

Bing Topology and Casson Handles

2013 Santa Barbara/Bonn Lectures

Michael H. Freedman & Team Freedman

Last update: June 18, 2013

Preface

In January and February 2013 Mike Freedman gave a series of 12 lectures at UC Santa Barbara with the goal to explain his proof of the 4-dimensional Poincaré conjecture. The lectures were broadcast live to the Max Planck Institute for Mathematics (MPIM) in Bonn as part of the *Semester on 4-manifolds and their combinatorial invariants* where Frank Quinn and Peter Teichner ran supplementary discussion sessions. Among the Santa Barbara audience was Bob Edwards who not only contributed various helpful remarks but also stepped in as a guest lecturer and presented his take on “the Design” which is probably the most confusing piece of the proof.

The lectures have been recorded and made available online at the MPIM homepage. In addition, Peter Teichner had the idea to create lecture notes. The original idea was to take handwritten notes and make them available directly after the lecture but after the first lecture it became clear to the designated note taker, Teichner’s Ph.D. student Stefan Behrens, that this was not feasible. Instead he decided to create L^AT_EX notes with the help of the video recordings. The result was a mixture of a word-by-word transcription of the lectures and more or less successful attempts to convert some of the more pictorial arguments into written form. Later on, the rough draft of the notes was revised and significantly improved in a collaborative effort of the MPIM audience. The following people were involved in this process:

Chapters 1 & 2	Stefan Behrens and Peter Teichner
Chapter 3	Henrik Rüping
Chapter 4	Xiaoyi Cui and Nathan Sunukjian
Chapter 5	Daniele Zuddas
Chapter 6	Matt Hogancamp and Ju A. Lee
Chapter 7	Thomas Vogel
Chapter 8	Wojciech Politarczyk and Mark Powell
Chapters 9 & 10	Stefan Behrens and Daniel Kasprowski

However, it goes without saying that most of the credit is due to Mike Freedman who did a sublime job at presenting this material which had become considered as almost impossible to understand. It was a pleasure listening to him explain the beautiful ideas involved in his proof and how they all came together. He managed to make things that had been considered almost impossible to understand seem accessible and not scary at all.

Hopefully, these notes along with the videos will make this high point in 4-manifold topology more accessible in the future.

Contents

Preface	2
I Bing Topology	5
1 The Schoenflies theorem after Mazur and Morse	5
1.1 The Schoenflies problem	5
1.2 Mazur's Schoenflies theorem	5
1.3 Removing the standard spot hypothesis	8
2 The Schoenflies theorem via the Bing shrinking principle	10
2.1 Shrinking cellular sets	10
2.2 Brown's proof of the Schoenflies theorem	13
2.3 The Bing shrinking criterion	13
3 Decomposition space theory and shrinking: examples	15
3.1 Some decomposition space theory	15
3.2 Applying the Shrinking criteria to shrink an "X"	17
3.3 Three descriptions of the Alexander gored ball	17
3.4 The Bing decomposition: the first shrink ever	20
3.5 Something that cannot be shrunk	23
3.6 The Whitehead decomposition	25
4 Decomposition space theory and shrinking: more examples	26
4.1 Starlike-equivalent sets and shrinking	27
4.2 A slam dunk for the Bing shrinking criterion	32
4.3 Mixing Bing and Whitehead	33
5 The ball to ball theorem	34
5.1 The main idea of the proof	35
5.2 Relations	36
5.3 Iterating the main idea: admissible diagrams	37
5.4 Proof of the ball to ball theorem	39
II Casson handles	41
6 From the Whitney trick to Casson handles	41
6.1 The Whitney trick in dimension 4	41
6.2 4-manifolds in the early 1970s: surgery and h-cobordism	43
6.3 Finding dual spheres	45
6.4 Casson handles and Kirby calculus	48
7 Exploring Casson handles	53
7.1 Picture Camp	53
7.2 The boundary of a Casson handle	59
7.3 An exercise in wishful thinking	60
7.4 A glimpse at reimbedding and the Design	61

8	Combinatorics: constructing infinite convergent towers	64
8.1	Gropes and transverse spheres	64
8.2	Height raising and reimbedding for Casson towers	69
8.3	Height raising for gropes	73
8.4	Where Casson left us	76
8.5	Reimbedding in the grope world: tower height raising with control	77
8.6	The disc embedding theorem	81
9	Geometric control and the Design	84
9.1	Where Casson left us off	84
9.2	Towers, CEQFAS handles and reimbedding	85
9.3	The Design in CEQFAS handles	88
9.4	Embedding the Design in the standard handle	92
9.5	Holes, gaps and the Endgame	94
10	Epilogue: Edwards' original shrink	98

Part I

Bing Topology

1 The Schoenflies theorem after Mazur and Morse

In the 1950s there was pervasive pessimism about the topological category because nobody knew how to tackle even the simplest problems without smooth or piecewise linear charts. The watershed moment was in 1959 when Mazur gave his partial proof of the Schoenflies conjecture [Maz59].

1.1 The Schoenflies problem

The Schoenflies problem is a fundamental question about spheres embedded in Euclidean space. We denote the d -dimensional Euclidean space by \mathbb{R}^d , the unit ball by $B^d \subset \mathbb{R}^d$ and the sphere by $S^d = \partial B^{d+1} \subset \mathbb{R}^{d+1}$. We phrase the Schoenflies problem as a conjecture.

Conjecture 1.1 (Schoenflies). *Any continuous embedding of S^d into the \mathbb{R}^{d+1} extends to an embedding of B^{d+1} .*

In 1914, the 1-dimensional case of the conjecture was proven in full generality by Caratheodory and Osgood-Taylor using elaborate methods from complex analysis. The 2-dimensional case was studied in the 1920s by Alexander who first circulated a manuscript claiming a proof but soon discovered a counterexample himself, the famous *horned sphere*. Later Alexander found an extra condition under which the Schoenflies conjecture holds in dimension 2. This extra condition, the existence of a *bicollar*, makes sense in arbitrary dimensions; it means that the embedding of S^d should extend to an embedding of $S^d \times [-1, 1]$ in which S^d sits as $S^d \times \{0\}$.

In the 40 years that followed almost no progress was made which led to much dismay about the topological category until Mazur gave his argument.

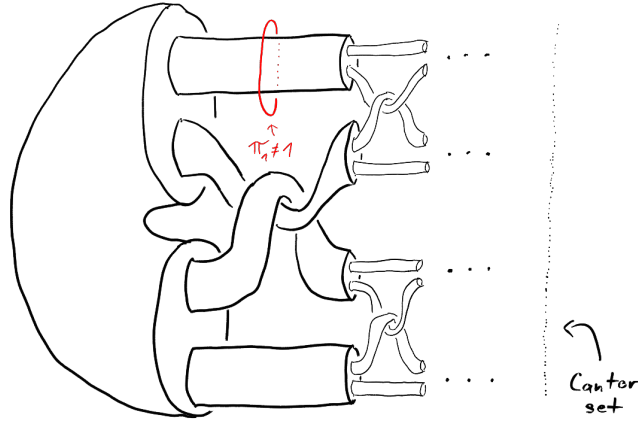


Figure 1: Alexander's horned sphere. (The red circle is not contractible in the exterior.)

1.2 Mazur's Schoenflies theorem

Consider a bicollared embedding $i: S^d \times [-1, 1] \rightarrow \mathbb{R}^{d+1}$ and fix a point $p \in S^d = S^d \times \{0\}$. Possibly after translation we can assume that $i(p) = 0$. We say that p is a *standard spot*

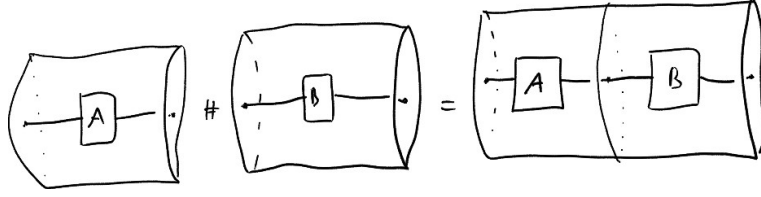


Figure 2: Adding knots.

of i if the following condition is satisfied. There is a disk $D \subset S^d$ around p such that, if we consider \mathbb{R}^{d+1} as $\mathbb{R}^d \times \mathbb{R}$, then

1. i maps $D \times 0$ to $\mathbb{R}^d \times 0$ and
2. for each $q \in D$ the interval $q \times [-1, 1]$ is mapped to $i(q) \times [-1, 1] \subset \mathbb{R}^d \times \mathbb{R}$.

Roughly, this means that i is “as standard as possible” around p .

Theorem 1.2 (Mazur [Maz59]). *Let $i: S^d \times [-1, 1] \rightarrow \mathbb{R}^{d+1}$ be a bicollared embedding which has a standard spot. Then i extends to an embedding of B^{d+1} .*

The strategy of Mazur’s proof – which is given in Chapter 1.2.2 below – was based on the *Eilenberg swindle* which is an observation in commutative algebra.

1.2.1 The Eilenberg swindle

Let A be a projective module (over some ring) written as a summand $A \oplus B = F$ of a free module F . Then on the one hand we have

$$(A \oplus B) \oplus (A \oplus B) \oplus (A \oplus B) \oplus \dots \cong F^\infty$$

while, on the other hand, a different grouping of the summands gives

$$A \oplus (B \oplus A) \oplus (B \oplus A) \oplus (B \oplus A) \oplus \dots \cong A \oplus F^\infty$$

since the direct sum is associative and commutative (up to isomorphism). Consequently, A becomes a free module after direct summing with an infinitely dimensional free module. In other words, a projective module is *stably free* in the infinite dimensional context.

Before going into Mazur’s proof we take a look at a warm up example of an application of this principle in topology.

Example 1.3 (Do knots have inverses?). Knots in \mathbb{R}^3 (or S^3) can be added by forming connected sums. The question is whether, given a knot A , there is a knot B such that $A \# B$ is isotopic to the trivial knot I , denoted by $A \# B \cong I$.

If we think of knots as strands in a cylinder connecting one end to the other, then the connect sum operation is realized by simply stacking cylinders next to each other. Note that this operation is also commutative and associative. Indeed, the associativity is obvious and to see the commutativity we can simply shrink B so that it becomes very small compared to A , then slide it along A and let it grow again.

Now let us assume that a knot A has an inverse B , ie $A \# B \cong I$. In this case the swindle works as follows. We sum infinitely many copies of $A \# B$ and think of them as living in a cone which, in turn, lives in a cylinder (see Figure 3). Then we clearly have

$$(A \# B) \# (A \# B) \# (A \# B) \# \dots \cong I^\infty \cong I$$

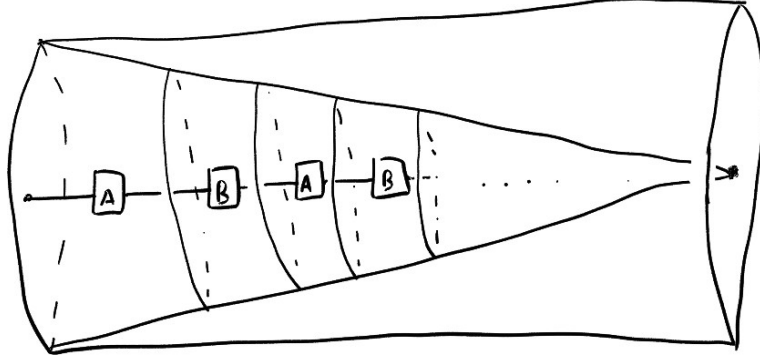


Figure 3: Stacking infinitely many copies of $A\#B$ in a cone.

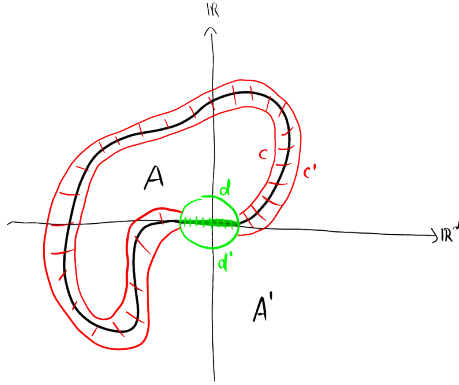


Figure 4: The decomposition used in Mazur's proof.

while a different grouping gives

$$A\#(B\#A)\#(B\#A)\#(B\#A)\#\dots \cong A\#I^\infty \cong A$$

which proves that A must be the trivial knot.

Remark 1.4. Besides the fact that there are easier proof for the fact that non-trivial knots don't have inverses, the above proof has another drawback in that it "looses category". For example, we might have started with smooth or piecewise linear knots but the conclusion holds only in the topological category since the isotopy we constructed will not be smooth or piecewise linear at the cone point.

1.2.2 The proof of Theorem 1.2

By passing to the one point compactification we can consider i as a flat embedding $S^d \hookrightarrow S^{d+1}$ and we cut out a small $(d+1)$ -ball around the standard spot. What is left is another $(d+1)$ -ball which is cut into two pieces A and A' by the image of i . The standard spot hypothesis implies that the boundary of A (resp. A') is a d -sphere which is decomposed into two standard d -balls c and d (resp. c' and d'), see Figure 4

By construction we have $A \cup_{c \sim c'} A' \cong B^{d+1}$ and, after noticing that gluing two balls along balls (of one dimension lower) gives another ball, it follows that

$$\begin{aligned} B^{d+1} &\cong (A \cup_{c \sim c'} A') \cup_{d' \sim d} (A \cup_{c \sim c'} A') \cup_{d' \sim d} (A \cup_{c \sim c'} A') \dots \\ &\cong A \cup_{c \sim c'} (A' \cup_{d' \sim d} A) \cup_{c \sim c'} (A' \cup_{d' \sim d} A) \cup_{c \sim c'} (A' \cup_{d' \sim d} A) \dots \end{aligned}$$

where the regrouping is justified by the associativity of gluing.

This is the setup for another swindle and in order to make it work we have to show that the identification $A' \cup_{d' \sim d} A$ is also a ball. To see this, note that c and d are isotopic in ∂A and, since ∂A is collared by the standard spot hypothesis, this isotopy can be extended to an isotopy of A in the standard way. Analogous results hold for c' and d' in $\partial A'$ and from this we can construct a homeomorphism from $A \cup_{c \sim c'} A'$ to $A \cup_{d \sim d'} A'$ and, since gluing is symmetric, $A' \cup_{d' \sim d} A$ is a ball and the swindle tells us that A is a ball. In fact, by reversing the roles of A and A' we also see that A' is a ball.

Finally, we can glue back in the neighborhood of the standard spot and we see that $S^{d+1} \setminus i(S^d)$ is the union of two open balls which proves the theorem.

1.3 Removing the standard spot hypothesis

After Mazur's work, there was a lot of interest in removing the standard spot hypothesis. This was done in 1960 in a paper by Marston Morse [Mor60].

Theorem 1.5 (Morse [Mor60]). *Any flat embedding $S^d \rightarrow \mathbb{R}^{d+1}$ has a standard spot after applying a homeomorphism of \mathbb{R}^{d+1} .*

Combining the results of Mazur and Morse we immediately get:

Theorem 1.6 (generalized Schoenflies theorem). *Any flat embedding of S^d into \mathbb{R}^{d+1} extends to an embedding of B^{d+1} .*

However, by the time Morse had completed Mazur's argument, Brown had already given an independent proof of Theorem 1.6 which will be discussed in Chapter 2.

Morse used a technique called *push-pull* which was common knowledge around that time. This technique is very general and is valuable in its own right.

1.3.1 Push-pull

We will introduce the technique by proving a theorem which uses it.

Lemma 1.7 (Application of push-pull). *Let X and Y be two compact metric spaces. If $X \times \mathbb{R}$ is homeomorphic to $Y \times \mathbb{R}$, then $X \times S^1$ is homeomorphic to $Y \times S^1$.*

Proof. Let $h: X \times \mathbb{R} \rightarrow Y \times \mathbb{R}$ be a homeomorphism. Then $Y \times \mathbb{R}$ has two product structures, the intrinsic one and the one induced from $x \times \mathbb{R}$ via h .

By compactness we can find $a, b, c, d \in \mathbb{R}$ such that

- $Y_a = Y \times a$, $Y_c = Y \times c$, $X_b = h(X \times b)$ and $X_d = h(X \times d)$ are disjoint in $Y \times \mathbb{R}$ and
- $X_b \subset Y \times [a, c]$ and $Y_c \subset h(X \times [b, d])$

as illustrated on the left side of Figure 5. The idea is to find a homeomorphism Π of $Y \times \mathbb{R}$ such that the composition $\Pi \circ h: X \times \mathbb{R} \rightarrow Y \times \mathbb{R}$ has enough periodicity to create a homeomorphism $X \times S^1 \rightarrow Y \times S^1$. We construct Π as a composition

$$\Pi = C^{-1} \circ P_Y \circ P_X \circ C$$

where the steps are illustrated in Figure 5. The maps P_X and P_Y will constitute the actual pushing and pulling while C , which we might call *cold storage*, makes sure that nothing is pushed or pulled unless it's supposed to be.

The maps are defined as follows:

- C rescales the intrinsic \mathbb{R} -coordinate of $Y \times \mathbb{R}$ such that $C(Y \times [a, c])$ lies below the level of X_b and leaves X_d untouched.

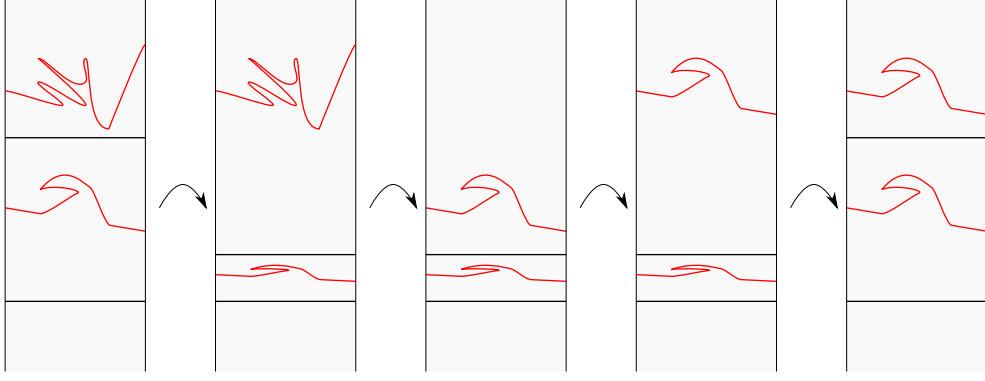


Figure 5: The push pull construction. The red lines indicate the X -levels.

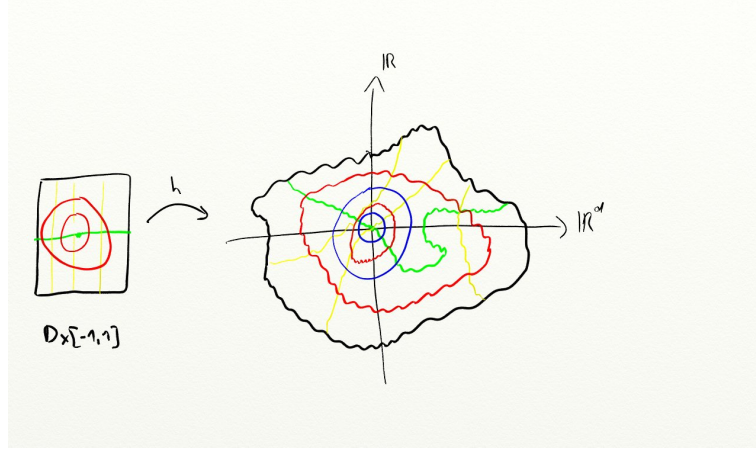


Figure 6: Creating a standard spot.

- P_X pushes X_d down to X_b along the \mathbb{R} -coordinate induced by h .
- P_Y pulls X_b up along the intrinsic \mathbb{R} -coordinate by the amount $c - a$.

Observe that Π leaves X_b untouched and that $\Pi(X_d)$ appears as a translate of X_b in the intrinsic \mathbb{R} -coordinate. By repeating this construction we can create periods with respect to both translations which proves the claim. \square

Exercise 1.8. (a) Fill in the details in the proof.

(b) Find spaces X and Y , such that $X \times S^1 \cong Y \times S^1$ but $X \times \mathbb{R} \not\cong Y \times \mathbb{R}$.

1.3.2 The proof of Theorem 1.5

Consider an embedding $h: S^d \times [-1, 1] \rightarrow \mathbb{R}^{d+1}$ and fix a point $p \in S^d \times 0$. By translation we can assume that $i(p) = 0$. We choose local coordinates on a disk $D \subset S^d$ containing p and observe that we get an induced local coordinate system on $h(D \times [-1, 1]) \subset \mathbb{R}^{d+1}$, see Figure 6.

In this new local coordinate system, the embedded sphere clearly has a standard spot, so it remains to extend it to a global coordinate system.

To achieve this we can use a push-pull argument. The idea is to compare the standard polar coordinates in \mathbb{R}^{d+1} with the ones induced by h . Again, by compactness we can find interlaced pairs standard spheres and curved spheres and, using push pull, we can find an isotopy such that transforms one of the curved sphere into a translate of the other along

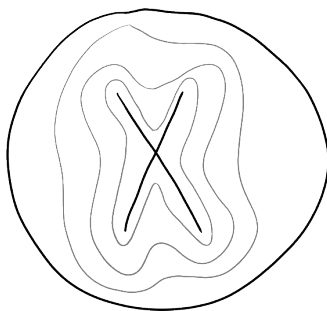


Figure 7: An example of a cellular set.

the standard radial coordinate and preserves a neighborhood of the origin. Moreover, now we can extend the local chart by periodicity to cover all of \mathbb{R}^{d+1} .

1.3.3 More about push-pull

A way to think of the push-pull argument is that it gives control over a homeomorphism in one linear direction. A major technical problem when working with topological manifolds is to gain control of a homeomorphism in many directions simultaneously. The culminating step in this direction was Kirby's work on the torus trick which we will take up later.

We end this lecture by giving some more applications, probably also due to Brown, of push-pull.

Theorem 1.9 (Brown). *A locally bicollared codimension one embedding is globally bicollared.*

Theorem 1.10 (Brown). *Collars for codimension one submanifolds are unique up to isotopy.*

2 The Schoenflies theorem via the Bing shrinking principle

In this lecture we are getting closer to the material that will be at the core of the 4-dimensional arguments, namely *decomposition space theory*. We will introduce these ideas through the notion of *cellular sets* (see Definition 2.1 below)

2.1 Shrinking cellular sets

Most of the following is due to Brown [Bro60]. We begin by introducing two central notions.

Definition 2.1. A subset $X \subset B^d$ is called *cellular* if it can be written as the intersection of countably many nested balls in B^d , that is, if there embedded d -balls $B_i \subset B^d$, $i = 1, 2, \dots$, such that $B_{i+1} \subset \text{int} B_i$ and $X = \bigcap_i B_i$.

As an example, Figure 7 illustrates that the letter X is a cellular subset of B^2 .

Exercise 2.2. Which capital letters are cellular sets?

Remark 2.3. Note that cellularity is not an intrinsic property of the space X but depends on the specific embedding.

We will see that the notion of cellular sets is closely related to maps whose point preimages are mostly singletons while some points have larger preimages. One can think of such maps as close to being homeomorphisms. The following terminology will be useful.

Definition 2.4. Let $f: X \rightarrow Y$ be a map. A point preimage $f^{-1}(y)$, $y \in Y$, is called an *inverse set* of f if it contains more than one element.

Eventually, we will obtain a characterization of cellular sets in terms of inverse sets. We start by showing that isolated inverse sets are cellular.

Lemma 2.5. Let $f: B^d \rightarrow \mathcal{B}^d$ be a continuous map between balls with a unique inverse set $X = f^{-1}(y)$, $y \in \mathcal{B}^d$ which is contained in the interior of B^d . Then X is cellular.

Note that we do not assume that f maps boundary to boundary.

Proof. We first observe that $y \in \mathcal{B}^d$ is an interior point and that we can find an $\epsilon > 0$ such that the standard ball $B_\epsilon(y) \subset \mathcal{B}^d$ around y is contained in the image of f .

Exercise 2.6. Convince yourself that this is true. (Hint: Invariance of domain)

Next we choose some homeomorphism $s_\epsilon: \mathcal{B}^d \rightarrow \mathcal{B}^d$ of the target ball which restricts to the identity on the smaller ball $B_{\epsilon/2}(y)$ and squeezes the rest of \mathcal{B}^d into $B_\epsilon(y)$. (Such a homeomorphism can be obtained, for example, by constructing an isotopy that moves y to 0 with which we then conjugate a suitable radial contraction.) Using this we now define a map $\sigma_\epsilon: B_s^d \rightarrow B_s^d$ from the source ball to itself by

$$\sigma_\epsilon(x) = \begin{cases} x & \text{if } x \in X \\ f^{-1} \circ s_\epsilon \circ f(x) & \text{if } x \notin X. \end{cases}$$

Note that σ_ϵ is well defined because f restricts to a injection $B^d \setminus X \rightarrow \mathcal{B}^d \setminus \{y\}$ and, by construction, s_ϵ does not map $f(x)$ to y . Moreover, one can show:

Exercise 2.7. Verify that σ_ϵ is injective, continuous and a homeomorphism onto its image.

The proof is finished by choosing a sequence $\epsilon = \epsilon_1 > \epsilon_2 > \dots > 0$ which converges to zero and the observation that the balls $B_i = \sigma_{\epsilon_i}(B^d) \subset B^d$ exhibit X as a cellular set. \square

The next result introduces the central idea of *shrinking* in the context of cellular sets.

Lemma 2.8 (Shrinking cellular sets). Let $X \subset B_d$ be a cellular set. For any $\epsilon > 0$ there exists a homeomorphism $h_\epsilon^X: B^d \rightarrow B^d$ such that

- (i) h_ϵ^X is the identity outside a ϵ -neighborhood of X .
- (ii) The diameter of $h_\epsilon^X(X)$ is less than ϵ .

Proof. Since X is cellular we can find a ball around X all of whose points have distance at most ϵ to X . Given such a ball, we can construct h_ϵ^X by a similar “radial squeeze” argument as in the previous proof. \square

Going one step further, we can not only shrink cellular sets but in fact crush them.

Lemma 2.9. If $X \subset B^d$ is cellular, then the quotient B^d/X is homeomorphic to B^d .

Proof. We choose a sequence $\epsilon_1 > \epsilon_2 > \dots > 0$ which converges to zero and define a map

$$g^X = \lim \left(\dots \circ h_{\epsilon_2}^X \circ h_{\epsilon_1}^X \right): B^d \rightarrow B^d$$

with h_ϵ^X given as in Lemma 2.8.

Exercise 2.10. Convince yourself that this limit exists in the space $C(B^d, B^d)$ of continuous self maps of B equipped with the sup-norm.

If $\pi: B^d \rightarrow B^d/X$ denotes the quotient map, then it is easy to see that the expression $\pi \circ g^{-1}$, which is a priori only a *relation*, actually defines a map. Moreover:

Exercise 2.11. $\pi \circ (g^X)^{-1}: B^d \rightarrow B^d/X$ is a homeomorphism.

□

Remark 2.12. Some remarks about the topology of quotients are in place. First of all, we will always equip quotient spaces with the *quotient topology* which means that the open sets in the quotient are exactly those sets whose preimages are open. Since all these lectures live in the world of compact metric spaces, it is important to note that it is not completely obvious how to obtain metrics on quotient spaces. What keeps us safe is the *Urysohn metrization theorem* which states that compact, second countable Hausdorff spaces are metrizable. Compactness and second countability are never an issue since these properties are preserved by dividing out closed sets, so all we have to check is the Hausdorff property which will be obvious in most cases we will encounter.

We want to point out two rather obvious properties of the map g^X which nevertheless have interesting consequences. On the one hand, it is easy to see that g^X maps X to a single point while it is injective in the complement of X . This gives another characterization of cellular sets.

Corollary 2.13. *A subset $X \subset B^d$ is cellular if and only if there is a map $f: B^d \rightarrow B^d$ as in Lemma 2.5 which has X as its unique inverse set. (In fact, one can assume that $B^d = B^d$, that f is surjective and that it restricts to the identity on a neighborhood of the boundary.)*

On the other hand, g^X is defined as the limit of homeomorphisms. Maps with this property are important enough to have a name.

Definition 2.14. A continuous map $f: X \rightarrow Y$ between complete metric spaces X and Y is called *approximable by homeomorphisms* or *ABH* if it can be written as the limit of homeomorphisms in the sup-norm.

Hence, from the above consideration we can deduce:

Corollary 2.15. *If $f: B^d \rightarrow B^d$ is the identity on the boundary and has a unique inverse set, then f is ABH.*

Remark 2.16. A more precise statement of Lemma 2.9 is that the projection $\pi: B^d \rightarrow B^d/X$ is ABH.

Proposition 2.17. *Let $f: B^d \rightarrow B^d$ be a map which is the identity on the boundary and has finitely many inverse sets X_1, \dots, X_N . Then:*

1. *Each X_i is cellular.*
2. *f is ABH.*
3. *The quotient $B^d/\{X_1, \dots, X_N\}$ is homeomorphic to B^d .*

Proof. We only treat the case $N = 2$, the rest is an easy induction. As in the proof of Lemma 2.5 we consider a conjugation of the form $f^{-1} \circ s_1 \circ f$ where s_1 is a squeeze for X_1 but this time we don't get a homeomorphism because of X_2 . Instead, we obtain a map which has X_2 as its unique inverse set. From this we see that X_2 is cellular and can thus be collapsed. Moreover, using Corollary 2.15 we can fix the above map to obtain a homeomorphism with which we can continue with the arguments in the proof of Lemma 2.5 eventually showing that X_1 is also cellular. Finally, the fact that f is ABH can easily be deduced from the local nature of the construction of g^X above. □

2.2 Brown's proof of the Schoenflies theorem

After this lengthy discussion we come back to the Schoenflies theorem (Theorem 1.6) which is almost turned into a one-line consequence by Proposition 2.17. For convenience we restate the Schoenflies theorem in a slightly different but equivalent form.

Theorem 2.18 (Brown [Bro60]). *Let $i: S^{d-1} \hookrightarrow S^d$ be an embedding which admits a bicollar. Then the closure of each component of $S^d \setminus i(S^{d-1})$ is homeomorphic to B^d .*

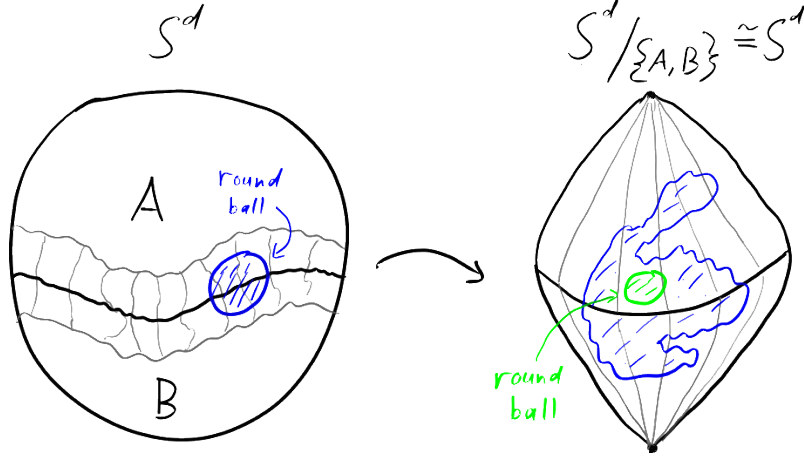


Figure 8: The setup of Brown's proof.

Proof. We extend i to an embedding $I: S^{d-1} \times [-1, 1] \hookrightarrow S^d$ and denote by $A, B \subset S^d$ the closures of the components of the complement of the image of I . Observe that the quotient $S^d / \{A, B\}$ is obviously homeomorphic to S^d since I essentially identifies it with the suspension of S^{d-1} which is homeomorphic to S^d (see Figure 8). We can thus write down the composition

$$S^d \xrightarrow{\pi} S^d / \{A, B\} \xrightarrow{\cong} S^d$$

which is a map with exactly two inverse sets, namely A and B . In order to apply Proposition 2.17 we have to reduce the situation to a map between balls. This is achieved by excising small, standard balls from the source and target spheres such that the ball in the source (blue in Figure 8) is contained in the image of I and the one in the target (green in Figure 8) is contained in the interior of the image of the one in the source. \square

2.3 The Bing shrinking criterion

The Bing shrinking criterion is a natural generalization of Lemma 2.8 where we showed how to shrink cellular sets. The idea of shrinking goes back to a paper of Bing in 1952 [Bin52] but was only formalized by Bob Edwards in his ICM talk in 1978 [Edw80] where he gave a very succinct statement of a shrinking criterion which had previously been absent in the literature.

Theorem 2.19 (Bing shrinking criterion à la Edwards 1978 [Edw80]). *Let $f: X \rightarrow Y$ be a map between complete metric spaces. Then f is ABH if and only if for any $\epsilon > 0$ there is a homeomorphism $h: X \rightarrow X$ such that*

- (i) $\forall x \in X: \text{dist}_Y(f(x), f \circ h(x)) < \epsilon$ and
- (ii) $\forall y \in Y: \text{diam}_X((f \circ h)^{-1}(y)) < \epsilon$.

The first condition roughly means that h is close to the identity as measure in the target space Y and the second condition controls the size of preimage sets. This is very similar to what we have seen in the process of shrinking cellular sets. The upshot of the theorem is that finding a coherent way of shrinking the preimage sets is equivalent to approximating by homeomorphisms.

Proof. The one direction is very easy. If we have a sequence of approximating homeomorphisms $h_n: X \rightarrow Y$, then the compositions of the form $h_n^{-1} \circ h_{n+k_n}$ will satisfy (i) and (ii) for a given $\epsilon > 0$ as long as n, k_n is large enough.

The other, more interesting direction can be proved by elementary methods but we will sketch a “Bourbaki style” proof due to Edwards. Consider the space $C(X, Y)$ with the sup-norm topology. This is well known to be a complete metric space. According to Edwards we consider the set

$$\{f \circ h \mid h: X \rightarrow X \text{ homeomorphism}\} \subset C(X, Y)$$

and denote by E its closure. Then E is also a metric space and, in particular, satisfies the *Baire category theorem* which states that the countable intersection of open and dense sets is still dense. For $\epsilon > 0$ we let E_ϵ be the subset of E of all maps whose inverse sets have diameter strictly less than ϵ . This is clearly an open set and the conditions (i) and (ii) eventually imply that E_ϵ is also dense in E . By the Baire category theorem the set $E_0 = \bigcap_{\epsilon > 0} E_\epsilon$ is dense in E , in particular, it is non-empty. But E_0 clearly consists of homeomorphisms because the inverse sets have to be points. \square

The relation to cellularity is given by the following easy exercise.

Exercise 2.20. If X, Y are compact metric spaces and $f: X \rightarrow Y$ is ABH, then all point preimages $f^{-1}(y)$ are cellular.

Remark 2.21. So far we have discussed the situation where we take a ball and a couple of subsets and we ask whether we can crush these sets to points and still get something homeomorphic to a ball. The general question in decomposition space theory is a little wilder. Usually we have some manifold and there may be infinitely many things to crush in it, possibly even uncountably many. To answer the question whether the quotient is homeomorphic to the original space there are two main tools. One of them is the Bing shrinking criterion with which we can try to shrink everything simultaneously in a controlled way. However, we will later see examples which indicate that this is a highly non-trivial problem. The problem is that, when we have infinitely many inverse sets, these might be linked in the sense that whenever we shrink some of them, some others will be stretched out, leading to a subtle and beautiful story which things can or cannot be shrunk.

