $F \cap S = F' \cap S$ and proceed with the proof of Theorem 13.3 as written. □

16 Isotopies of guiding spheres

Definition 16.1. *Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 and S, S' are spheres transverse to the defining surface $F = F(\mathbb{C})$. If any $\vec{\mathbb{C}} \in \overrightarrow{(\mathbb{C}, S)}$ and $\vec{\mathbb{C}}' \in \overrightarrow{(\mathbb{C}, S')}$ cocertify, then we say $\overrightarrow{(\mathbb{C}, S)}$ and $\overrightarrow{(\mathbb{C}, S')}$ cocertify and write $\overrightarrow{(\mathbb{C}, S)} \sim \overrightarrow{(\mathbb{C}, S')}$.*

Our goal in this section is to prove:

Theorem 16.2. *Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 that is not tiny and supports the Heegaard splitting $S^3 = A \cup_T B$. Suppose further that S_0 and S_1 are balanced or almost balanced spheres for $\mathbb{C}$ that are isotopic in S^3 through balanced or almost balanced spheres $S_t, 0 \le t \le 1$. Then $\overrightarrow{(\mathbb{C}, S_0)} \sim \overrightarrow{(\mathbb{C}, S_1)}$.*

As usual, let F denote the defining surface $F = F(\mathbb{C})$ and put the isotopy in general position so that generically S_t is transverse to F, but for certain discrete values of t, S_t is tangent to F at a single point, a non-degenerate critical point of index zero (F has a local max at v in a collar of S), two (a local min) or one (a saddle tangency).

Lemma 16.3. *Theorem 16.2 is true if the isotopy S_t passes through only a single critical point, of index zero or two (a max or a min).*

Proof. The cases of index zero or index two are symmetric. Let $S_{m(iddle)}$ be the sphere on which the critical point v occurs, and, in a collar of S_m, we may suppose F has a minimum. Call the bottom of the collar $S_{b(elow)}$ and the top $S_{a(bove)}$. We can assume that other than the local minimum, F intersects the collar in vertical cylinders. Then $F \cap S_a$ is the union of $F \cap S_b$ and a single circle c, where c bounds a disk D_F in $F - S_a$ (the disk in F containing v) and a parallel disk D_S in $S_a - F$. Let $\vec{\mathbb{C}}_b$ (resp $\vec{\mathbb{C}}_a$) be the sequence of flagged chamber complex decompositions determined by $\mathbb{C}$ and S_b (resp. S_a). Thus $\vec{\mathbb{C}}_b \in \overrightarrow{(\mathbb{C}, S_1)}$ (say) and $\vec{\mathbb{C}}_a \in \overrightarrow{(\mathbb{C}, S_0)}$. See Figure 42. Augment the collection of circles $F \cap S_b$ with the single (ghost) circle c' parallel in S_b to $c \subset S_a$ and call the resulting sphere with ghost circle S_c. Let $\vec{\mathbb{C}}_c$ be a sequence of flagged chamber complex decompositions determined by $\mathbb{C}$ and S_c. Then by Corollary 15.10 $\vec{\mathbb{C}}_b \sim \vec{\mathbb{C}}_c$.

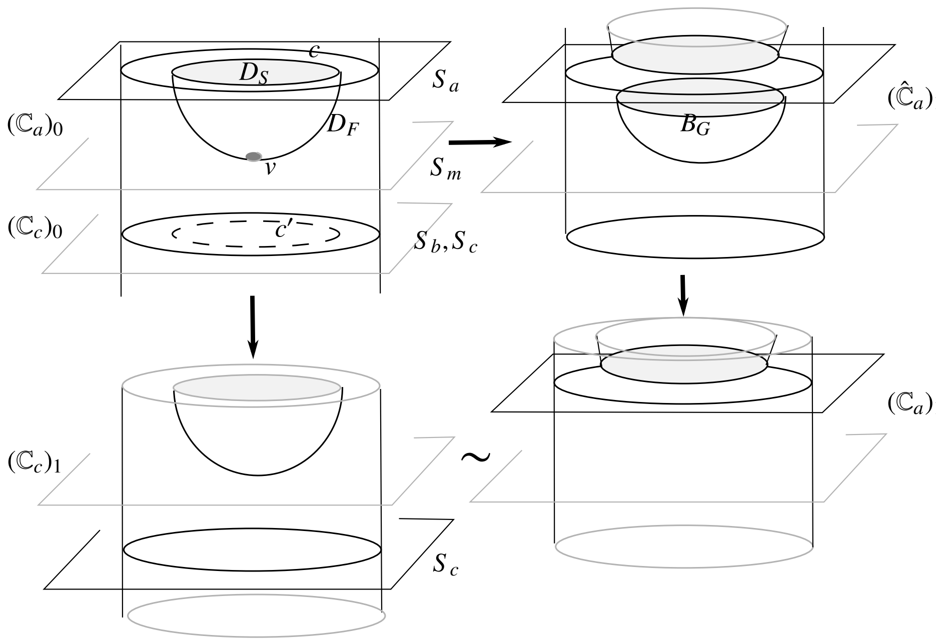

Figure 42: Passing S through a minimum

Now compare $\vec{\mathbb{C}}_a$ with $\vec{\mathbb{C}}_c$. The pattern of circles in $F \cap S_a$ is the same as in $F \cap S_c$ except that the F-circle c in S_a appears as a ghost circle c' in S_c. Thus the decomposition $\mathbb{C}_0 \xrightarrow{\mathcal{D}_0} \mathbb{C}_1$ is the same in each sequence, except in $\vec{\mathbb{C}}_c$ the ghost circle c' is just removed, whereas in $\vec{\mathbb{C}}_a$, F is surgered by D_S, creating an extra ball chamber B_G bounded by the sphere $D_S \cup D_F$. But our convention for flagging declares the handlebody B_G to be empty and therefore a goneball, so the result of the flagged chamber complex decomposition $\mathbb{C}_1$ is the same in both cases, as is the remaining pattern of circles $S_a \cap F(\mathbb{C}_1) = S_c \cap F(\mathbb{C}_1)$. It follows that the entire ensuing flagged chamber complex decomposition sequence is the same, so $\vec{\mathbb{C}}_a \sim \vec{\mathbb{C}}_c$. Combining, we have $\vec{\mathbb{C}}_a \sim \vec{\mathbb{C}}_c \sim \vec{\mathbb{C}}_b$ so $\overrightarrow{(\mathbb{C}, S_0)} \sim \overrightarrow{(\mathbb{C}, S_1)}$ as required. □

Next suppose the isotopy S_t has a single critical point that is a saddle tangency with F at a point $v \in F$. As in the proof of Lemma 16.3 we consider how F intersects a collar of S_m, the sphere of the tangency at

v, as shown in Figure 43. A neighborhood in F of $S_m \cap F$ (a figure eight) is a pair of pants with three disjoint circles: c_b, whose parallel in S_m is incident to v in two points, and c_1, c_2 each of whose parallels in S_m is incident to v in a single point. With S_a and S_b as in Lemma 16.3, we can assume that c_1 and c_2 both lie in S_a, and c lies in S_b.