In this discussion the Alexander’s horned sphere makes another prominent appearance and, in fact, it was main the focus of Bing’s paper [Bin52]. Bing’s motivation was the following. In the 1930s Wilder had constructed an interesting space and had asked whether it was homeomorphic to the 3-sphere. He considered the exterior of Alexander’s horned ball and took its double. Wilder’s interest was the fact that the doubled object has an obvious involution which merely exchanges the two halves. However, if this space turned out to be homeomorphic to S^3 , then this would give a very interesting involution on the 3-sphere whose fixed point set is a very wild 2-sphere. As a consequence, it would be a topological involution which is not conjugate to a smooth involution. While he could gather some evidence that his space was the 3-sphere, Wilder was unable to produce a conclusive proof and his question remained unanswered until Bing’s paper.

3 Decomposition space theory and shrinking: examples

3.1 Some decomposition space theory

We begin by setting up some basic terminology from *decomposition space theory*. An extensive account is given in Daverman's book [Dav86] although the terminology used therein differs slightly from ours.

Definition 3.1. A *decomposition* of a space X is a collection $\mathcal{D} = \{\Delta_i\}_{i \in I}$ of pairwise disjoint, closed subsets $\Delta_i \subset X$, the *decomposition elements*, indexed by a (possibly uncountable) index set I .

Given a decomposition \mathcal{D} of X , the *quotient space* or *decomposition space* of \mathcal{D} is the space X/\mathcal{D} obtained by crushing each set $\Delta \in \mathcal{D}$ to a point¹ and endow it with the quotient topology as in Remark 2.12. We generically denote quotient maps by $\pi: X \rightarrow X/\mathcal{D}$.

Remark 3.2. You may notice that our notion of decomposition is a slight abuse of language. Strictly speaking, a decomposition should be a partition of the whole space into pairwise disjoint sets. However, any decomposition as in Definition 3.1 can be completed to an honest decomposition by adding singletons to the decomposition for each point which was not contained in any decomposition element. Clearly, these singletons do not change the decomposition space or the quotient map so that they can essentially be ignored. **Equivalently we can think of it as an equivalence relation with closed equivalence classes.**

We will usually start with X being a compact, second countable, Hausdorff space (and thus metrizable by Urysohn's theorem) and we need a condition to guarantee that the X/\mathcal{D} stays within this nice class of spaces. As mentioned earlier, the only problem is the Hausdorff property.

Definition 3.3. Let X be a space with a decomposition \mathcal{D} . Given any subset $S \subset X$ we define its *(\mathcal{D} -)saturation* as

$$\pi \in v(\pi(X)) = (S \setminus \bigcup_{\Delta \in \mathcal{D}} \Delta) \cup (\bigcup_{\Delta \subset S} \Delta)$$

and say that S is *(\mathcal{D} -)saturated* if $S = \pi^{-1}(\pi(S))$. **The saturation is the smallest, saturated subset of X that contains S . We can also consider the largest saturated subset of X , defined by**

$$S^* := X \setminus \pi^{-1}\pi(X \setminus S).$$

In other words, a subset is saturated if it is the union of decomposition elements and points outside decomposition elements.

Definition 3.4. A decomposition \mathcal{D} of X is *upper semi-continuous* if each $\Delta \in \mathcal{D}$ has a saturated neighborhood system or, equivalently, if $U \subset X$ is an open subset, then U^* is also open.

Figure 9 illustrates some typical upper semi-continuous behavior of a decomposition and a failure thereof.

Exercise 3.5. Show that the two conditions in Definition 3.4 are in fact equivalent.

Exercise 3.6. Let \mathcal{D} be an upper semi-continuous decomposition of X . Show that, if X is second countable, then so is X/\mathcal{D} . Show the same for the Hausdorff property

¹Note that X/\mathcal{D} is different from $X/(\Pi_i \Delta_i)$ where all the Δ_i are crushed to a single point!

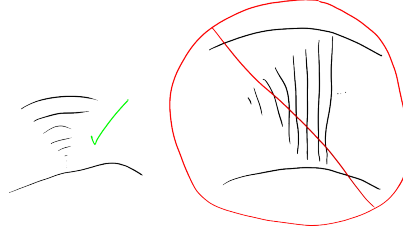


Figure 9: Upper semi-continuity and its failure

The latter exercise implies that, if we consider upper semi-continuous decompositions of compact metric spaces, then we never have to leave this nice class of spaces. However, there is usually no canonical metric on the quotient. It's tempting to try to “see” a metric on the quotient without going through any metrization theorems but it's usually not that easy.



Figure 10: The “middle third” construction of the Cantor set.

Example 3.7. An interesting 1-dimensional decomposition is to take the “middle third” construction of the Cantor set in the unit interval (see Figure 10) and to think of the closed middle third regions as the elements of the decomposition. It turns out that the quotient is again homeomorphic to the interval and one might try to measure length on the quotient by imagining being a taxi cab driving through the interval and turning the meter off inside decomposition elements. But the inconvenient thing is that the Cantor set has measure zero, so the naive taxi cab idea doesn't work.

Now that we have singled out the appropriate class of spaces and decompositions for our purposes, we introduce the important concept of *shrinkability*. With the Bing shrinking criterion (Theorem 2.19) in mind we make the following working definition.

Definition 3.8 (Shrinkability, working definition). An upper semi-continuous decomposition \mathcal{D} of a compact metric space X is *shrinkable* if the quotient map $\pi: X \rightarrow X/\mathcal{D}$ is ABH.

This definition certainly hides the actual shrinking but it avoids the (ultimately inessential) ambiguity of choosing a metric on the decomposition space on the other end of Theorem 2.19. We will usually appeal to the following shrinking criterion which easily follows from Theorem 2.19.

Corollary 3.9. Let \mathcal{D} be an upper semi-continuous decomposition of a compact metric space X . Assume that for any $\epsilon > 0$ there exists a homeomorphism $h: X \rightarrow X$ such that

- (i) h is supported in an ϵ -neighborhood of the decomposition elements and
- (ii) for each $\Delta \in \mathcal{D}$ we have $\text{diam}(h(\Delta)) < \epsilon$.

Then \mathcal{D} is shrinkable.

Remark 3.10. In its most general form, shrinkability can be defined for an arbitrary decomposition \mathcal{D} of an arbitrary space X by requiring that for any saturated open cover \mathcal{U} and any arbitrary open cover \mathcal{V} of X there exists a homeomorphism $h: X \rightarrow X$ such that

- (i) $\forall x \in X \exists U \in \mathcal{U}: x, h(x) \in U$ and
- (ii) $\forall \Delta \in \mathcal{D} \exists V \in \mathcal{V}: h(\Delta) \in V$.

What happens to the old assumptions, like the quotient is metrizable etc.

3.2 Applying the Shrinking criteria to shrink an “X”.

Maybe its really just a remark or an example.

Example 3.11. Let $A = \{(x, \pm x) \mid -\frac{1}{2} \leq x \leq \frac{1}{2}\}$ and let $\mathcal{D} = \{A\}$ be the decomposition of $X := B_1(0) \subset \mathbb{R}^2$. We already verified that $\pi : X \rightarrow X/\mathcal{D}$ is ABH, but it will be illuminating to verify the conditions of the Shrinking principle. Let us verify the conditions appearing in the most general form in remark 3.10.

So let \mathcal{U} be any saturated cover of X and let \mathcal{V} be any open cover of X . We now have to produce a homeomorphism with certain properties. By saturation we can find a open set $U \in \mathcal{U}$ that contains A . We will define h in such a way that $\text{supp}(h) \subset U$. Thus we get

$$\forall x \in X \exists U \in \mathcal{U}: x, h(x) \in U.$$

The statement is trivial for $x \notin \text{supp}(h)$. Otherwise x and $h(x)$ both lie in U . Pick some point $x \in X$ and an open set $V \in \mathcal{V}$ containing x . Now choose a homeomorphism of U fixing its boundary that maps A into a small ball around x that is still contained in $V \cap U$. This will ensure the second shrinking condition. The picture looks like this: A bit fishy

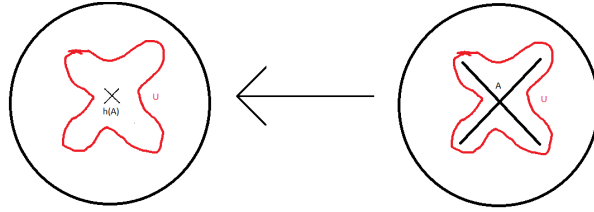


Figure 11: The homeomorphism that shrinks the X .

but I hope its OK.

3.3 Three descriptions of the Alexander gored ball

We now come back to Bing’s discussion of the Alexander horned sphere, more precisely, its exterior as shown in Figure 1. By turning the latter picture inside out we obtain a ball where the horns poke into the inside. Following a suggestion of Bob Edwards we call this creature the *Alexander gored ball*, denoted by \mathcal{A} . We will give three descriptions of \mathcal{A} .

3.3.1 The usual picture: an intersection of balls

The first picture is the above mentioned inside out version of Figure 1. Starting with the standard 3-ball, we drill two pairs of holes in it creating two almost-tunnels which are “linked” as in Figure 12 and repeat this construction indefinitely as indicated.

Remark 3.12. Note that this construction is easily modified to show that \mathcal{A} is a cellular subset of B^3 .

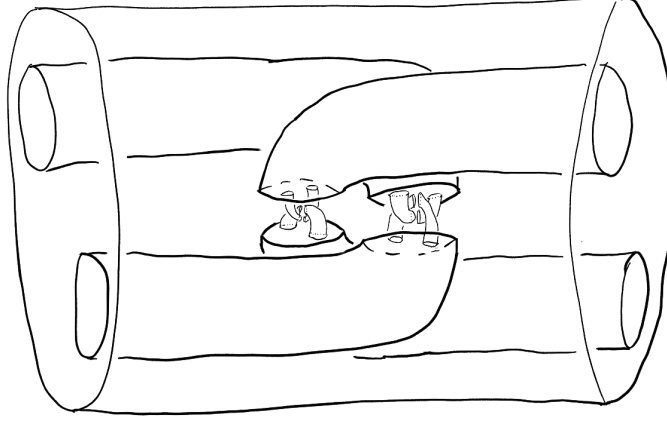


Figure 12: \mathcal{A} as a countable intersection of 3-balls.

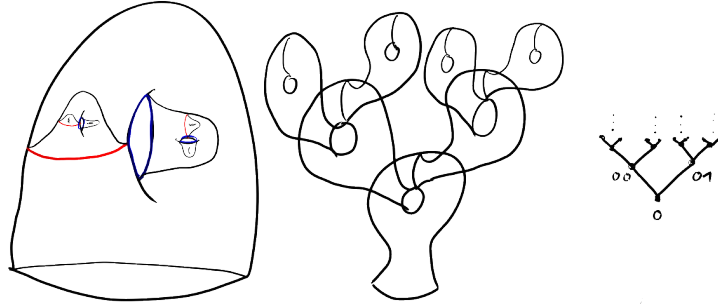


Figure 13: \mathcal{A} as a grope.

3.3.2 The grope picture

There's an equally productive picture of \mathcal{A} as an infinite union of thickened, punctured tori with some limit points added. The construction goes as follows. We denote by T_0 the 2-torus with an open disk removed and let $\mathbf{T} = T_0 \times [0, 1]$. We also fix a meridian-longitude pair of curves $\mu, \lambda \subset T_0$. We start with a single copy \mathbf{T}_0 of \mathbf{T} and attach two extra copies \mathbf{T}_{00} and \mathbf{T}_{01} along $\mu_0 \times \{0\}$ and $\lambda_0 \times \{1\}$, respectively, where meridian and longitude are indexed according to the corresponding copy of \mathbf{T} (note that there's a canonical framing swept under the rug). Next, we attach four more copies \mathbf{T}_{000} , \mathbf{T}_{001} , \mathbf{T}_{010} and \mathbf{T}_{011} along $\mu_{00} \times \{0\}$, $\lambda_{00} \times \{1\}$, $\mu_{01} \times \{0\}$ and $\lambda_{01} \times \{1\}$ and so on. The result is an infinite union of copies of \mathbf{T} indexed by a tree as indicated in Figure 13. This process can be done carefully so that the resulting space is embedded in 3-space and which forces the longitudes and meridian to get smaller and smaller in the successive stages such that they will ultimately converge to points in the limit. It is clear from the construction that these limit points form a Cantor set in 3-space; indeed, they correspond to the limit points of the tree in Figure 13 which, in turn, correspond to infinite dyadic expansions.

We claim that the infinite union of tori together with these limit points is homeomorphic to \mathcal{A} . To see this we first give a complicated description of B^3 . Figure 14 shows a picture of $B^2 \times I \cong B^3$ which can be interpreted as relative handle decomposition built on B^2 with a canceling 1- and 2-handle pair (the neighborhoods of the yellow arcs). In the intermediate level we see a copy of T_0 with a longitude-meridian pair given by the belt circle of the 1-handle and the attaching circle of the 2-handle each of which naturally bounds a disk; we call these disks *caps* for T_0 , anticipating some future language. We now take this picture of $B^2 \times I$ and replace the neighborhood of each yellow arc, which is

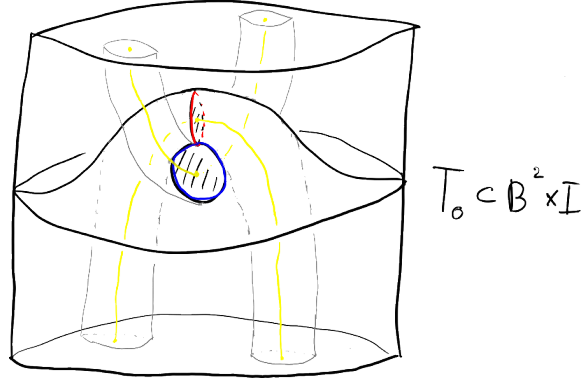


Figure 14: A “capped torus” thickened in 3-space

clearly homeomorphic to $B^2 \times I$, with a copy of the entire picture and repeat this process indefinitely. Note that, with each step the caps will appear smaller and smaller and, again, limit to a Cantor set in B^3 . Moreover, through each of these limit points there will be precisely one yellow arc with endpoints on either the top or the bottom of $B^2 \times I$ giving rise to a Cantor set worth of arcs.

The relation between the two constructions of the gored ball now becomes obvious. On the one hand, we can modify the construction slightly to the extend that, in each step we drill out the neighborhoods of the yellow arcs (parametrized by $B^2 \times [-1, 1]$, say), but instead of gluing the model along the whole of $S^1 \times [-1, 1]$ we only glue along $S^1 \times [-\epsilon, \epsilon]$ for some small $\epsilon \in (0, 1)$. This modified construction matches precisely with Figure 12. On the other hand, it is easy to see that, if we remove the neighborhood of the yellow arcs from $B^2 \times I$ in Figure 14, then we are left a copy of $T = T_0 \times I$. Removing the neighborhoods of all yellow arcs in each step of the construction exhibits the infinite union of tori in Figure 13 embedded in B^3 and what’s missing from the gored ball is exactly the limit Cantor set of the longitude meridian pairs.

Remark 3.13. The “union of (capped) tori”-construction is a first example of a (*capped*) *grope*. More general versions of this constructions (where the torus can be replaced with a different orientable surface with one boundary component in each step and the thickening takes place in 4-space) will play a central role in the 4-dimensional arguments. Forgetting the embedding into 3-space in our toy example, we can think of the grope as a 2-complex made by assembling a countable collection of punctured tori as in the middle of Figure 13. This kind of picture, although not accurate in three dimensions, will appear frequently in the 4-dimensional context.

Remark 3.14. I figured this out by hand. Hopefully it is correct. If there is a better argument, we should give it. Let $\partial\mathcal{A}$ be the boundary of the Alexander gored-ball. If we add a collar to it, ie. consider the mapping cylinder of the inclusion of the boundary we get a 3-ball. But the mapping cylinder construction does not change the homotopy type. Thus \mathcal{A} is simply connected. Maybe we should place a reference to the collar statement here (IF there is none Schoenflies will also do the job).

3.3.3 The gored ball as a decomposition space

Finally, we will exhibit \mathcal{A} as a decomposition space B^3/\mathcal{D} where the decomposition \mathcal{D} of B^3 is given by the Cantor set of yellow arcs in our complicated picture of $B^3 \cong B^2 \times I$ from the previous section^{ref}. Moreover, these yellow arcs will exhibit some iterated clasping as in Figure 15.

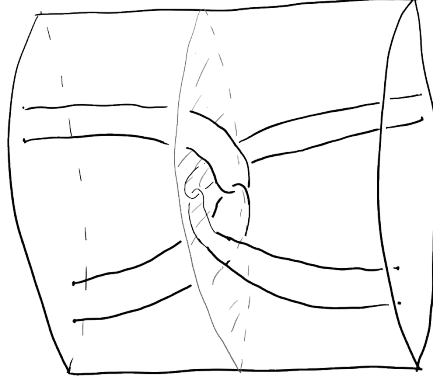


Figure 15: The decomposition of B^3 giving rise to \mathcal{A} .

To see that the quotient space of \mathcal{D} is homeomorphic to \mathcal{A} we take a closer look at the first description. There we had written $\mathcal{A} = \bigcap_{k=0}^{\infty} B_k$ as a countable intersection of 3-balls B_k such that $B_0 = B^3$ and $B_{k+1} \subset B_k$. The important observation is that, in each step, B_{k+1} can be obtained from B_k by performing an ambient isotopy of 3-space, in particular, we have homeomorphisms “retractions” instead of “homeomorphisms”? Otherwise I do not see some of the statements... $h_k: B_k \rightarrow B_{k+1}$. Taking the limit we obtain a map

$$h_{\infty} := \lim_k h_k: B^3 \rightarrow \bigcap_k B_k = \mathcal{A}$$

whose union of inverse sets consist of all points that are moved by h_k for infinitely many k and it is easy to see that the h_k can be chosen such that these points agree with the yellow arcs in \mathcal{D} . \mathcal{D} is given by the preimages of points of h_{∞} . Thus $B^3/\mathcal{D} \xrightarrow{\cong} \mathcal{A}$ is a bijection which is, in fact, a homeomorphism given that h_{∞} is continuous. Let us have a look at the decomposition induced by h_{∞} . We can further collapse everything that between the gaps in the tori in the n -th stage we get a map $B^3 \rightarrow \mathcal{A} \rightarrow \mathcal{A}/\sim_n$. The induced decomposition on B^3 consists of 2^n cylinders linked as in ref to a picture. If we pass to the limit we have to take the intersections of those sequences of cylinders and we end up with the decomposition given by the Cantor set of yellow arcs from the previous section.ref. Especially note that by compactness $\mathcal{A} = \lim_{\leftarrow} \mathcal{A}/\sim_n$.

Exercise 3.15. Show that $h_{\infty}: B^3 \rightarrow \mathcal{A}$ is continuous.

Remark 3.16. This construction of \mathcal{A} brings up an interesting question about the Bing shrinking criterion. We have described \mathcal{A} as the result of crushing certain arcs in B^3 to points and this crushing could be done by homeomorphisms of 3-space and it can be done so as not to move points far in the quotient. So one might think that crushing this decomposition should not change the topology of B^3 . However, the point is that the Bing shrinking criterion requires homeomorphisms of B^3 and not of the ambient space. It only tells us that the topology of 3-space does not change when \mathcal{D} is crushed to points. In fact, for fundamental group reasons the topology of B^3 has to change which tells us that \mathcal{D} cannot be shrunk. I understand this as: The interior of the Alexander gored ball has nontrivial fundamental group, while the interior of a Ball has trivial fundamental group. If they were homeomorphic, the interior would be mapped to the interior (argument!). So they can not be homeomorphic.

3.4 The Bing decomposition: the first shrink ever

In the following we will see the first decomposition that was ever shrunk and how it was shrunk. Interestingly enough, it was quite a non-trivial shrink.

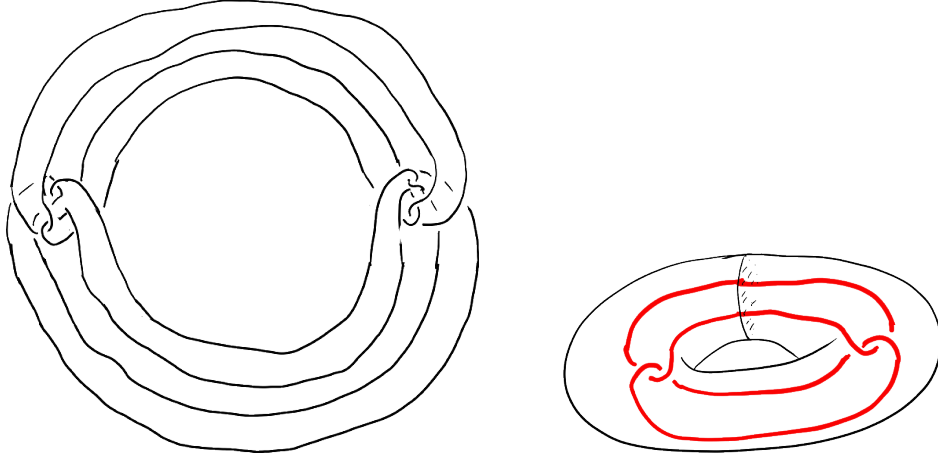


Figure 16: The second stage of the Bing decomposition (left) and its defining pattern (right).

We take the double DA of the Alexander gored ball and also double the corresponding decomposition \mathcal{D} of B^3 to obtain a decomposition $\mathcal{B} = D\mathcal{D}$ of S^3 known as the *Bing decomposition*. The picture of the Bing decomposition is what nowadays would be called the infinitely iterated untwisted *Bing double* of the unknot in S^3 (the left of Figure 16 shows second iteration). More precisely, the Bing decomposition is the countable intersection of nested solid tori where the nesting pattern is shown in the right of Figure 16.

Remark 3.17. When we draw a picture of a decomposition what we really draw is usually only a *defining sequence*, that is, we describe a system of closed sets nested in each other and the decomposition element are the components of the intersection. In many cases it is enough to indicate a nesting pattern as we did above for the Bing decomposition.

Theorem 3.18 (Bing, 1952 [Bin52]). *The Bing decomposition is shrinkable. In particular, the quotient map $D\pi: S^3 \rightarrow DA$ is ABH.*

As mentioned earlier, this answered the question of Wilder whether the obvious involution on DA was an exotic involution on S^3 instead of just an involution on some pathological metric space.

So how do we shrink this thing? Recall that the decomposition elements are the countable intersection of nested solid tori where each torus contains two successors which form a neighborhood of the Bing double of the core of the original torus. Note that the n -th stage of this construction adds 2^n solid tori and in each stage the tori get thinner and thinner while their “length”² roughly remains the same. Now, given some $\epsilon > 0$ we have to come up with a homeomorphism of S^3 which shrinks each decomposition element to size less than ϵ and whose support has distance less than ϵ from the decomposition elements. The basic idea is as follows: We focus on a stage far enough in the construction, call it n_ϵ , where the 2^{n_ϵ} tori are thinner than ϵ and in each of these tori we produce an isotopy that shrinks the decomposition elements. Since all tori within a given stage are isometric it is enough to describe such an isotopy on a single torus.

To construct such an isotopy, the first thing to note is that the normal directions in the torus will not cause any problem because we can choose the torus as thin as we like. The only problem is the linear extent and it doesn’t really matter what coordinate system we use to make that small. Bing’s idea was to lay the torus out along the real line, to just measure diameter in terms of that direction and to make the tori of the subsequent

²More precisely: the length of the core curves.

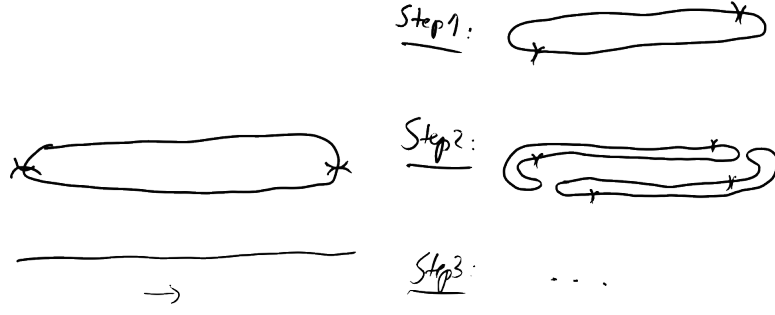


Figure 17: Bing's construction.

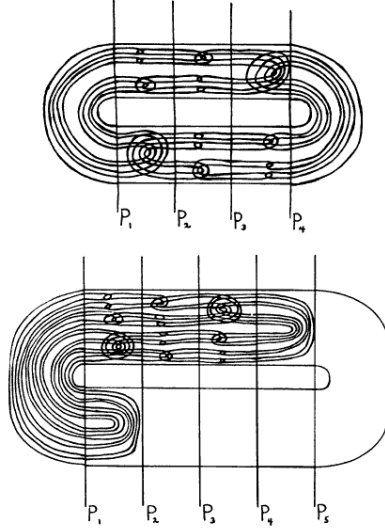


Figure 18: Bing's construction.

stages have small diameter. This can be achieved by successively rotating the clasps in the subsequent stages as indicated in Figure 17.

Remark 3.19. **I think the following argument is clearer:** Let us apply the shrinking criterion from remark 3.10. Given any saturated open cover \mathcal{U} of B^3 and any open cover \mathcal{V} of B^3 . By compactness we can assume without loss of generality that \mathcal{U} is finite. Since $\mathcal{A} \cong \lim_{\leftarrow} B^3 / \sim_n$ there is some N such that \mathcal{U} arises as a pullback of a cover of B^3 / \sim_N . So if we manage to define a self-homeomorphism of B^3 whose support is contained in the 2^N decomposition elements of \sim_N , it would fulfill the first shrinking condition.

A decomposition element of \sim_N is a solid torus. Now we have to find a homeomorphism of the torus relative boundary that moves each decomposition element of DD into a open set of \mathcal{V} . By compactness \mathcal{V} has a positive Lebesgue number. Thus it would also suffice to move it into some ε -Ball. We will even move the larger decomposition elements of \sim_M for some $M \gg N$ into such balls. The idea is to cut the torus into $M - N$ equally sized pieces and arrange the decomposition elements from the M -th stage in such a way that each decomposition element is contained in at most two of them. Furthermore we can isotope everything as close to the meridian as needed. This will then ensure the second shrinking condition if we pick M large enough. Figure 18 is taken from Bings '52 paper. It shows exactly how to arrange the tori. **hope there is no copyright problem.**

Remark 3.20. Why the obvious rotation doesn't seem to work at first sight but does

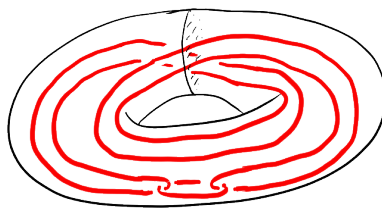


Figure 19: A defining pattern for \mathcal{B}_2

eventually. (Coming soon...)

An interesting question in point set topology, similar in spirit to Wilder's original question, is whether the Bing involution on S^3 is topologically conjugate to a Lipschitz homeomorphism (with respect to the round metric). What makes this interesting is that it has to do with dihedral group symmetry. The basic element in the shrinking process above is to rotate within sub-solid-tori in a big solid torus and this rotation creates a tremendous stretching, because the tori are very thin and we have to rotate rather long distances around them. Moreover, each rotation can be made clockwise or counterclockwise and this choice seems to be reversed by the Bing involution. So to say that the involution is Lipschitz means that the top and the bottom copy of \mathcal{A} have roughly the same shape. But if they have the same shape, then the cavities between the rotated layers in the tori should be mirror images of each other which seems very unrealistic. Unfortunately, it is not clear how to make this rigorous because the tori can have very bad shapes and cannot be assumed to be round.

This very specific question motivates a more abstract question which we state as a conjecture.

Conjecture 3.21. *Any finite, bi-Lipschitz group action on a compact 3-manifolds is conjugate to a smooth action.*

3.5 Something that cannot be shrunk

I didn't get why punctured Meridian discs are the right notion. Until now we have seen both simple and complicated things that shrank and later in these lectures we're going through extreme efforts to show that certain other things also shrink. But not everything shrinks as the following example, also due to Bing, shows.

We consider the decomposition \mathcal{B}_2 of S^3 defined by a similar defining sequence as the original Bing decomposition. Again, the decomposition consists of a Cantor set worth of continua which are defined as a nested intersection of solid tori and the solid tori are nested in such a way that, at each stage, two of them are placed inside one in the previous stage. The only difference is in how the two are placed and, basically, the "2" in \mathcal{B}_2 means that they go around twice instead of once (see Figure 19).

In order to see that \mathcal{B}_2 doesn't shrink, we call the two nested tori in the defining pattern P and Q and consider two meridional disks A and B in the ambient solid torus. The idea is to show that some decomposition element has to intersect both A and B in the first stage, no matter what happens in the subsequent stages. More precisely, we will show that in each stage n at least one of the 2^n tori has to intersect A and B . If that's true, then the decomposition can't be shrunk.

We take a 2-fold cover of this picture and take lifts of P , Q , A and B which we continue to denote by the same letters for convenience. Although each of them has two lifts, it will be enough to consider only one of them for P and Q .

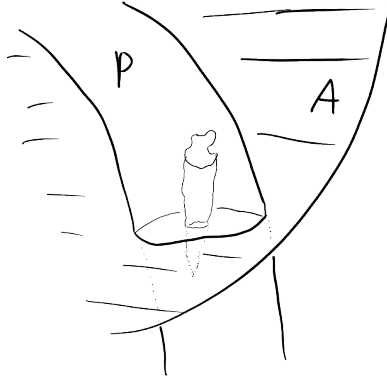


Figure 20: A “substantial” intersection.

Lemma 3.22. *In the lifted picture, either P or Q has “substantial” intersection with A and B*

The precise meaning of “substantial” is as follows. Ideally, the intersections should be in meridian disks for P and Q but this is too much to hope for, so we have to come up with a slightly weaker notion. The right notion turns out to be intersections in *punctured* meridian disks, that is, meridian disks punctured by loops that are trivial in the boundary of P or Q (see Figure 20).

Proof. We look at the intersections $\partial P \cap A$, $\partial P \cap B$, $\partial Q \cap A$ and $\partial Q \cap B$, each will be a collection of circles. Since these circles lie in planar disks, are unknotted and have framing zero in space, each of them must be either a longitude, a meridian or trivial in the boundary of P or Q . Moreover, all these possibilities can occur and we have to think about them.

First, suppose there is some longitudinal intersection in $\partial P \cap A$, say. We claim that either $\partial Q \cap A$ or $\partial Q \cap B$ contains no longitude. The reason for this is that such an intersection would not be consistent with the linking in the picture. Observe that the cores or any longitudes of P and Q together with either boundary of an A or B disk form a copy of the Borromean rings which are known to be a non-trivial link as detected by the Milnor invariant $\bar{\mu}_{123}$, for example. But if Q had a longitudinal intersection with either A or B , then the link would have to be trivial. Moreover, we claim that Q must intersect all lifts of A and B in punctured meridian disks for if either of them contained only trivial intersections, then standard arguments in 3-manifold topology would allow us to remove all intersections and make Q disjoint from that lift and the link would have to be trivial or at least have trivial $\bar{\mu}_{123}$. So each lift must contain at least one meridian and an innermost one will then bound a punctured meridian disk.

Now, if there is no longitude in the intersections, then we focus on P and Q one at a time. For example, if P has no substantial intersections (and thus only trivial intersections) with both lifts of A , then we can deform it into the complement of those lifts. Similarly, if Q has no substantial intersection with either both lifts of A or B , then we can deform the whole link into either half or three quarters of the solid torus and in neither case can this create a non-trivial $\bar{\mu}_{123}$. \square

Based on this lemma we can run an inductive argument. If in some stage there is a P or Q which has substantial intersections with a lift of A or B , then within it we see a P and a Q and one of those has to have a substantial intersection with A and B . Thus we get a nested sequence of solid tori which all have intersections with A and B . Their intersection is a decomposition element of \mathcal{B}_2 intersecting A and B . We have shown

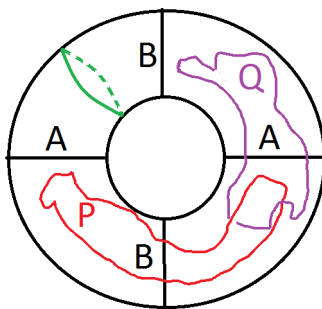


Figure 21: One of P or Q intersects both A and B , otherwise adding a meridian of the Big Torus won't give the Borromean rings.

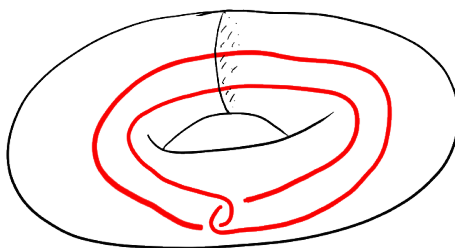


Figure 22: The nesting pattern for the Whitehead decomposition.

that such a thing always exists no matter how we exactly choose to embed the 2-Bing Decomposition. Especially there is always one which is not contained in some ε -Ball if $\varepsilon < d(A, B)$. Thus \mathcal{B}_2 cannot be shrunk.

3.6 The Whitehead decomposition

Another prominent example is the *Whitehead decomposition* \mathcal{W} of S^3 . Just as the two other examples it can be described as an infinite intersection of nested solid tori with nesting pattern as in Figure 22. In other words, in each stage a solid torus is embedded into its predecessor as a neighborhood of a *Whitehead double* of the core. This decomposition is clearly not shrinkable because its elements are not even cellular. But there's a very interesting story when the situation is moved into four dimension by crossing with a line which will be discussed in the next chapter.

Remark 3.23. Actually, one should be a little more precise. In order to explain the nesting pattern we should not only describe the core of the nested solid torus but also a framing number which determines how the core is thickened. Although this affects how subsequent stages are embedded, the shrinking properties of the decomposition will actually be independent of the framings. So for the point set topology we don't have to pay attention to the framings. But in the context of smooth constructions it will be important that the framing number is zero, that is, that we take *untwisted* doubles.

Remark 3.24. From the construction it is clear that the Whitehead decomposition has only a single element, also known as the *Whitehead continuum*, whereas the Bing decomposition had a Cantor set worth of elements. But in practice, this distinction is fleeting because there will be multiplicities. For example, the nesting patterns might involve several copies of Whitehead or Bing doubles in each step. The meaning of multiple Bing double is easy to understand.

Exercise 3.25. Repeat the “union of tori” construction of the gored ball with a genus g surface Σ with one boundary component: fix a standard basis of curves for $H_1(\Sigma)$, start

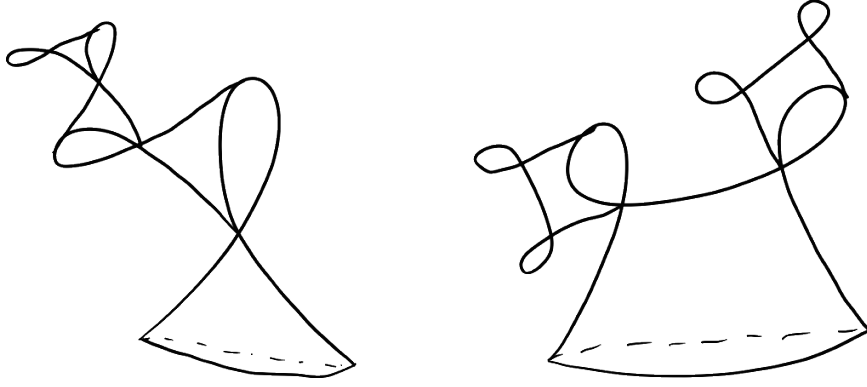


Figure 23: Two Casson handles. More about this later...

with a copy of $\Sigma \times [0, 1]$, successively attach further copies of $\Sigma \times [0, 1]$ along the basis curves such that everything embeds in 3-space and, finally, add the limit set of the sequence of basis curves and take the double. Show that the corresponding decomposition of S^3 has a nesting pattern where g parallel Bing doubles of the core are embedded in a solid torus.

Similarly, we will see that multiple Whitehead doubles are related to some kind of 4-dimensional diagrams (Figure 23) which we will learn all about. The left picture, where there is no branching, corresponds to a single Whitehead double while the right hand side would have two Whitehead doubles which would also lead to a Cantor set worth of decomposition elements.

I also tried to apply the most general shrinking criterion to the Whitehead decomposition cross the real line. However Without compactness It is not true that a saturated open cover comes from a finite stage and so it is getting a bit tricky. The idea in Bings paper to consider crossing with S^1 first, where we have compactness, and then pass to the cover seems to fix this but it feels so tricky.

4 Decomposition space theory and shrinking: more examples

In the last section we focused on the idea of shrinking cellular decomposition, that is, decompositions whose elements are cellular sets (Definition 2.1). It is natural to ask if the converse is true: Is a cellular decomposition necessarily shrinkable?

In Proposition 2.17 we have seen that the answer is yes for finite decompositions. So the first real problem occurs when there are infinitely many cellular sets. We saw two examples of this case, one where the answer was yes (\mathcal{B}) and one where the answer was no (\mathcal{B}_2). So cellularity alone is not enough. Note that both examples have a further property in common. Namely, we can arrange that the decompositions only have a countable number of elements and further that for any $\epsilon > 0$ there are only finitely many elements that have diameter larger than ϵ . The latter condition is called *null*. More abstractly:

Definition 4.1. A collection of subsets $\{T_i\}_{i \in I}$ of a metric space X is called *null* if for any $\epsilon > 0$ there are only finitely many $i \in I$ such that $\text{diam}(T_i) > \epsilon$.

An immediate consequence of nullity is that the number of subsets in the collection must be countable; we simply have to consider a countable sequence of ϵ 's converging to zero.

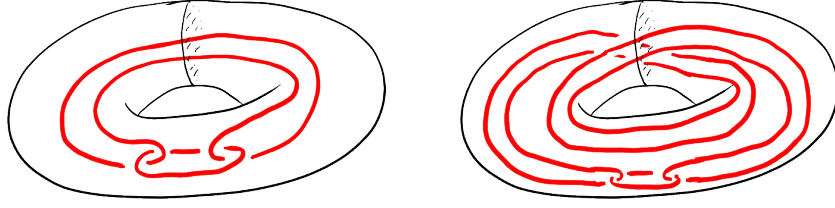


Figure 24: Models for defining sequences for \mathcal{B} and \mathcal{B}_2 .

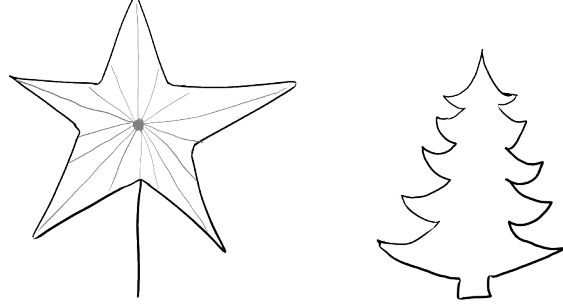


Figure 25: A starlike set (left) and a starlike-equivalent set (right),

To realize \mathcal{B} and \mathcal{B}_2 as null (and thus countable) decompositions, we can choose the defining sequences of each asymmetrically. We still use a series of tori, but this time choose one torus to be much longer than the other (see Figure 24).

We can organize the defining sequence as a binary tree with one branch for each torus, a long branch for each long tori, and a short branch for each short tori. In this scheme, a decomposition element corresponds to a path in the tree. Decomposition elements only have non-zero size when we take the short branch only finitely many times and, moreover, we can estimate the size of the decompositions with non-trivial diameter: If we always take the small torus to have $\frac{1}{10}$ the diameter of the previous one, then the diameter of this decomposition element is less than $\frac{1}{10^i}$ where i is the deepest stage at which we took a short branch. This implies that the sequence is null.

Remark 4.2. Note that the homeomorphism types of the decomposition elements of \mathcal{B} and \mathcal{B}_2 , are not well defined. In fact, they depends on the defining sequence. Using the symmetric defining sequence of \mathcal{B} we found that each of the uncountably many (!) decomposition elements was an arc. But in the asymmetric picture, all decomposition elements that take the short branch infinitely many times must have diameter zero and are thus points. It turns out, however, that the topology of the quotient is well defined.

To sum up, we should be disappointed that even null, cellular decompositions do not shrink. In fact, later we will see that the proof of the Poincaré conjecture requires us to shrink such a decomposition, so we will need to look at stronger properties than cellularity.

4.1 Starlike-equivalent sets and shrinking

Consider a subset $S \subset \mathbb{R}^d$ of Euclidean space. We say that S is *starlike* if it has an “origin” $0_S \in S$ and is a union of closed rays emanating from this origin. More generally, we call a subset $S \subset X$ of an arbitrary space *starlike-equivalent* if a neighborhood of S embeds in Euclidean space such that S is mapped onto a starlike set.

Exercise 4.3. Let $S \subset \mathbb{R}^d$ be a starlike set. We define its *radius function* as the map $\rho_S: S^{d-1} \rightarrow [0, \infty)$ given by $\rho_S(\xi) = \max\{t \geq 0 \mid 0_S + t\xi \in S\}$. Show that S is closed if and only if its radius function is upper semi-continuous.

Obviously, starlike sets are contractible. Moreover, it is easy to see that closed, starlike sets are cellular. In fact, they are a little more than that: not only can they be approximated by arbitrary balls but by starlike balls which just amounts to approximating the radius function by a continuous function.

Exercise 4.4. If $\rho: S^{d-1} \rightarrow (0, \infty)$ is a continuous function, then the starlike set given by $\{t\xi \mid t \leq \rho(\xi)\} \subset \mathbb{R}^d$ is homeomorphic to a ball.

This extra bit of regularity that starlike-equivalent sets have over cellular sets turns out to be strong enough to guarantee shrinkability.

Theorem 4.5 (Bean [Bea67]). *Any null, starlike-equivalent decomposition of \mathbb{R}^d is shrinkable.*

Most of the work is in proving the following lemma.

Lemma 4.6. *Let $T \subset B^d$ be starlike and let $\{T_i\}$ be a null collection of closed sets disjoint from T . Then for any $\delta > 0$ there is a radial homeomorphism $k: B^d \rightarrow B^d$ which is the identity on the boundary and satisfies*

- (i) *the diameter of $k(T)$ is less than δ and*
- (ii) *either $k(T_i)$ has diameter less than δ or for each point $x \in T_i$ the distance of x and $k(x)$ is less than δ .*

In other words, the lemma tells us that we can shrink T without doing too much damage to the other T_i . However, it does not say, for example, that none of the T_i gets stretched bigger than it was before. Even if something had diameter less than $\delta/1000$, say, all we can say is that it will stay smaller than δ .

Proof of Theorem 4.5. We fix some $\epsilon > 0$ and consider all the starlike-equivalent sets which are larger than ϵ . By nullity there are only finitely many of those and we can thus find coordinates in which, for instance, the largest such set appears starlike. We can then use the lemma to shrink the largest decomposition element. Then repeat this on the next largest decomposition element, and so on a finite number of times until the decomposition elements have been shrunk sufficiently. The only tricky element to this is to remember to use the modulus of continuity to make sure that the shrink in the starlike coordinates is sufficient to make the sets small enough in the original coordinates.

□

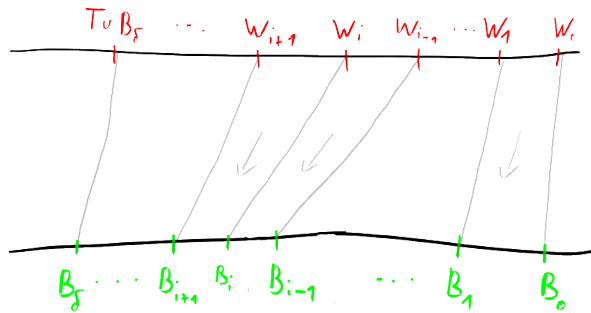


Figure 26: Defining k on a ray.

Proof of Lemma 4.6. For convenience, we assume that the origin of T is the actual origin of B^d .

We start by choosing a nice neighborhood system of the star T , that is, we write $T = \bigcap_{i=0}^{\infty} V_i$, $V_{i+1} \subset \text{int} V_i$, where the V_i are starlike balls with the following properties:

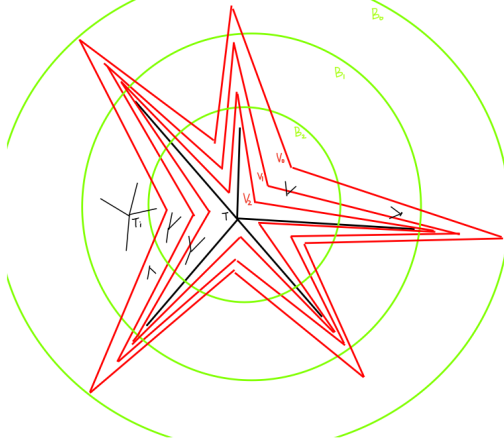


Figure 27: Setup for the starlike shrink

- V_0 intersects only those T_i with diameter less than $\delta/2$.
- Each T_i intersects the frontier of at most one V_j (“no spanning”).

This can be achieved by the null condition and upper semi-continuity. See Figure 27.

Next, we define B_i to be a sequence of round balls of radii r_i around the origin of T such that the $r_0 = 1$ (in other words, $B_0 = B^d$) and the radii decrease as slowly as $0 < r_i - r_{i+1} < \frac{\delta}{8}$ and converge to δ . As a final step in the setup we let $W_i = V_i \cup B_i$.

We then define a map $k: B^d \rightarrow B^d$ by requiring that, for each ray R emanating from the origin, k maps the segment $R \cap (W_i \setminus W_{i+1})$ affinely onto $R \cap (B_i \setminus B_{i+1})$. This is illustrated in Figure 26.

Case 1: T_i is disjoint from V_0 . In this region, points are moved less than $\delta/8$. This is in fact true for all points which lie in some $B_i \setminus V_i$. (More specifically, if x is in $B_i \setminus B_{i+1}$, then $k(x)$ will also be in $B_i \setminus B_{i+1}$, and the difference of the radii of these balls is less than $\delta/8$.)

In all of the remaining cases, T_i will be shrunk smaller than δ . Choose $x, y \in T_i$. We’ll show that $d(k(x), k(y)) < \delta$.

Case 2: $x, y \in B_i \setminus V_i$, but T_i intersects V_0 . (see Figure 28). Points in this region move less than $\delta/8$ for the same reason as in Case 1. Note moreover that since T_i intersects V_0 , its diameter is less than $\delta/2$. Therefore $d(k(x), k(y)) < d(x, y) + d(x, k(x)) + d(y, k(y)) < \delta/2 + \delta/8 + \delta/8 < \delta$.

Case 3: x is not in any $B_i \setminus V_i$. Then x lies between V_j and V_{j+2} for some j . In this case, x and y will be pushed into B_j . Consider the ray between x and the origin, and let x' be the point this ray intersects B_{j+2} . Define y' similarly. There are two subcases.

Case 3a: y is not in B_{j+2} . (see Figure 29). Note that $d(x, y) < \delta/2$, since T_i intersects V_0 . Since the x' and y' are pulled in radially from x and y , we also have that $d(x', y') < \delta/2$. The key step is to notice that $k(y)$ and $k(x)$ will both lie between B_j and B_{j+2} . That is, $d(x', k(x)) < 2\frac{\delta}{8}$ and $d(y', k(y)) < 2\frac{\delta}{8}$. Consequently, by the triangle inequality, $d(k(x), k(y)) < \delta/2 + \delta/4 + \delta/4 < \delta$.

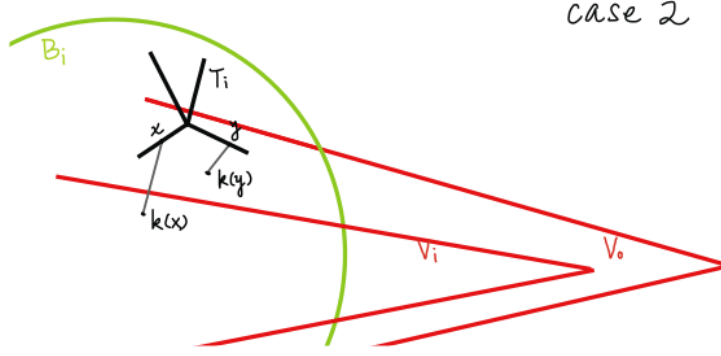


Figure 28: Case 2

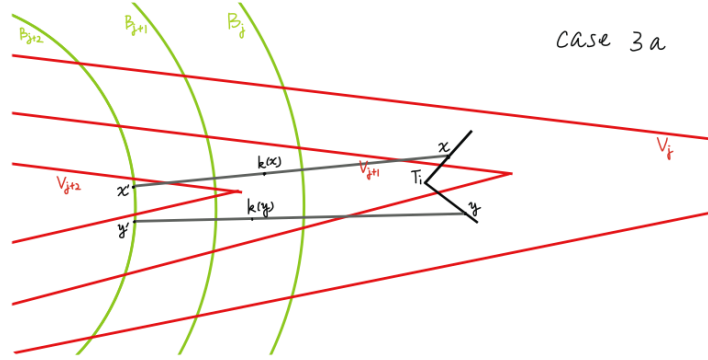


Figure 29: Case 3a

Case 3b: y is in B_{j+2} . (see Figure 30). In this case, $d(y, k(y)) < \delta/8$. So $d(x, k(y)) < d(x, y) + k(y, k(y)) \leq \delta/2 + \delta/8 < 5\delta/8$. However, because of the radial nature of the shrink, $k(x)$ will be closer to $k(y)$ than x is. “This is because $k(x)$ is closer to the origin than x , but still outside the B_{j+2} , and $k(y)$ is inside B_{j+2} ”.

□

Remark 4.7. The fact that the lemma is true is probably not too surprising, but it is somewhat remarkable that there does not seem to be a soft, qualitative proof. Although the Bing school is probably as far from analysis as one gets in mathematics, one actually has to put pencil to paper and make a calculation involving a $\delta/8$ -argument.

Remark 4.8. A useful generalization of Bean’s theorem is due to Mike Starbird and Richard Denman [DS83]. Instead of stars one can consider birds (Figure 31) and ask: what is the difference between a star and a bird? While a star is starlike, meaning that a single radial crush reduces it to a point, a bird is *recursively starlike* or *birdlike* which means that a finite number of starlike crushes are enough to turn the bird into a point. All of the arguments above generalize immediately to birdlike decompositions, and it turns out that this is exactly what we will need in the proof of the Poincaré conjecture; we will need to shrink 2-stage birds instead of stars.

Specifically, we’ll encounter sets which are a union of $S^1 \times B^3$ with a disk D^2 where ∂B^2 is attached to $S^1 \times \{p\}$ where $p \in \partial B^3$ (Figure 32). If such a set is embedded in a standard way in Euclidean space then it has birdlike equivalent structure. Roughly, it becomes starlike after crushing the central B^2 (a precise argument would require keeping track of collared neighborhoods). In [Fre82] these sets are called “red blood cells”.

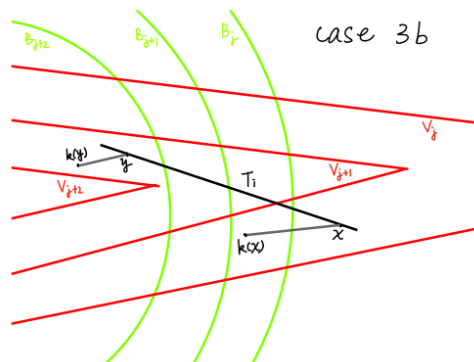


Figure 30: Case 3b

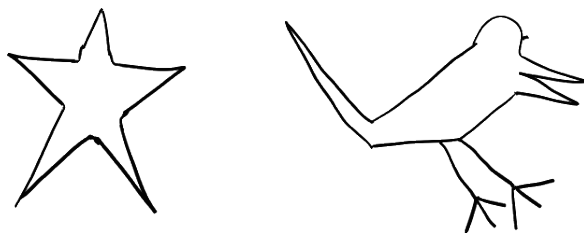


Figure 31: A star and a bird.

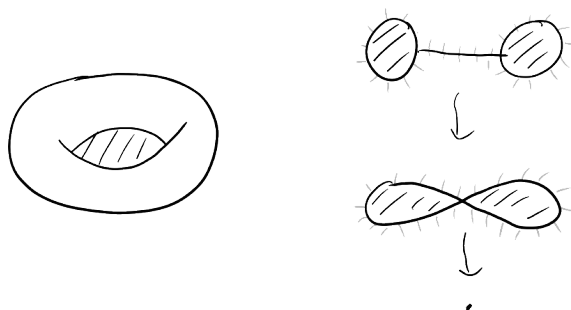


Figure 32: A red blood cell ($S^1 \times B^3 \cup B^2$) from [Fre82]

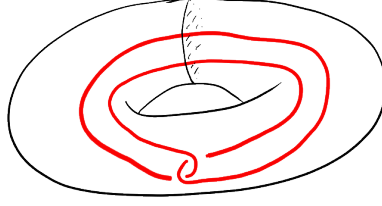


Figure 33: The defining pattern for the Whitehead decomposition.

4.2 A slam dunk for the Bing shrinking criterion

Recall that the Whitehead decomposition \mathcal{W} was constructed as the intersection of nested solid tori where the basic embedding pattern was as shown in Figure 33. It's quite easy to see that the elements of the Whitehead decomposition are not cellular because it is impossible to embed a ball around the embedded torus in the big torus. It is also known that its complement is not simply connected at infinity. So the Whitehead decomposition is definitely not shrinkable. Surprisingly, if we cross with \mathbb{R} and consider the uncountable decomposition

$$\mathcal{W}_{\mathbb{R}} = \{W \times \{t\} \mid W \in \mathcal{W}, t \in \mathbb{R}\}$$

of $S^3 \times \mathbb{R}$, then this becomes shrinkable!

Remark 4.9. This example leads to the notion of *manifold factor*: while S^3/\mathcal{W} is not a manifold (this also follows from local fundamental group considerations), the Bing shrinking criterion shows that the product $S^3/\mathcal{W} \times \mathbb{R} \cong (S^3 \times \mathbb{R})/\mathcal{W}_{\mathbb{R}}$ is homeomorphic to $S^3 \times \mathbb{R}$.

The fact that $\mathcal{W}_{\mathbb{R}}$ is shrinkable was first proved by Andrews and Rubin in 1965 [AR65] although the argument seems to have been known to Shapiro several years earlier who probably mentioned it to Bing. We will present the same argument that Bob Edwards decided to explain on a napkin during some southern California topology meeting around 1977. At that time it had been known that the Whitehead link and Whitehead doubles were intimately connected with Casson handles and the fact that the Whitehead decomposition was nearly a manifold was very encouraging for the attempt to prove the Poincaré conjecture since it indicated that the extra direction that is available in four dimension might make some of the wildness disappear.

Theorem 4.10. *The decomposition $\mathcal{W}_{\mathbb{R}}$ of $S^3 \times \mathbb{R}$ is shrinkable.*

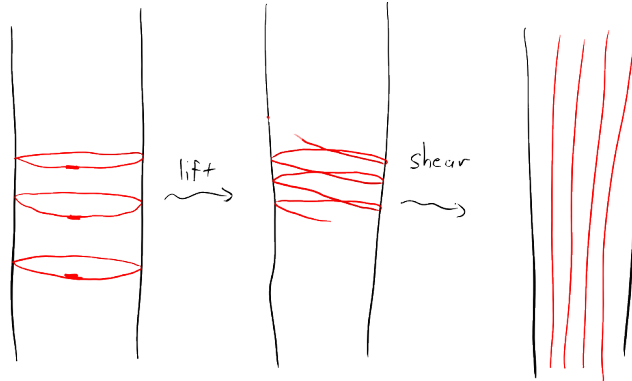


Figure 34: Shrinking $\mathcal{W}_{\mathbb{R}}$.

We will only present a sketch of the proof. A more detailed account is given in Kirby's lecture notes [Kir89, p.87], for example.

Proof. Schematically, going into a very deep stage of the Whitehead decomposition has the effect of squeezing down the radial coordinate in the outer solid torus of that stage and the stage appears as a “circle with an arc inside” whose ends slightly overlap. Crossed with the real line this picture sits in every level.

We know that we can do nothing to shrink the “arc” inside the “circle” but using the fourth dimension we can shift the embeddings in a spiral fashion so that there is a holonomy, meaning that going once around the big torus raises the extra direction by an arbitrarily small amount (see Figure 34). If we re-embed the subsequent stages in that way, the tori don’t link themselves but a different copy and the totality of tori in the next stage appear as spirals. So instead of a line worth of circles we are left with a circle worth of spirals.

But now we can shear the picture to straighten the spirals and the interval segments of which they are made are now very short vertical segments. So every one of the subsequent tori has now been shifted and sheared so that it is concentrated mostly in the vertical direction in which it is very short. Looking back at what we have done so far, we realize that we have met the requirements of the Bing shrinking criteria because we did not have to begin in the outer stage. We can start in an arbitrarily deep stage, which means that we can define the homeomorphism to be supported on a set of arbitrarily small radius (taking care of the support condition), and by shifting/shearing we can make the subsequent stages arbitrarily small. \square

Remark 4.11. The proof of the Poincaré conjecture which we will present follows more closely the approach of [FQ90] and will not directly use the above theorem but it did play a central role in the original proof [Fre82].

4.3 Mixing Bing and Whitehead

So far we have considered decompositions of S^3 obtained by iterating either Bing or Whitehead doubling. We can also mix the two in the sense that, in each stage, we can either use the Bing or the Whitehead double. Roughly, it turns out that, if we put in enough Bing, Bing always wins over Whitehead and the decomposition will be shrinkable.

We have by now become used to the idea that, in the end, Bing doubling somehow makes things shorter while Whitehead doubling doubles the length. So if we build a mixed Bing-Whitehead decomposition, we should think that the Bing doubles are helping us while the Whitehead doubles are hurting us.

Precise results in this direction were worked out by Starbird and Ancel [AS89] in the late ’80s. One very concise result addresses decompositions of the form

$$\mathcal{W} \mathcal{B}^{b_1} \mathcal{W} \mathcal{B}^{b_2} \mathcal{W} \mathcal{B}^{b_3} \dots \quad (4.1)$$

which means that in the first stage we start with a Whitehead double followed by Bing doubling in the next b_2 stages and so on.

Theorem 4.12. *The decomposition defined a sequence as in (4.1) is shrinkable if and only if the series $\sum_i \frac{b_i}{2^i}$ diverges.*

The proof is a tour-de-force in matching up the two techniques that were highlighted in Section 3. The first technique was Bing’s rotation trick to show that something does shrink and the other was to track intersections with certain meridional disk to show that something doesn’t.

While the precise bound is probably not very intuitive, it should be clear that the decomposition shrinks if the sequence of Bing doubles grows sufficiently fast. Although

each Whitehead double sets us back in terms of shrinking, we can make up for it with enough layers of Bing.

While it is possible to write down shrinkability criteria for more general series of Bing and Whitehead doubles, interestingly enough Theorem 4.12 is exactly what will encounter in four manifold topology.

5 The ball to ball theorem

It is time to discuss another main ingredient in the proof of the Poincaré conjecture: the *ball to ball theorem*. The formulation we will give is taken from [FQ90, p.80] but the proof will be closer to the original one given in [Fre82] where we actually prove a sphere to sphere version. As pointed out by Bob Edwards, there is another writeup by Fredric Ancel [Anc84]. To put the theorem into context, recall that we have seen a decomposition that was cellular and null but still did *not* shrink. The ball to ball theorem is a tool for shrinking such decompositions when the extra information is available that the quotient is a manifold. In the basic case, the quotient is a ball.

Theorem 5.1 (Ball to ball Theorem). *Let $f: B^4 \rightarrow B^4$ be a surjective map such that*

- (a) *the collection of inverse sets is null,*
- (b) *the singular image of f is nowhere dense³ and*
- (c) *for a closed set $E \subset B^4$ containing ∂B^4 the map $f^{-1}(E) \xrightarrow{f} E$ is a homeomorphism.*

Then f can be approximated by homeomorphisms that agree with f on $f^{-1}(E)$.

Here, the *singular image* of f , henceforth denoted by $\text{sing}(f)$, is the set of points in the target whose preimages are not singletons.

Some remarks about this theorem are in place. First, the restriction to dimension four is completely irrelevant, the proof actually works in all dimensions. Moreover, there is a much more general result of Siebenmann which states that, under weaker assumptions than in Theorem 5.1, any map between manifolds of dimension five or higher can be approximated by homeomorphisms. This was actually known before Theorem 5.1 which would have been an easy special case in high dimensions. So, it was only the 4-dimensional case that required a special argument.

Second, there is a general principle called *majorant shrinking* due to Bob Edwards which was developed in the context of proving the local contractibility of the space of homeomorphisms. It addresses the question of approximating a given map by homeomorphisms which already is a homeomorphism on the preimage of a closed set. If one understands the annulus conjecture, the torus trick and related material towards the local contractibility of the space of homeomorphisms of a manifold, then the hypothesis (c) in Theorem 5.1 is completely redundant. However, this involves working with non-compact spaces, leading to a rather elaborate theory. This was actually central in the original arguments in [Fre82] but it is not used in [FQ90].

It is also worth noting that the hypothesis (c) implies that f is proper and of degree one. So, the assumption on f to be surjective is not really needed, becoming a consequence of (c).

³Recall that a subset is *nowhere dense* if its closure has empty interior.

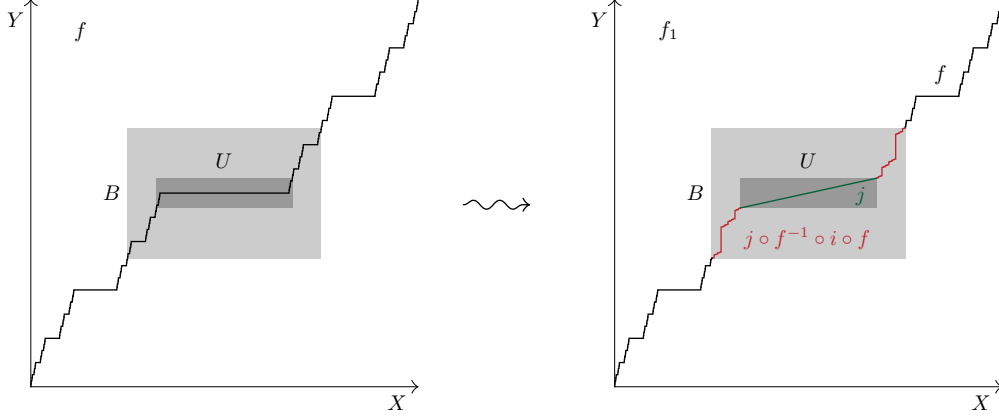


Figure 35: The modification of the Cantor function.

5.1 The main idea of the proof

Before going into the details we will outline the main idea of the proof. First, to avoid confusion, we relabel the source and the target balls to X and Y , that is, we put $X = Y = B^4$ and consider $f: X \rightarrow Y$. Note that f fails to be a homeomorphism exactly when there are inverse sets and, obviously, the ones with largest diameter are the most offensive. Since we are assuming that the collection of inverse sets is null, there are at most finitely many largest inverse sets and we can work on each separately.

The idea is to look at a small neighborhood of such a largest inverse set $f^{-1}(y)$ in X , to implant it photographically in a tiny ball U around its image point $y \in Y$ and then to taper the original f into this little photograph by similar tricks as in Brown's proof of the Schoenflies theorem. What this step accomplishes is to get rid of one largest inverse set, but it turns out that there is some price to pay: although we start with a function what we will end up with is only a relation. Thinking in terms of the graph of the function, we have traded a large horizontal spot for some vertical steps.

A good example to keep in mind is the graph of the Cantor function (Figure 35) which has a large horizontal spot in the middle. What is going to happen is that a neighborhood of this horizontal spot is changed by a homeomorphism which tilts the horizontal spot but introduces some vertical cliffs. Moreover, we can control the tilting such that the vertical spots are as short as we want. Of course, the Cantor function can be approximated by homeomorphisms without such complications, but in dimension four is not so simple.

Eventually, the theorem will follow from an iteration of this idea once we have recast it in the language of relations.

Now we will put the above ideas into precise formulas, to illustrate the very first step of the proof. For simplicity we assume that a largest inverse set of f maps to $0 \in Y$. We take two concentric round balls $U \subset B \subset Y$ around $0 \in Y$, where U has very small radius (for example, we can take $U = B_\delta(0)$ and $B = B_{2\delta}(0)$ for some small $\delta > 0$) and define an “inverse squeeze” homeomorphism

$$i: B \xrightarrow{\cong} Y$$

which is the identity on U and sends $B - U$ onto $Y - U$ (for example by taking a radial expansion, but later we will need more flexibility). We note that the composition $i^{-1} \circ f$ is a homeomorphism near ∂X , so it can be extended to a homeomorphism

$$j: X \xrightarrow{\cong} B$$

(this extension always exists, we can take the cone for instance). Finally, we define $f_1: X \rightarrow Y$ by

$$f_1 = \begin{cases} f & \text{on } X - f^{-1}(B) \\ j & \text{on } f^{-1}(\text{Int } U) \\ j \circ (f^{-1} \circ i \circ f) & \text{on } f^{-1}(B - \text{Int } U). \end{cases} \quad (5.1)$$

Exercise 5.2. Check that f_1 is a closed *relation*.

Now, why is f_1 not a function? The reason are the points $y \in B - U \subset Y$ such that $i(y)$ lies in the singular image of f while y itself does not. In that case, $f^{-1}(y)$ must be a single point $x \in f^{-1}(B - \text{Int } U) \subset X$ and going through the definition it is easy to see that $f_1(x)$ is, in fact, a subset of Y with more than one element. So, f_1 is only a relation. But note that $f_1(x)$ is by definition contained in B which we could choose arbitrarily small. So, f_1 is close to being a function in some sense. Before we continue we recall some elementary facts about relations.

5.2 Relations

Recall that a *relation* from X to Y is just a subset $R \subset X \times Y$. As indicated, we like to think of a relation as a multivalued function $X \xrightarrow{R} Y$ which assigns to $x \in X$ the subset

$$R(x) = \{y \in Y \mid (x, y) \in R\} \subset Y.$$

Functions can be identified with their graphs, and so, by abusing terminology, we consider functions as special relations. Clearly, a relation $R: X \rightarrow Y$ is a function if and only if $R(x)$ is a singleton for all $x \in X$. So, the set of functions naturally embeds in the set of relations.

In analogy with functions, we define the composition of relations $X \xrightarrow{R} Y \xrightarrow{S} Z$ as

$$S \circ R = \{(x, z) \in X \times Z \mid \exists y \in Y: (x, y) \in R, (y, z) \in S\} \subset X \times Z.$$

Also, any relation $X \xrightarrow{R} Y$ has an inverse given by

$$R^{-1} = \{(y, x) \in Y \times X \mid (x, y) \in R\} \subset Y \times X.$$

Obviously, we have $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$ and it is easy to see that these definitions reduce to the corresponding ones for functions.

Next, we want to bring topology into the picture, as usual in the context of compact metric spaces. First of all, it is an easy exercise in point set topology that for compact metric spaces X and Y a function $f: X \rightarrow Y$ is continuous if and only if its graph is closed in $X \times Y$ (in fact, compact Hausdorff is enough). So, the appropriate generalization of continuous functions are closed relations. Clearly, the failure of a relation $R: X \rightarrow Y$ to be a function is measured by the quantity

$$\text{vd}(R) \stackrel{\text{def}}{=} \max_{x \in X} \text{diam}_Y(R(x))$$

which we call the *vertical defect* of R . We also define the *horizontal defect* by

$$\text{hd}(R) \stackrel{\text{def}}{=} \max_{y \in Y} \text{diam}_X(R^{-1}(y)) = \text{vd}(R^{-1})$$

which is the obstruction for R^{-1} to be a function. In particular, a closed relation R is a homeomorphism if and only if

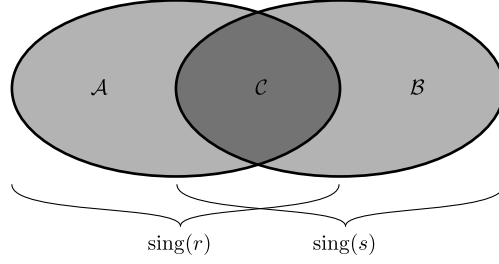


Figure 36: The sets \mathcal{A} , \mathcal{B} and \mathcal{C} .

- (i) $R(x) \neq \emptyset \neq R^{-1}(y)$ for all $x \in X$ and for all $y \in Y$ and
- (ii) $\text{hd}(R) = \text{vd}(R) = 0$.

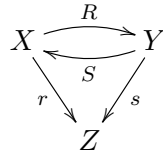
As a final piece of notation we define the *singular image* of a relation R by

$$\text{sing}(R) = \{y \in Y \mid \# R^{-1}(y) > 1\}.$$

So far our strategy has been to stay within the world of functions and to approximate f by homeomorphisms by reducing the horizontal defect. However, we could also go into the world of closed relations and try to reduce both the horizontal and vertical defects and this is exactly how we will prove Theorem 5.1.

5.3 Iterating the main idea: admissible diagrams

As mentioned before, the idea is to iterate the construction of f_1 out of f and, in order to do this, we have to extend the construction from functions to relations where we now have to worry about the singular images of both the relation and its inverse. A simple way of doing this (which was suggested by Ric Ancel and Jim Cannon) is to use a third ball $Z = B^4$ to keep track of where the singular sets lie. The formalism involves commutative diagrams of closed relations of the form



denoted altogether by $\mathfrak{A} = (R, S; r, s)$, where r and s are surjective functions, and $X = Y = Z = B^4$ are balls. Such a diagram \mathfrak{A} is called *admissible* if

- (a) $S = R^{-1}$;
- (b) the collections of inverse sets of r and s are null;
- (c) the singular images $\text{sing}(r)$ and $\text{sing}(s)$ are nowhere dense in Z ;
- (d) R and r are homeomorphisms over a neighborhood E of ∂B^4 ;
- (e) the sets $\mathcal{A} = \text{sing}(r) - \text{sing}(s)$, $\mathcal{B} = \text{sing}(s) - \text{sing}(r)$ and $\mathcal{C} = \text{sing}(r) \cap \text{sing}(s)$ (see Figure 36) are *mutually separated*, that is, each is disjoint from the closures of the others, and
- (f) R restricts to a homeomorphism $R: r^{-1}(\mathcal{C}) \xrightarrow{\cong} s^{-1}(\mathcal{C})$.