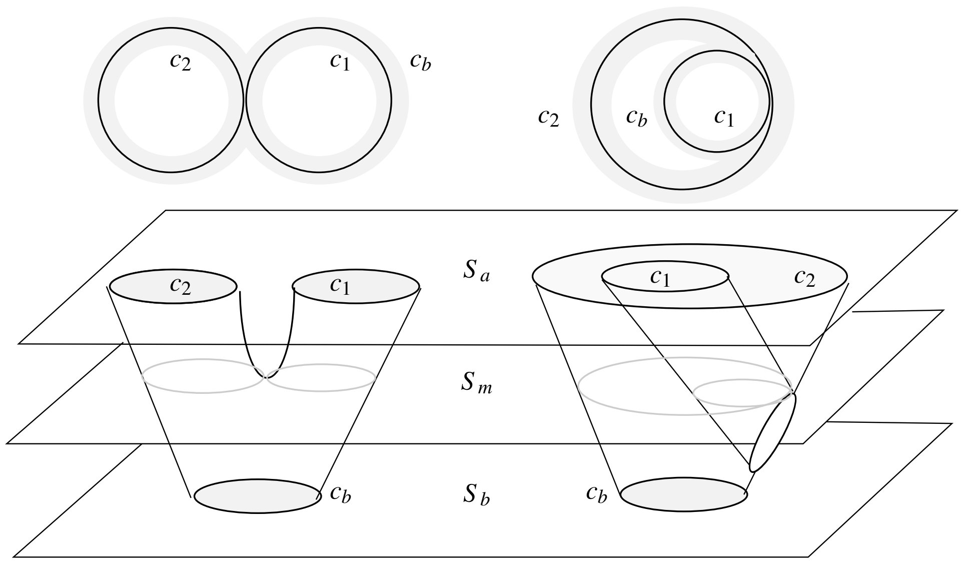

Figure 43: A neighborhood of critical point v: two viewpoints

As in Section 9 let Y_a (resp Y_b) be the tree defined by the collection of circles $F \cap S_a$ (resp $F \cap S_b$), e_1, e_2 be the edges in Y_a corresponding to c_1, c_2 and e_b be the edge in Y_b corresponding to c_b. As defined in 9.1, let $\rho_i = \rho_{Y_a}(e_i), i = 1, 2$ and $\rho_b = \rho_{Y_b}(e_b)$. Put another way, in the decomposition sequences of $\mathbb{C}$ guided by S_a and S_b, each of the two circles $c_i, i = 1, 2$ bounds a disk $D_i \in \mathcal{D}_{\rho_i}$ in the decomposition $\mathbb{C}_{\rho_i} \xrightarrow{\mathcal{D}_{\rho_i}} \mathbb{C}_{\rho_i+1}$ and c_b bounds a disk $D_b \in \mathcal{D}_{\rho_b}$ in the decomposition $\mathbb{C}_{\rho_b} \xrightarrow{\mathcal{D}_{\rho_b}} \mathbb{C}_{\rho_b+1}$. With no loss of generality, assume $\rho_1 \leq \rho_2$ (i. e. $\mathcal{D}_{\rho_1}$ does not appear after $\mathcal{D}_{\rho_2}$ in the decomposition sequence). There are two possibilities for the disks $D_1, D_2 \subset S_a$, and the two viewpoints in Figure 43 are chosen so that in either case each D_i is on the bounded side of c_i in S_a: either D_1 and D_2 are disjoint, as shown in the left side of the figure, or $D_1 \subset D_2$ as shown in the right side. That is, the two viewpoints in the figure are chosen so that in each the D_i appear as bounded regions. Similarly, D_b can either lie on the side of c_b that contains v (the bounded side in the left of the figure and unbounded side in the right) or the other side. We have:

Lemma 16.4. *If D_b lies on the side of c_b that contains v, then $\rho_1 \leq \rho_b$.*

Proof. Recall that, aside from the component that contains v, F intersects the collar of S only in vertical annuli. Hence the interior of the bounded side of c_b on the left of Figure 43 (or the the interior of the unbounded side of c_b on the right) contains a parallel copy of each circle of $F \cap \text{int}(D_1)$. □

Lemma 16.5. *Theorem 16.2 is true if the isotopy S_t passes through only a single critical point, of index one (a saddle tangency).*

Proof. We make two claims:

Claim 1: With no loss of generality, we can assume that $\rho_1 \leq \rho_b$.

Proof of Claim 1: By Lemma 16.4 this is true when D_b lies on the side of c_b that contains v, so suppose D_b lies on the other side. That is, suppose D_b lies on the outside (unbounded side) of c_b in the left of Figure 43 or on the inside (bounded side) of c_b on the right. Suppose $\rho_b < \rho_1$ and consider what happens if we add $\rho_1 - \rho_b$ ghost circles to D_b, parallel to ∂D_b. This raises ρ_b so that $\rho_b = \rho_1$. Since there is a copy of the pattern of circles $F \cap D_1$ on the other side of c_b in S_b, that is in $S_b - D_b$, (see left panel in Figure 44), it remains true that the maximal distance in Y_b from e_b to a leaf of $Y_\pm$ is still minimized in D_b, so we do not change the side of c_b on which D_b lies. Let S_b^+ denote S_b with these ghost circles added, and denote by $\vec{\mathbb{C}}_b^+$ be the sequence of flagged chamber complex decompositions on $\mathbb{C}$ determined by S_b^+. By Corollary 15.10 it suffices to show that $\vec{\mathbb{C}}_b^+$ and $\vec{\mathbb{C}}_a$ cocertify. This proves Claim 1.

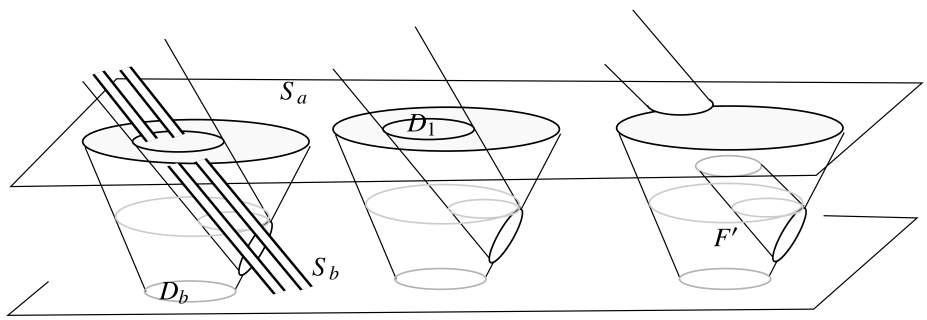

Figure 44: When D_b lies on the side of c_b not containing v

Since, per the claim, $\rho_1 \leq \rho_b$ and by definition $\rho_1 \leq \rho_2$ there is no difference between the decomposition sequences for $\vec{\mathbb{C}}_a$ and $\vec{\mathbb{C}}_b$ until we reach the decomposition $\mathbb{C}_{\rho_1} \xrightarrow{\mathcal{D}_{\rho_1}} \mathbb{C}_{\rho_1+1}$. $F_{\rho_1} = F(\mathbb{C}_{\rho_1})$ still appears as in Figure 43, but with the added information that $\text{int}(D_1)$ contains neither ghost circles nor F-circles. In particular, it is possible to define a new chamber complex $\mathbb{C}'$ obtained from $\mathbb{C}_{\rho_1}$ by decomposing along D_1 alone. That is

$$\mathbb{C}_{\rho_1} \xrightarrow{D_1} \mathbb{C}'.$$

Claim 2: Both S_a and S_b are still balanced or almost balanced for $\mathbb{C}'$.

Proof of Claim 2: Let F' be the defining surface $F(\mathbb{C}')$. See Figure 44. We name 4 regions and 6 genera:

- X is the region below S_b and g_X is the genus of $F_{\rho_1} \cap X = F' \cap X$.
- $X^+ \supset X$ is the region below S_a, g_X^+ is the genus of $F_{\rho_1} \cap X^+$, and $g_X'^+$ is the genus of $F' \cap X^+$.
- Y is the region above S_a, g_Y is the genus of $F_{\rho_1} \cap Y$, and g'_Y is the genus of $F' \cap Y$
- $Y^+ \supset Y$ is the region above S_b, and $g_Y'^+$ is the genus of $F' \cap Y^+$

Surgery on $D_1 \subset S_a$ will not affect the genus of the part of F_{ρ_1} lying below or above S_a, so (0) $g'^+_X = g^+_X$, $g'_Y = g_Y$ and S_a is balanced or almost balanced for $\mathbb{C}'$.

To prove the claim, we then need only focus on S_b. Here are two further observations:

1. Since $F_{\rho_1} \cap X^+$ is obtained from $F_{\rho_1} \cap X$ by attaching a (genus 0) pair of pants on a single boundary component, $g^+_X = g_X$.

2. Since $F' \cap Y^+$ is obtained from $F' \cap Y$ by attaching a (genus 0) annulus on a single boundary component, $g'^+_Y = g'_Y$.

By definition S_b is balanced or almost balanced for $\mathbb{C}'$ if and only if the pair of numbers $\{g_X, g'^+_Y\} \neq \{0,k\}, k \geq 2$. Indeed, by definition, the pair is $\{0,0\}$ if and only if S_b is planar balanced, it is $\{0,1\}$ if and only if S_b is almost balanced, and the pair is $\{j,k\}$, $j,k \geq 1$ if and only if S_b is non-planar balanced. Similarly, since S_a is balanced or almost balanced for $\mathbb{C}_{\rho_1}$, we have $\{g^+_X, g_Y\} \neq \{0,k\}, k \geq 2$. From (1) we have $g_X = g^+_X$ and combining (0) and (2) gives $g'^+_Y = g_Y$. In particular, $\{g_X, g'^+_Y\} = \{g^+_X, g_Y\} \neq \{0,k\}, k \geq 2$, completing the proof of Claim 2.