Note that an admissible diagram is determined by the (admissible) pair (R, r) .

Exercise 5.3. Check that $s(\text{sing}(R)) = \mathcal{A}$ and $r(\text{sing}(S)) = \mathcal{B}$.

The starting data for our inductive scheme is the diagram

$$\begin{array}{ccc} & f & \\ X & \xrightarrow{\quad} & Y \\ & f^{-1} & \\ & f \searrow & \swarrow \text{id} \\ & Z & \end{array}$$

which is clearly admissible. So, when we start s is the identity so that \mathcal{B} and \mathcal{C} are empty and we only have \mathcal{A} . But when we do the first step and pass to f_1 we erase some singular image of f by implanting the photograph and we create singular image of the inverse relation which means that \mathcal{B} and \mathcal{C} come alive. Thereafter, we will have to deal with non-empty \mathcal{A} , \mathcal{B} and \mathcal{C} . Roughly, \mathcal{C} is where we have already solved the problem in the sense that, even though there is some singular image, we have arranged the map above them to be a homeomorphism.

Lemma 5.4. *Let $\mathfrak{A} = (R, S; r, s)$ be an admissible diagram. For any neighborhood $\mathcal{N}(R) \subset X \times Y$ of R and for any $\epsilon > 0$ there is an admissible diagram $\mathfrak{A}' = (R', S'; r', s')$ with $r' = r$, $R'|_E = R|_E$, $R' \subset \mathcal{N}(R)$ and $\text{hd}(R') < \epsilon$.*

In other words, there is a relation which is arbitrarily close to R whose horizontal spots have arbitrarily small size and we can also control the vertical defect.

In the proof of the lemma we use a notion of *general position* which is very different from the usual one in manifold theory. We state it as an exercise.

Exercise 5.5. Let X be a compact manifold and let $C \subset X$ be countable and $N \subset X$ be nowhere dense. Then for all $\epsilon > 0$ there exists a homeomorphism h of X which is ϵ -close to the identity and is supported in an arbitrary neighborhood of C , such that $h(C) \cap N = \emptyset$.

Proof of Lemma 5.4. As the first step, by following the construction of f_1 above, we consider two small round balls U and B centered at a point $a \in \mathcal{A}$ such that

- (i) $r^{-1}(a)$ has maximal diameter,
- (ii) $U \subset B \subset Z - (\text{Cl}(\text{sing}(s) \cup E) \cap Z)$ (this is possible because \mathcal{A} is separated from $\text{sing}(s) = \mathcal{B} \cup \mathcal{C}$) and
- (iii) $\partial U \cap \text{sing}(r) = \emptyset$ (possible because $\text{sing}(r)$ is countable).

Then, s is non-singular over B , and so $s|_B : s^{-1}(B) \rightarrow B \cong B^4$ is a homeomorphism. Let $i : B \rightarrow Z$ be suitable homeomorphism such that $i|_U = \text{id}_U$. By Exercise 5.5 we can assume that $\text{sing}(r) \cap \partial U = i(\text{sing}(r)) \cap (\text{sing}(r) - U) = \emptyset$. Now, choose a homeomorphism $j : X \rightarrow s^{-1}(B) \subset Y$ such that $j = s^{-1} \circ i^{-1} \circ r$ on ∂X .

The relation R' can be constructed by iteratively applying the following formula finitely many times

$$R' = \begin{cases} R & \text{on } X - r^{-1}(\text{Int}(B)) \\ j & \text{on } r^{-1}(\text{Int}(U)) \\ j \circ (r^{-1} \circ i \circ r) & \text{on } r^{-1}(\text{Int}(B) - \text{Int}(U)). \end{cases}$$

This is almost the same as the formula (5.1) for f_1 , the only thing that is slightly different is the definition of j . Again, it is easy to check that R' is well defined. To complete the new admissible diagram we simply take $r' = r$ and for s' we are forced to choose

$$s' = \begin{cases} s & \text{on } Y - s^{-1}(B) \\ i^{-1} \circ r \circ j^{-1} & \text{on } s^{-1}(B) \end{cases}$$

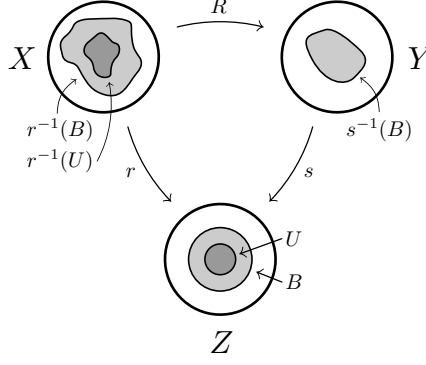


Figure 37: The admissible diagram.

as can be shown by a simple computation (hint: first compute $(s')^{-1} = R' \circ r^{-1}$).

It remains to check the mutual separation of \mathcal{A}' , \mathcal{B}' and \mathcal{C}' . We have $\text{sing}(s') = \text{sing}(s) \cup i^{-1}(\text{sing}(r))$, and so

$$\mathcal{A}' = \mathcal{A} - U, \quad \mathcal{B}' = \mathcal{B} \cup (i^{-1}(\text{sing}(r)) - U) \quad \text{and} \quad \mathcal{C}' = \mathcal{C} \cup (\mathcal{A} \cap U).$$

So, $R' : r^{-1}(\mathcal{C}') \rightarrow (s')^{-1}(\mathcal{C}')$ is a homeomorphism because over \mathcal{C} it coincides with the old R , and over $\mathcal{A} \cap U$ we have $R' = j$.

Now the proof is almost done. By choosing B small enough we can guarantee that R' is contained in $\mathcal{N}(\mathcal{R})$ and the photographic implanting trick using j wipes out one large horizontal spot of R (by Exercise 5.3). Since the collection of inverse sets of r is null, by repeating the construction finitely many times we can eliminate the inverse sets of R' of diameter bigger than ϵ . \square

5.4 Proof of the ball to ball theorem

The final step of the proof was not addressed in detail in the lectures. We follow the basic idea from Ancel's paper [Anc84].

We start with our function $f : B^4 \rightarrow B^4$. Let us choose a closed neighborhood $\mathcal{N}(f) \subset B^4 \times B^4$. Our goal is to construct a homeomorphism $h : B^4 \rightarrow B^4$ which is close to f , meaning that $h \subset \mathcal{N}(f)$ (we are still considering functions as subsets of $B^4 \times B^4$).

The idea is to construct in $\mathcal{N}(f)$ a sequence $(R_n)_{n \in \mathbb{N}}$ of closed relations of decreasing horizontal and vertical defects, and take the limit. However, Lemma 5.4 only allows us to make the horizontal defect arbitrarily small, while the vertical defect can increase just a little bit, being controlled by the given neighborhood of the original relation. So in order to gain more control on the vertical defect, we will construct R_n from $R_{n-1}^{-1} = S_{n-1}$, by using the lemma. In this way, we can reduce, alternately, both the horizontal and the vertical defects. Moreover, we prefer to obtain the homeomorphism h as the intersection of a telescopic sequence of compact sets in $\mathcal{N}(f)$, instead of considering the limit of a sequence of relations.

Now we give the formal construction. For a subset $A \subset X$ of a metric space (X, d) we denote by $\mathcal{N}_\epsilon(A)$ the ϵ -neighborhood of A in X . Let $(\epsilon_n)_{n \in \mathbb{N}}$ be a suitably chosen positive sequence converging to zero. By induction we define a sequence of admissible diagrams $\mathfrak{A}_n = (R_n, S_n; r_n, s_n)$ such that $\mathfrak{A}_0 = (f, f^{-1}; f, \text{id})$ and, for $n \geq 0$, \mathfrak{A}_{n+1} is obtained from $\mathfrak{A}_n^{-1} = (S_n, R_n; s_n, r_n)$ by Lemma 5.4 under the assumption that $R_{n+1} \subset \text{Int}(\mathcal{N}_{\epsilon_n}(S_n))$ and $\text{hd}(R_{n+1}) < \epsilon_n$. Here S_n and $\mathcal{N}_{\epsilon_n}(S_n)$ play the role of R and $\mathcal{N}(R)$ in the lemma. Note that, by construction we have $R_{n|E} = \text{id}_E$.

Moreover, ϵ_n is chosen so that $\mathcal{N}_{\epsilon_n}(S_n) \subset \text{Int}(\mathcal{N}_{\epsilon_{n-1}}(R_{n-1}))$, and $\lim_{n \rightarrow \infty} \epsilon_n = 0$, starting from a sufficiently small ϵ_0 satisfying $\mathcal{N}_{\epsilon_0}(f) \subset \text{Int}(\mathcal{N}(f))$. So, $\text{vd}(R_{n+1}) < \epsilon_{n-1} + 2\epsilon_n$. It follows that, roughly speaking, R_n converges to a homeomorphism h for even n , and to h^{-1} for odd n . To formalize this, we consider the relation $\mathcal{H}_n = \mathcal{N}_{\epsilon_{2n}}(R_{2n})$. We have $\mathcal{H}_{n+1} \subset \mathcal{H}_n \subset \mathcal{N}(f)$ for all n , $\text{hd}(\mathcal{H}_n) < \epsilon_{2n-1} + 2\epsilon_{2n}$, and $\text{vd}(\mathcal{H}_n) < \epsilon_{2n-2} + 2\epsilon_{2n-1} + 2\epsilon_{2n}$.

Finally, we put

$$h = \bigcap_{n=0}^{\infty} \mathcal{H}_n \subset B^4 \times B^4$$

and it is now obvious that $h : B^4 \rightarrow B^4$ is a homeomorphism such that $h|_E = \text{id}_E$. This concludes the proof of the ball-to ball theorem.

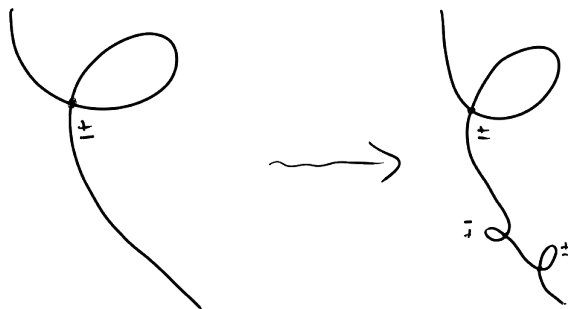


Figure 38: Adjusting the algebraic self-intersection number.



Figure 39: The Whitney trick in the plane.

Part II

Casson handles

6 From the Whitney trick to Casson handles

6.1 The Whitney trick in dimension 4

After all this work in compact metric spaces, we now step into the smooth world. An important geometric construction in smooth manifold is the *Whitney trick* which was introduced by Whitney [Whi44] in order to prove that any n -dimensional smooth manifolds, $n \geq 3$, embeds in the Euclidean space \mathbb{R}^{2n} of twice the dimension, which is one better than what we get for free(=by general position). However, this result is of little consequence compared to the method used to prove it.

Remark 6.1. Conspicuously, Whitney's proof does not work for $n = 2$ although, by example, the fact that any surface embeds in \mathbb{R}^4 is still true since it holds for the real projective plane.

Whitney looked at the manifold M^m as it is mapped into \mathbb{R}^{2n} and he realized that by general position there are only transverse double points to think about to which one can assign signs and he also realized the algebraic sum of these points can be controlled by putting in small kinks (see Figure 38). In particular, one can make it zero. In the case of algebraically zero self intersection he was led to a local picture as in Figure 39 where there are two sheets of the manifold (two different manifolds would work, too) with two intersection points of opposite signs. To improve the situation to an embedding he noticed that there must be a disc W as in the picture (which he probably didn't call W) and that it looks like it should be possible to push one sheet over the other along the disk to cancel the two intersection points.

While this looks pretty good in the plane, how should it work in high dimensions? Some aspects of the picture are conserved when passing to higher dimension, for example, the intersection points stay 0-dimensional (since the dimensions of the two intersecting



Figure 40: The Hopf link near a transverse intersection.

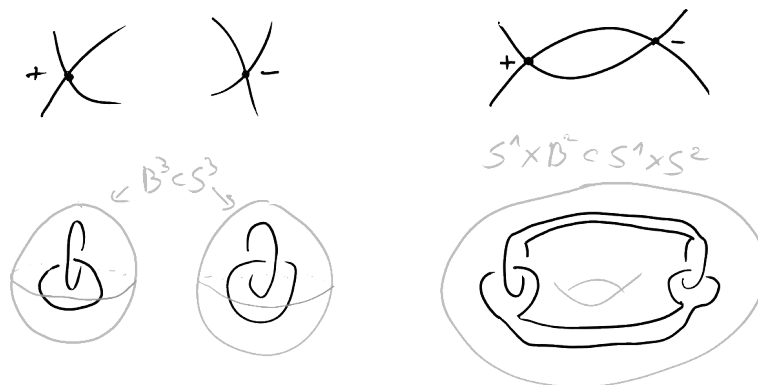


Figure 41: From two Hopf links to a Bing double

sheets add up to the dimension of the ambient space) and the disk W is going to stay 2-dimensional. The idea is to find two arcs connecting the two points, one in each sheet, which will make up two halves of the boundary of W . In addition, some normal data is necessary in order to do the push without introducing new intersections. Something that's favorable in high dimensions ($n \geq 3$) is that 2-dimensional objects are generically disjoint from everything $< n - 2$ -dimensional, so we don't have to worry about making (the interior of) W embedded and disjoint from the image of the manifold; in contrast, in four dimensions ($n=2$) W will intersect itself and the manifold. Another advantage is that the space of $(n - 1)$ -planes in \mathbb{R}^{2n-2} , whose fundamental group controls the ambiguity in normal data, is simply connected for $n \geq 3$ while for $n = 2$ the fundamental group is \mathbb{Z} . However, since we are ultimately interested in 4-manifolds we have to make a careful analysis of the 4-dimensional situation.

A local model for a transverse intersection of surfaces in 4-manifolds are the planes $\{z = 0\}$ and $\{w = 0\}$ in \mathbb{C}^2 . If we look at a small 3-sphere around the origin, then it is an extremely important observation that the two planes intersect the 3-sphere in a Hopf link (see Figure 40). Let's build up to the Whitney trick from this starting point. Suppose that we have two intersection points of opposite sign. That means we should have two Hopf links in two separate 3-spheres (Figure 41). This picture comes from thinking of neighborhoods of the points and asking what's going on the boundary, but for the Whitney trick we have to look a little more globally at a neighborhood of the circle made up from the two arcs connecting the points. Assuming that the ambient 4-manifold is simply connected, the neighborhood of such a circle is $S^1 \times B^3$ and on its boundary $S^1 \times S^2$ it turns out that we see a Bing double of $S^1 \times \{\text{south pole}\}$ in the solid torus $S^1 \times \text{southern hemisphere} \subset S^1 \times S^2$. Note that the opposite signs of the two intersection points account for the fact that we see a link with linking number zero, if signs agreed, then we would get linking number plus

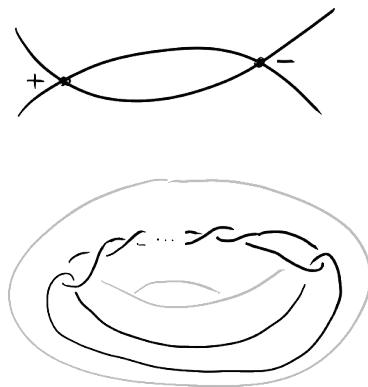


Figure 42: The wrong framing produces twisted Bing doubles.

or minus two.

Figure 41 is also great for understanding the “normal data” mentioned above which is nothing but a choice of framing for the circle. Note that we have drawn an untwisted Bing double and the reason for this is that we have implicitly chosen the correct framing. A different framing will give a twisted Bing double as in Figure 42. In any case, this picture really only lives in the copy of $S^1 \times S^2$ embedded in the our ambient 4-manifold and not in a 3-sphere. But if there’s an embedded Whitney disk W bounded by the circle and we attach it to the picture, then we create a bordism from $S^1 \times S^2$ to a 3-sphere inside the 4-manifold. In other words, then W provides a core for surgery which takes this $S^1 \times S^2$ to a 3-sphere and it’s highly relevant what link we end up with in this 3-sphere. Because if we end up with the unlink, then we can easily see how to separate the two surfaces; we can think of each circle as representing a hole in a surface which needs to be repaired and if we can fill these holes with disjoint disks then we obtain new surfaces without intersection points and it’s not hard to see that they are isotopic to the original surfaces.

6.2 4-manifolds in the early 1970s: surgery and h -cobordism

The 1950s and 1960s saw the development of surgery theory and the h -cobordism theorem which allowed to translate classification problems for manifolds of dimension five and higher completely into questions in homotopy theory and algebra. In the beginning of the 1970s surgery theory had become very mature; at the frontier were things like equivariant or controlled surgery which were very technical subjects. On the other hand, almost nothing was known in dimension four. The main reason was the failure of the Whitney trick and the goal of this section is to pin down how this failure prevents us from using surgery and the h -cobordism theorem.

Surgery theory addresses the existence problem whether we can find a manifold within some previously identified homotopy type. (This might take place in the smooth, PL, topological or any other category.) The h -cobordism theorem is related to the corresponding uniqueness problem. When we have found two manifolds M_1 and M_2 of dimension d , say, in the same homotopy type and we want to find an isomorphism between them. The strategy, which was developed by Smale, is to first find some cobordism N of dimension $d + 1$ between the two and then try to find the simplest possible one. Preferably this simplest N should be in the same homotopy type as the M_i , in other words, the inclusions $M_i \hookrightarrow N$ should be homotopy equivalences, and in this case N is called an h -cobordism. In a favorable situation, for example, if all fundamental groups are trivial and the dimensions are high enough, then Smale would conclude that N is a product and, in particular, M_1 and M_2 are isomorphic. But in the end, it turns out that the existence and uniqueness side

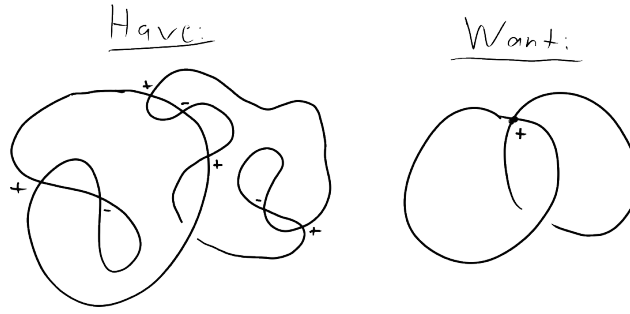


Figure 43: Trying to surger out a hyperbolic pair.

involve essentially the same methods because in the h -cobordism story the main problem is to take turn some cobordism into an h -cobordism using surgery.

In dimension four, Milnor had investigated the simply connected homotopy types of closed manifolds or, more abstractly, *Poincaré duality spaces* which are certain cell complexes that obey the algebraic condition that manifolds obey in terms of the duality between homology and cohomology. Milnor was able to classify the (homotopy types of) 4-dimensional Poincaré duality spaces that are simply connected and it turned out that they are in one-to-one correspondence with (equivalence classes of) \mathbb{Z} -unimodular quadratic forms or, in less fancy terms, symmetric integer matrices with determinant plus or minus one up to similarity. So in 1973, the existence and uniqueness questions in dimension four that people were thinking about were which of these forms are realized by smooth (or equivalently PL) manifolds and whether two homotopy equivalent 4-manifolds are h -cobordant and thus diffeomorphic (or PL homeomorphic). It turns out that these two questions are very similar.

An interesting example of a simply connected 4-manifold known as the *K3 surface* is given by a quartic $K = \{W^4 + X^4 + Y^4 + Z^4 = 0\}$ in \mathbb{CP}^3 . It was known that its quadratic form is isomorphic to the matrix

$$E_8 \oplus E_8 \oplus H \oplus H \oplus H$$

where E_8 is the famous even, 8-by-8 integer matrix known as the Cartan matrix of the exceptional Lie algebra and $H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is a so-called *hyperbolic pair* which corresponds to the intersection form of $S^2 \times S^2$. Now, we know that there is a Poincaré duality space corresponding to $E_8 \oplus E_8$ and there is the obvious projection

$$E_8 \oplus E_8 \oplus H \oplus H \oplus H \xrightarrow{\text{pr}} E_8 \oplus E_8.$$

The question was whether one could realize this projection geometrically. More precisely, is it possible to do surgery on K with the effect of removing the hyperbolic pairs from the intersection form? We know by now that this won't work in the smooth category as a consequence of Donaldson theory. One might wonder why we didn't just take one copy of E_8 , but it was already known by Rokhlin's theorem that the intersection form of a smooth, spin 4-manifold must have signature divisible by 16. It turns out that one copy of E_8 does occur for a topological 4-manifold which is a concrete instantiation of something called the Kirby-Siebenmann invariant which we'll get to later on. So since back in 1973 we knew Rokhlin's theorem and we were always trying to work in the smooth category (and only desperation forced us out of it) we're going to try to make a smooth $E_8 \oplus E_8$ -manifold out of the K3 surface K .

So what's the picture for doing this? We know from the Hurewicz theorem that the hyperbolic pairs in the homology of K are represented by maps $S^2 \rightarrow K$ which we can

take as smooth and in general position. One hyperbolic pair is shown schematically on the left of Figure 43. In general, there will be excess intersections, after all we only know the algebraic intersection numbers. If we could use the Whitney trick to remove the extra intersections to get to the picture on the right then we could do surgery on either of the embedded spheres, that is, we could cut out a neighborhood diffeomorphic to $S^2 \times B^2$ of one sphere and replace it with $B^3 \times S^1$, with the effect of removing the hyperbolic pair from the intersection form. (The other sphere is also important since it makes sure that the surgery does not change the fundamental group.)

Remark 6.2. All this is completely analogous to dimension two. If we have a surface which has an extra genus that we don't want, then we identify a dual pair of simple closed curves, cut out an annular neighborhood of one of them and fill in the two resulting boundary components with disks.

Next, let's try to prove the h -cobordism theorem in dimension four. Given an h -cobordism N^5 between two simply connected 4-manifolds M_1 and M_2 , we look at a handle decomposition of N^4 and try to simplify it as much as possible. The handle decomposition induces a chain complex computing $H_*(N, M_1) = 0$ which can be simplified algebraically such that there are only 3-chains and 2-chains and the boundary map is an isomorphism represented by the identity matrix in suitable bases. The whole idea of the proof is to match the geometry with the algebra of the chain complex. We can easily remove handles of index 0 and 5 and with just a little more work, since everything is simply connected, we remove handles of index 1 and 4. What is left are only handles of index 2 and 3 and the level in between the 2- and 3-handles is another 4-manifold $M_{2/3}$ which is obtained from either M_1 or M_2 by a sequence of surgeries on embedded circles. But since the manifolds are simply connected all circles are null-homotopic and in dimension 4 homotopy implies isotopy, so the surgeries happen on trivial, standard circles and it's an easy exercise that this results in $S^2 \times S^2$ summands. It follows that we have diffeomorphisms

$$M_{2/3} \cong \begin{cases} M_1 \# k(S^2 \times S^2) \\ M_2 \# k(S^2 \times S^2) \end{cases}$$

where k is the number of 2- (or 3)-handles. The spheres we see in $M_{2/3}$ are the k belt spheres of the 2-handles and the k attaching spheres of the 3-handles and each of them has a dual sphere which intersects it in a single point. So the situation is very clean separately, but the problem is that the belt spheres and attaching spheres together don't look so nice. However, the algebra forces that for each belt sphere there is an algebraically dual attaching sphere. So the picture is very similar to the one we encountered when we tried to surger out hyperbolic pairs (see).

This time we have a pair of embedded spheres which are algebraically dual and if we can get rid off all extra intersection points, then we can also cancel the remaining handles of index 2 and 3, hence N must be the trivial cobordism.

6.3 Finding dual spheres

The solution to finding Whitney disks is virtually the same for surgery and the h -cobordism problems and the first two steps toward the solution were taken by Andrew Casson in 1974 [Cas86]. Let us put Casson's ideas into a broad brush before going into the details.

The first idea is that the solving local problem is impossible but there is global information that might help. So what is the local problem? Since the 1950s Fox and Milnor had developed the idea of *slice knots*, that is, knots in S^3 which bound a smooth or locally

⁴More precisely, we take a relative handle decomposition of N built on M_1 .

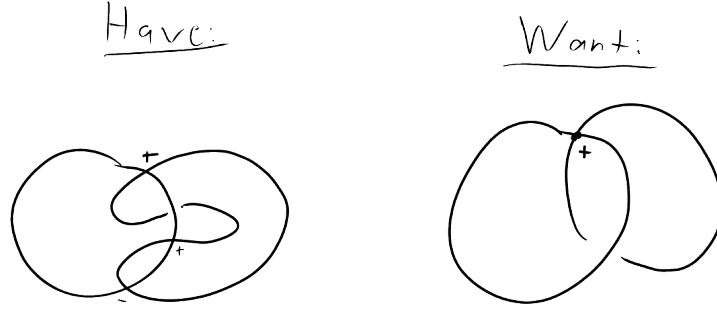


Figure 44: Algebraically dual spheres in the middle level of an h -cobordism.

flat disks in B^4 . Slice knots arise, for example, as cross sections of knotted 2-spheres in 4-space. By now, a tremendous amount is known about the obstruction theory for slice knots. For example, Peter Teichner and his collaborators [COT03, COT04] have pushed this very far and before Casson and Gordon [CG78] had developed some higher order obstructions. But even in 1973 it was known that there were obstructions, for example, the Seifert matrix has to obtain a certain form for a knot to be slice. This is discouraging for our attempts to remove self-intersections of immersed disks in 4-manifolds because any knot bounds an immersed disk in B^4 for non-slice knots it's impossible to remove the intersection points. But Casson realized that there was global information that could be exploited, namely the fact that the spheres in the surgery and h -cobordism problem come in dual pairs.

The second idea is that there is a trade-off the fundamental group of the complement, which is bad, and intersections which, in some sense, are also bad. This can be explained in the local model of a transverse double point (see Figure 40 on page 42). We already saw that the intersecting planes appear as a Hopf link in the boundary of a small ball around the intersection. Another interesting gadget that appears in this picture is the *Clifford torus* $\{(z, w) \in \mathbb{C}^2 \mid |z| = |w| = \frac{1}{\sqrt{2}}\}$ which is the common boundary of tubular neighborhoods of the two components of the Hopf link. This Clifford torus exemplifies the relation between intersection points and the fundamental group of the complement because the complement of the intersecting planes actually deformation retracts onto the Clifford torus and thus has free Abelian fundamental group of rank two (generated by linking circles of the planes). In contrast, the complement of two disjoint planes in \mathbb{C}^2 has free fundamental group of rank two, so the intersection point must account for a relation in the fundamental group. This is just the most microscopic instance of a concept that we'll be working with all the time.

So how are Casson's ideas used to make progress on the surgery and h -cobordism problems? In both problems we have an algebraically dual pair of spheres $a, b \subset M$ in a 4-manifold M and we'd like to remove algebraically canceling pairs of (self-)intersection points by finding Whitney disks. For simplicity, we take M simply connected most of the time. For starters, we're going to try to fix up the fundamental group of the complement. We'd like to arrange that $\pi_1(M \setminus (a \cup b))$ is also trivial; more generally, if M is not simply connected, we'd like the map $\pi_1(M \setminus (a \cup b)) \rightarrow \pi_1(M)$ to be an isomorphism. In that case we call the pair a, b π_1 -negligible. The simple reason is that a Whitney move, for example, on b across a Whitney disk that intersects a , will not help us in reducing the number of intersection points. So if a, b are π_1 -negligible, then at least we have a chance to find useful Whitney disks although, even if we can achieve π_1 -negligibility, the Whitney disks will still cross themselves and we still have work to do. But this is just a first step in what will turn out to be an infinite construction and before we proceed with the construction

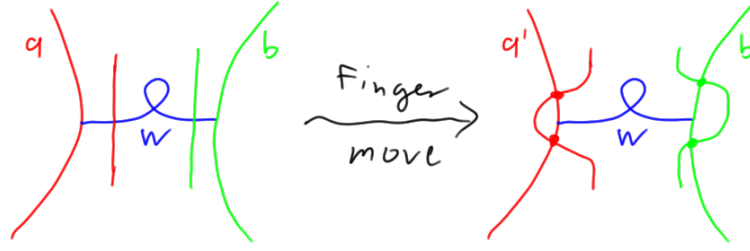


Figure 45: Trading intersections for self-intersections.

we want to keep firm control over the fundamental group.

The key thing in arranging something to be π_1 -negligible is to make all linking circles null-homotopic in the complement as can be seen immediately by van Kampen's theorem. Geometrically, such a null-homotopy would create an immersed disk in the complement which, together with the meridian disk bounding the linking circle, gives a sphere which intersects the original object in one point. So π_1 -negligibility is basically equivalent to the existence of what we call *geometric duals*, that is, spheres that meet the important pieces in one point.

Remark 6.3. In [FQ90] the terminology *transverse spheres* is used for geometric duals with self-intersection zero which is sometimes useful. In these lectures we will also never run into geometric duals with nontrivial self-intersection although they can arise in certain situations.

In the surgery situation we need geometric duals \hat{a} and \hat{b} for spheres a and b which form a hyperbolic pair in the intersection form of M , in particular, a and b are algebraically dual. The idea is to modify a and b simultaneously by regular homotopies until they become geometrically dual. We take some framed, immersed Whitney disk W for a pair of algebraically canceling intersection points between a and b which will meet both a and b in general. The first step is to push a and b off W as indicated in Figure 45 by so-called *finger moves* resulting spheres a' and b' intersecting each other just as a and b but they don't meet W and have more self-intersections than the original spheres. Next we do a Whitney trick across W on either a' or b' and we get new spheres that are regularly homotopic to a and b , have two fewer intersection point but possibly more self-intersections each. Repeating this process finitely many times we eventually obtain a geometrically dual pair \hat{a} and \hat{b} .

In the h -cobordism problem the setup was slightly different. Here a and b are not only algebraically dual but each is embedded and comes with a geometric dual \hat{a} and \hat{b} , respectively. These dual spheres already tell us that a and b are π_1 -negligible individually but they might not be simultaneously, the problem is that \hat{a} might intersect b , for example. Our task is to arrange \hat{a} to be disjoint from b with the additional difficulty of that we're only allowed to move a and b by isotopies (instead of regular homotopies) in order to keep them embedded. The first step is to make the intersection number $\hat{a} \cdot b$ zero which is easily achieved by "adding" copies of a to \hat{a} , that is, connect summing a and \hat{a} inside M along a suitable arc. This is also known as the *Norman trick*. Next we take some framed, immersed Whitney W disk for intersections between \hat{a} and b and, as in the surgery case, the idea is to cleanse W from bad things and then do the Whitney trick on \hat{a} . However, we have to pay close attention to a and b in the process. If we did the Whitney trick right away we'd be running danger of creating new intersections with whatever W intersects which a priori might be either of a , b , \hat{a} and \hat{b} . But we don't care about intersections of \hat{a} with \hat{b} or itself so that the only problems are caused by intersections of W with a or b . We can remedy the b intersection by Norman tricks adding copies of \hat{b} to W so that the new W

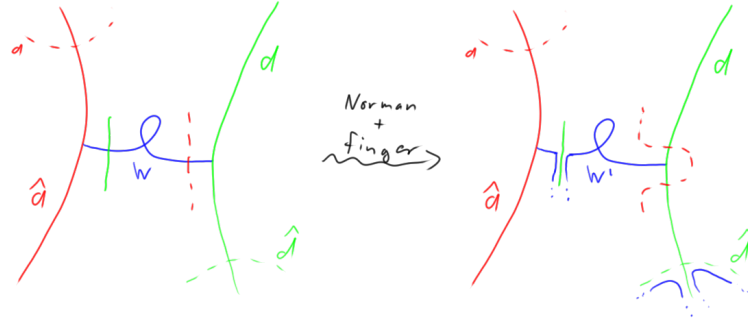


Figure 46: Dual spheres for the h-cobordism problem.

has only problematic intersections with a which, in turn, can be removed by isotoping a off W by finger moves in the direction of b . This will possibly make the new Whitney disk more singular (if \hat{b} meets W) and create new a, b -intersections and but that's okay. Now that the Whitney disk is clean, a singular Whitney trick on \hat{a} produces a (most likely immersed) geometric dual for the isotoped version of a .

Remark 6.4. You probably noticed that finger moves are exactly the inverses of Whitney tricks. It is ironic that the first step in order to do Whitney tricks is to do their inverses, we somehow have to make the situation worse before we can start to make it better.

As mentioned, this is only the first step in Casson's construction. The following lemma, which can be proved using similar tricks as above, leads the way to the further steps.

Lemma 6.5. *Let M be a simply connected 4-manifold and let $\alpha: (D^2, \partial D^2) \rightarrow (M, \partial M)$ be a properly immersed disk such that there exists a 2-cycle β with $\alpha \cdot \beta = 1$ (homologically). Then α is regularly homotopic (rel ∂) to a π_1 -negligible $\alpha': (D^2, \partial D^2) \rightarrow (M, \partial M)$.*

Exercise 6.6. Prove Lemma 6.5.

6.4 Casson handles and Kirby calculus

We have figured out that for both the surgery and the h -cobordism problem we can assume that there's a pair of algebraically dual spheres whose union is π_1 -negligible and we can start throwing Whitney disks for unwanted intersection points into their complement. That's still not perfect because the disks will cross themselves but Casson noticed that the process can be iterated. Indeed, any Whitney disk meets the Clifford tori of its intersection points exactly once and thus we can use Lemma 6.5 to arrange the Whitney disks to be disjoint and π_1 -negligible in the complement of a and b so that we can map a second collection of Whitney disks into the complement of a, b and the union of the first collection of Whitney disks and so on (see Figure 47). This leads to an infinite process for building a 2-dimensional spine which, in the end, turns out to be contractible (it is not homeomorphic to the disk in any obvious way, though). Indeed, we're building some infinite, non-compact 2-complex going into the 4-manifold which has lots of free fundamental group at every stage but at each stage that free fundamental group is annihilated by the immersed disks of the next stage so the direct limit will be a contractible space. So we can build this thing that's somehow "like a disk" and it bounds the Whitney circle. This is the first conceptual core of the idea of what Casson called *flexible handles* and what everyone else calls *Casson handles*.

Remark 6.7. Note that the Whitney disks in the higher stages of Figure 47 don't really look like the ones we know in that they only involve one intersection point each (see also

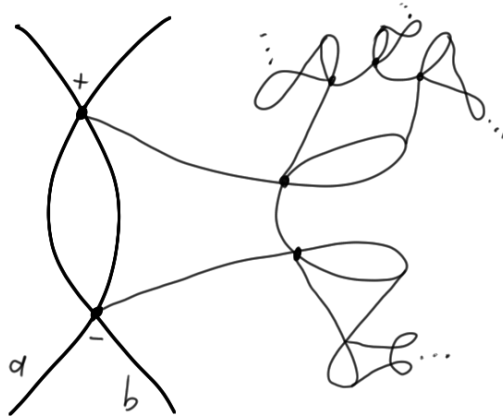


Figure 47: Building a Casson handle.