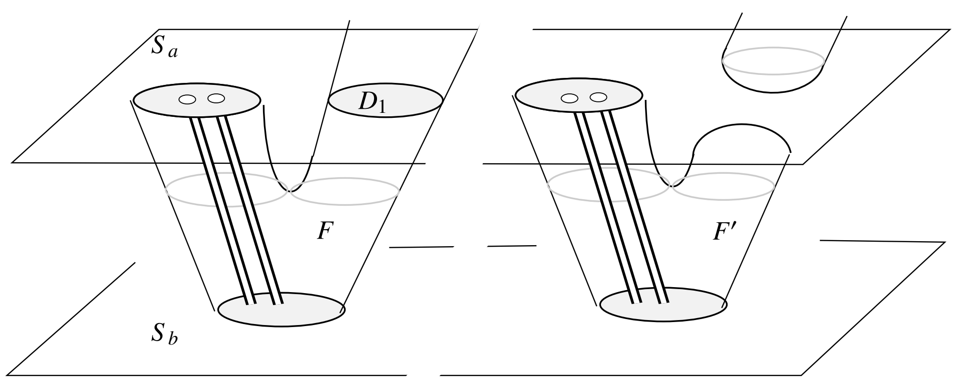

Figure 45: Only annuli between the planes

F' intersects the collar between S_a and S_b entirely in spanning annuli (see Figure 45 and right panel of Figure 44), so the decomposition sequence on $\mathbb{C}'$ determed by S_a and S_b is the same; we denote it $\vec{\mathbb{C}}'$. As in the proof of Lemma 16.3 let $\vec{\mathbb{C}}_b$ (resp $\vec{\mathbb{C}}_a$) be the sequence of flagged chamber complex decompositions determined by $\mathbb{C}$ and S_b (resp. S_a). Thus $\vec{\mathbb{C}}_b \in \overrightarrow{(\mathbb{C}, S_1)}$ (say) and $\vec{\mathbb{C}}_a \in \overrightarrow{(\mathbb{C}, S_0)}$. Now apply, for both S_a and S_b, Theorem 13.3 as augmented by Corollary 15.11 to conclude

$$\vec{\mathbb{C}}_a \sim \vec{\mathbb{C}}' \sim \vec{\mathbb{C}}_b \quad \text{so} \quad \overrightarrow{(\mathbb{C}, S_0)} \sim \overrightarrow{(\mathbb{C}, S_1)}$$

□

Proof of Theorem 16.2. As noted before Lemma 16.3 a generic isotopy S_t from S_0 to S_1 consists of a sequence of isotopies, each containing a single critical point and hence one to which either Lemma 16.3 or Lemma 16.5 applies. □

17 All balanced or almost balanced spheres cocertify

In this section we show that the requirement in Theorem 16.2 that S_0 and S_1 are isotopic through balanced or almost balanced spheres is superfluous. In particular we show:

Proposition 17.1. *Suppose $\mathbb{C}$ is a flagged chamber complex in S^3. Then any pair of balanced or almost balanced spheres for $\mathbb{C}$ are isotopic in S^3 through balanced or almost balanced spheres for $\mathbb{C}$.*

We defer the proof in order to quickly observe some consequences: Following Theorem 16.2 this immediately implies:

Corollary 17.2. *Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 that supports the genus g Heegaard splitting (S^3, T) and is not tiny. Let S^x and S^y be balanced or almost balanced spheres for $\mathbb{C}$. Then $\overrightarrow{(\mathbb{C}, S^x)} \sim \overrightarrow{(\mathbb{C}, S^y)}$.* □

In particular, any such flagged chamber complex $\mathbb{C}$ in S^3 determines a unique certificate. That is, the homeomorphism $h_{(\mathbb{C},S)} : (S^3, T) \to (S^3, T_g)$, for S a balanced or almost balanced sphere (provided, say, by Corollary 10.7), does not in fact depend on the choice of S, up to eyeglass equivalence. We will denote (a choice of) such a homeomorphism simply $h_{\mathbb{C}} : (S^3, T) \to (S^3, T_g)$. This expands the use of the notation, first introduced following Corollary 8.3, beyond just those flagged chamber complexes that themselves certify, to all flagged chamber complexes that are not tiny. Following Proposition 5.18, it applies to flagged chamber complexes obtained from a Heegaard surface by weak reduction. It also allows the following natural definition and notation:

Definition 17.3. *Suppose $\mathbb{C}, \mathbb{C}'$ are flagged chamber complexes in S^3 that support the genus g Heegaard splitting (S^3, T) and are not tiny. Then $\mathbb{C}$ and $\mathbb{C}'$ cocertify (written $\mathbb{C} \sim \mathbb{C}'$) if the homeomorphisms $h_{\mathbb{C}}, h_{\mathbb{C}'} : (S^3, T) \to (S^3, T_g)$ (as just defined above) are eyeglass equivalent.*

Corollary 17.4. *With this expanded definition of $h_{\mathbb{C}}$, Corollary 8.4 remains true. That is, if $\mathbb{C}$ certifies and $\tau \in G(S^3, T)$ then $h_{\tau(\mathbb{C})}\tau \sim h_{\mathbb{C}}$.*

Proof. This follows immediately from Corollary 9.6. □

The proof of Proposition 17.1 is brief (it appears just before Proposition 17.7) but the background needed is extensive. Since it makes use of elementary Morse and Cerf Theory [Mi], [Ce], we will work in the smooth category. That is, we will take the defining surface $F = F(\mathbb{C})$ to be a smooth submanifold of S^3, and the pair of spheres S^x and S^y, to be smooth submanifolds, smoothly transverse to F. Let $p : S^3 \to [-1, 1]$ be the standard height function, denote by S^2 the standard equator $p^{-1}(0)$, and call the points $p^{-1}(1), p^{-1}(-1)$ the north and south poles of S^3. Denote by B_ϵ a pair of small ball neighborhoods of the poles, that is $B_\epsilon = p^{-1}([-1, -1 + \epsilon] \cup [1 - \epsilon, 1])$.

Choose a pair of points q_n, q_s (to correspond to north and south poles) disjoint from $F \cup S^x \cup S^y$ so that q_n and q_s lie on opposite sides of each of the spheres S^x and S^y. This is easily done: If S^x and S^y are disjoint, choose a generic point in the interior of each of the two ball components of $S^3 - (S^x \cup S^y)$. If, on the other hand, S^x and of S^y intersect, observe that a regular neighborhood of a curve c of intersection intersects $S^x \cup S^y$ in a copy of $\mathsf{X} \times c$. (Here X means two crossed lines, e. g. the x- and y- axes in $\mathbb{R}^2$.) Then set q_n, q_s to be generic points in opposite quadrants of some $\mathsf{X} \times \{point\}$.

By the Schoenflies Theorem there are diffeomorphisms $\phi^x, \phi^y : (S^3; q_n, q_s) \to (S^3; p^{-1}(1), p^{-1}(-1))$ so that $\phi^x(S^x) = \phi^y(S^y) = S^2 \subset S^3$. Choose a small neighborhood B_ϵ of the poles in S^3 that is disjoint from $\phi^x(F) \cup \phi^y(F)$ and adjust ϕ^x, ϕ^y near q_n, q_s so that p_x and p_y coincide on a neighborhood $U = (p^x)^{-1}(B_\epsilon) = (p^y)^{-1}(B_\epsilon)$ of q_n, q_s. It is elementary in this case (or use [Lau] and the fact that the space of orientation preserving diffeomorphisms of the ball is connected, indeed contractible) that there is an isotopy $\theta : (S^3, U) \times I \to (S^3, B_\epsilon)$ between ϕ^x and ϕ^y, fixed on U. That is, for $\theta_t : (S^3, U) \to (S^3, B_\epsilon)$ defined by $\theta_t(z) = \theta(z, t)$, we have $\theta_0 = \phi^x$ and $\theta_1 = \phi^y$ and for all t, and each $u \in U$, $\theta_t(u) = \phi^x(u) = \phi^y(u)$.

Now consider the function $p^x : F \to [-1, 1]$ defined by $p^x = p\phi^x|F$ and similarly $p^y = p\phi^y|F$. The functions p^x, p^y are homotopic via the homotopy $p_t = p\theta_t|F$. Morse and Cerf theory tell us that for a generic copy of $F \subset S^3$ arbitrarily near the original F (a copy which we henceforth take as F), the functions p^x, p^y are Morse functions and the homotopy p_t between them is Cerf. Here are the salient points of what is meant by this:

p^x (and p^y) are Morse: The function $p^x : F \to [-1, 1]$ has only a finite number of critical values $\{s_i\} \subset (-1, 1)$ and each critical value is the image of a single non-degenerate critical point in F. Non-degenerate means that the Hessian of the function is non-singular (see [Mi]). Put another way, each sphere $S^x_s = (p\phi^x)^{-1}(s)$ is transverse to F for each regular value of s, and, for each critical value s_i, $S^x_{s_i}$ is transverse to F except at a single point of tangency, where F intersects a neighborhood of the tangency point in $S^x_{s_i}$ as either a maximum, a minimum, or a saddle point.