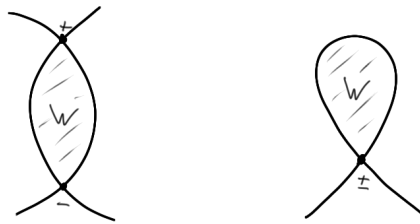


Figure 48: Two-sided (left) and one-sided (right) Whitney disks.

Figure 48). Such *one-sided* Whitney disks were also discussed in Whitney’s paper [Whi44] and there is a corresponding Whitney trick just as in the two-sided case. The difference is that, in the case of self-intersections, the two-sided trick produces regular homotopies while the one-sided version only gives homotopies. (For intersections of two different surfaces the two-sided trick actually produces isotopies of one of the surfaces.) To keep the pictures as simple as possible we only draw one-sided Whitney disks in higher stages. For two-sided Whitney circles we’ve seen earlier that the two intersecting sheets appear on the boundary of a tubular neighborhood as a Bing double. Similarly, in the one-sided case we see a Whitehead double (Figure 49) which might be twisted if we choose the wrong framing. (Note that this is intimately related to our discussion in Remark 3.23.)

As the term “handle” indicates, there’s still something missing from our discussion: we’ve just focused on the 2-dimensional spine and we have to find a way to thicken the construction a little bit in order to get a good 4-dimensional neighborhood. This is related to the framing issue mentioned before and disasters lurk if we don’t carefully select framing conditions for all the Whitney disks. But it turns out that this is not a big issue in the end. We will define the framing conditions idiosyncratically in terms of the so-called *Kirby calculus*. There is an equivalent way of understanding this in terms of 3-manifold topology which fits perfectly with our discussion of Bing and Whitehead doubling and decomposition space theory. It turns out that one of the most important things in the infinite construction is what the frontier looks like, which is a 3-manifold. Different framings can actually produce different frontiers and we can understand if we got the framings right in terms of the 3-manifold topology of the frontier. We’ll say more about this later.

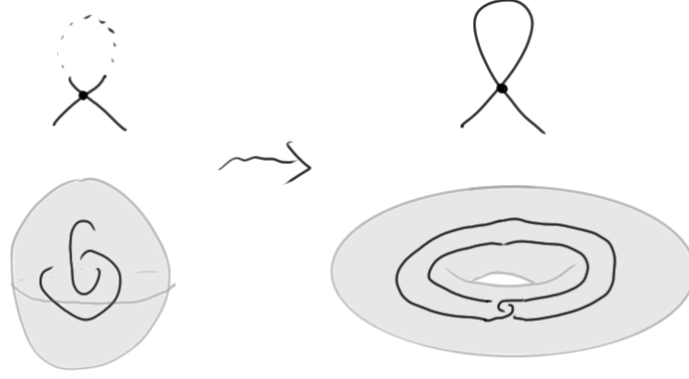


Figure 49: Self-intersections and Whitehead doubles.

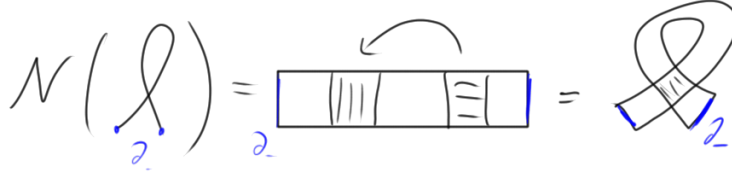


Figure 50: Self-plumbing $B^2 \times B^2$. (Imagine the extra dimensions!)

6.4.1 Kinky handles and Kirby calculus

We want to understand how to thicken immersed Whitney disks, so we have to know what a neighborhood of a disk crossing itself in a 4-manifold look like. By standard results in differential topology, a generic immersion of B^2 can be extended (uniquely up to isotopy) to an immersion of $B^2 \times B^2$ – in other words, a 4-dimensional 2-handle⁵ – which factors through an embedding of a self-plumbing of $B^2 \times B^2$ into the 4-manifold (see Figure 50). (Recall that self-plumbing means to identify two interior sub-disks and glue them together interchanging horizontal and vertical coordinates.⁶) Note that the plumbings don't affect the solid torus $\partial_- = S^1 \times D^2$ – the *attaching region* of the 2-handle – which is indicated in blue; self-plumbed 2-handles are commonly called *kinky handles* and ∂_- serves as their attaching region. In summary, the correct notion of a thickened (or framed) immersed Whitney disk is an embedding of a kinky handle and any Whitney disk has a canonical thickening, that is, whenever we can find a Whitney disk we can also find a kinky handle and vice versa.

Exercise 6.8. Show that $B^2 \times B^2$ becomes homeomorphic to $S^1 \times B^3$ after one self-plumbing.

To see how we get to Figure 49, we should take the sphere that contains the Hopf link around the crossing and surger it to itself by cutting out a 3-ball on each component and declaring that their boundary 2-spheres are identified by reflection (Figure 51, left). That gives a Whitehead double embedded in $S^1 \times S^2$ and its 0-framed neighborhood is just the attaching region ∂_- of the corresponding kinky handle.

The pictures in Figure 51 actually show the full glory of the 4-dimensional situation. In fact, we should think of the whole situation as a 4-ball around the intersection point

⁵ Some terminology: the disks $B^2 \times \{0\}$ and $\{0\} \times B^2$ are called *core* and *cocore*, their boundaries $S^1 \times \{0\}$ and $\{0\} \times S^1$ are the *attaching circle* and *belt circle* and $\partial_- = S^1 \times B^2$ is called the *attaching region*. (For some reason, $B^2 \times S^1$ doesn't seem to have a standard although the term *belt region* suggests itself.)

⁶ More precisely, we take disjoint embeddings $\phi, \psi: D^2 \rightarrow \text{int} D^2$ and identify $\phi(D^2) \times D^2$ with $\psi(D^2) \times D^2$ via $(\phi(x), y) \sim (\psi(y), x)$.

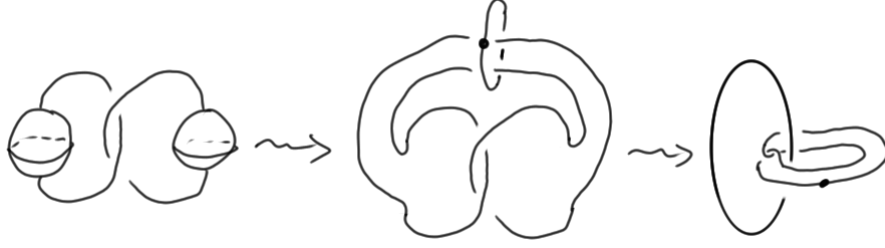


Figure 51: Kirby diagrams of a single (negative) self-plumbing of $D^2 \times D^2$.

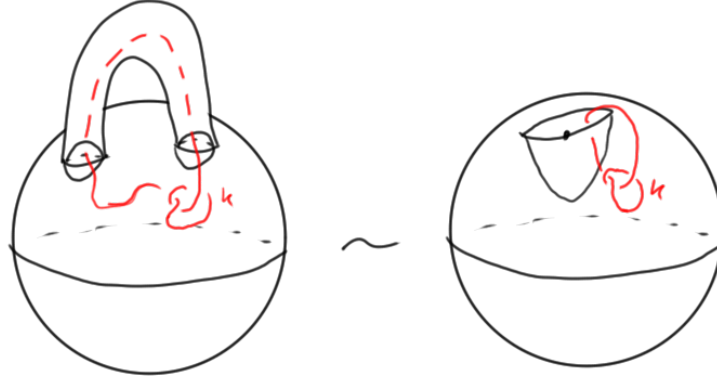


Figure 52: How to read Kirby diagrams.

to which a 4-dimensional 1-handle is attached and its attaching region are the two balls in the left picture. So we simply have to imagine a 4-dimensional 1-handle connecting these balls. Schematically, this is shown in Figure 52 and there's another notation which is part of the Kirby calculus to represent links in simple 3-manifolds such as $S^1 \times S^2$. In that notation we record the fundamental group generator of $S^1 \times S^2$ dually by drawing a circle that links both strands in the middle of Figure 51 and putting a dot on it. The meaning of this notation is that anytime we see an unlink with dots on it in S^3 , we should think of removing the standard slices in B^4 beneath the dotted components as indicated in the right of Figure 52. If we just have one dotted component, then the complement of the standard slice is homeomorphic to a meridian of the dotted circle times a 3-ball, so it is sort of the Alexander dual space of $S^1 \times B^3$. Moreover, by general position we can always arrange links in the boundary of the Alexander dual picture to be actually out on the boundary of B^4 . So a general link in $S^1 \times S^2$ can then be drawn as a link in S^3 which might link the dotted unknot.

The other thing we have to know about Kirby calculus is that whenever we see a link component that doesn't have a dot on it, then that means that a 4-dimensional 2-handle has been attached to the 4-ball (instead of dug out in the case with dots) and then it better have a number next to it in order to determine a framing by adding that number of twists to the canonical framing of the knot in S^3 .⁷

6.4.2 The definition of Casson handles

We can finally describe what Casson handles actually look like in the simply connected context. The process of building Casson handles in a π_1 -negligible situation is then as follows:

⁷Note that there's \mathbb{Z} worth of choices for the framing coming from $\pi_1(SO(2))$.

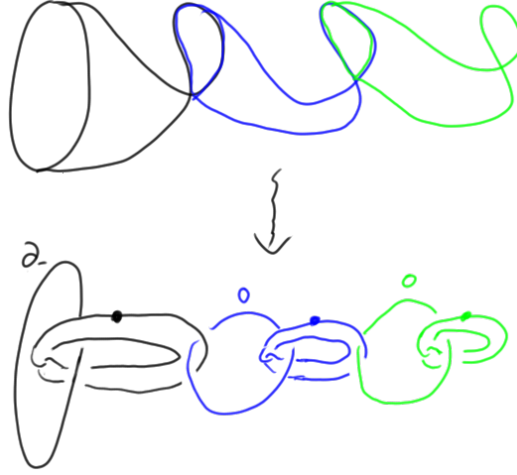


Figure 53: The first three stages of the simplest Casson handle.

- Locate Whitney disks for unwanted intersections. (Lather.)
- Make them π_1 -negligible and thicken to kinky handles. (Rinse.)
- Repeat this process for the new π_1 -negligible situation. (Repeat.)

Note that each thickening process induces a canonical framing on the corresponding Whitney circle and we have to use precisely this framing to attach the kinky handles in each stage. This infinite process replaces each original Whitney disk with a contractible 4-dimensional gadget called a *Casson handle*, generically denoted by CH , and inside CH we find the attaching region ∂_- of the kinky handle in the first stage which now serves as attaching region for CH . In other words, a Casson handle (CH, ∂_-) has been attached to each original Whitney circle. After scraping off all frontier of CH except for ∂_- we obtain an *open Casson handle* which we denote by $\mathring{\text{CH}}$. Our main goal is to prove the following theorem.

Theorem 6.9. *The pair $(\mathring{\text{CH}}, \partial_-)$ is homeomorphic to $(B^2 \times \mathbb{R}^2, S^1 \times \mathbb{R}^2)$.*

Once we know this result we can make surgery and h -cobordism work in the simply connected context in dimension four because everything we wished to have done with Whitney disks we can now do with Casson handles.

In order to draw Kirby diagrams of Casson handles, the first thing we do is to make the pattern in the middle of Figure 51 more symmetrical by passing to the right picture using the symmetry of the Whitehead link so that the attaching region ∂_- of the kinky handle appears as an unknot and the dotted circle has the clasp on it. Next we note that the diagram generalizes immediately to kinky handles with several self-plumbings, we simply have to add several dotted circles to (untwisted) Whitehead doubles of meridians of the attaching circle. For the inductive process we have to located the attaching regions for the higher stages of kinky handles and it is easy to see that, in each stage, the kinky handles are attached to meridians of the dotted circles of the previous stage and the framing condition is that we have to use the zero framing of these meridians. Figure 53 shows the first three stages of the simplest Casson handle where each Whitney disk has only one self-intersection and Figure 54 shows an example with more complicated branching of self-intersections.



Figure 54: A more complicated Casson handle.

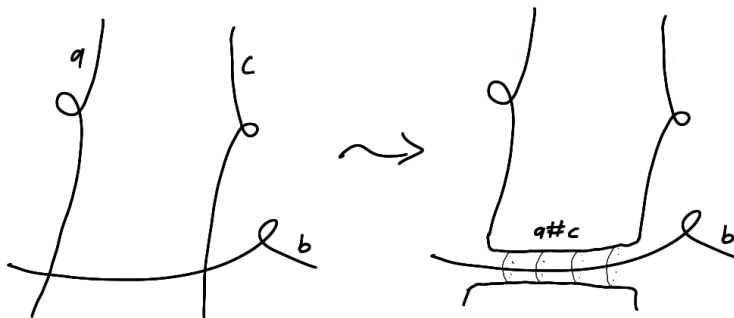


Figure 55: Piping.

7 Exploring Casson handles

7.1 Picture Camp

Welcome to Picture Camp! We're going to draw a lot of pictures and get familiar with them. The first set of pictures are kind of schematics – although they can be made rigorous by regarding them as cross section of 4-dimensional picture – where arcs represent bits of surfaces.

Piping. We might have a configuration of bits of surfaces a , b and c as in the left of Figure 55 where a and c both intersect b in one point. (Note that it wouldn't make a difference if the self intersection of b were in the middle because we're looking at 2-dimensional sheets.) Then we can add a tube along some arc on b to unite the sheets of a and c – that is, we take an ambient connected sum of a and c – so that the result has no intersection with b . This is called *piping*.

Here's an example of how we will use piping. Suppose we have an immersed disk attached to some lower stage stuff and it has a dual sphere which might have self-intersections, too. And suppose we know that the loops in the dual associated with its self-intersection points are trivial in the fundamental group while the double point loops in the disks are essential. Then we can pipe the dual onto the disk and what we get is still a disk. But we gain something because we see that the new disk has trivial double point loops.

Pushing down. Another picture which we call *pushing down* is shown in Figure 57. Lots of our constructions have a surface coming into another surface sort of forming a T-junction (for example, when we glue in Whitney disks). What we see is that one point of intersection between a and b is traded for two points of intersection of opposite sign



Figure 56: Piping to improve double point loops.

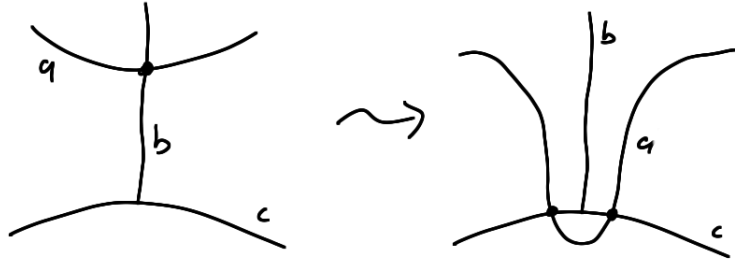


Figure 57: Pushing down.

between a and c .

Spinning. The last of the schematic pictures is called *spinning* which we have already used to adjust framings of Whitney disks at the expense of creating intersections. Suppose we have one edge of a surface a and another thing W coming in. Then spinning means that we change the normal picture by first pushing the arcs behind and then bring the back arc through to create an intersection point at some moment and then after creating that intersection point the arc stays on top and we push it back to the original position.

The Whitehead link. Next we point out some symmetric links that we have to deal with. We've already encountered the *Whitehead link* (Figure 59) as the beginning of a Casson handle. Recall that in Kirby calculus it represents a neighborhood of a disk with a single self-intersection. The symmetry allows us to draw either component as a trivial unknot. (Technically, this link appears in two forms depending on the sign of the claps, but this seems to be irrelevant for the decomposition space theory.)

There's a quite useful generalization of this symmetry. Let $\text{Wh}_{j,k}(H)$ be the link

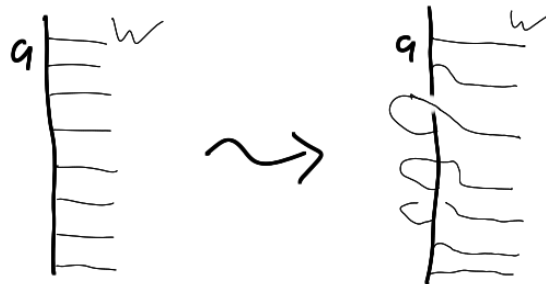


Figure 58: Spinning.



Figure 59: Symmetry of the Whitehead link.

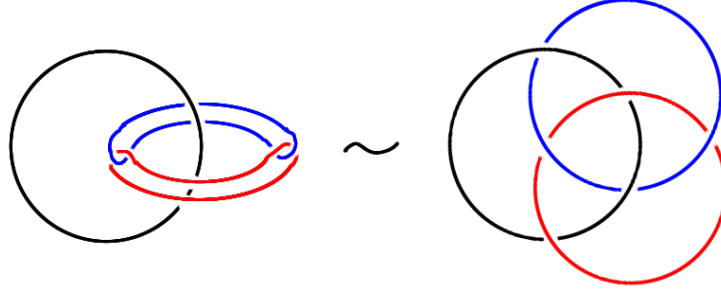


Figure 60: A symmetry of the Borromean rings

obtained from the Hopf link H by taking the untwisted Whitehead double of the first and the second component j and k times, respectively. Note that the Whitehead link is either $\text{Wh}_{0,1}(H)$ or $\text{Wh}_{1,0}(H)$ and the symmetry says that these two are isotopic.

Exercise 7.1. Show that $\text{Wh}_{j,k}(H)$ and $\text{Wh}_{j',k'}(H)$ are isotopic if and only if $j+k = j'+k'$.

The Borromean rings. Another link arose when we studied the double of the Alexander gored ball. There we encountered a link in a solid torus and referred to it as the Bing double of the core circle. But it's often convenient to represent the solid torus dually by drawing the core of its complement in the 3-sphere (which is also a solid torus) and if we do that we get the *Borromean rings* which have a threefold symmetry as indicated in Figure 60.

While we're drawing the Borromean rings, there's also a 4-dimensional picture associated with them. If we put dots on two components, then we claim that we see a Kirby diagram of a thickened punctured torus $T_0 \times B^2$ where the third component with the 0-framing represents $\partial_- = \partial T_0 \times B^2$ (see Figure 61). To see this, note that T_0 deformation retracts onto a wedge of circles and as an abstract manifold $T_0 \times B^2$ can't be anything

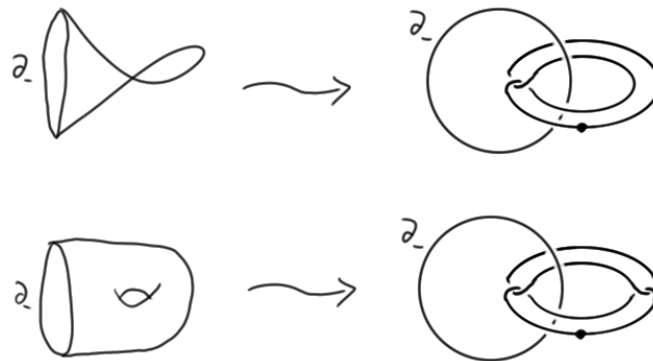


Figure 61: From immersed disks and surfaces to Whitehead and Bing doubles.



Figure 62: The Borromean rings and $T_0 \times B^2$.

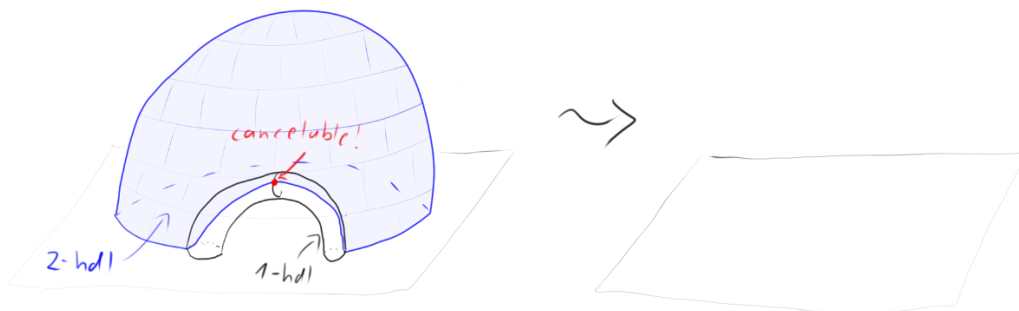


Figure 63: An igloo melting back into the arctic.

else than $S^1 \times B^3 \natural S^1 \times B^3$. In Kirby calculus this is represented by a dotted 2-component and the two circles appear as meridians of the dotted circles. Furthermore, the core of ∂_- should be representable as a curve in the complement of the dotted circles and it has to represent the commutator of the meridians in the fundamental group. And since this is a very simple picture it must be the simplest possible such curve and therefore it's the third component of the Borromean rings.

If that proof doesn't satisfy you entirely, here's a more math style derivation. When we analyzed the Alexander gored ball we sat that $T_0 \times I$ can be obtained from $B^2 \times I$ by drilling out two 3-dimensional 1-handles. Now, if we cross with another I -factor to get $T_0 \times B^2$, then we simply have to cross the 3-dimensional picture with I and on the boundary we see what's shown in Figure 62.

Some more Kirby calculus. Some of the basic ingredients in manifold topology is manipulating Morse functions and the most basic manipulation is to cancel critical points of adjacent indices when they're geometrically in a configuration where they can be canceled; the way Smale would have said this is that the descending manifold of the index $k + 1$ critical point meets the ascending manifold of the index k point transversely in one point. The paradigm picture for this is the igloo melting back into the arctic (see Figure 63) where a 3-dimensional 1-handle (the archway) and 2-handle (the roof) are canceled. The cancellation criterion is satisfied since the roof edge passes over the archway exactly once and the cancellation proceeds by melting the igloo back down into the ice.

So how does this look in Kirby calculus? In the case of index 1 and 2 (which is all we'll need) there'll be a dotted circle representing a 1-handle and the attaching circle of the canceling 2-handle linking the dotted circle geometrically once (see Figure 64). Canceling them means that they disappear from the picture but we can't just erase them, we have to pay attention to other things that might pass through (indicated by the tubes). If

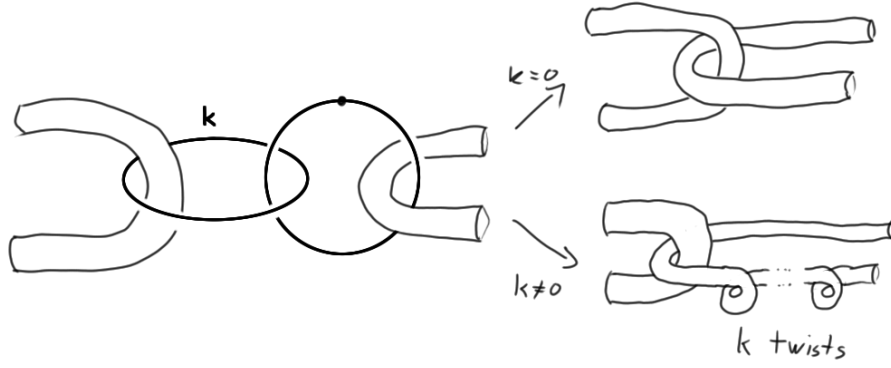


Figure 64: Cancellation of 1- and 2-handles in Kirby calculus.

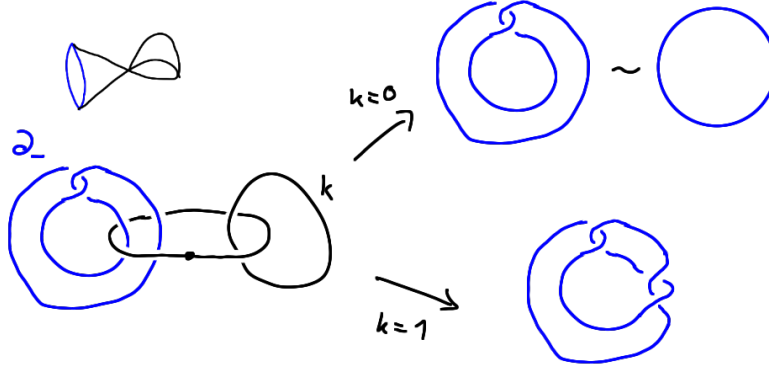


Figure 65: Cancellation in a simple Casson tower.

nothing passes through the dotted circle, then erasing works fine, but if something passes through, then we have to modify the tubes as in Figure 64. Note that the framing of the two handles doesn't matter for the cancellation criterion but it does affect the way the cancellation appears in the diagram.

Actually, in the diagrams for Casson towers and handles all the framings will be zero if we do it correctly, so we never have to do the calculation with a nonzero twist. In order to fully appreciate why we need zeros let's look at a simple example and see what happened if there was a non-zero twist.

Example 7.2. We consider a 2-stage Casson tower with one disk in each stage such that the first stage has one self-intersection, the second is embedded and its framing produces framing coefficient k in the Kirby diagram (see Figure 65). By construction, the 2-handle of the second stage cancels the 1-handle of the first stage and in order to make the cancellation more transparent we have used the symmetry of the Whitehead link to draw the 1-handle as an unknot and the attaching region ∂_- as a Whitehead double. According to the rules, when we cancel we have to twist what goes through the 1-handle side k -times, in this case that's ∂_- .

Suppose first that we were lucky and the 2-handle had zero framing. Then we can simply erase the pair and we don't have to twist anything. What's left would be ∂_- represented by a zero framed unknot. But that's just the picture of a 2-handle

$$(B^2 \times B^2, S^1 \times B^2) \cong (B^4, 0\text{-framed unknot}).$$

So this would be great it's perfectly consistent with the schematic picture (top left) because it says that the little kink in the first stage wasn't necessary after all.

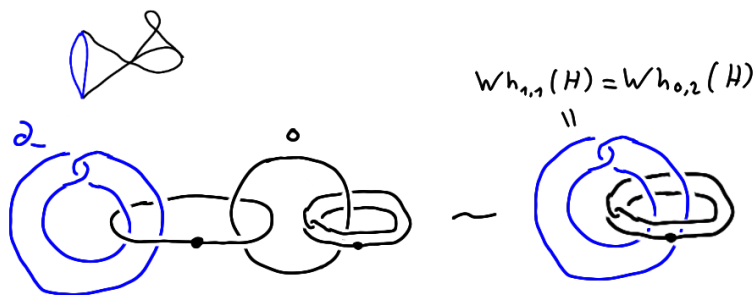


Figure 66: Cancellation in a more complicated Casson tower.

However, if k wasn't zero but $k = 1$, say, then we'd have to put one full twist in ∂_- and we'd either get a trefoil or a figure eight knot (depending on the sign of the clasp). In any case, neither of them is slice and it doesn't look like there's a smoothly embedded core associated with this construction. So this is not an auspicious start for a 4-dimensional 2-handle and we've kind of gotten off on the wrong foot. This is one of the many places where it's necessary to keep zero framings propagating in Casson's construction.

Example 7.3. Now let's consider a general Casson tower with only zero framed 2-handles. Again, by construction, all 1-handles except for those in the last stage can be canceled with the 2-handles of the subsequent stages. What would that look like? To keep things simple let's first assume that each stage has only one disk with a single self-intersection; Figure 66 contains shows the case of two stages. Using the symmetry of the generalized Whitehead links (Exercise 7.1) it is easy to see that after canceling we see the link $Wh_{0,n}(H)$ where n is the number of stages, that is, we see ∂_- as a trivial unknot and a 1-handle which appears as an iterated Whitehead double of a meridian.

From here on, it is probably obvious that the general case, with an arbitrary branching pattern in the stages of the tower, results in ∂_- as a trivial unknot together with ramified, iterated Whitehead doubles of a meridian.

Exercise 7.4. Use cancellation in Kirby calculus to show that

- (a) as an absolute space, a Casson tower is diffeomorphic to $\mathbb{H}^r S^1 \times B^3$ for some r .
- (b) the interior of a Casson handle is diffeomorphic to \mathbb{R}^4 .

Remark 7.5. As we keep Whitehead doubling links, in some sense they become more and more fragile with respect to 4-dimensional topology. It's not link the link would become the unlink or anything like that because each doubling gives an amalgamation with a free group in the link group but, for example, if we start with a link that has non-trivial Milnor invariants, then after we double a couple of times we get a boundary link whose Milnor invariants vanish, and if we double further, we get something called a *good boundary link* which, for all our knowledge, looks like it was smoothly slice. So it look like if we can double things repeatedly, we're making the link closer and closer to the trivial link. That philosophy is intimately related to the idea that it's good to go to higher and higher Casson towers. Somehow the fact that the boundary region of a Casson handle is tangled up with a more and more highly doubled 1-handle as we increase the tower should intuitively make us think that the 1-handle is influencing the attaching region less and less and it's looking more like an actual 2-handle.

7.2 The boundary of a Casson handle

A Casson handle CH is a non-compact 4-dimensional manifold improperly embedded in 4-space and its boundary has two pieces

$$\partial\text{CH} = \partial_-\text{CH} \cup_{T^2} \partial_+\text{CH}$$

where $\partial_-\text{CH}$ is just a solid torus, the attaching region, and we want to explain what the more complicated piece $\partial_+\text{CH}$ looks like. We assume that CH is given to us as a Kirby diagram living in S^3 , which we consider as a union of two solid tori

$$S^3 \cong S^1 \times B^2 \cup_{T^2} B^2 \times S^1 (= \partial(B^2 \times B^2)) \quad (7.1)$$

where $\partial_-\text{CH}$ shows up as $S^1 \times B^2$, and the idea is to keep track of what's happening to the other solid torus $B^2 \times S^1$ throughout the construction. It turns out that the Kirby diagram gives us a precise way of understanding this.

Remember that when we discussed the Alexander horned sphere as the boundary of the gored ball it was very helpful to focus on the portion of the boundary which was there permanently. In the construction of the gored ball as an infinite union (with some limit set thrown in), in each finite stage of the union there were always some annulus regions in the boundary that were there only temporarily, they weren't in the boundary anymore after we attached the next stage. Similarly, in one higher dimension, as we go out in the Casson handle attaching stages of kinky handles, the attaching regions of the kinky handles are only temporarily in the boundary; when we attach the k -th stage it has some solid tori in it which disappear from the boundary when we attach the $(k+1)$ -th stage.

Let's see how this works in detail. For convenience we write $T_n \subset \text{CH}$ for the Casson tower in the n -th stage and note that the boundary decomposes as $\partial T_n = \partial_-\text{CH} \cup_{T^2} \partial_+ T_n$. Since $\partial_-\text{CH}$ is not affected by going to higher stages it is enough to focus on the ∂_+ -parts. At the very beginning, we just have $T_0 = B^4$ and $\partial_+ T_0$ is the solid torus $B^2 \times S^1$ from (7.1). In the first stage we attach 1-handles along dotted circles in $\partial_+ T_0$ which appear as untwisted Whitehead doubles of the core $\{0\} \times S^1$. Recall that the dotted circle notation for 1-handles means that a neighborhood of a standard slice for the dotted circle is to be removed from B^4 . This has the following effect on the boundary.

Exercise 7.6. Show that carving out the standard slice underneath a dotted circle from B^4 changes the boundary by a zero framed surgery on the dotted circle.

So we obtain $\partial_+ T_1$ from $\partial_+ T_0$ by zero framed surgery on the dotted circles, that is, we remove a neighborhood of each dotted circle and replace it with a solid torus such that the longitudinal push off of the dotted circle bounds a disk. Thus $\partial_+ T_1$ has two parts, $\partial_+ T_0$ minus neighborhoods of the dotted circles and a collection of solid tori which are slightly more impossible to draw; the former is a cobordism from $T^2 = \partial(\partial_+ T_0)$ to tori around the Whitehead curves given by the dotted circles. The important observation is that the impossible to draw tori are isotopic to the attaching regions of the 2-handles in the second stage and will thus disappear when we pass to $\partial_+ T_2$, so we don't actually have to draw them. Moreover, the dotted circles of the second stage can be isotoped into the attaching regions of their corresponding 2-handles so that the cobordism part of $\partial_+ T_1$ remains untouched when passing to $\partial_+ T_2$, in fact, it will stay permanently in the boundary and gives the first piece of $\partial_+\text{CH}$.

With this understood, it's very easy to figure out how to keep going. In each stage we're supposed to put in the attaching regions of the 2-handles of the next stage, which would fill in all the holes in S^3 again, and then remove neighborhood of the dotted circles. In particular, if there were no extraneous 1-handles hanging off in the next stage, the

attaching regions of the 2-handles would repair the damage in S^3 and then we'd see that ∂_-CH bounded a disk and we'd be done. But realistically, there will always be a proliferating number of 1-handles and that means that the next stage is not a solid torus but it's a solid torus minus some Whitehead curves sitting at its core. So ∂_+CH is a cobordism from a torus to a (perhaps iterated) Whitehead double, followed by another such cobordism and so on.

What we see is that ∂_+CH is something that we've already run into when we were studying decompositions. If there was no branching, it would just be the solid torus $B^2 \times S^1$ minus a standard Whitehead continuum and, if there was branching, it would be a solid torus minus a Cantor set worth of Whitehead continua which correspond to the endpoints of the tree which describes the branching pattern.

7.3 An exercise in wishful thinking

Later we will develop some combinatorics involving a careful study of dual spheres that will enable us to achieve a lot of additional control over Casson's construction. But let's jump ahead a little bit and imagine that

- (a) such control has been achieved and
- (b) there was a variant of the construction that also involved surface stages

to see what this would say about the boundary.

Casson's original construction gives some non-compact thing which might fill up the whole manifold which Casson himself analyzed to a certain extent. In particular, he showed that an open neighborhood of the construction was proper homotopy equivalent to a standard handle, but that was a bit of a dead-end result, it wasn't possible to do much with that statement. But imagine that the construction was geometrically controlled – for example, like our construction of the Alexander gored ball where we went smaller and smaller in each step – and proceeded to a well defined limit. Then ∂_- would still be a solid torus but ∂_+ now would be the decomposition space of a ramified Whitehead decomposition instead of its complement. Indeed, the farther and farther out stages would get smaller and smaller which is exactly what would happen if we had taken that infinite intersection and crushed it to a point. So where we wanted to have a handle we at least see something with a manifold factor (remember Theorem 4.10?) as its frontier and if this were topologically collared, we'd see a solid torus cross an interval in there; everything would resolve and that would mean that we could put in a topological Whitney disk and we'd really be cooking but, of course, we have no way of knowing it's collared. But at least it looks like we're moving in the right direction.

Now, as long as we're imagining, suppose when we build this construction, instead of always using disks with double points, we were able to replace such a disk with an embedded surface. In fact, that's not such a far fetched thing because double points of surfaces can be resolved locally with the result of increasing the genus; in a ball around the double point we can simply replace the two intersecting sheets by the annulus connecting the two components of the Hopf link on the boundary (see Figure 67).

At the end of the day it turns out that, once we've found a Casson handle, then we can replace it with a construction that has any number of surface stages alternating in any fashion we like with any number of disk stages. For convenience we'll still call such a modified construction a Casson handle although Casson didn't talk about surfaces.

Now suppose we can do the modified construction in a controlled way; that's going to be a winning combination. Instead of just getting successive layers of disks we could intersperse an arbitrary number of surface stages whenever we like and at the end everything would be getting smaller and smaller and converge to a Cantor set. But from what we have learned in picture camp we know that the frontier of such a construction

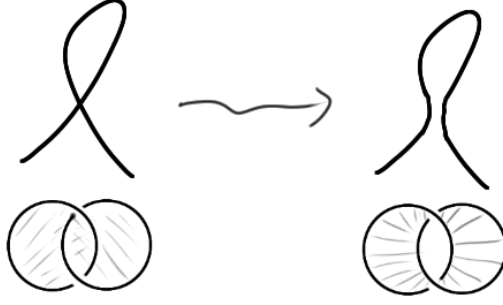


Figure 67: Resolving double points of surfaces.

would be the decomposition space of a mixed Bing-Whitehead decomposition; each disk stage gives a Whitehead double while surface stages result in Bing doubling (Figure 61). Now, remember that we had this result of Starbird and Ancel (Theorem 4.12) which tells us exactly when such decompositions shrink. Roughly, we need lots of Bing doubles to compensate for a few Whitehead doubles, in other words, we need lots of surface stages to make up for each disk stage. To sum up, we can expect

$$\begin{aligned}\partial_+ &= (B^2 \times S^1) \setminus \mathcal{D}_{\mathcal{W}/\mathcal{B}} && \text{without control and} \\ \partial_+ &= (B^2 \times S^1)/\mathcal{D}_{\mathcal{W}/\mathcal{B}} && \text{with control.}\end{aligned}$$

In particular, if there are enough surface stages in the controlled case, then $\mathcal{D}_{\mathcal{W}/\mathcal{B}}$ is shrinkable and get $\partial_+ \cong B^2 \times S^1$ so that the frontier of this infinite construction looks exactly like the boundary of a 2-handle! That's considerable progress because if we think of goal as trying to produce a 2-handle where we wish to find a Whitney disk, at this point the boundary of whatever we have found is find, except we don't know if it's collared.

Our strategy will be to produce not only one frontier but to fill this kind of Casson handle up with many frontiers, each of which is recognized to be the standard thing, and at the end of the day very little of the Casson handle will be left unexplored, so little that it can ultimately be related to the standard handle by the sphere-to-sphere theorem (Theorem 5.1).

Remark 7.7. One might wonder: why we don't take just surface stages? After all, this would always give shrinkable decompositions. The problem with this approach is that we will actually need some disk stages to achieve geometric control. In turns out that we really need to intersperse with some disks to keep local information about the fundamental group.

7.4 A glimpse at reimbedding and the Design

To give more motivation for the combinatorics to come we continue to plow ahead and take a look at what we like to call “the design”. One of the conclusions of these combinatorics is the idea of *reimbedding* which roughly says that, if we have a sufficiently complicated finite construction, then we can produce arbitrarily larger finite constructions inside.

Theorem 7.8 (Reimbedding for Casson towers). *Inside any Casson tower T_m^0 of sufficiently large height m (that is, there are m stages of disks) there is another tower $T_n^1 \subset T_m^0$ of arbitrary height n such that T_n^1 and T_m^0 have the same bottom stage.*

Remark 7.9. In [Fre82] we used the constant $m = 6$ which is not optimal. In fact, Gompf and Singh [GS84] lowered the bound to $m = 5$ which may or may not be optimal. But at the end of the day it doesn't really matter what the constant is as long as there is one.

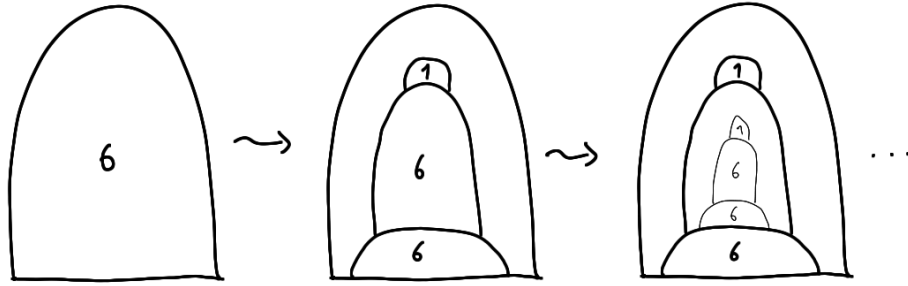


Figure 68: Reimbedding, first attempt: $13 = 6 + 6 + 1$

Let's see what we can do with this. For example, we can take something of height 6 and locate something of height 13 inside which we consider as something with six stages, six further stages and one more stage on top (just to make sure that nothing is tangled up in the fundamental group). Then we do the same thing in the middle six stages and repeat this process ad infinitum as shown schematically in Figure 68.

This picture already shows how to obtain control. We can modify each step by using the top stages to show that the middle six stages are null-homotopic inside the original 6-stage tower. After all, as an absolute manifold the middle six stages are just a boundary sum of copies of $S^1 \times B^3$ (and thus have a 1-dimensional spine), the final layer then provides null-homotopies for each $S^1 \times \{0\}$ and in four dimension homotopy implies isotopy for 1-dimensional things. So this whole middle 6-stage tower can be moved into a little ball somewhere inside the original tower and only then we focus on its internal structure. By construction this modified process forces the towers to converge to points and if there was branching in the construction we'd see a Cantor set again.

This controlled construction gives us a single Casson handle and its frontier in the tower we started with – remember that our goal is to fill up most of this handle with, ultimately, a Cantor set worth of frontiers. In order to do that we need another construction that bifurcates in another direction. This means that, every time we do this reimbedding, we want to have two choices, sort of an inner and an outer way to proceed. So we start over with a 6-stage tower but now we take a tower with $19 = 6 + 6 + 6 + 1$ stages inside. Now there are two middle 6-stage towers in each we find another 19-stage tower and so on. Then the frontier exhibits a dyadic branching pattern as indicated in Figure 69 and it's easy to explicitly see what we're building: in the 3-dimensional context each piece would be $B^2 \times S^1$ minus a neighborhood of a ramified Whitehead link⁸ and the thick red lines indicate that we take a small neighborhood so that each piece is of the form $(B^2 \times S^1 \setminus \nu L) \times I$ for some link L .

Now since these pieces are explicitly parametrized, we can take the standard 2-handle $B^2 \times B^2$ and start fitting them into it (see Figure 70). What we end up seeing is, first of all, the attaching region (shaded in grey), and then the red pieces will appear as indicated. In particular, they will get smaller and smaller and converge to some limit set (which may or may not be a Cantor set, this depends on whether we work with disks only or allow surfaces). What we have found is some common closed set in both the Casson handle and the standard handle which we call *the Design*.

Remark 7.10. The path to the proof that we're going to explain – before we double back and give the original proof whose high point is a fantastic shrink due to Bob Edwards [Fre82, Chapter 8] – is to use surface stages because we can lean rather heavily on what we've already learned about shrinking decompositions that have adequate Bing

⁸Remember that we're only using disks at this point, if we were also using surfaces, we'd see mixed Bing-Whitehead links.

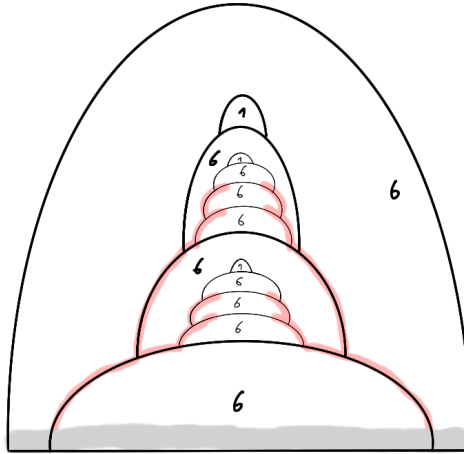


Figure 69: Reimbedding, second attempt: $19 = 6 + 6 + 6 + 1$ (aka “the Design”, part 1).

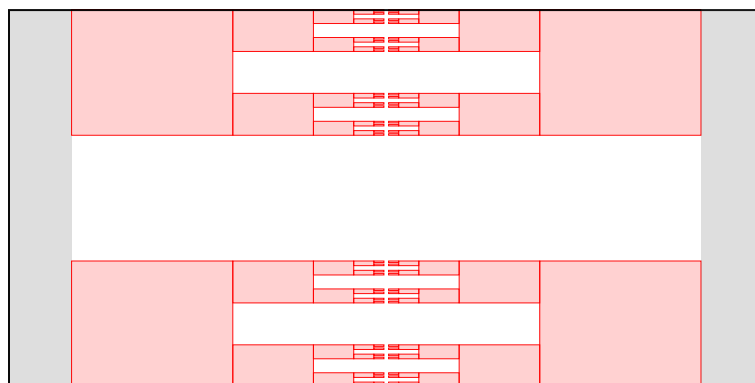


Figure 70: Filling up the standard 2-handle (aka “the Design”, part 2).

components to them. The point of going through enough combinatorics to work with surface stages – a technology that was unavailable in 1981 – is that we can really arrange the limit set in the standard 2-handle to be a Cantor set.

The complement of the Design in both the exotic and standard situation is a countable number of regions and, although they’re a tiny bit mysterious in the exotic guy, back in the standard handle all this is explicitly coordinate driven and they’re just product solid tori. The proof will finish by studying a common quotient Q

$$\begin{array}{ccc}
 B^2 \times B^2 & \xrightarrow[\cong]{} & CH \\
 \downarrow \scriptstyle (ABH) \alpha & & \downarrow \scriptstyle \beta (ABH) \\
 & Q &
 \end{array}$$

where α and β are decomposition maps and the argument will be that we can first shrink the α -decomposition using the bird-like equivalent techniques so that α turns out to be ABH and that will tell us that Q is a ball. Then we have a map from CH to a ball and, although we don’t know that CH is a ball, we’ll understand that it is contained in a ball (in a similar way that the Alexander gored ball is contained in the standard 3-ball) and we can use the ball-to-ball theorem to deduce that β is also ABH so that CH must be homeomorphic to the standard handle.

8 Combinatorics: constructing infinite convergent towers

Original notes for this section expanded upon by Wojciech Politarczyk and Mark Powell.

We are first going to explain some of the original combinatorics in working with honest Casson handles [Fre82] and then go into more detail of the combinatorics involved when a Casson handle is modified so as to also have surface stages, as in [FQ90]. The latter turns out to be much simpler, and is really quite an improvement. We begin by explaining gropes, then we do reimbedding for Casson towers, making use of some grope technology to simplify the original argument. Then we explain height raising for capped gropes, before proving the main result of this section on controlled reimbedding for towers, which are formed from a mixture of surface and disk stages. The controlled reimbedding for towers construction will be iterated in order to construct the “design” in the next chapter.

8.1 Gropes and transverse spheres

Time in lecture: 0:00 until 43:00.

We have already met a grope in Section 3.3.2 when we discussed the Alexander gored ball as a union of thickened, punctured tori. The term grope was introduced by Jim Cannon and we are going to use the word to mean something that might be either 2-, 3- or 4-dimensional. The dimension should always be clear from the context. A grope is determined by a 2-complex; a 3- or 4-dimensional grope are thickenings of this 2-complex.

To construct a *2D grope*, start with an oriented surface with one boundary component called the *base* or *1st stage*, take a symplectic basis⁹ of curves on the surface, attach further oriented surfaces with a single boundary component – the *2nd stage* – and repeat this process (see Figure 71). Note that each surface in the construction can have arbitrary genus but since the pictures get exponentially complicated anyway we will draw only genus one surfaces. Abstractly, a 2D grope is a 2-complex, but we can embed it in 3-space by

⁹A *symplectic basis* is a basis of first homology consisting of simple closed curves $a_i, b_i \subset$ such that a_i and b_i intersect transversely in one point and there are no further intersections.

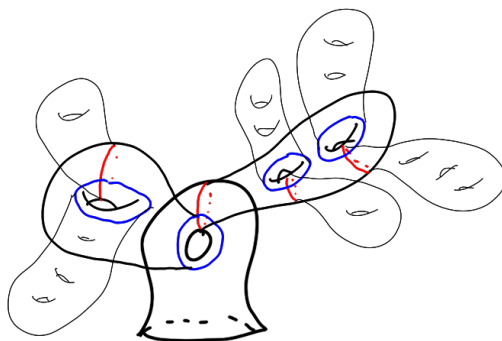


Figure 71: A grope as a 2-complex.

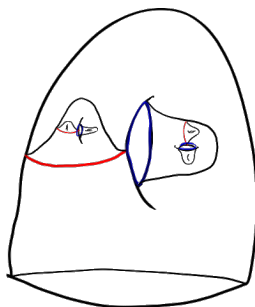


Figure 72: A 2-dimensional grope in 3-space.

attaching half of the surfaces in each stage on the outside and the others to the inside as in Figure 72. In order to get a *3D grope* we simply take a thickening of this picture in 3-space. (This is fine for any finite stage but if we do an infinite construction and want a limiting object, then it is important that the stages shrink and limit to a Cantor set.) Finally, to get *4D gropes* we just take the 3-dimensional model and cross it with an interval. The 4D version can also be described by a handle picture in the spirit of the pictures for Casson towers. The ∂_- attaching circle is specified, and then the pictures look almost the same, the only difference is that we see Bing doubles (one for each genus) instead of Whitehead doubles (Figure 73). Note that Bing doubling an unknot turns a meridian of that unknot into a commutator in the meridians of the two new components.

We generically denote gropes by G and call the boundary of the base surface the *attaching circle* and its thickening the *attaching region* $\partial_- G$ of the grope. Moreover, any 3D or 4D grope naturally contains a 2D grope as a deformation retract, this will be called the *body*.

Remark 8.1. Group theoretically gropes are related to the *commutator series* (or *derived series*) of the fundamental group because a grope displays its attaching circle as a commutator in the first stage, then as a commutator of commutators and then a commutator of commutators of commutators and so on.

At this point in Lecture 8, from 6:30 - 14:20, Mike gave a discussion about half gropes and their relation to the lower central series, string puzzles, and a curve which bounds a capped grope with embedded caps bounds an embedded disc, but the number of times which the caps are used by the disc is exponential in the class of the half grope: the complexity of a symmetric grope is measured by its height whereas the complexity of a half gropes is measured by its class. This was technically unnecessary for developing the theory but important to understand, the reader is recommended to watch the appropriate

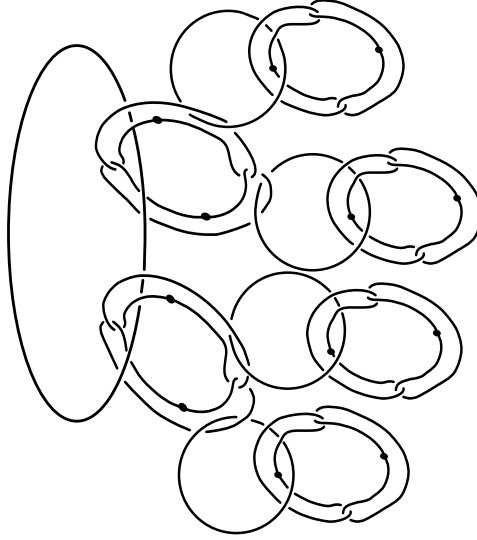


Figure 73: A Kirby diagram of a 4-dimensional grope.

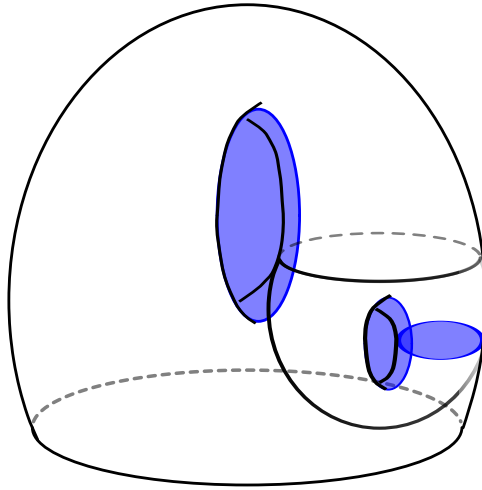


Figure 74: A capped grope.

clip.

From now on we mostly focus on 4D gropes. Note that in our models in 4-space the top stage surfaces of a grope have a symplectic basis of curves which bound embedded disks (already visible in the 3D picture). If we add them to the picture, then we obtain an example of a *capped grope*. More generally, we call any immersed disk in 4-space which bounds a basis curve in a top stage surface and is otherwise disjoint from the body of the grope a *cap* and a *capped grope* G^c is a grope G together with caps for all basis curves in the top stage. Note that we allow arbitrary intersections and self-intersections among the caps.

Remark 8.2. As usual, there is a framing issue that we have swept under the rug. The basis curves have a natural framing in 4-space¹⁰ and this framing must extend across the caps. In Kirby calculus language this means that we attach 0-framed kinky handles to meridians of the 1-handles in the top stage of a grope. However, since we allow the caps to intersect each other it is not completely straightforward to draw the pictures.

¹⁰Up to orientations, they have a canonical framing inside the surface which canonically extends to a framing in 3-space and the latter is canonically framed in 4-space.

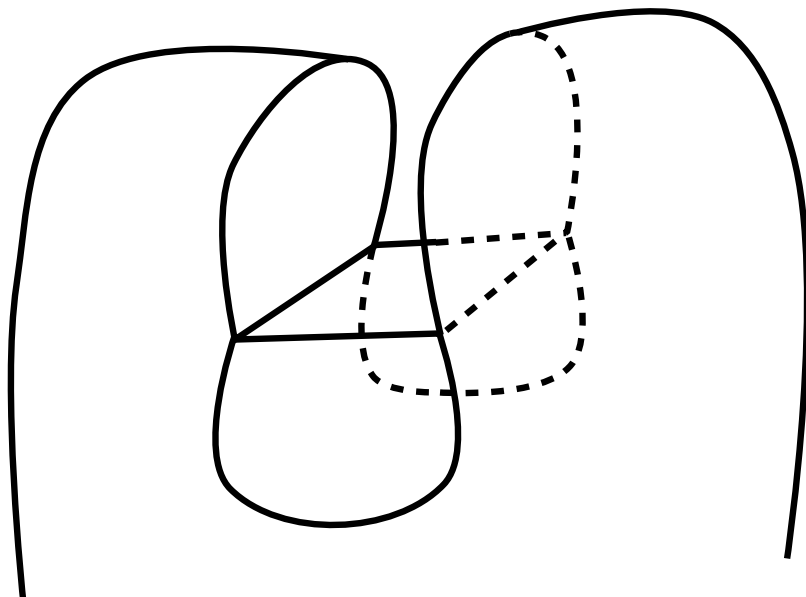


Figure 75: Symmetric surgery aka contraction.

The correct way to deal with framing issues is to introduced the immersions into the framed model, and then define a capped grope in a 4-manifold to be a framed embedding of this model, as in the book [FQ90].

Exercise 8.3. Figure out how intersections among caps appear in the handle pictures.

Exercise 8.4. Show that a capped grope with embedded caps is diffeomorphic to a standard 2-handle (relative to their attaching regions).

We have a lot to learn about 1-stage capped gropes. The discussion by Mike in the lectures here is useful, Lecture 8, 14:20 -20:13. A 1-stage capped grope is simply an oriented surface with a single boundary component with disks attached along a symplectic basis, Figure 75 shows the genus one case. Let us first look at the picture in three dimensions. If we have embedded caps as in the picture, then we can certainly use either one of them to turn the punctured torus into a disk by surgery. But there is another interesting option when both disks are available which is called *symmetric surgery* or *contraction* [FQ90]: we remove cylinders around both curves but we think mod 2 and leave the square where the cylinders intersect in the space, and then we put in four disks (one pair for each cap) back in. The virtue of this construction is that it allows us to trade intersections with caps for self-intersections of whatever was intersecting the caps. Indeed, say something called X intersects the caps as in Figure 76, then the observation if that, after the contraction we can push one arc up and the other down. Note that this process of *pushing off the contraction* creates two self-intersections of X for each intersection with the caps.

Remark 8.5. Lecture 8, 20:13 – 24:00 Why is this an important technology? Recall that one of the key points in Casson’s constructions was to clean things up by promoting homological duals to geometric duals which we called transverse or dual spheres. Recall that a transverse sphere to a surface enables us to see that the surface is π_1 -negligible using finger moves. Thus transverse spheres are what makes these constructions run; they enable us to attach higher stages in Casson towers and eventually to build Casson handles. It turns out that, if we understand it correctly, contraction is a very efficient machine for producing dual spheres. In fact, a test of how efficient it is, is that it has enabled Quinn to prove the annulus conjecture in dimension four [Qui82] (see also [Edw84]). He thought of

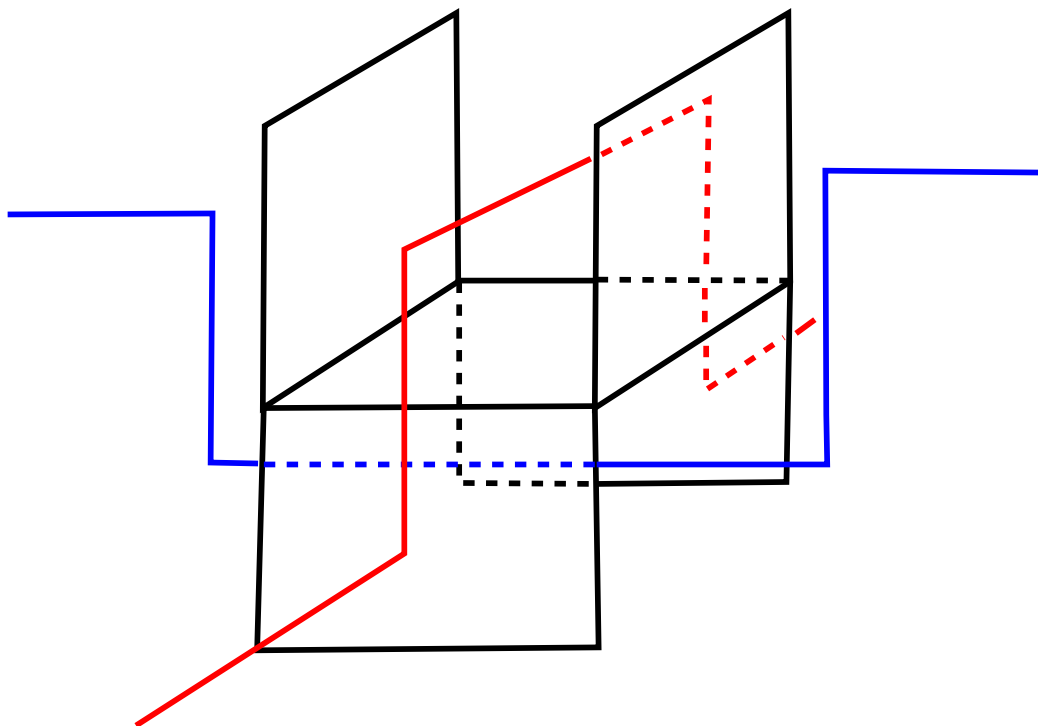


Figure 76: Pushing things off the contraction.

the annulus conjecture as a controlled h -cobordism theorem and in order to do h -cobordism with geometric control you need ready access to dual spheres everywhere, you couldn't go looking for them far away.

In our original approach [Fre82] we were basically making dual spheres “by hand” and, although there was an algebraic presence of contraction picture in the background, we first learned about it from Bob Edwards at the 1982 Durham conference when Quinn had just announced his proof.

The technique of pushing off contractions has had a lot of influence in using gropes as a variant of Casson towers and that is the way that the book [FQ90] is written. Basically, wherever we could we used a capped surface instead of a disk, which is much more convenient in terms of combinatorics.

We are going to try to explain the bare bones the old and the new combinatorics, especially the new grope-based combinatorics, as these gives the route to the technically simplest proof of the disk embedding theorem. But before this, we want to give the most interesting corollary of the contraction trick, which allows us to produce arbitrarily many disjoint duals in one fell swoop. With the old technology we had to find duals one at a time, which was one reason why the original combinatorics were so hard.

To see why having many disjoint duals is useful, suppose we have a surface S intersected once by another surface T . If there is a dual sphere \hat{S} for S we can pipe T and \hat{S} to remove the intersection of T and S . That is good, but suppose we do not have just T but another surface T' and we want both T and T' disjoint from S . Moreover suppose that we only have one dual sphere \hat{S} . Of course, even if we have only one dual sphere \hat{S} we can pipe both T and T' with parallel copies of \hat{S} to achieve disjointness from S , but this is not entirely satisfactory because we add intersections of T and T' , unless \hat{S} happens to be embedded with trivial normal bundle. It would be much more convenient to have not only one dual but an arbitrary number of disjoint ones. So how do capped gropes and contraction solve this? The next lemma is called the “infinity lemma,” even though it

only really guarantees an arbitrarily large finite number of disjoint duals. See lecture 8, 24:30 – 37:50.

Lemma 8.6 (∞ -Lemma). *A (4D) (disk- or sphere-like) capped grope G^c of height n with immersed caps (disjoint from the body) contains arbitrarily many disjoint height $n - 1$ capped gropes with bodies parallel to the first $n - 1$ stages of G^c .*

Proof. For the proof we think of G^c as being 2-dimensional, and perform operations in a neighbourhood of it. Take a parallel copy of G^c , contract the parallel copy to obtain a height $n - 1$ capped grope. Then push the caps of the original off the contraction.

We get a new capped grope of height n G' which is disjoint from the contraction of the parallel copy. Each intersection of the caps of the original gives rise to two intersections between the caps of G' .

We repeat this process. A parallel copy of G' is disjoint from the first parallel copy. Contract this parallel copy and push the caps of G' off the contraction, to obtain a capped grope G'' . We now have two height $n - 1$ capped gropes which are parallel to the first $n - 1$ stages of G^c and disjoint from G'' .

The procedure, which as promised occurs in a neighbourhood of the original, may be iterated as many times as required. \square

Back to the situation described before the lemma, if we remove a disk D from \hat{S} where it meets S , we are left with an immersed disk. Now, if we had a capped grope instead of the immersed disk, then we could cap off parallel copies of D with disks from the lemma and get infinitely many disjoint dual spheres and we can pipe arbitrarily many things off S without making them cross themselves!

8.2 Height raising and reimbedding for Casson towers

Time in lecture: 43:00 until 59:00. The presentation in the lecture contained mistakes (which hopefully we have rectified here), so proceed with caution.

Let us see how we can raise height in Casson towers with the help of the ∞ -Lemma 8.6. In the next section we will talk about the more general towers interlacing surface and disk stages used in [FQ90]. The combinatorics for Casson towers is slightly messier and we could skip this part, but it is the way the original proof [Fre82] worked.

For an excellent account of this reimbedding, the reader is recommended to consult the write up of Gompf and Singh [GS84]. The method we will discuss now, however, is supposed to be improved from that in that we use a transverse capped grope to more easily produce multiple disjoint transverse spheres.

Suppose we are looking at a 5-stage Casson tower. The argument of this section all occurs in a neighbourhood of this tower. The aim is to find a 6 stage tower with the same boundary embedded in this neighbourhood.

Theorem 8.7. *A 5 stage Casson tower has a 6 stage Casson tower with the same boundary embedded in a neighbourhood.*

There is a computation in topology, using Kirby pictures of a Casson tower, that tells you that the second stage has a transverse sphere.

Exercise 8.8. Prove this. Hint: Lemma 4.1 in [Fre82]. To understand the proof of that Lemma one needs to know that dashed curves indicate that the associated stages of the Casson tower have had their interiors removed from the Casson tower, so we should just consider the boundaries of those stages. It is also worth pointing out that the first and last arrows in the picture, Diagram 4.3 of [Fre82], are not Kirby moves but are moves which

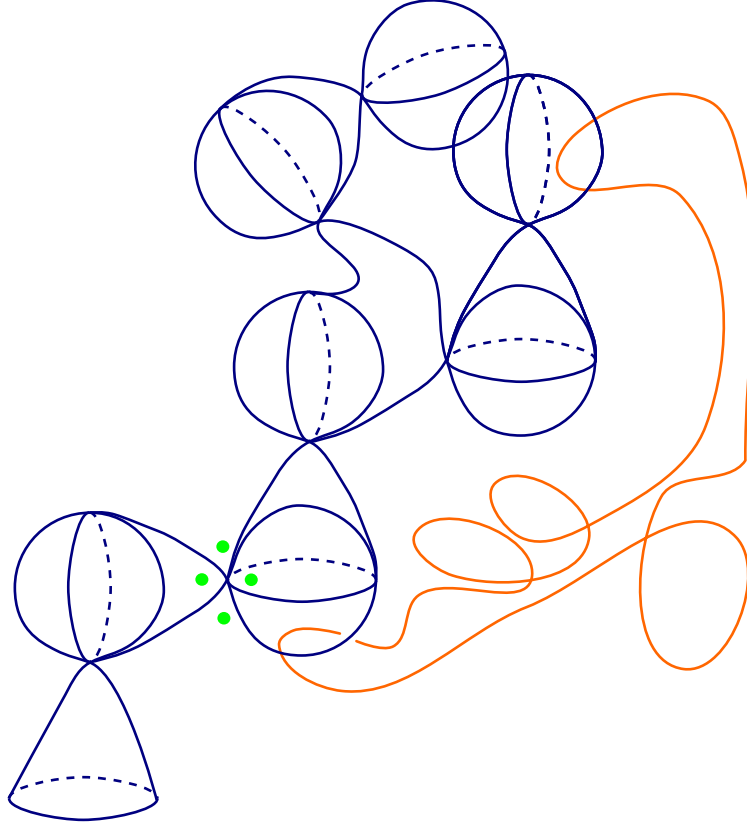


Figure 77: 5-stage Casson tower with orange transverse sphere for the second stage and green Clifford torus.

preserve the fundamental group. The proof in the paper only deals with the unramified case, but the ramified case just involves repeating the calculation for each branch.

This fact is proven, in more generality and with the ramified case dealt with, in [GS84, Lemma 3.1.0].

This transverse sphere for the second stage has two problems: it is immersed and crosses the top stage see figure 77.

What we shall do is to turn this sphere into a dual torus for the third stage which has caps, and the caps cross themselves and the top stage see figure 78.

In the light of Lemma 8.6 this is an improvement since it is more powerful to have a dual capped surface than just a dual sphere. The way this works is that the Clifford tori of double points in the second stage give us the body of the transverse surface for the third stage. The longitude and meridian of such a Clifford torus are linking circles for the second stage, so bound discs which intersect the second stage. Tube these discs into the transverse sphere for the second stage constructed above. We now have a capped surface which is transverse to the third stage, although the caps intersect the fifth stage. By taking a parallel copy, contracting this capped surface back to a sphere and pushing off the contraction we can trade its intersections with the fifth stage for new intersections amongst the fifth stage discs. So now we have a transverse sphere for the third stage which is otherwise disjoint from the tower as in figure 79.

Then we use the Clifford tori in the third stage to obtain dual capped surfaces for the fourth stage, just as above. The Clifford tori in the third stage have caps which intersect the third stage but we tube them into transverse spheres for the third stage.

We can further make all intersections between fifth stage discs into self intersections,

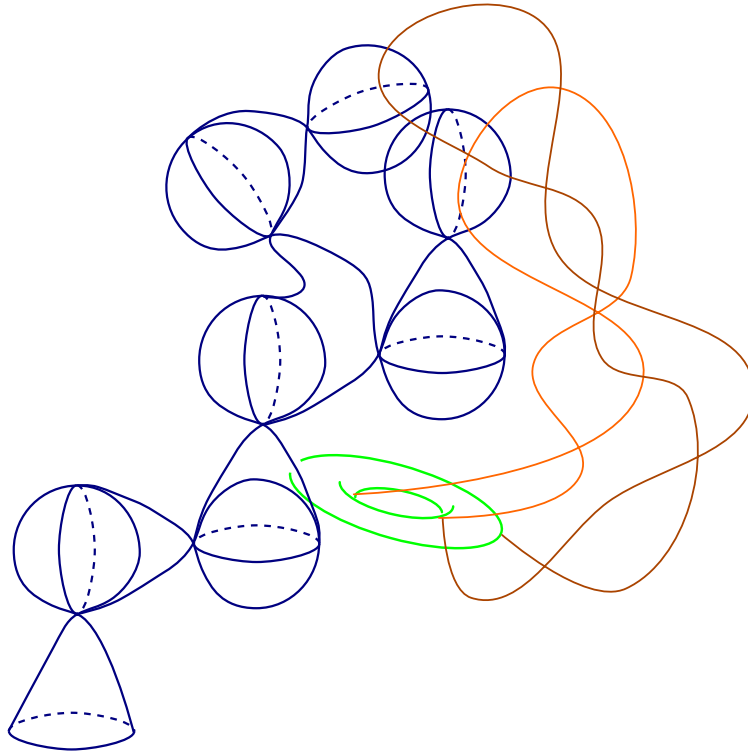


Figure 78: 5-stage Casson tower with transverse capped torus for the third stage.

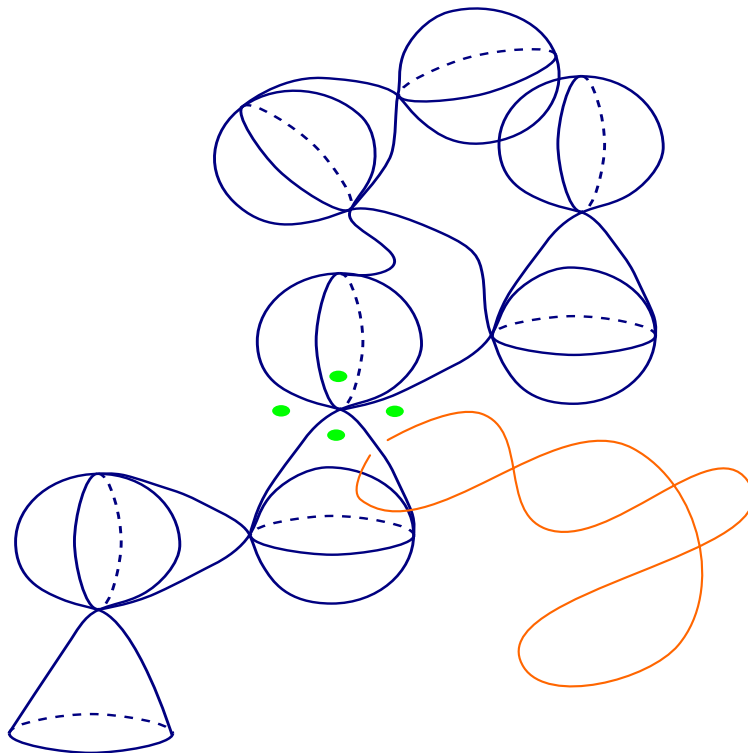


Figure 79: Casson tower with a transverse sphere for the third stage which is otherwise disjoint from the tower.