The homotopy p_t is Cerf: For generic t the function $p_t : F \to [-1, 1]$ is Morse. At a finite set $\tau = \{t_j\} \subset I$ the function p_t is not Morse, solely because either

- near a single exceptional critical point on $z \in F$, p_t has a "birth-death" singularity given by the local model $p_t(z_1, z_2) = p_t(z) + z_1^3 \pm tz_1 \pm z_2^2$ or
- there are two non-degenerate critical points of p_t on F with the same critical value

Beyond this, and perhaps clarifying it, is a description of the “Cerf graphic" in the square $(t, s) \in I \times [-1, 1]$ (see, for example, introductory remarks in [GK], from which Figure 46 is taken): There are a finite collection Γ of curves in $I \times [-1, 1]$ so that each $\gamma \in \Gamma$ is the graph of a smooth function to $[-1, 1]$ on a closed interval in I whose end points are among the set $\tau \cup \partial I$. See green in Figure 46. Each generic vertical arc $t \times [-1, 1]$ is transverse to the graphs Γ and intersects each curve $\gamma \in \Gamma$ in at most one point. Each vertical arc $t_i \times [-1, 1], t_i \in \tau$ has the same property, with a single exceptional point $s \in [-1, 1]$ where either

- (t_i, s) is a left end point of exactly two such graphs, a birth point. (So $t_i \times [-1, 1] \in$ red vertical lines in Figure 46)
- (t_i, s) is a right end point of exactly two such graphs, a death point. (So $t_i \times [-1, 1] \in$ red vertical lines in Figure 46)
- At (t_i, s) two graphs cross transversally. (So $t_i \times [-1, 1] \in$ blue vertical lines in Figure 46)

The collection Γ of graphs constitute all points where the functions p_t have critical points. That is, $(t, s) \in I \times [-1, 1]$ lies in Γ if and only if s is a critical value for p_t. In particular, each complementary

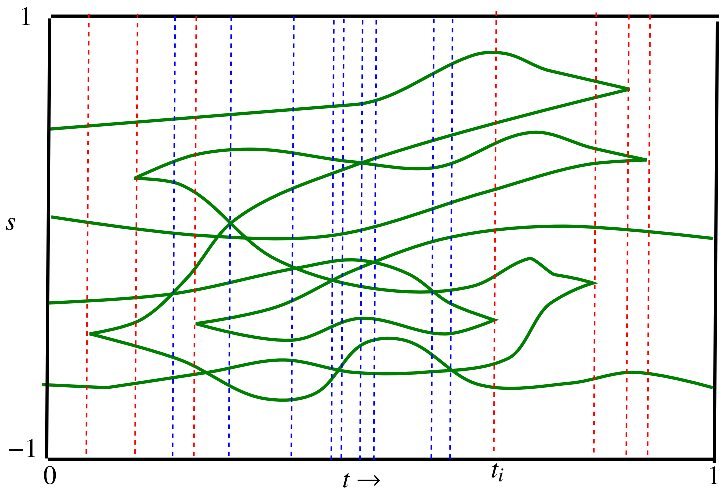

Figure 46: Graphic (green) with $t_i \in \tau$

component R of Γ in $I \times [-1,1]$ (called a *region* of the graphic) has the property that for any $(t,s) \in R$, s is a regular value for p_t.

Given this as background, observe that the natural sweep-out of $S^3 - \{poles\}$ by the 2-spheres $S_s = p^{-1}(s), s \in (-1,1)$, in which S_0 is the equator S^2 induces a t-parameterized smooth family of sweep-outs $S^t_s = \theta_t^{-1}(S_s)$ beginning with a sweep-out $(\phi^x)^{-1}(S_s)$ that contains $S^x = (\phi^x)^{-1}(S^2)$ and ending with a sweep-out $(\phi^y)^{-1}(S_s)$ that contains $S^y = (\phi^y)^{-1}(S^2)$. Taking this sweep-out point of view, we can now incorporate the ideas of Section 10, in particular Definition 10.5 and Lemma 10.6, as we now describe.

Suppose $(t,s) \in R$, where R is a region of the graphic. Since s is a regular value of p_t, the sphere S^t_s is transverse to F, and, since R is connected, any other sphere $S^{t'}_{s'}, (t',s') \in R$ is isotopic to S^t_s through spheres transverse to F. Following Definition 10.5 we can then assign a label $\frac{a_R}{b_R}$ to R, where the non-negative integer a_R is the genus of the part of F lying *a*bove S^t_s and similarly b_R is the genus of the part of F lying *b*elow S^t_s. Here are some elementary observations:

1. S^t_s is balanced or almost balanced for $\mathbb{C}$ if one of these is true:

   - $0 \le a_R, b_R \le 1$
   - Both $a_R, b_R \ge 1$

   In this case we say that the region itself is balanced or almost balanced.

2. Hence S^t_s is *not* balanced or almost balanced for $\mathbb{C}$ (say S^t_s is *akilter*) if one of these is true:

   - $a_R = 0$ and $b_R \ge 2$ (say then +akilter) or
   - $b_R = 0$ and $a_R \ge 2$ (say then −akilter).

   In this case we say that the region itself is respectively +akilter or −akilter.

3. Suppose R and R' are adjacent regions in the graphic, both incident to a birth singularity (t_0, s_0). That is, the regions are separated near (t_0, s_0) by the pair of curves in Γ that end at (t_0, s_0). Then a

generic sweep-out to the right of (t_0, s_0) (that is via $S_s^{(t_0+\epsilon)}, s \in (-1,1)$), passing briefly through R', say, differs from a sweep-out just to the left of (t_0, s_0), passing through R, by the brief introduction of a minimum (or maximum), followed by a cancelling saddle tangency. The consequent brief transfer of a disk in F from one side of S_s^t to the other has no effect on the genus of F on each side, so $\frac{a_R}{b_R} = \frac{a_{R'}}{b_{R'}}$.

4. The same is true for R and R' adjacent regions in the graphic, both incident to a death singularity (t_0, s_0).

We now interpret Lemma 10.6 in this context.

Lemma 17.5. *Suppose, in the setting described above,* genus$(F) \geq 2$ *and t_0 is a generic value of t, that is $t_0 \in I - \tau$. The vertical arc $\mathfrak{v} = t_0 \times [-1,1]$ has the following properties:*

1. *There is a unique curve $\gamma_- \subset \Gamma$, intersecting $\mathfrak{v}$ in a point (t_0, s_-), so that the region just below γ_- is $-$akilter and the region just above γ_- is balanced or almost balanced.*
2. *There is a unique curve $\gamma_+ \subset \Gamma$ intersecting $\mathfrak{v}$ in a point (t_0, s_+), so that the region just below γ_+ is balanced or almost balanced and the region just above γ_+ is $+$akilter.*
3. *$s_- < s_+$ and the subinterval $t_0 \times [s_-, s_+]$ of $\mathfrak{v}$ intersects at least three different regions of the graphic.*

Proof. We will apply Lemma 10.6 to the sweepout of F by the spheres $S_s^{t_0}$. (The lemma is applicable because it is equivalent to considering the sweep-out of $\theta_{t_0}(F)$ by the spheres S_s.) Note first that the bottom regions of the graphic (those incident to $I \times \{-1\}$) are all labeled $\frac{genus(F)}{0}$ and the top regions are all labeled $\frac{0}{genus(F)}$. Since genus$(F) \geq 2$ the regions at the bottom are $-$akilter and those at the top are $+$akilter.

As noted in Lemma 10.6 as s rises in $\mathfrak{v}$, passing from region to region, a_R cannot rise and b_R cannot fall. To progress under these rules from the label $\frac{a}{0}, a \geq 2$ to the label $\frac{0}{b}, b \geq 2$ there is a first transition from a region R that is $-$akilter to a region R' that is balanced or almost balanced, either because $\frac{a_R}{b_R} = \frac{2}{0}$ and $\frac{a_{R'}}{b_{R'}} = \frac{1}{0}$ or because $\frac{a_R}{b_R} = \frac{a}{0}, a \geq 2$ and $\frac{a_{R'}}{b_{R'}} = \frac{a}{1}$. Moreover, once that transition is made, there is no way to transition back to $-$akilter without either lowering the denominator in $\frac{a}{1}$ or raising the numerator in $\frac{1}{0}$, neither of which is allowed. So this transition defines the curve γ_- and so s_-. The symmetric argument identifies the curve γ_+ and shows that it lies above γ_-, that is $s_+ > s_-$. See Figure 47.

More subtly, notice that according to Lemma 10.6 as s passes upward through a curve γ from region R to region R' the difference $a_R - b_R \in \mathbb{Z}$ does not increase (for this would require either a_R to rise or b_R to fall) and can decrease by at most one (because at most one of these occurs: a_R drops by 1 or b_R rises by 1). That is, $(a_{R'} - b_{R'}) \leq (a_R - b_R) \leq (a_{R'} - b_{R'}) + 1$. So it takes at least 4 transitions to go from a $-$akilter region, where $a_R - b_R \geq 2$, to a $+$akilter region, where $a_R - b_R \leq -2$. So, not counting the end points, there are at least two graphs in Γ that cross the interior of $t_0 \times [s_-, s_+]$ and so 3 distinct regions through which $t_0 \times [s_-, s_+]$ passes. □

Lemma 17.6. *Suppose, in the setting described above,* genus$(F) \geq 2$. *Then there are continuous (indeed, piecewise smooth) functions $f_-, f_+ : I \to [-1,1]$ so that*

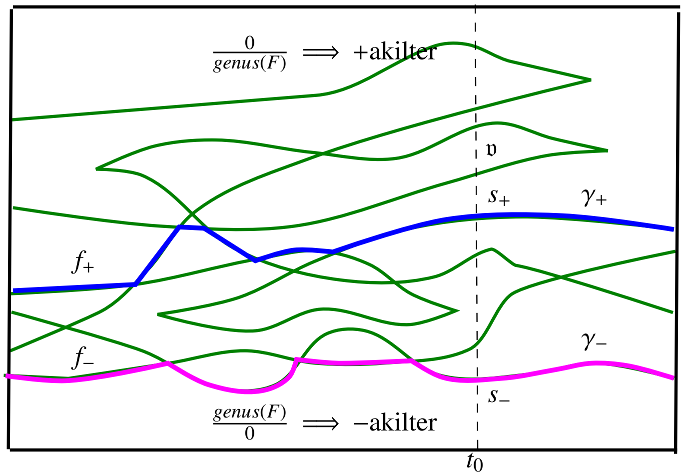

Figure 47: Spanning arcs γ_+, γ_- and the gap between

- *For all t, $f_-(t) < f_+(t)$.*
- *The graphs of both $f_\pm$ lie in Γ.*
- *Regions below the graph of f_- are all −akilter.*
- *Regions above the graph of f_+ are all +akilter.*
- *Regions between the two graphs are all balanced or almost balanced.*
- $f_-(0) < 0 < f_+(0)$ *and* $f_-(1) < 0 < f_+(1)$.