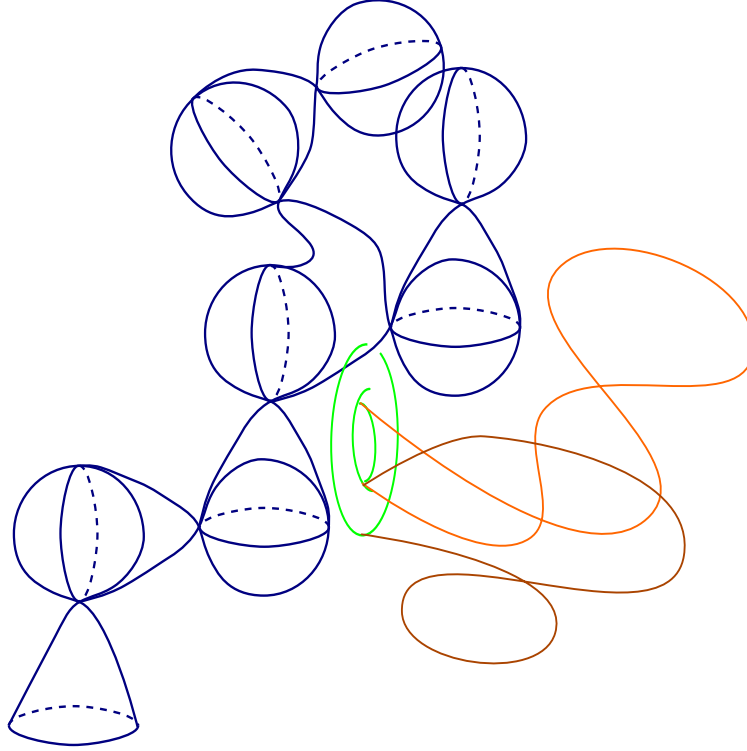


Figure 80: Casson tower with a transverse capped torus for the fourth stage.

by pushing intersections down and tubing into multiple disjoint transverse spheres for the fourth stage from taking the capped surface and applying the ∞ -lemma, see figure 81.

Now we take all self-intersection points of the fifth stage and tube them down into the transverse spheres for the fourth stage. The effect of this is that double point loops become trivial in π_1 , since the double point loops in the transverse spheres came from a lower stage, and the lower stage was part of a higher tower and therefore π_1 null. So the tower we have is π_1 null (see more discussion below).

The next thing is to use that our tower is π_1 -null so that all the double point loops in the fifth stage bound null-homotopies. These null-homotopies may cross the caps of our fourth stage transverse capped surface.

So, contract the fourth stage capped surface and push the null-homotopies off the contraction. This gives us a transverse sphere for the fourth stage which is disjoint from the rest of the tower and the null-homotopies, at the cost of introducing new intersections in the null-homotopies.

Take the Clifford tori for the fourth stage to obtain dual capped surfaces for the fifth stage and, similarly to the construction performed twice above, tube the caps into the transverse spheres for the fourth stage to obtain caps which are disjoint from the rest of the tower.

Now remove intersections between the null-homotopies (which will become the 6th stage of the tower) by pushing down and tubing into the parallel transverse spheres for the fifth stage which come from the ∞ lemma.

Next, the null-homotopies may intersect one another. We trade these for self intersections by pushing intersections down and tubing into multiple disjoint transverse spheres for the fifth stage from the ∞ -lemma. Thus we have a 6 stage tower.

Now, as above, we also take all self-intersection points of the sixth stage and tube them down into the transverse spheres for the fifth stage. Once again the effect of this is

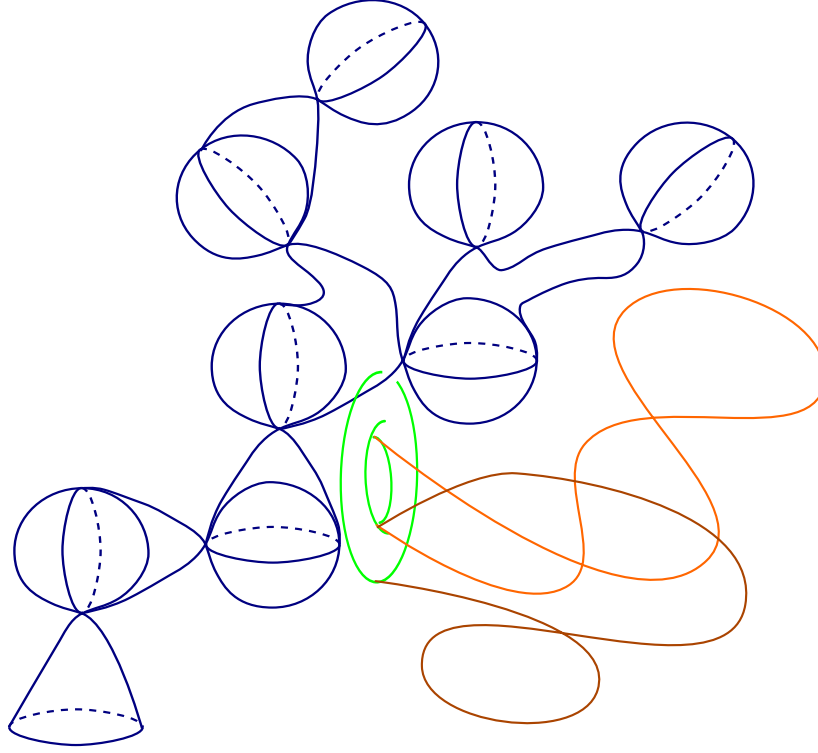


Figure 81: Exchanging intersection between discs in the top stage for selfintersections.

that double point loops become trivial in π_1 , since the double point loops in the transverse spheres came from a lower stage. So the tower we have is again π_1 null.

Now there is a potential problem. There is the possibility that the π_1 -nullity is ruined if one of the caps of a transverse grope intersects the bottom stage. So we add one more stage at the bottom of the whole tower, and say that all along we were just talking about the top 5 stages of a 6 stage tower.

Therefore we see that within any 6-stage Casson tower we can find a 7-stage tower and then we can keep going, applying this to the top 6 stages of the 7 stage tower, and so on. Eventually we prove the reimbedding theorem 7.8, that there is an 13 stage Casson tower inside a 6 stage tower.

8.3 Height raising for gropes

Lecture 8, 1:14:40 to end.

We are over the most painful combinatorics from now on we use capped gropes to height raise instead of all disk stages. We will see that the combinatorics is much more pleasant. Here is the key result.

Lemma 8.9 (Height raising for capped gropes). *Inside a neighbourhood of any 3-stage capped grope G_3^c there exists an n -stage capped grope $G_n^c \subset G_3^c$ with the same attaching region.*

Proof. When building the grope G_3^c we can divide each stage of surface into a left (−) and right branch (+) because of the dual curves in the symplectic basis. (In the picture in 3-space this would be the inside surfaces versus the outside.) On top we have the caps which are immersed and cross all through each other. Schematically, this is shown in the left picture in Figure 82.

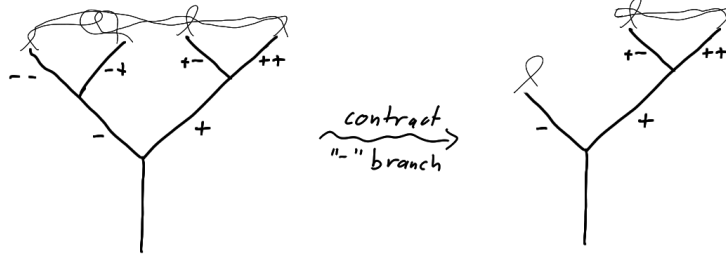


Figure 82: Shortening the left branch to clean up cap intersections.

The first thing we can do is to take the left branch and to sacrifice one of its bits of height in order to get all the intersections off. More precisely, we contract the third level of the left branch to push all intersections with the caps from the right branch off and we get the right side of Figure 82. The left branch of the second stage is then just a capped surface.

In the next step we use the shortened left branch to create a capped surfaces dual to the surfaces in the right branch of the second stage. To keep things simple we assume that everything has genus one; the general case is only more complicated from the point of view of notation. Denote the base surface by Σ and let Σ_{\pm} be the left ($-$) and right ($+$) branches of the second stage which are attached along a symplectic basis $C_{\pm} \subset \Sigma$. If we take the unit normal bundle of Σ and restrict it to C_- , then we get a torus T and it is easy to see that T intersects Σ_+ in one point. So we have found a homological dual to Σ_+ which is already a good sign but, of course, we want a higher quality dual. Note that T meets the left branch Σ_- in a circle (in fact, a parallel push off of C_-) so that we can remove a neighborhood of this circle from T and fill the resulting two boundary components with two parallel copies of the capped surface Σ_-^c . This “surgery” turns T into a capped surface and since the caps of Σ_- did not interfere with the right branch, we really get a dual capped surface for Σ_+ .

At this point the right branch still has all three stages and caps on top. Now we use piping to convert all intersection among the right branch caps into intersections with Σ_+ (recall that the conversion ratio is 1:4) and for each of these intersections we throw in a disjoint copy of the dual capped surface for Σ_+ . This turns each cap from the right branch into a capped surface. This recreates all kinds of intersections between right branch caps and left branch caps, so in a sense we are back to where we started. But notice that the multiplicities have changed. The left branch now has height one over the base surface but the right branch has three stages on top of the base surface.

Now we simply reverse the rôles of the right and left branches. We do a contraction on the right branch to get the caps disjoint, create something dual to the left branch and pipe away. The important observation is that this time we not only get a dual capped surface but actually a dual 2-stage capped grope. So the height of the right branch goes down from 3 to 2, but the left branch grows taller from height 1 to 3. By iterating this process further and alternating between the left and right branch we see that the height of the branches evolves as in the table in Figure 83. In summary, we can first raise the height of both branches asymmetrically and then use contraction to cut the heights down symmetrically. \square

Remark 8.10. Having to work alternately on the left and the right side seems to be a general feature of using gropes. The reason is that, although we can use a copy of the left side to make a dual for the right side and vice versa, these dual are not disjoint, they intersect in two points. That usually spoils more homogeneous arguments so that we have

left	right
$1 + \textcolor{red}{2}$	$1 + 2$
$1 + 1$	$1 + \textcolor{red}{3}$
$1 + \textcolor{red}{3}$	$1 + 2$
$1 + 2$	$1 + \textcolor{red}{4}$
$1 + \textcolor{red}{5}$	$1 + 3$
$1 + 4$	$1 + \textcolor{red}{7}$
\vdots	\vdots

Figure 83: The heights of both branches progress in Fibonacci sequences. (The red numbers indicate on which branch we are working.)

to go back and forth between the two sides.

Remark 8.11. Lemma 8.9 is kind of a baby reimbedding lemma. But we will actually have to work with gropes that have not only one but two layers of caps.

Remark 8.12. A transverse grope for a capped grope G^c is a properly immersed capped grope $(G^c)^t$ whose first stage surface intersects a first stage surface of G^c in exactly one point and this point is the only intersection point of the bodies. Additional intersections are allowed between caps of the two capped gropes. Using similar arguments as in the proof of Lemma 8.9 one can prove that each component of the $+$ side and each component of the $-$ side have transverse capped gropes. Transverse gropes for the components of the same side are always disjoint. For more details see [FQ90, Section 2.6].

We can actually do better than the previous height raising result. Namely it is really only necessary to start with a height 1.5 grope.

Lemma 8.13 (Height raising for capped gropes mark II). *Inside a neighbourhood of any 1.5-stage capped grope G_3^c there exists an n -stage capped grope $G_n^c \subset G_3^c$ with the same attaching region.*

Proof. The proof will be presented as an adaptation of the proof above. To begin, we have a $+$ -side capped surface, while the $-$ side is just caps.

First we remove intersections between the $+$ and $-$ caps. Let F_- denote the dual capped surface to the $-$ side, which is constructed from two parallel copies of the $+$ side capped surface and an annulus which joins them together in a neighbourhood of the attaching circle for the $+$ side.

Take a parallel copy of F_- , contract and push everything off the contraction as in Lemma 8.6. We obtain a sphere F'_- , and the capped surface F_- is modified so as to be disjoint from F'_- . Repeat this procedure: we want one of these dual spheres for each intersection of $+$ and $-$ caps. Now, for every intersection of $+$ and $-$ caps, tube the $+$ cap into one of the F'_- dual spheres. The duals F'_- may intersect the $-$ caps, so we may obtain new caps $- \cdot$ caps $-$ intersections by this operation. That is acceptable.

Now tube each intersection of the $-$ side caps into a parallel copy of F_+ . This turns the $-$ side caps into capped surfaces.

We have now raised the height by one on the $-$ side. We started with $(+, -)$ heights being $(1, 0)$ and now we have $(1, 1)$. We now raise height on the $+$ side. To achieve this we just repeat the above process, with $+$ and $-$ reversed. That is, we form a dual capped surface for the $+$ side, F_+ . Use the ∞ -lemma to create sufficiently many disjoint parallel copies F'_+ of F_+ , and use these to remove intersections between $+$ and $-$ side caps. Then tube intersections of $+$ side caps into parallel copies of F_+ . This creates a grope with height $(2, 1)$.

Now repeat as many times as desired, alternating $+$ and $-$ sides. From $(2, 1)$ we go to $(2, 3)$, then $(5, 3)$, then $(8, 5)$. We really see the Fibonacci numbers occurring with this method. Once we get both sides of height at least 2, then we can apply the argument in Lemma 8.9, which raises height a little slower but creates fewer intersections, which then have to be cleaned up, on the way. \square

8.4 Where Casson left us

Lecture 9, 0:00 to 12:05

We start where Casson left us and do it with the new capped surface technology. We are going to recall the original setup, and move on to reimbedding for towers, which is the main tool for the construction of the design.

We will see how this leads to rather easy shrinking arguments. Then in the next chapter we will go back and revisit Edwards' original shrink.

Recall that the original motivation for doing all this came from the surgery and h -cobordism programs for 4-manifolds, which we continue to take simply connected for simplicity. Both programs led to the same kind of problem. We ended up in some (smooth, simply connected) 4-manifold M that contained possibly immersed spheres $a, b \subset M$ which were homologically dual and we somehow had to make them geometrically dual using the Whitney trick. Casson's early preparation made the complement $M \setminus (a \cup b)$ simply connected as well, so that we could put in Whitney disks W into the complement, spanning all Whitney circles for cancelling intersection points. With a little bit of work we also made the further complement of $a \cup b \cup W$ simply connected by finding transverse spheres to W so that we could iterate this process to build Casson towers and eventually Casson handles.

This was the pre-grope technology. Let us see how the same situation plays out with gropes. With $a, b, W \subset M$ as above, notice that we have the Clifford torus for a transverse crossing of a and b sitting around; this is an algebraic dual to the disk W and its meridian and longitude are linking a and b respectively. But we already fixed up that a and b have transverse spheres which means that this dual object to W actually is a capped surface. So what can we do with this? We can immediately start building a grope version of W by piping the double points of W into parallel copies of the base surface which turns W into an embedded surface with caps. We can actually keep going and pass to a 2-stage capped grope. Right now we have a surface with caps, and the caps may intersect one another. All the intersection points of the caps can be pushed down to the base and then piped with parallel copies of the dual torus, so that they become surfaces. Then we can iterate that and produce a grope of arbitrary height with one layer of caps which are all disjoint from the body but the caps are crossing themselves. So this is really rapid progress compared to Casson towers.

Remark 8.14. Notice that we managed to use a shortcut for the grope height raising, using the dual classes. The previous grope height raising was a little more laborious, since in that situation we did not have the dual sitting around, so we had to use the left and right branches to create each other's duals.

Remark 8.15. The key open question in the topological category is whether capped gropes G_n^c of some height n with one layer of caps are already enough to finish the game, that is, whether a neighbourhood of this structure already contains an embedded flat disk. As we will see, two layers of caps do the trick. The A-B-slice problem or the conjecture that surgery and h -cobordism (or rather s -cobordism) work for free groups are all related to this central question.

8.5 Reimbedding in the grope world: tower height raising with control

Lecture 9, 12:05 - 32:20.

Now we are going to use a second layer of caps to get a suitable analogue for the reimbedding theorem for Casson towers (Theorem 7.8). First we have to say what the analogue of Casson towers should be.

Definition 8.16. A *CEQFAS*¹¹-tower with $k \in \mathbb{N}$ stories of heights $N \in \mathbb{N}^k$, is a 4-manifold pair (T_N, ∂_-) obtained in the following inductive way. For $k = 1$ let $G_{N_1}^c$ be a model capped grope of height N_1 . We introduce intersections between caps by performing finger moves. These produce Whitney disks which pair together new intersections. Apart from them there are also *accessory disks*. If D is one of the caps on which the finger move was performed. Then the accessory disk is a disk whose boundary lies on D and passes through one of the self-intersection points, see figure 84.

Grope $G_{N_1}^c$ with Whitney disks and accessory disks constitutes a capped tower. Suppose we defined $(k - 1)$ -story capped tower. Then a k -story capped tower is constructed from the $(k - 1)$ -story tower by replacing the top layer caps by one-story capped towers.

Remark 8.17. At this moment we would like to emphasise the difference between one-story capped towers and capped gropes. The difference lies in the additional layer of caps. Every one-story capped tower consists of a capped grope union tower caps.

In notation we could write such a tower as a “composition”

$$T_N = c \circ c \circ G_{N_k} \circ \cdots \circ c \circ G_{N_1}$$

where each G_{N_i} is a grope of height N_i and c is a layer of caps.

Definition 8.18. Using the notation as above a *body* of the tower T_N consists of capped gropes

$$\text{body} = c \circ G_{N_k} \circ \cdots \circ c \circ G_{N_1}.$$

The the remaining caps are called *tower caps*.

Definition 8.19. A *proper immersion* of a capped tower T_N into 4-manifold M is a map which is an embedding of the body of T_N , immersion on a caps such that caps do not intersect the image of the body.

Theorem 8.20 (Reimbedding for towers). *Let G_3^c be a properly immersed capped grope in a 4-manifold M which is π_1 -null, that is all double point loops of caps are null-homotopic in M . Suppose also that G_3^c possess a transverse sphere. Then for any $N \in \mathbb{N}^k$, $k \in \mathbb{N}$, there is a CEQFAS-tower $T_N \subset G_3^c$ such that $\partial_- T_N = \partial_- G_3^c$.*

Moreover if $N \in \mathbb{N}^k$ and $\epsilon_2 \geq \dots \geq \epsilon_k > 0$ are real numbers, then T_N can be constructed in such a way that for each $1 \leq i \leq k$ there exists a collection of disjoint balls of radius less than ϵ_i such that each component of the n_i -th story is contained (apart from perhaps a collar neighbourhood of its attaching circle) in one ball from this collection.

Before proceeding with the proof we will stop for a little while to discuss transverse gropes in the setting of towers. Suppose we have a capped grope G^c which is a part of a capped tower T^c . If $(G^c)^t$ is a transverse grope for G^c , then its caps may intersect the caps of G^c . Since the caps of G^c belong to the body of the tower this makes the transverse grope useless for eliminating intersections with G^c . In our situation it is desirable to have transverse grope $(G^c)^t$ for the tower stage with the caps intersecting tower caps. Next lemma solves this problem.

¹¹CEQFAS=Casson-Edwards-Quinn-Freedman-Ancel-Starbird

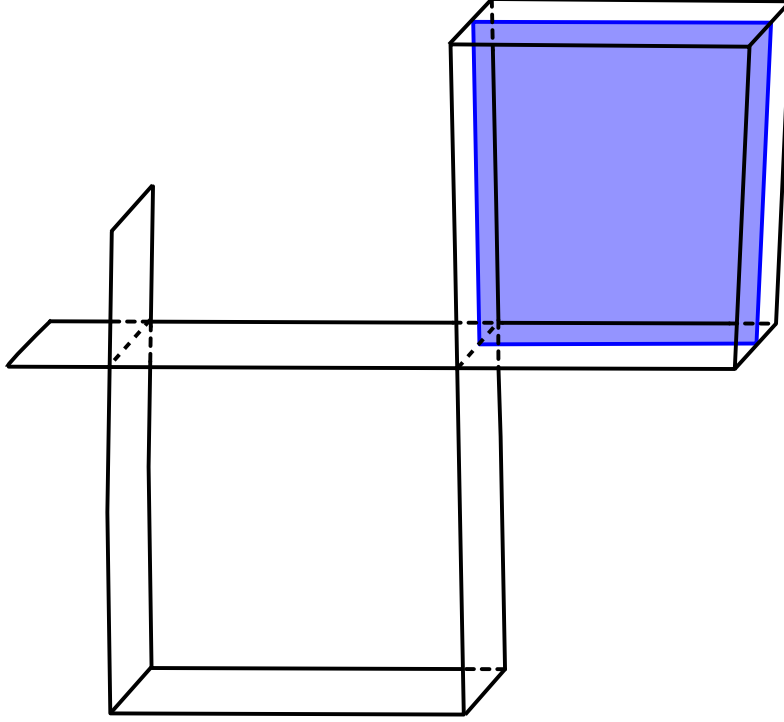


Figure 84: An accessory disk.

Lemma 8.21. *Suppose that T_n^c is an n -story capped tower. Let G_m^c denote the first story of T_n^c and assume that $m \geq 2$. Divide the second stage surfaces of G_m^c into two sets E_+ and E_- exactly as in Lemma 8.9. Then E_+ and E_- possess transverse capped gropes whose caps intersect only tower caps.*

Proof. Exactly as in the proof of Lemma 8.9 construct a transverse grope E_+^t for the $+$ side. This grope was constructed from the two copies of the $-$ side. This implies that the caps of E_+^t are parallel to the caps of the first story grope. The idea now is to eliminate intersections between caps of G_m^c and caps of E_+^t using disk-like gropes provided by the part of the second story by performing a grope version of the Whitney trick.

By the definition of a tower, the self-intersections of the caps of G_m^c are paired by Whitney disks. Each such pair gives two pairs of intersections between the caps of G_m^c and the caps of E_+^t . Parallel copies of the second story grope provide disk-like gropes which pair these intersection points like a Whitney disk. We will call them *Whitney gropes*. For each pair of intersection points paired in such a way, cut out a neighbourhood of the Whitney circle and then glue in two parallel copies of the Whitney grope and an annulus which is parallel to a neighbourhood of the arc on the cap of G_m^c . This operation yields a new capped grope which is transverse to E_+ and whose caps intersect second story caps. Repeating this procedure an appropriate number of times gives a transverse capped grope for E_+ whose caps intersects the caps of the top story and which lies in the neighbourhood of the body of T_n^c . This new grope is also π_1 -null.

The last step is the application of the ordinary Whitney move to exchange the intersections with the top story caps with the intersections with the tower caps. \square

The proof of Theorem 9.8 is divided into two steps. The first step consists of constructing a capped tower from a given capped grope. This is done in the following lemma.

Lemma 8.22. *Suppose that $G_m^c \rightarrow M$ is a proper immersion of a disk-like capped grope of height at least three which is π_1 -null and possesses a transverse sphere. Then the*

embedding of the body can be extended to a proper immersion of a one-story capped tower with arbitrarily many grope stages.

Proof. Suppose we would like to obtain a capped tower with n grope stages. The first thing to do is to perform an ordinary grope height raising to get a capped grope G_{n+1}^c of height $n + 1$. This transformation takes place in the neighbourhood of the original grope which guarantees the π_1 -nullity of the new grope.

The second step deals with the intersection of caps. First enumerate the components of the top stage of G_{n+1}^c . Contract the first component and push all other components off the contracted one. Repeat this process for other components to get a new capped grope G_n^c of height n and disjoint caps. Notice now that since all self-intersections of the caps were introduced when pushing caps off the contractions of the other caps, they are paired by Whitney disks.

The third step is concerned with producing the tower caps. From the previous step, self-intersection points come equipped with the immersed Whitney disks. Since everything is placed in a neighbourhood of the original grope, and the original grope was π_1 null, one can find immersed accessory disks for the existing Whitney disks. The Whitney disks and accessory disks may intersect the capped grope G_n^c . To eliminate these intersections push them down to the lowest grope stage and pipe with the transverse sphere. These disks are now the required tower caps. \square

The next lemma describes the procedure which increases the number of stories of the capped tower.

Lemma 8.23 (Tower height raising). *Let T^c be a properly immersed n -story capped tower with grope height at least 4 in the first story. Then the tower caps can be changed so that the embedding of the body extends to a proper immersion of a $(n + 1)$ -story capped tower with arbitrary grope height in the top story.*

Proof. The first step in the proof is to construct transverse gropes F_+^t and F_-^t for T^c as in Lemma 8.21. Contract these and push off the tower caps to obtain transverse spheres for the second stage of the first story of T^c .

Next, delete a neighbourhood of the first stage surfaces from the ambient manifold. This yields a new ambient 4-manifold M' and a union of a disk-like capped towers which will be denoted by $(T^c)'$. Each component has a transverse sphere for the bottom stage. These will be used later to construct caps for the new tower.

Divide the surfaces of the second stage of the first story of $(T^c)'$ (which was the third surface stage of T^c) into $+$ and $-$ sides as in Theorem 8.9. Construct transverse gropes E_+^t for the $+$ side. Caps of these gropes may intersect the tower caps of the $-$ side. Contract the top stage of E_+^t and push off the $-$ side caps. This new transverse grope $(G^c)^t$ is π_1 -null because it lies in the neighbourhood of the body of $(T^c)'$.

In order to get the “arbitrary” grope height claimed in the statement of the lemma, now raise the height of $(G^c)^t$ using Theorem 8.9 to the desired arbitrary height plus one.

By the ∞ -lemma 8.6 it is possible to construct any finite number of disjoint parallels of $(G^c)^t$. These copies will be used to turn tower caps of the $+$ side into gropes of arbitrary height. Intersections between tower caps of the $+$ side can be pushed down to the second stage of the first story. To each intersection point with the second stage add a disjoint parallel of $(G^c)^t$. This transforms $+$ tower caps into capped gropes which are disjoint from the body.

At this point, we have changed the tower caps into a capped grope on the $+$ side. Thus on the $+$ side, we now have a second story. There are still some problems. The new capped gropes do not have tower caps. We need to arrange the caps of this new grope

to be π_1 -null, and to have only self intersections, and then we will be able to deduce the existence of tower caps.

Above procedure yields a new tower T'' which is not a capped tower. T'' consists of the original $+$ and $-$ part of the tower $(T^c)'$ with modified caps on the $-$ side and with the addition of the newly constructed capped gropes on the $+$ side. These new capped gropes are not equipped with a second layer of caps, which are required for a capped tower. The new tower is π_1 -null. For the $-$ side this follows from the π_1 -nullity of $(T^c)'$. For the $+$ side π_1 -nullity follows from the fact that the intersection of the caps come from the intersections of the caps of the disjoint parallels of $(G^c)^t$ and this transverse grope is π_1 -null by the construction.

To eliminate intersections between different caps from the top story raise grope height by one and perform a contraction and push off on each surface of the top stage of the new top story one at a time. The π_1 -nullity is preserved since everything takes place in the neighbourhood of the original tower.

The last step is the construction of the tower caps for the $+$ stages. The method is exactly the same as in the final paragraph of the proof of Lemma 8.22. Note in particular that the hypothesis of Lemma 8.22 require transverse spheres. For this we need to use the transverse sphere F_+^t constructed in the first paragraph: these transverse spheres are used to ensure that the tower caps are disjoint from the body.

The described procedure raises the tower height of the $+$ side of $(T^c)'$ by one. Repeating the same procedure raises the $-$ side. After this, reinstate the bottom surface stages of the first story, and their regular neighbourhoods. \square

In the proof of lemma 8.23 the construction of a new story was done such that this new capped grope was winding around the body of the tower. Later we will need certain convergence condition for towers, which force us to separate higher stories from the lower ones. This issue is dealt with in the next lemma. This can work since tower height raising can be performed on just the top n stories of a tower.

Lemma 8.24. *Let T_1^c be a proper immersion of a one-story capped tower with grope height at least 4, and $\epsilon > 0$. Then the embedding of the grope body extends to a proper immersion of a two-story capped tower T_2^c such that the components of the second story (minus neighbourhoods of their boundary circles) lie in disjoint balls of radius less than ϵ .*

Proof. First step of the proof is focused on extending the proper immersion of the tower T_1^c to a proper immersion of a two-story tower satisfying two additional properties: first the second story grope has to possess arbitrary many grope stages and secondly distinct components of the second story have disjoint images and are π_1 -null in the complement of the first story.

This is achieved in the following way. Use the tower height raising 8.23 to extend the given immersion to an immersion of a 3-story capped tower T_3^c . Discard the tower caps and contract the top story grope. This yields disks which are disjoint tower caps for the resulting two-story tower $(T_2^c)'$. The disjointness is a consequence of the disjointness of the components of the top story of T_3^c . The original tower caps of T_3^c assure that the components of the top story of $(T_2^c)'$ are π_1 -null in the complement of the bottom story.

Observe that a regular neighbourhood of any tower is also a regular neighbourhood of a 1-complex. To see this note that the lower surface stage of a tower deformation retracts onto the attaching curves together with some arcs joining curves from the same components of the surface. Iterating this construction one obtains a deformation retraction of a capped grope onto its caps union a tree. In a similar way caps can be retracted to a graph consisting of loops passing by only one self-intersection points union some arcs.

Performing this construction story by story and then retracting caps of the tower yields the desired graph.

Consider parallel copies of the attaching circles of the components of the second story. Observe now that every capped tower with one boundary component is in fact a mapping cylinder of certain map from these boundary circles to a 1-complex which is a deformation retract of the tower.

Let K denote a one complex which is a spine of the second story of $(T_2^c)'$ union tower caps, let L denote parallel copies of the boundary circles of the second story and let M_L denote the aforementioned mapping cylinder of certain mapping from L to K .

In order to move the top story of $(T_2^c)'$ into a union of small balls it is sufficient to move K into some small ball since the story is located in the regular neighbourhood of K . Moving K into a union of balls is possible since K is null-homotopic 1-complex and in dimension 4 homotopy implies isotopy for 1-complexes.

In more detail this construction works as follows. For every component of K choose a ball of radius less than ϵ such that it is disjoint from $(T_2^c)'$ and from other balls. The π_1 -nullity condition implies that K is null-homotopic in the complement of the bottom story of $(T_2^c)'$. For each component of K choose a homotopy which moves it into the prescribed ball.

This yields a map $K \times I \rightarrow M$. This map can be approximated rel. $K \times 0$ by an immersion in general position in such a way that $K \times 1$ stays inside the balls. The image of this immersion intersects the grope caps and the mapping cylinder M_L in a discrete set of points. Use finger moves to push all surfaces intersecting the image off the $K \times 1$ end. This changes the tower $(T_2^c)'$ by an isotopy and yields an embedding of $K \times I$ disjoint from the bottom story of the tower, M_L . Furthermore $K \times 1$ end remains inside the balls.

Use the constructed embedding of $K \times I$ to define an isotopy of the regular neighbourhood of K . Extend this isotopy to the isotopy of $(T_2^c)'$ by defining it to be constant on the bottom story. This isotopy moves the top story of the tower inside the chosen balls. This finishes the proof of the lemma. \square

Finally we indicate how to put everything together to prove Theorem 8.20.

Proof of Theorem 8.20. First use Lemma 8.22 to extend the embedding of the body of the grope to a proper immersion of a one-story tower with as many surface stages as desired. Next use Lemma 8.23 to raise height of the tower (i.e. the number of stories) by 1. The next step is to utilise Lemma 8.24 to assure that components of the second story are placed in small disjoint balls. From now on the procedure can be repeated as many times as necessary, just applied to the top story of the tower each time. \square

8.6 The disc embedding theorem

By Theorem 8.20, and the remark that it is actually sufficient to begin with a height (1.5) capped grope if we can find such a grope with a transverse sphere, then we can find a convergent tower. From this tower, we will construct the design and apply the shrinking arguments to obtain a locally flat embedded disc.

There remains the problem of getting from the hypotheses of the disc embedding theorem to the starting point of $G_{1.5}^c$ -with-transverse-sphere.

Theorem 8.25 (Disc Embedding Theorem hypothesis to grope starting point). *Suppose that A is an immersed disc in a 4-manifold M with an algebraically transverse sphere B ($\lambda(A, B) = \pm 1$) such that $\mu(B) = 0$. Then there is a properly immersed height 1.5 capped grope in M with a transverse sphere for the body and the same boundary as A .*

Proof. We divide the proof up into steps.

1. Make $A \cdot B = \pm 1$ geometrically. To achieve this, pair up the other intersections of $A \cdot B$ with immersed Whitney discs - we know this can be done by the algebraic condition $\lambda(A, B) = \pm 1$. These Whitney discs W_1 may have intersections $W_1 \cdot A$ and $W_1 \cdot B$. Push the intersections with A off onto A and those with B off onto B , using finger moves. This creates new $A \cdot A$ and $B \cdot B$ self intersections, which occur paired up with Whitney discs, since they came from finger moves. It makes the W_1 discs disjoint from A and B .

Then perform the immersed Whitney move, pushing A across W_1 . This removes all but one of the $A \cdot B$ intersections. It does not change $\mu(B)$. It creates new $A \cdot A$ intersections, from the self intersections of W_1 . These new self intersections of A are also paired up with Whitney discs. To see this we take a disc D and compute the self intersection number of two parallel copies of D , one with reversed orientation:

$$\begin{aligned}\mu(D - D') &= \mu(D) + \mu(-D') + \lambda(D, -D') = \mu(D) + \mu(D') - \lambda(D, D') \\ &= \mu(D) + \mu(D') - \mu(D) - \overline{\mu(D')} = \mu(D') - \overline{\mu(D')},\end{aligned}$$

since:

$$\lambda(D, D') = \mu(D) + \overline{\mu(D')}.$$

Then recall that the indeterminacy of the self-intersection form is such that $\mu(D') - \overline{\mu(D')} = 0$. Thus $\mu(D - D') = 0$ and so we see that the self intersections of A after the immersed Whitney move can be paired up algebraically. Thus we can remove pairs of intersection points from $A \cdot B$ without changing $\mu(A)$ or $\mu(B)$.