Proof. For $t_0 \notin \tau$ define $f_\pm(t_0) = s_\pm$ as given in Lemma 17.5. The functions $f_\pm$ are smooth near t_0 because each curve $\gamma \in \Gamma$ is smooth. The same argument applies for $t_0 \in \tau$ with some (t_0, s) a birth-death point since, as we have seen above, $\frac{a_R}{b_R}$ does not change as s passes through such a point.

The situation is only a bit more complicated for $t_0 \in \tau$ with two curves in Γ crossing at (exactly one) point (t_0, s_0) in $t_0 \times [-1, 1]$. In this case, the region below (t_0, s_0) could be −akilter and the region above (t_0, s_0) either balanced or almost balanced, and then it is natural to define $f_-(t_0) = s_0$. By examining generic nearby values of t we see that the interval $t_0 \times (s_0, s_+)$ must still pass through at least two distinct regions, so $s_0 < s_+$ and $f_-(t_0) < f_+(t_0)$ as required. The graph of f_- may not be smooth at t_0 since values of $f_-(t)$ for t just less than t_0 may lie on one of the two crossing curves in Γ and values of $f_-(t)$ for t just more than t_0 lie on the other. Still the graph of f_- is piecewise smooth at t_0. Symmetric comments apply when the region above (t_0, s_0) is +akilter and the region below (t_0, s_0) either balanced or almost balanced.

Finally, our initial definition was that the balanced or almost balanced sphere $S^x = S^0_0$ so $f_-(0) < 0 < f_+(0)$ and the balanced or almost balanced sphere $S^y = S^1_0$ so $f_-(1) < 0 < f_+(1)$. □

Proof of Proposition 17.1. If genus$(F) \le 1$ let $f : [0, 1] \to [-1, 1]$ be any smooth function whose graph is transverse to the curves Γ in the graphic and for which $f(0) = f(1) = 0$. If genus$(F) \ge 2$ pick such a function so that, in addition, $f_- < f < f_+$ Then the 1-parameter family of spheres $S^t_{f(t)}, 0 \le t \le 1$ is an isotopy of S^x to S^y. Moreover each sphere $S^t_{f(t)}$ is balanced or almost balanced for $\mathbb{C}$, since the graph of f passes only through balanced or almost balanced regions of the graphic. □

There is a related proposition about single sweep-outs, whose proof is similar to that of Lemma 17.5:

Proposition 17.7. *Suppose $\mathbb{C}$ is a flagged chamber complex in S^3 that is not tiny and $S_s, s \in [-1,1]$ is a generic sweep-out of S^3 by spheres. Suppose, E is a properly embedded disk in a chamber of $\mathbb{C}$ so that, for some generic $-1 < e < 1$, ∂E lies entirely in S_e. That is, ∂E is a circle component of $S_e \cap F(\mathbb{C})$. Put E in preferred alignment and let $\mathbb{C}'$ be the flagged chamber complex obtained from $\mathbb{C}$ by decomposition along E. Then $\mathbb{C}' \sim \mathbb{C}$.*

Proof. Since S_s is a generic sweep-out, there is a finite set $\sigma \subset [-1,1]$ so that for $s \notin \sigma$, S_s is transverse to $F = F(\mathbb{C})$, and for each $s \in \sigma$, F and S_s have a single non-degenerate point of tangency. Apply the notational convention for regions introduced before Lemma 17.5 to the (interval) components of $[-1,1]-\sigma$. That is, for each such interval $\mathfrak{i} \subset [-1,1]-\sigma$ pick $s \in \mathfrak{i}$ and let $a_\mathfrak{i} \geq 0$ be the genus of the part of F lying above S_s and $b_\mathfrak{i}$ be the genus of the part of F lying below S_s, and then assign to $\mathfrak{i}$ the symbol $\frac{a_\mathfrak{i}}{b_\mathfrak{i}}$. Following Lemma 10.6, and similar to the argument used for a generic value of t in the proof of Lemma 17.5, we have:

- The lowest interval in $[-1,1]-\sigma$ is assigned $\frac{genus(F)}{0}$
- The highest interval $[-1,1]-\sigma$ is assigned $\frac{0}{genus(F)}$
- As s rises through each critical level $s \in \sigma$ the integer $a_\mathfrak{i} - b_\mathfrak{i}$ does not increase and can decrease by at most 1.

Since, as we ascend through the interval components of $[-1,1]-\sigma$, the difference $a_\mathfrak{i} - b_\mathfrak{i}$ begins at genus(F) and ends at $-genus(F)$ it follows that there is an interval $\mathfrak{i}_0$ for which the difference is zero, that is $a_{\mathfrak{i}_0} = b_{\mathfrak{i}_0}$. In particular, for $s \in \mathfrak{i}_0$, S_s is balanced for $\mathbb{C}$, either planar balanced (if $a_{\mathfrak{i}_0} = b_{\mathfrak{i}_0} = 0$) or non-planar balanced (if $a_{\mathfrak{i}_0} = b_{\mathfrak{i}_0} \geq 1$). Pick a generic $s_0 \in \mathfrak{i}_0$ ($s_0 \neq e$) so that S_{s_0} is transverse to E.

Claim: The sphere S_{s_0} is balanced or almost balanced for $\mathbb{C}'$ as well as for $\mathbb{C}$.

Proof of Claim:. Define the pair of whole numbers a', b' in direct analogy to the pair $a_{\mathfrak{i}_0}, b_{\mathfrak{i}_0}$: Namely, a' is the genus of the part of F' lying above S_{s_0} and b' is the genus of the part of F' lying below S_{s_0} We will now show by construction that there is a close relation between the pairs a', b' and $a_\mathfrak{i}, b_\mathfrak{i}$.

We can assume, with no loss, that $s_0 > e$. (If $s_0 < e$ reverse the rolls of a and b in the following argument.). Let $\mathfrak{c} \subset S_{s_0}$ be the collection of circles $S_{s_0} \cap E$. Decomposing $\mathbb{C}$ along E to obtain $F' = F(\mathbb{C}')$ can be viewed as a 3-stage process:

- first remove a small annular neighborhood $\eta(\partial E)$ of ∂E from F, then
- attach a pushed-off copy of E to each boundary component of $F - \eta(\partial E)$; call this added pair of disks $2E$. The resulting surface F_η intersects the level sphere S_{s_0} in the curves $(F \cap S_{s_0}) \cup 2\mathfrak{c}$, where $2\mathfrak{c}$ denotes two parallel copies of $\mathfrak{c}$, one copy for each copy of E in $2E$.
- Finally, remove any resulting sphere components of F_η that bound goneballs.

Observe the behavior of the genus through each of these three stages of the construction. We consider the stages in reverse order: Removing spheres from F_η does not change the genus of the part of the surface lying above S_{s_0} or below, since each component removed is a subsurface of spheres and is therefore planar. Similarly, adding $2E$ does not affect the genus of the surface above or below, since each component added is a subsurface of $2E$ and therefore planar, and the ends of the annulus $\eta(\partial E)$ are attached to different planar subsurfaces of $2E$. Finally, removing $\eta(\partial E)$ from F does not affect the genus of the part of F above S_{s_0}, and, though it may lower the genus of the part of F below S_{s_0} (if it is non-separating in the component in which it lies), it lowers it by at most one.

We conclude that $a' = a_{i_0}$ and $b_{i_0} - 1 \le b' \le b_{i_0}$. In particular

- If $\frac{a_{i_0}}{b_{i_0}} = \frac{0}{0}$ then $\frac{a'}{b'} = \frac{0}{0}$.
- If $\frac{a_{i_0}}{b_{i_0}} = \frac{1}{1}$ then $\frac{a'}{b'} = \frac{1}{1}$ or $\frac{1}{0}$.
- If $\frac{a_{i_0}}{b_{i_0}} = \frac{m}{m}, m \ge 2$ then $\frac{a'}{b'} = \frac{m}{m}$ or $\frac{m}{m-1}$.

In any case, we see that the sphere S_{s_0} is balanced or almost balanced for $\mathbb{C}'$, proving the Claim.

The proof now proceeds by induction on $|\mathfrak{c}|$, the number of curves of intersection of E with S_{s_0}. If E is disjoint from S_{s_0}, so $|\mathfrak{c}| = 0$ the result follows from Theorem 13.3, as augmented by Corollary 15.11.

Assume then that $|\mathfrak{c}| \ge 1$ and let

$$\overrightarrow{(\mathbb{C}, S_{s_0})}: \quad \mathbb{C} \xrightarrow{\mathcal{D}_0} \mathbb{C}_1 \xrightarrow{\mathcal{D}_1} \dots \xrightarrow{\mathcal{D}_{n-1}} \mathbb{C}_n$$

and

$$\overrightarrow{(\mathbb{C}', S_{s_0})}: \quad \mathbb{C}' \xrightarrow{\mathcal{D}'_0} \mathbb{C}'_1 \xrightarrow{\mathcal{D}'_1} \dots \xrightarrow{\mathcal{D}'_{n-1}} \mathbb{C}'_n$$

be complete flagged chamber complex decomposition sequences guided by S_{s_0} for $\mathbb{C}$ and $\mathbb{C}'$ respectively. (First add the circles $\mathfrak{c}$ as ghost circles for the decomposition sequence for $\mathbb{C}$, so that both sequences use the same set of guiding disks throughout, and so run in parallel. By Corollary 15.10 the addition of these ghost circles does not change how $\mathbb{C}$ certifies.) Focus attention on the first decomposition $\mathbb{C}'_i \xrightarrow{\mathcal{D}'_i} \mathbb{C}'_{i+1}$ in which a circle in $\mathfrak{c}$ appears as the boundary of a disk in $\mathcal{D}'_i$. Let $\mathfrak{c}_0 \in \mathfrak{c}$ be such a circle, chosen among all candidates to be innermost in $2E$; let $D_0 \in \mathcal{D}'_i$ be the disk that $\mathfrak{c}_0$ bounds in S_{s_0}; and let $E_0 \subset 2E$ be the disk that $\mathfrak{c}_0$ bounds in $2E$.

Although E_0 may intersect S_{s_0} in other components of $\mathfrak{c}$, by construction none bound disks in any $\mathcal{D}'_j, j \le i$, so E_0 remains intact as part of the defining surface $F'_i = F(\mathbb{C}'_i)$. Consider whether D_0 is an essential disk in the chamber of $\mathbb{C}'_i$ in which it lies:

If D_0 is an essential disk, then the ball $B_0 \subset S^3$ that the sphere $S = D_0 \cup E_0$ bounds (on the side in S^3 not containing ∂E) contains components of F'_i. By construction, B_0 is disjoint from $2E$ so these components of F'_i must also be components of the defining surface $F_i = F(\mathbb{C}_i)$. Thus S is an incompressible sphere in a chamber of both $\mathbb{C}'_i$ and $\mathbb{C}_i$. This implies that $\mathbb{C}'_i \sim \mathbb{C}_i$, hence $\overrightarrow{(\mathbb{C}, S_{s_0})} \sim \overrightarrow{(\mathbb{C}', S_{s_0})}$ and so, following Definition 17.3, $\mathbb{C} \sim \mathbb{C}'$ as required.

On the other hand, if D_0 is an inessential disk, then $\text{int}(B_0)$ is disjoint from F'_i so B_0 is disky in $\hat{\mathbb{C}}'_{i+1}$ and so (perhaps with an appropriate choice of sibling) becomes a goneball in $\mathbb{C}'_{i+1}$. In other words, after decomposition by $\mathcal{D}'_i$ (and the similar elimination of the parallel copy of E_0 in $2E$ at the next stage of the decomposition sequence) we obtain the same chamber complexes as if E_0 had been replaced by D_0 in the disk E. But this replacement lowers $|\mathfrak{c}|$ by at least 1. By our inductive assumption on $|\mathfrak{c}|$, $\mathbb{C} \sim \mathbb{C}'$, as required.

To provide more detail in this last case, when D_0 is inessential, let E' be the disk that results from replacing the disk $E_0 \subset E$ by D_0 as just described. Recall the following diagram for the inductive step in the proof of Proposition 13.2, which in turn enables the proof of Theorem 13.3:

$$
\begin{array}{ccccccc}
\dots \xrightarrow{\mathcal{D}_{i-1}} & \mathbb{C}_i & \xrightarrow{\mathcal{D}_i} & \mathbb{C}_{i+1} & \xrightarrow{\mathcal{D}_{i+1}} \dots & & = \vec{\mathbb{C}} \\
 & \Big\downarrow & & \Big\downarrow E & & & \\
 & E\ \Big\downarrow & & \mathbb{C}^{i+1}_{i+1} & \xrightarrow{\mathcal{D}^{i+1}_{i+1}} \dots & & = \vec{\mathbb{C}}^{i+1}_E \\
 & \Big\downarrow & & \Updownarrow & & & \\
 & \mathbb{C}^i_i & \xrightarrow{\mathcal{D}^i_i} & \mathbb{C}^i_{i+1} & \xrightarrow{\mathcal{D}^i_{i+1}} \dots & & = \vec{\mathbb{C}}^i_E
\end{array}
$$

As the argument there would ultimately be applied here, the decomposing disks are all guided by S_{s_0}. (One consequence is that the notation for the decomposing disks in the diagram can all be simplified to those for the guiding disks in the top row.) The problem with further use of that inductive step here is that for $j > i$ the decomposing disks $\mathcal{D}_j$ might pass through E. This means that their interiors may intersect $F(\mathbb{C}'_j)$, making later inductive steps potentially nonsensical. However, since B_0 is a goneball in the decomposition $\mathbb{C}'_i \xrightarrow{\mathcal{D}_i} \mathbb{C}'_{i+1}$ nothing is lost by replacing E with E' to get the simplified diagram:

$$
\begin{array}{ccccccc}
\dots \xrightarrow{\mathcal{D}_{i-1}} & \mathbb{C}_i & \xrightarrow{\mathcal{D}_i} & \mathbb{C}_{i+1} & \xrightarrow{\mathcal{D}_{i+1}} \dots & & = \vec{\mathbb{C}} \\
 & E'\ \Big\downarrow & & & & & \\
 & \mathbb{C}^i_i & \xrightarrow{\mathcal{D}^i_i} & \mathbb{C}^i_{i+1} & \xrightarrow{\mathcal{D}^i_{i+1}} \dots & & = \vec{\mathbb{C}}^i_E
\end{array}
$$

Then observe that since $|E' \cap S_{s_0}| < |E \cap S_{s_0}| = |\mathfrak{c}|$ the inductive hypothesis on $|\mathfrak{c}|$ implies that $\overrightarrow{(\mathbb{C}_i, S_{s_0})} \sim \overrightarrow{(\mathbb{C}^i_i, S_{s_0})}$. This eliminates the need for the further inductive steps used in the proof of Proposition 13.2. □

Because Proposition 17.7 requires that $\mathbb{C}$ not be tiny, it cannot be directly applied to the flagged chamber complex defined by a Heegaard surface T, since both chambers are empty handlebodies. But it does tell us something crucial about flagged chamber complexes that are obtained from T by weak reduction:

Proposition 17.8. *Suppose $S^3 = A \cup_T B$ is a Heegaard splitting and $S_s, s \in [-1, 1]$ is a generic sweep-out of S^3 by spheres. Suppose $\mathcal{D}$ is a weakly reducing collection of disks for T and E is a properly embedded disk disjoint from $\mathcal{D}$, lying in either A or B. Suppose further that, for some generic $-1 < e < 1$, ∂E lies entirely in S_e. Let $\mathbb{C}$ be the flagged chamber complex obtained from T by decomposition along $\mathcal{D}$ and $\mathbb{C}'$ be the flagged chamber complex obtained from T by decomposition along $\mathcal{D} \cup E$. Then $\mathbb{C}' \sim \mathbb{C}$.*

Proof. Let $\mathbb{C}_{DE}$ denote the flagged chamber complex obtained from $\mathbb{C}$ by decomposition along E. Proposition 17.7 does apply to this decomposition since, by Proposition 5.18, $\mathbb{C}$ is not tiny. Thus $\mathbb{C} \sim \mathbb{C}_{DE}$. It remains to prove a similar relation between $\mathbb{C}'$ and $\mathbb{C}_{DE}$.

We are in a position to apply Proposition 12.12, where the Heegaard splitting here plays the role of $\mathbb{C}$ in Proposition 12.12; $\mathbb{C}'$ here corresponds to $\mathbb{C}_{D+E}$ there; and the notation $\mathbb{C}_{DE}$ here was chosen to be consistent with the same notation there. Then, according to Proposition 12.12, either $\mathbb{C}_{DE} = \mathbb{C}'$ or $\mathbb{C}_{DE} \dashrightarrow \mathbb{C}'$. In the former case we are done, so we consider the case $\mathbb{C}_{DE} \dashrightarrow \mathbb{C}'$, and turn to Corollary 11.12.

Following Corollary 10.7 let S_{s_0} be a sphere in the sweep-out that is balanced for $\mathbb{C}'$ and let

$$\vec{\mathbb{C}}' : \quad \mathbb{C}' \xrightarrow{\mathcal{D}_0} \mathbb{C}'_1 \xrightarrow{\mathcal{D}'_1} \dots \xrightarrow{\mathcal{D}'_{n-1}} \mathbb{C}'_n$$

be a complete flagged chamber complex decomposition sequence guided by S_{s_0}. Similarly, let $\vec{\mathbb{C}}_{DE}$ be a complete flagged chamber complex decomposition sequence for $\mathbb{C}_{DE}$ guided by S_{s_0}. Since $\mathbb{C}_{DE} \dashrightarrow \mathbb{C}'$ their defining surfaces F_{DE} and F' differ by at most the insertion of a bullseye. Since the components of the bullseye are all spheres, the insertion does not affect the genus of the parts of the surface above and below S_{s_0}, so S_{s_0} is balanced for $\mathbb{C}_{DE}$ as well. In particular, by Proposition 10.2, both $\vec{\mathbb{C}}'$ and $\vec{\mathbb{C}}_{DE}$ certify.

Apply Corollary 11.12 to $\mathbb{C}_{DE} \dashrightarrow \mathbb{C}'$: Let $\vec{\mathbb{C}}^m_{DE}$ be the maximal deflationary sequence of $\mathbb{C}_{DE}$. If $\vec{\mathbb{C}}^m_{DE} = \vec{\mathbb{C}}_{DE}$ so the entire sequence is a deflationary sequence, then by the second outcome of Corollary 11.12, $\vec{\mathbb{C}}'$ and $\vec{\mathbb{C}}_{DE}$ cocertify. If, on the other hand, $\vec{\mathbb{C}}^m_{DE}$ is not the entire sequence $\vec{\mathbb{C}}_{DE}$ then by the first outcome of Corollary 11.12, $\vec{\mathbb{C}}^m_{DE}$ certifies, so by the second outcome, again $\vec{\mathbb{C}}'$ and $\vec{\mathbb{C}}_{DE}$ cocertify. In other words, $\overrightarrow{(\mathbb{C}_{DE}, S_{s_0})} \sim \overrightarrow{(\mathbb{C}', S_{s_0})}$ or, following Definition 17.3, $\mathbb{C}_{DE} \sim \mathbb{C}'$ as required. □

18 The Goeritz group is the eyeglass group

Suppose (S^3, T_g) is the standard genus $g \geq 2$ Heegaard splitting of Section 3 and τ is an element of the Goeritz group $G(S^3, T_g)$, as described in [JM]. Let $T_\theta \subset S^3, 0 \leq \theta \leq 2\pi$ be a representative of τ in $\pi_1(\mathrm{Img}(S^3, T))$, with $S^3 = A_\theta \cup_{T_\theta} B_\theta$ and $T_0 = T_{2\pi} = T_g$.

Put another way, τ is represented by an isotopy $\Theta_\theta : S^3 \to S^3, 0 \leq \theta \leq 2\pi$ with Θ_0 the identity and $\Theta_{2\pi}(T_g) = T_g$. T_θ then denotes the surface $\Theta_\theta(T_g) \subset S^3$ and $\Theta_{2\pi}$ represents $\tau : (S^3, T_g) \to (S^3, T_g)$. Similarly A_θ, B_θ denote respectively $\Theta_\theta(A), \Theta_\theta(B)$. Finally, for any $\theta, \theta' \in [0, 2\pi]$, denote the composition $\Theta_\theta \Theta^{-1}_{\theta'}$ by $\Theta^{\theta'}_\theta : (S^3, T_{\theta'}) \to (S^3, T_\theta)$. This definition implies that, for any $\theta'' \in [0, 2\pi]$, $\Theta^{\theta''}_\theta \Theta^{\theta'}_{\theta''} = \Theta^{\theta'}_\theta$.

We briefly review some of the results of [FS1], where S^3 is swept out by level spheres S_s of the standard height function $p : S^3 \to [-1, 1]$.

There are values $0 < \theta_1 < \theta_2 < \dots < \theta_n < 2\pi$ so that for each $\theta \notin \{\theta_i, 1 \leq i \leq n\}$ there is a pair of weakly reducing disks (a_θ, b_θ) associated to T_θ so that:

- $a_\theta \subset A_\theta$ and $b_\theta \subset B_\theta$
- The isotopy class of the pair (a_θ, b_θ) does not change throughout each interval in $[0, 2\pi] - \{\theta_i, 1 \leq i \leq n\}$. By this we mean, if θ, θ' are in the same interval of $[0, 2\pi] - \{\theta_i, 1 \leq i \leq n\}$ then the pair of disks $(\Theta^{\theta'}_\theta(a'_\theta), \Theta^{\theta'}_\theta(b'_\theta))$ is properly isotopic to the pair (a_θ, b_θ) in (S^3, T_θ).

A shorthand notation for this is $(a_\theta, b_\theta) = (a_{\theta'}, b_{\theta'})$.

- $(a_{2\pi}, b_{2\pi}) = (a_0, b_0)$. That is, the pair of disks $(a_{2\pi}, b_{2\pi})$ is properly isotopic to the pair (a_0, b_0) in (S^3, T_g).
- For every $0 \leq \theta \leq 2\pi$ each of the disks a_θ and b_θ have level boundaries. That is, each the circles ∂a_θ and ∂b_θ lies in a single sphere of the sweep-out by S_s.

The last property is used for the application of Proposition 17.8 in the proof of Lemma 18.2 below. It follows from the construction described in [FS1, Appendix], as applied in [FS1, Subsection 4.5]. The disks a_θ and b_θ are chosen in [FS1] by how their boundaries lie in $T_\theta \cap S$, for S one of the level spheres S_s. (The argument in [FS1] requires that genus$(T) \geq 2$.) Note that, unlike the collection of disks used in disk decomposition sequences that we have been long discussing, it is not part of the construction for the sweep-outs in [FS1] that the entire disk a_θ or b_θ lies in a single sphere of the sweep-out, only that the boundary of each disk does.

The third property above, that $(a_{2\pi}, b_{2\pi}) = (a_0, b_0)$ (the notation here differs somewhat from that in [FS1]) follows from the fact that $T_{2\pi} = T_0 = T_g$. Indeed, we might as well take the sweep out by copies of the Heegaard surface used in the construction to be the same in both cases, which means that the weakly reducing pair of disks arising from the construction will be given by the same rule.

For each $\theta \notin \{\theta_i, 1 \leq i \leq n\}$, define the flagged chamber complex $\mathbb{C}_\theta$ as that obtained from T_θ by weak reduction along the pair of disks (a_θ, b_θ).

Observe some properties:

- By Proposition 5.18 no chamber complex $\mathbb{C}_\theta$ is tiny.
- Suppose θ, θ' are in the same interval of $[0, 2\pi] - \{\theta_i, 1 \leq i \leq n\}$. Since the pair of disks $(\Theta_\theta^{\theta'}(a'_\theta), \Theta_\theta^{\theta'}(b'_\theta))$ is properly isotopic to the pair (a_θ, b_θ), the chamber complex $\Theta_\theta^{\theta'}(\mathbb{C}_{\theta'})$ is also properly isotopic to the chamber complex $\mathbb{C}_\theta$ in (S^3, T_θ).
- Since $(a_{2\pi}, b_{2\pi}) = (a_0, b_0)$ we have $\mathbb{C}_{2\pi} = \mathbb{C}_0$.

Finally, for each $\theta \notin \{\theta_i, 1 \leq i \leq n\}$ define $h_\theta : (S^3, T_\theta) \to (S^3, T_g)$ to be $h_{\mathbb{C}_\theta}$ as defined following Corollary 17.2. (Technically, h_θ is only defined up to eyeglass equivalence, since that is the case for $h_{\mathbb{C}_\theta}$.) We have

Lemma 18.1. *Suppose θ, θ' are in the same interval of $[0, 2\pi] - \{\theta_i, 1 \leq i \leq n\}$. Then $h_\theta \sim h_{\theta'} \Theta_{\theta'}^{\theta}$.*

Proof. As just observed the chamber complex $\Theta_\theta^{\theta'}(\mathbb{C}_{\theta'})$ is properly isotopic to the chamber complex $\mathbb{C}_\theta$ in (S^3, T_θ) so $h_\theta = h_{\mathbb{C}_\theta} = h_{\Theta_\theta^{\theta'}(\mathbb{C}_{\theta'})}$. Now apply Corollary 17.4: $h_{\Theta_\theta^{\theta'}(\mathbb{C}_{\theta'})} \Theta_\theta^{\theta'} \sim h_{\mathbb{C}_{\theta'}}$ so $h_\theta \sim h_{\mathbb{C}_{\theta'}} (\Theta_\theta^{\theta'})^{-1} = h_{\theta'} \Theta_{\theta'}^{\theta}$ □

Further following [FS1], for each $1 \leq i \leq n$ there are three disjoint disks, either

- $a_{\theta_i} \subset A_{\theta_i}, b_{\theta_i}^{\pm} \subset B_{\theta_i}$ with the pairs $(a_{\theta_i}, b_{\theta_i}^{+})$ and $(a_{\theta_i}, b_{\theta_i}^{-})$ each weakly reducing or
- symmetrically $a_{\theta_i}^{\pm} \subset A_{\theta_i}, b_{\theta_i} \subset B_{\theta_i}$ with the pairs $(a_{\theta_i}^{+}, b_{\theta_i})$ and $(a_{\theta_i}^{-}, b_{\theta_i})$ each weakly reducing.

with the property that (respectively for the two cases) for small ϵ

- $\Theta^{\theta_i}_{\theta_i\pm\epsilon}(a_{\theta_i}) = a_{\theta_i\pm\epsilon}$ and $\Theta^{\theta_i}_{\theta_i\pm\epsilon}(b^{\pm}_{\theta_i}) = b_{\theta_i\pm\epsilon}$
- symmetrically $\Theta^{\theta_i}_{\theta_i\pm\epsilon}(a^{\pm}_{\theta_i}) = a_{\theta_i\pm\epsilon}$ and $\Theta^{\theta_i}_{\theta_i\pm\epsilon}(b_{\theta_i}) = b_{\theta_i\pm\epsilon}$

Lemma 18.2. *Suppose θ,θ' are any two points in $[0,2\pi]-\{\theta_i, 1\le i\le n\}$. Then $h_\theta \sim h_{\theta'}\Theta^{\theta}_{\theta'}$.*

Proof. It suffices to prove the case in which θ,θ' are in adjacent intervals, for then we can just proceed around the circle. (See Figure 48 for a highly schematic picture.) So suppose that they are on adjacent intervals sharing the end point $\theta_i \in (0,2\pi)$. Following Lemma 18.1 it suffices to prove the case in which $\theta = \theta_i - \epsilon$ and $\theta' = \theta_i + \epsilon$, for small ϵ.

With no loss of generality assume that the three weakly reducing disks in (S^3,T_i) are $a_{\theta_i} \subset A_{\theta_i}, b^{\pm}_{\theta_i} \subset B_{\theta_i}$ and let $\mathbb{C}_i$ be the (not tiny) chamber complex obtained by weakly reducing T_{θ_i} along the three disks $a_{\theta_i}, b^{\pm}_{\theta_i}$. Let $\mathbb{C}^+_i$ be that obtained by weakly reducing along just the pair $a_{\theta_i}, b^+_{\theta_i}$ and $\mathbb{C}^-_i$ be that obtained by weakly reducing along just $a_{\theta_i}, b^-_{\theta_i}$. Apply Proposition 17.8 to deduce that $\mathbb{C}^-_i \sim \mathbb{C}_i$. Similarly, $\mathbb{C}^+_i \sim \mathbb{C}_i$, so as a result $\mathbb{C}^+_i \sim \mathbb{C}^-_i$.

The argument of Lemma 18.1 further shows that $h_{\mathbb{C}^+_i} \sim h_{\theta_i+\epsilon}\Theta^{\theta_i}_{\theta_i+\epsilon}$ and $h_{\mathbb{C}^-_i} \sim h_{\theta_i-\epsilon}\Theta^{\theta_i}_{\theta_i-\epsilon}$ so, combining, $h_{\theta_i-\epsilon}\Theta^{\theta_i}_{\theta_i-\epsilon} \sim h_{\theta_i+\epsilon}\Theta^{\theta_i}_{\theta_i+\epsilon}$. Hence $h_{\theta_i-\epsilon} \sim h_{\theta_i+\epsilon}\Theta^{\theta_i}_{\theta_i+\epsilon}\Theta^{\theta_i-\epsilon}_{\theta_i} \sim h_{\theta_i+\epsilon}\Theta^{\theta_i-\epsilon}_{\theta_i+\epsilon}$ as required. □

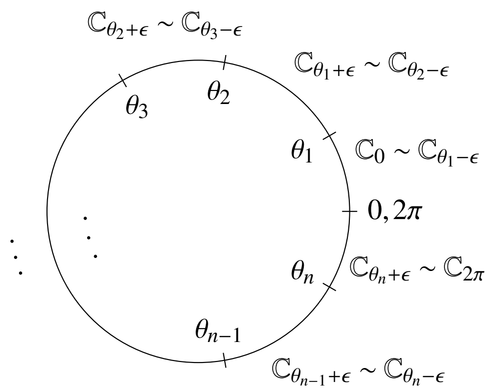

Figure 48: Schematic for Lemmas 18.1 and 18.2

Corollary 18.3. $\tau \in \mathcal{E}$

Proof. Choose, in Lemma 18.2 the values $\theta = 0, \theta' = 2\pi$. Then, per that lemma, $h_0 \sim h_{2\pi}\Theta_{2\pi}\Theta_0^{-1}$. Furthermore, since $\mathbb{C}_0 = \mathbb{C}_{2\pi}$, Corollary 8.3 has $h_0 \sim h_{\mathbb{C}_0} \sim h_{\mathbb{C}_{2\pi}} \sim h_{2\pi}$, and so $\tau = \Theta_{2\pi} \sim \Theta_0 = id_{(S^3,T_g)}$. □

Remark: It is perhaps not surprising that it is possible to position T_g with respect to the height function on S^3 so that h_0 and $h_{2\pi} : (S^3,T_g) \to (S^3,T_g)$ are not just eyeglass equivalent, but in fact are both the identity (under the standard inductive Assumption 3.4. We do not need this here.

Corollary 18.4. $G(S^3,T) = \mathcal{E}$.

Index

Acknowledgments

Michael Freedman has been an avid source of support for this project; his notion of a 'cycle of weak reductions' underpins it all. Aoife McCormick inspired the notion of 'chamber complex' - the method here for translating Freedman's idea into a recipe for Heegaard surface reduction.

References

[Bo] F Bonahon, *Cobordism of automorphisms of surfaces*, Ann. Sci. École Norm. Sup.(4) **16** 1983) 237–270

[Ce] J Cerf, *La stratification naturelle des espaces de fonctions différentiables rélles et le théorème de la pseudo-isotopie*, Publications mathématiques de l'I.H.É.S., tome 39 (1970) 5-173.

[CG] A. Casson and C. Gordon, *Reducing Heegaard splittings*, Topology and its applications, **27** (1987) 275-283. 210, 220, 233

[EE] C. Earle and J. Eells, *The diffeomorphism group of a compact Riemann surface*, Bull. Amer. Math. Soc. **73** (1967) 557–559 216

[FS1] M. Freedman and M. Scharlemann, *Powell moves and the Goeritz group*, Arxiv preprint 1804.05909 209, 210, 211, 213, 214, 225, 330, 331

[FS2] M. Freedman and M. Scharlemann, *Uniqueness in Haken's Theorem*, Mich. Math. J. **74** (2024) 119-142. 213, 219, 222, 223, 225, 234, 249, 251, 252, 253, 261, 262, 263, 269

[GK] D. Gay and R. Kirby, *Fiber-connected, indefinite Morse 2-functions on connected n-manifolds*, Proc. Natl. Acad. Sci. USA **108** (2011), no. 20, 8122–8125. 323

[GP] V. Guillemin and A. Pollack, *Differential Topology*, Prentice-Hall, 1974. 215

[Ha] A Hatcher, *Homeomorphisms of sufficiently large P^2-irreducible 3-manifolds*, Topol- ogy **15** (1976), 343–347.

[JM] J. Johnson and D. McCullough, *The space of Heegaard splittings*, J. Reine Angew. Math **679** (2013) 155-179 209, 212, 330

[La] M. Lackenby, *An algorithm to determine the Heegaard genus of simple 3–manifolds with nonempty boundary*, Alg. Geom. Top. **8** (2008) 911-934. 219, 220, 233, 235, 237, 245

[Lau] F. Laudenbach, *Sur les 2-sphères d'une variété de dimension 3*, Ann. of Math. (2) **97** (1973) 57–81. 323

[Mi] J. Milnor, *Morse Theory (Based on lecture notes by M. Spivak and R. Wells)*, Annals of Mathematics Studies, **51** Princeton University Press, Princeton, N.J., 1963. 323

[Moi] E. E. Moise., *Geometric topology in dimensions 2 and 3*, Graduate Texts in Mathematics, **47** Springer-Verlag, New York - Heidelberg - Berlin, 1977. 211

[Pa] R. Palais, *Local triviality of the restriction map for embeddings*, Comment. Math. Helv. **34** (1960) 305-312. 215

[Po] J. Powell, *Homeomorphisms of S^3 leaving a Heegaard surface invariant*, Trans. Amer. Math. Soc. **257** (1980) 193–216. 209, 211

[Sc1] M. Scharlemann, *A Strong Haken's Theorem*, to appear in Alg. Geom. Top, Arxiv preprint 2003.08523. 213, 219, 221, 223, 224, 234, 242, 243, 245, 246, 251, 252, 258, 281

[Sc2] M. Scharlemann, *One Powell generator is redundant*, Proc. Amer. Math. Soc., Series B **7** (2020) 138-141. 209

[Sc3] M. Scharlemann, Powell's Conjecture on the Goeritz group of S^3 is stably true, to appear in Alg. Geom. Top, Arxiv preprint arXiv:2210.13629. 210

[Sp] M. Spivak, *Calculus on manifolds*, Avalon Publishing, 1965. 215

[Wa] F. Waldhausen, *Heegaard-Zerlegungen der 3-Sphäre*, Topology **7** (1968) 195–203. 209, 211, 218, 222, 251, 261

[Wh] J. H. C. Whitehead, *On equivalent sets of elements in a free group*, Ann. of Math. **37** (1936), 782–800.

AUTHOR

Martin Scharlemann
Mathematics Department
University of California
Santa Barbara, CA 93106-3080 USA

mgscharl@math.ucsb.edu
https://web.math.ucsb.edu/~mgscharl