2. Make $\mu(A) = 0$. To do this, recall that $\mu(B) = 0$. Tube $A \cdot A$ intersections into parallel copies of B . This creates new $A \cdot A$ and $A \cdot B$ intersections but they are now all paired up, by parallel copies of the Whitney discs that pair up self-intersections of B .
3. Repeat Step (1) for new $A \cdot B$ intersections which arose from (2). This makes $A \cdot B$ into one point again, but now we also have that $\mu(A) = \mu(B) = 0$.
4. Make A into a capped surface. Since $\mu(A) = 0$, intersection points can be paired up and then one of the sheets can be tubed along. The Whitney disk becomes one cap and a disc bounding a meridian of the tube becomes the other. See [FQ90, Page 32] and Figure 85. We have caps $\cdot A$, caps \cdot caps and caps $\cdot B$ intersections.
5. Tube caps $\cdot A$ intersections into B . This creates new caps \cdot caps and caps $\cdot B$ intersections.
6. For each caps $\cdot B$ intersection, push B down so it intersects A twice, and then tube B back into a parallel copy of itself. The advantage of this is that the new caps $\cdot B$ intersections this creates are paired up, since $\mu(B) = 0$. We also get new $B \cdot B$ intersections, which are again paired up.
7. We do similar moves to caps \cdot caps intersections, in order to also arrange that they are paired up. For each such intersection, push down one cap, tube into parallel copies of B . There will be new caps \cdot caps and caps $\cdot B$ intersections but all caps \cdot caps intersections are paired up as are all caps $\cdot B$ intersections.
8. Now we make caps $\cdot B = \emptyset$ using the immersed Whitey move: pair up caps $\cdot B$ with Whitney discs W_2 . The Whitney discs may have intersections, which we need to fix with finger moves. For each $W_2 \cdot$ caps intersections push off the caps onto the caps, create new caps \cdot caps intersections but obtain $W_2 \cdot$ caps $= \emptyset$. For each $W_2 \cdot A$

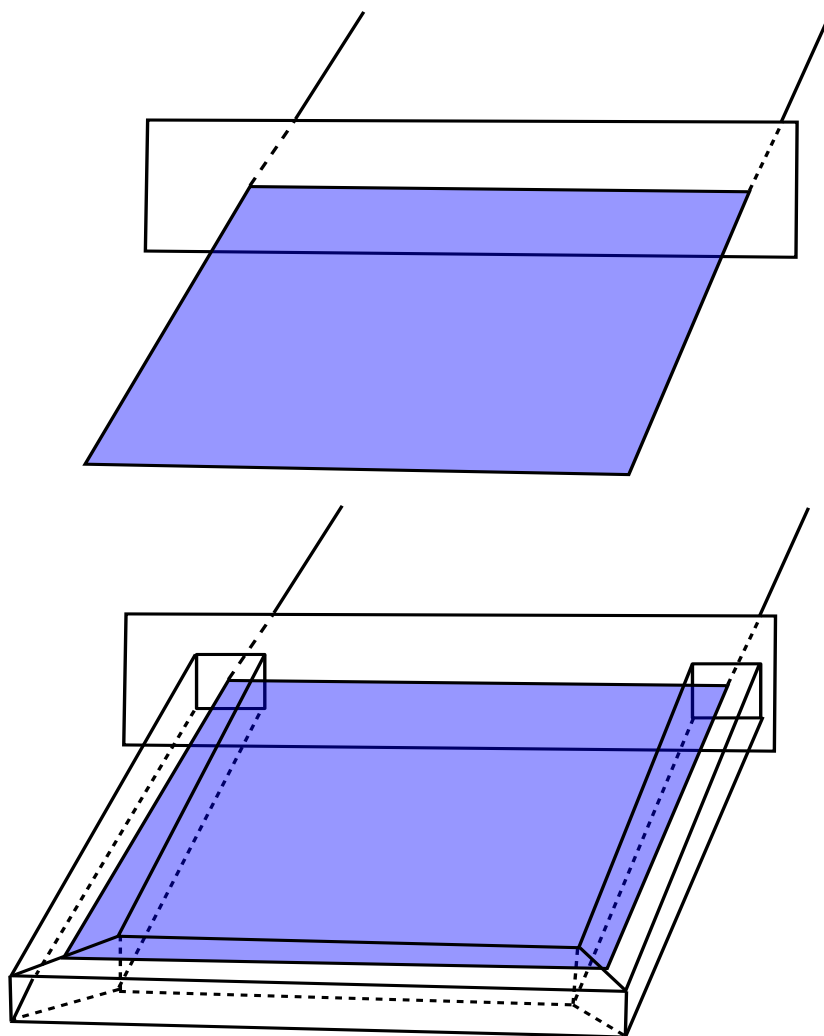


Figure 85: Converting an immersed disc to a capped surface.

intersection, tube W_2 to B . We get new $W_2 \cdot B$ intersections and new $W_2 \cdot W_2$ intersections (there were likely already such intersections). The latter we do not mind about, the former we sort out next. For each $W_2 \cdot B$ intersections push off B onto B , create new $B \cdot B$ intersections but obtain $W_2 \cdot B = \emptyset$. Now we can do the immersed Whitney move to arrange that $\text{caps} \cdot B = \emptyset$, at the expense of more $\text{caps} \cdot \text{caps}$ and $B \cdot B$ intersections.

9. Taking stock, we have a properly immersed capped surface A with a disjoint transverse sphere B , such that $\mu(\text{caps}) = \mu(B) = 0$.
10. Take the caps on one half of a symplectic basis for the surface and convert them into capped surfaces A_2 as in Step (4), with the caps of the second stage surfaces denoted caps_2 . The secondary caps, caps_2 , come from Whitney discs pairing up $\text{caps} \cdot \text{caps}$ intersections and discs bounding the added tubes. There are $\text{caps}_2 \cdot \text{caps}_2$, $\text{caps}_2 \cdot A_2$, $\text{caps}_2 \cdot A$, $\text{caps}_2 \cdot \text{caps}$ and $\text{caps}_2 \cdot B$ intersections.
11. To remove $\text{caps}_2 \cdot A_2$, $\text{caps}_2 \cdot A$ and $\text{caps}_2 \cdot \text{caps}$ intersections, push down caps_2 into A if necessary and then tube into parallel copies of B . This creates new $\text{caps}_2 \cdot \text{caps}_2$ intersections, and new $\text{caps}_2 \cdot B$ intersections, all of which are paired up.
12. Tube $\text{caps}_2 \cdot \text{caps}_2$ intersections into B too. This creates further new $\text{caps}_2 \cdot \text{caps}_2$ and $\text{caps}_2 \cdot B$ intersections, but now $\mu(\text{caps}_2) = 0$.
13. We need to make $\text{caps}_2 \cdot B = \emptyset$. Then the proof is finished, for we will have a properly immersed height (1.5) capped grope with a transverse sphere. To achieve this last step, just repeat Step 8. That is, perform the immersed Whitney move, using discs W_3 which pair up $\text{caps}_2 \cdot B$ intersections. There may be intersections with W_3 . Push down W_3 and tube it into B for each intersection of the form $W_3 \cdot A$, $W_3 \cdot \dots A_2$ or $W_3 \cdot \text{caps}$. This will create, as in Step 8, new $W_3 \cdot W_3$, $W_3 \cdot \text{caps}_2$ and $W_3 \cdot B$ intersections. For each $W_3 \cdot \text{caps}_2$ intersection push caps_2 off onto caps_2 . For each $W_3 \cdot B$ intersection push B off onto B . Then perform the immersed Whitney move. We obtain new $\text{caps}_2 \cdot \text{caps}_2$ intersections, maintaining $\mu(\text{caps}_2) = 0$ by the computation in Step (1). Note that we have also arranged that caps_2 intersections are paired up, and that $\text{caps}_2 \cdot \text{caps} = \emptyset$. Furthermore all caps are disjoint from B . This completes the proof.

□

9 Geometric control and the Design

We start where Casson left us and do it with the new capped surface technology. We'll see how this leads to rather easy shrinking arguments. In the next chapter we will go back and revisit Edwards' original shrink.

9.1 Where Casson left us off

Recall that the original motivation for doing all this came from the surgery and h -cobordism programs for 4-manifolds which we continue to take simply connected for simplicity. Both programs led to the same kind of problem. We ended up in some (smooth, simply connected) 4-manifold M that contained possibly immersed spheres $a, b \subset M$ which were homologically dual and we somehow had to make them geometrically dual using the Whitney trick. Casson's early preparation made the complement $M \setminus (a \cup b)$ simply connected as well so that we could put in Whitney disks W into the complement spanning all Whitney circles for canceling intersection points. With a little bit of work we also

made the further complement of $a \cup b \cup W$ simply connected by finding transverse spheres to W so that we could iterate this process to build Casson towers and eventually Casson handles.

This was the pre-grope technology. Let's see how the same situation plays out with gropes. With $a, b, W \subset M$ as above, notice that we have the Clifford torus for a transverse crossing of a and b sitting around; this is an algebraic dual to the disk W and its meridian and longitude are linking a and b , respectively. But we already fixed up that a and b have transverse spheres which means that this dual object to W actually is a capped surface. So what can we do with this? We can immediately start building a grope version of W by piping the double points of W into parallel copies of the base surface which turns W into an embedded surface with caps. And we can actually keep going and pass to a 2-stage capped grope. Indeed, all the intersection points of the caps can be pushed down to the base and then piped with further copies of the original dual. Then we can iterate that and produce a grope of arbitrary height with one layer of caps which are all disjoint from the body but the caps are crossing themselves. So this is really rapid progress compared to Casson towers.

Remark 9.1. Notice that we kind of shortcut the grope height raising which was a little more laborious. But in that situation we did not have the dual sitting around so we had to use the left and right branches to create each others duals.

Remark 9.2. The key open question in the topological category is actually whether capped gropes G_n^c of some height n with one layer of caps are already enough to finish the game, that is, whether this structure already contains an embedded flat disk. As we'll see, two layers of caps do the trick. Things like the A-B-slice problem or the conjecture that surgery and h -cobordism (or rather s -cobordism) work for free groups are all related to this central question.

9.2 Towers, CEQFAS handles and reimbedding

Our goal is to carry out the strategy outlined in Section 7.4 using the capped grope technology. First we have to say what the analogue of Casson towers and handles should be. The following definition is roughly the same as the one given in [FQ90, Chapter 3.2].

Definition 9.3. A *tower* with $k \in \mathbb{N} \cup \{\infty\}$ *stories of heights* $N \in \mathbb{N}^k$, is a 4-manifold pair $(\mathcal{T}_N, \partial_-)$ obtained from $(B^2 \times B^2, \partial_-)$ by adding N_1 layers of surface stages, followed by one layer of caps with only self-intersections, then another N_2 surface stages, another layer of disks and so on.

Remark 9.4. Recall that there's a dual way to think of towers. Instead of attaching things to $B^2 \times B^2$ we can also remove things from it. One way to see this is Kirby calculus where a tower is appears as an unknot representing $\partial_- \mathcal{T}$ and the core curve of the complementary solid torus in S^3 is Bing and Whitehead doubled (possibly infinitely many times) according to the structure of the tower. Recall further that each component of the iterated Bing-Whitehead link has a dot on it which means that we should take slices in the interior of $B^2 \times B^2$ and remove neighborhoods; of course, the slices and their neighborhoods are supposed to be disjoint.

It's probably not surprising that we're mostly interested in infinite towers. Note that such a gadget is non-compact smooth 4-manifold; in fact, by Remark 9.4 we can think of them as open subsets of $B^2 \times B^2$ if we like. However, the more abstract approach taken in the definition has its advantages. For example, the following compactification construction is more natural without a fixed embedding into $B^2 \times B^2$. We take a quick break to review some more point-set topology.

Remark 9.5 (Ends and endpoint compactifications). Let X be a topological space and let's assume that it's metrizable, just to be safe. Intuitively, an “end” of X should be something connected that lives “far out” in X (that is, in the complement of a large compact subset). For example, compact spaces should have no ends, the real line should have two, and \mathbb{R}^n for $n \geq 2$ should have only one end. A way to make this precise is to take an ascending sequence $K_1 \subset K_2 \subset \cdots \subset X$ of compact subsets whose interiors cover X . An *end* of X is then defined to be a descending sequence $U_1 \supset U_2 \supset \cdots$ where each U_i is a connected component of $X \setminus K_i$. It is an easy exercise to check that this notion of ends is independent of the choice of the K_i . This leads to the notion of *endpoint compactification* \widehat{X} of X whose underlying set is the disjoint union of X and the collection of its ends, and a topology is defined in a similar way as for the one-point compactification.

Here's a useful exercise to get acquainted with the notion of ends.

Exercise 9.6. Let \mathcal{T} be an infinite tower. Show that there is a bijection between the ends of \mathcal{T} and those of $\partial\mathcal{T}$.

Taking the endpoint compactification of an infinite tower \mathcal{T} we obtain a compact space $\widehat{\mathcal{T}}$ which we call a *compactified tower*. Although we gain compactness by passing to $\widehat{\mathcal{T}}$, there's a price to pay. For one thing, there's no obvious reason that $\widehat{\mathcal{T}}$ should be a manifold (In fact, most of the work that lies ahead would become superfluous if we knew that certain compactified towers are manifolds.) In addition, the embedding of \mathcal{T} into $B^2 \times B^2$ doesn't extend to an embedding of $\widehat{\mathcal{T}}$.

Anyway, let's take a closer look at the anatomy of infinite towers and their endpoint compactifications. As in the case of Casson handles, the boundary of an infinite tower \mathcal{T} has two parts

$$\partial\mathcal{T} = \partial_-\mathcal{T} \cup_{T^2} \partial_+\mathcal{T}$$

where the attaching region $\partial_-\mathcal{T}$ is naturally identified with $S^1 \times B^2$ in $B^2 \times B^2$. Of course, the compactification $\widehat{\mathcal{T}}$ doesn't have a boundary in the manifold sense, but a good replacement is suggested by Exercise 9.6 above. We simply define the *frontier* $\partial_+\widehat{\mathcal{T}}$ as the endpoint compactification of $\partial_+\mathcal{T}$.

Elaborating on Remark 9.4 we can reconcile

Proposition 9.7. *Let \mathcal{T} be an infinite tower with endpoint compactification $\widehat{\mathcal{T}}$.*

(a) *The boundary part $\partial_+\mathcal{T}$ embeds into $B^2 \times S^1$ as an open subset such that*

- *$\partial_+\mathcal{T} \cap \partial_-\mathcal{T}$ goes to $S^1 \times S^1$ via the standard identification,*
- *the complement is a mixed Bing-Whitehead decomposition $\partial\mathcal{D}$ of $B^2 \times S^1$, and*
- *the frontier $\partial_+\widehat{\mathcal{T}}$ is homeomorphic to $(B^2 \times S^1)/\partial\mathcal{D}$.*

(b) *Any embedding of $\partial_+\mathcal{T}$ as above extends to an embedding of \mathcal{T} into $B^2 \times B^2$ such that*

- *$\partial_-\mathcal{T}$ goes to $S^1 \times B^2$ via the standard identification,*
- *the complement is a decomposition \mathcal{D} of $B^2 \times B^2$, and*
- *the compactification $\widehat{\mathcal{T}}$ is homeomorphic to $(B^2 \times B^2)/\mathcal{D}$.*

Moreover, the embedding of \mathcal{T} can be arranged so that \mathcal{D} is the cone on $\partial\mathcal{D}$, in the sense that each element of \mathcal{D} is homeomorphic to the cone of an element of $\partial\mathcal{D}$.

Proof. As explained in Remark 9.4 a Kirby diagram for a fixed finite stage of \mathcal{T} is a Hopf link where one component has been Bing or Whitehead doubled finitely many times, possibly with ramifications. We choose thickened slices of the doubles inductively.

Say we're in the n th stage and we've chosen slice disks for the Bing-Whitehead double and thickened them to disjoint copies of $B^2 \times B^2$. We denote the union of these thickened slices by H_n . Passing to stage $n + 1$ adds another layer of (ramified) doubling and we can assume that this process is confined to the interior H_n so that we can choose thickened slices for the new doubles whose union H_{n+1} is contained in the interior of H_n . By definition, the infinite intersection $\cap_n H_n$ is complementary to the image of the embedding of \mathcal{T} in $B^2 \times B^2$ and gives rise to a decomposition $\tilde{\mathcal{D}}$. Moreover, **since the endpoint compactification of the complement of a decomposition is homeomorphic to the decomposition space**, we immediately see that the compactified tower $\hat{\mathcal{T}}$ is homeomorphic to $(B^2 \times B^2)/\tilde{\mathcal{D}}$.

Add lemma in DST chapters!



Similarly, on the boundary we obtain a decomposition $\partial\mathcal{D}$ of $B^2 \times S^1$ whose quotient is the frontier $\partial_+\mathcal{T}$. Since $\partial\mathcal{D}$ is given by the infinite intersection $\cap_n(H_n \cap B^2 \times S^1)$ and the $H_n \cap B^2 \times S^1$ form a nested sequence of mixed Bing-Whitehead tori, we see that $\partial\mathcal{D}$ is, in fact, a mixed Bing-Whitehead decomposition.

□

Now there's some good news. If there are enough surface stages in an infinite tower \mathcal{T} , then the associated decomposition of $B^2 \times S^1$ shrinks and the frontier $\partial_+\hat{\mathcal{T}}$ is homeomorphic to a solid torus. In particular, the compactified tower $\hat{\mathcal{T}}$ has a manifold as its frontier. Note that this can never happen for a Casson handle! According to Ancel and Starbird's theorem (Theorem 4.12) we can quantify what it means to have “enough surface stages”. This is the case whenever the series $\sum_i \frac{N_i}{2^i}$ diverges where (N_1, N_2, \dots) is the sequence of heights of a given tower \mathcal{T} . In that case we say that \mathcal{T} has the *Ancel-Starbird property*.

At this point we have a reasonable understanding of infinite towers. So far, so good. Next we need a replication device in the spirit of Theorem 7.8 to get going with the construction of the Design. The following result is a first approximation.

Proposition 9.8 (Reimbedding for towers). *Let G_3^c be a capped grope immersed in a 4-manifold M such that all double point loops of the caps are trivial in $\pi_1(M \setminus \text{base})$. Then for any $N \in \mathbb{N}^k$, $k \in \mathbb{N} \cup \{\infty\}$, there is a tower $\mathcal{T}_N \subset G_3^c$ with $\partial_-\mathcal{T}_N = \partial_-G_3^c$.*

In applications the base usually has a dual to it so the additional fundamental group concern is completely unwarranted. For example, we can prove surgery and h -cobordism without this hypothesis.



Proof. See Chapter 3.5 in [FQ90], **more details to follow...**

□

From this proposition we take away two things. One, it would be good to have infinite towers whose stories are all high enough so that we can reimbed in whatever story we like. Two, it shouldn't be too hard to build such towers using Proposition 9.8.

So instead of working with arbitrary towers we restrict our attention to a nice class of towers which is tailor-made for the Design construction.

Definition 9.9. An *open CEQFAS handle*¹², generically denoted by $C \cdot H$, is an infinite tower such that

¹²CEQFAS=Casson-Edwards-Quinn-Freedman-Ancel-Starbird

- (i) all stories of $C \cdot H$ have height at least three.
- (ii) $C \cdot H$ has the Ancel-Starbird property and

A (compact) CEQFAS handle $\widehat{C \cdot H}$ is the endpoint compactification of an open CEQFAS handle.

As mentioned above, the first condition ensures that the frontier of $\widehat{C \cdot H}$ is homeomorphic to S^3 , which certainly doesn't hurt. The second one gives a good amount of flexibility for reimbedding, which we'll use in the following upgraded version of Proposition 9.8.

Theorem 9.10 (Mitosis). *Let G_3^c be a capped grope as in Proposition 9.8. Then we can find a CEQFAS handle $\widehat{C \cdot H}$ inside G_3^c with $\partial_- \widehat{C \cdot H} = \partial_- G_3^c$. Furthermore, we can choose $\widehat{C \cdot H}$ such that everything above its bottom story is contained in an arbitrarily small ball inside G_3^c .*

Proof. Coming soon...



□

The rest of this chapter is devoted to the proof of the following theorem.

Main Theorem. *Any CEQFAS handle $(\widehat{C \cdot H}, \partial_- \widehat{C \cdot H})$ is homeomorphic to the standard 2-handle $(B^2 \times B^2, \partial_-)$ as a pair.*

9.3 The Design in CEQFAS handles

Remember that the Design is supposed to be a common closed subset of $\widehat{C \cdot H}$ and $B^2 \times B^2$. So how do we build the Design in the grope context? The key idea is to do an infinite construction using the controlled reimbedding theorem that produces an infinite collection of nested CEQFAS handles inside a given one. In fact, we'll actually get a Cantor set worth of CEQFAS handles as a result of inherent dyadic branching in the construction. We'll go through the construction twice, we first give a blueprint in order to convey the main ideas and then comes the fine print with all the technicalities.

But before going into the construction we set up some notation. By a *dyadic expansion* we mean a finite or infinite sequence $I = (i_1 i_2 \dots)$ of 0s and 2s, that is, an element of $\{0, 2\}^k$ where $k \in \mathbb{N} \cup \infty$. The number k is called the *length* of I which we also denote by $|I|$. For finite k and possibly infinite l we have the juxtaposition operation

$$\{0, 2\}^k \times \{0, 2\}^l \rightarrow \{0, 2\}^{k+l}, \quad (i_1 \dots i_k)(j_1 \dots j_l) \mapsto (i_1 \dots i_k j_1 \dots j_l)$$

and we use it to define a partial order of the set of all dyadic expansions by saying that J *contains* I , denoted by $I \subset J$, if $J = II'$ for some I' .

Sometimes it will be useful to make a notational distinction between finite and infinite expansions and, whenever appropriate, we do this by denoting the infinite ones by small letter $r = (r_1 r_2 \dots) \in \{0, 2\}^\infty$. Note that the set of infinite dyadic expansion can be identified with the middle-third construction of the Cantor set $C_3 \subset [0, 1]$ via the map

$$\{0, 2\}^\infty \longrightarrow C_3, \quad r = (r_1 r_2 \dots) \mapsto \sum_{j=1}^{\infty} \frac{r_j}{3^j}.$$

It is easy to see that the expansions that eventually contain only 0s or 2s correspond to the lower and upper endpoints of the outer-third intervals that survive to some finite stage of the construction, respectively. At times we want to consider finite dyadic expansions as infinite ones and we do so by adding a tail of zeros or, in other words, by juxtaposing with $(00 \dots)$.



Figure 86: The first three steps in the preliminary construction of the design.

9.3.1 The blueprint

Let $\widehat{C \cdot H}$ be a CEQFAS handle. By definition all stories of $\widehat{C \cdot H}$ have height at least 3 which means that we can apply the controlled reimbedding theorem (Theorem 9.10) in any story and we will exploit this to devise an infinite construction whose first three steps of the construction are schematically indicated in Figure 86.

In the *first step* of our construction we take $\widehat{C \cdot H}$ and locate another CEQFAS handle in its *first story*. For reasons that will become apparent momentarily we call the new handle $\widehat{C \cdot H}_{(2)}$ and also relabel the original handle $\widehat{C \cdot H}$ as $\widehat{C \cdot H}_{(0)}$, that is, we label the CEQFAS handles we produce in the first step by dyadic expansions of *length one*.

Now assume that we have done k steps and we have found a collection of 2^k CEQFAS handles $\widehat{C \cdot H}_{(i_1 \dots i_k)} \subset \widehat{C \cdot H}$, $(i_1 \dots i_k) \in \{0, 2\}^k$, labeled by the dyadic expansions of length k . Then the $(k+1)$ -*st step* is to take each $\widehat{C \cdot H}_{(i_1 \dots i_k)}$ and to locate new CEQFAS handles inside each capped grope of its $(k+1)$ -*st story*. These handles naturally attach to the first k stories of $\widehat{C \cdot H}_{(i_1 \dots i_k)}$ and we denote the union of these k first stories and the newly found handles by $\widehat{C \cdot H}_{(i_1 \dots i_{k+1})}$, which is a single CEQFAS handle. In addition, we let $\widehat{C \cdot H}_{(i_1 \dots i_k 0)} = \widehat{C \cdot H}_{(i_1 \dots i_k)}$ so that we end up with 2^{k+1} CEQFAS handles indexed by the dyadic expansions of length $k+1$.

As usual, we repeat this process indefinitely (with a certain amount of control, more about this later) and we end up with a collection of CEQFAS handles labeled by the infinite dyadic expansions $\{0, 2\}^\infty$ or, alternatively, the middle third Cantor set C_3 . So as promised we have found a Cantor set worth of handles inside $\widehat{C \cdot H}$ which itself corresponds to the expansion $(00 \dots)$ or $0 \in C_3$.

So apart from producing a lot of CEQFAS handles, what does this construction do for us? Let's be wasteful and forget about most parts of the handles, we only consider their frontiers and especially the union

$$D = \bigcup_{r \in C_3} \partial_+ \widehat{C \cdot H}_r$$

which is the first approximation of what's going to be the Design. According to Proposition 9.7, each frontier $\partial_+ \widehat{C \cdot H}_r$ is homeomorphic to $B^2 \times S^1$ which means that we can find an embedding

$$B^2 \times S^1 \times C_3 \xrightarrow{\cong} D \subset \widehat{C \cdot H}$$

such that $B^2 \times S^1 \times \{0\}$ goes to $\partial_+ \widehat{C \cdot H}$. On the other hand, we can also embed $B^2 \times S^1 \times C_3$ into the standard 2-handle via

$$B^2 \times S^1 \times C_3 \rightarrow B^2 \times B^2, \quad (z, e^{i\theta}, r) \mapsto (z, (1 - \frac{r}{3})e^{i\theta}),$$

so that we can consider D as a common closed subset of $\widehat{C \cdot H}$ and $B^2 \times B^2$.

Remark 9.11. Note that D looks very close to a collar for $\partial_+ \widehat{C \cdot H}$ in $\widehat{C \cdot H}$ as well as for $B^2 \times S^1$ in $B^2 \times B^2$. Further moral support that $\partial_+ \widehat{C \cdot H}$ might be collared – which it would have to be if $\widehat{C \cdot H}$ were a manifold – comes from our previous observation that the closure of the complement of C_3 in $[0, 1]$ gives a shrinkable decomposition (Example 3.7).

So what can we do with D ? Naively, we could try to shrink the decompositions of $\widehat{C \cdot H}$ and $B^2 \times B^2$ given by the closures of the components of the complement of D , but it turns out that in order to successfully shrink things we have to enlarge D a little bit. This enlarged version of D is the actual Design \mathbf{D} . Roughly, $\mathbf{D} \subset \widehat{C \cdot H}$ is going to be the union of neighborhoods $\nu \partial_+ \widehat{C \cdot H}_r$ over all $r \in C_3 \cap [0, \frac{1}{3}]$. However, in order to actually make this work we have to do the whole construction a little more carefully. So we start over again.

9.3.2 The fine print

We first have to take a small detour and discuss collars in towers. For that purpose consider an arbitrary infinite tower \mathcal{T} . Before passing to the endpoint compactification \mathcal{T} is a non-compact, smooth 4-manifold and thus its boundary, in particular $\partial_+ \mathcal{T}$ has a collar, that is, we can find an embedding $c: \partial_+ \mathcal{T} \times [0, 1] \hookrightarrow \mathcal{T}$ which maps $\partial_+ \mathcal{T} \times \{0\}$ to $\partial_+ \mathcal{T}$. Unfortunately, we can't just take any collar for our construction, we need some extra properties. We denote the stories of \mathcal{T} by S_n and let $\partial_n^+ \mathcal{T} = \partial_+ \mathcal{T} \cap S_n$. Then we require that c maps $\partial_n^+ \mathcal{T} \times [0, 1]$ into S_n and $(\partial_+ \mathcal{T} \cap \partial_- \mathcal{T}) \times [0, 1]$ into $\partial_- \mathcal{T}$.

Exercise 9.12. Convince yourself that we can always find such a collar!

From now on we will always assume that our collars have this property.

Now let's do the reembedding construction again with a little more precision. Along the way we will define certain *puzzle pieces* that will eventually make up the Design. As before we start with a CEQFAS handle $\widehat{C \cdot H}$ and fix some metric on it. The first step begins with relabeling the handle to $\widehat{C \cdot H}_{(0)}$ and choosing a collar $c_{(0)}: \partial_+ \widehat{C \cdot H}_{(0)} \times [0, 1] \hookrightarrow \widehat{C \cdot H}_{(0)}$ as above. For convenience we introduce the short hand notation

$$\nu_{(0)}[s, t] = c_{(0)}(\partial_+ \widehat{C \cdot H}_{(0)} \times [s, t])$$

and we define the first puzzle piece $\mathbf{D}_{(0)}$ as the intersection of the first story of $\widehat{C \cdot H}_{(0)}$ with $\nu_{(0)}[0, 1]$.

Next assume that we have done k steps and found CEQFAS handles $\widehat{C \cdot H}_I$, where $I \in \{0, 2\}^k$ is a finite dyadic expansion of length k with $i_1 = 0$, together with collars

$$c_I: \partial_+ \widehat{C \cdot H}_I \times [0, 1] \longrightarrow \widehat{C \cdot H}_I.$$

For the $(k + 1)$ -st step we take the first $(k + 1)$ stories from each handle $\widehat{C \cdot H}_I$ and truncate them by removing the part $\nu_I[0, \frac{2}{3})$ of the collar. The truncated $(k + 1)$ -st story consists of some number of capped gropes and inside each we can use Theorem 9.10 to locate a CEQFAS handle whose higher stories have diameter less than $\frac{1}{k}$ – this is the control we alluded to in the blueprint. (Note that we can't assume the whole handle to be small because its first story will be at least as big as its attaching region.) After adding the truncated first k stories of $\widehat{C \cdot H}_I$, to which the newly found handles naturally attach, we obtain a single CEQFAS handle which we call $\widehat{C \cdot H}_{I(2)}$. Note that the first k stories of $\widehat{C \cdot H}_{I(2)}$ overlaps with the image of the collar c_I in $\nu_I[\frac{2}{3}, 1]$ and we construct a collar $c_{I(2)}$ for $\widehat{C \cdot H}_{I(2)}$ that is compatible with c_I in the sense that

$$c_{I(2)}(p, t) = c_I(p, \frac{2}{3} + \frac{1}{3}t)$$

for all $p \in \partial_+ \widehat{C \cdot H}_{I(2)}$ the overlap. In addition, we rename $\widehat{C \cdot H}_I$ to $\widehat{C \cdot H}_{I(0)}$ and define a rescaled collar $c_{I(0)}(p, t) = c_I(p, \frac{1}{3}t)$. Finally, we define further puzzle pieces $\mathbf{D}_{I(i)}$, $i \in \{0, 2\}$, as the intersection of the $(k+1)$ -st story of $\widehat{C \cdot H}_{I(i)}$ with $\nu_{I(i)}[0, 1]$.

As before we iterate this process infinitely many times. We immediately see that we obtain a countable collection of CEQFAS handles $\widehat{C \cdot H}_I \subset \widehat{C \cdot H}$, one for each *finite* dyadic expansions; all of them are equipped with collars c_I . As far as the handles are concerned, this list contains a lot of repetition since for all I we have

$$\widehat{C \cdot H}_I = \widehat{C \cdot H}_{I(0)} = \widehat{C \cdot H}_{I(00)} = \dots$$

However, note that the corresponding collars are different – they become thinner and thinner as we attach more and more zeros to I – and the puzzle pieces $\mathbf{D}_I, \mathbf{D}_{I(0)}, \mathbf{D}_{I(00)}, \dots$ are derived from higher and higher stories of $\widehat{C \cdot H}_I$ (story $|I|$, $|I| + 1$ and so on). So what about the *infinite* dyadic expansions? After all, the blueprint promised a Cantor set worth of handles. For those expansions with an infinite tail of zeros, $r = I(00\dots)$ say, we simply take the “limit” of the finite case and let $\widehat{C \cdot H}_r = \widehat{C \cdot H}_I$. In this limit the collars $c_I, c_{I(0)}, \dots$ collapse onto $\partial_+ \widehat{C \cdot H}_I$ so we don’t equip $\widehat{C \cdot H}_r$ with a collar at all. Also, \mathbf{D}_r should be derived from the infinite story of $\widehat{C \cdot H}_r$ and the best way to make sense of this is to define \mathbf{D}_r as the endpoints of $\widehat{C \cdot H}_r$. For truly infinite expansions $r \in \{0, 2\}^\infty$, whose tail contains infinitely many 2s, we also get handles $\widehat{C \cdot H}_r \subset \widehat{C \cdot H}$ but this is slightly less obvious. A priori, the construction only gives *open* handles $\widehat{C \cdot H}_r \subset \widehat{C \cdot H}$ but similar control arguments as in the proof of Theorem 9.10 show that the closure of $\widehat{C \cdot H}_r$ in $\widehat{C \cdot H}$ is homeomorphic to its endpoint compactification. Again, we don’t need any collar and simply define \mathbf{D}_r as the endpoints.

At this point we can finally define the Design $\mathbf{D} \subset \widehat{C \cdot H}$ as the union of our puzzle pieces \mathbf{D}_I over all finite and infinite dyadic expansions starting with zero. If you like, you can think of the Design as a completed puzzle in $\widehat{C \cdot H}$. Since this is so important, here’s the definition once more in symbols

$$\mathbf{D} = \bigcup_{k \in \mathbb{N} \cup \{\infty\}} \bigcup_{\substack{|I|=k, \\ i_1=0}} \mathbf{D}_I. \quad (9.1)$$

Remark 9.13. The way we have defined the Design looks slightly different than the blueprint suggested. But we can easily make the connection by considering the neighborhoods $\nu \partial_+ \widehat{C \cdot H}_r = \bigcup_{I \subset r} (\mathbf{D}_I \cap \widehat{C \cdot H}_r)$ of the frontiers. A quick look at the definitions reveals that $\mathbf{D} = \bigcup_r \nu \partial_+ \widehat{C \cdot H}_r$. We will stick to the description in (9.1), though, since it is easier to work with.

Note that the design has a finite and an infinite part coming from the finite and infinite expansions; we call these $\mathbf{D}^{<\infty}$ and \mathbf{D}^∞ . The infinite part \mathbf{D}^∞ consist of the endpoints of all handles $\widehat{C \cdot H}_I$ with infinite expansions, in particular it is a “Cantor set worth of Cantor sets” inside $\widehat{C \cdot H}$. Clearly, the finite part seems more accessible and the good news is that, in some sense, it is enough to know.

Exercise 9.14. Show that the Design \mathbf{D} is (a) the closure of $\mathbf{D}^{<\infty}$ in $\widehat{C \cdot H}$ and (b) homeomorphic to the endpoint compactification of $\mathbf{D}^{<\infty}$.

Remark 9.15. Strictly speaking we shouldn’t speak of *the* Design since our building instructions involve so many choices that it is hard to imagine that two people would end up constructing the same set.

9.4 Embedding the Design in the standard handle

Now that we have found the Design \mathbf{D} in our CEQFAS handle $\widehat{\mathbf{C} \cdot \mathbf{H}} = \widehat{\mathbf{C} \cdot \mathbf{H}}_{(0)}$ we want to embed it into the standard handle $B^2 \times B^2$. We first focus on the finite part

$$\mathbf{D}^{<\infty} = \bigcup_{|I|<\infty} \mathbf{D}_I.$$

We denote by $\partial_+^k \widehat{\mathbf{C} \cdot \mathbf{H}}_I$ and $\partial_+^{\leq k} \widehat{\mathbf{C} \cdot \mathbf{H}}_I$ the intersections of $\partial_+ \widehat{\mathbf{C} \cdot \mathbf{H}}_I$ with the k -th story and the first k stories of $\widehat{\mathbf{C} \cdot \mathbf{H}}_I$, respectively. With this notation the collars give diffeomorphisms

$$\partial_+^{|I|} \widehat{\mathbf{C} \cdot \mathbf{H}}_I \times [0, 1] \xrightarrow[\cong]{c_I} \mathbf{D}_I \subset \widehat{\mathbf{C} \cdot \mathbf{H}}. \quad (9.2)$$

Moreover, if $I \subset J$ and $k \leq |I|$, then by construction $\widehat{\mathbf{C} \cdot \mathbf{H}}_J$ is contained in $\widehat{\mathbf{C} \cdot \mathbf{H}}_I$ and $\partial_+ \widehat{\mathbf{C} \cdot \mathbf{H}}_J$ meets the first k stories of $\widehat{\mathbf{C} \cdot \mathbf{H}}_I$ in $c_I(\partial_+^{\leq k} \widehat{\mathbf{C} \cdot \mathbf{H}}_I \times \{\rho_J^I\})$ for some $\rho_J^I \in [0, 1]$ so that we also have diffeomorphisms

$$\partial_+^{\leq k} \widehat{\mathbf{C} \cdot \mathbf{H}}_I \xrightarrow[\cong]{} \partial_+^{\leq k} \widehat{\mathbf{C} \cdot \mathbf{H}}_J \quad (9.3)$$

induced by the collars.

In the light of the above, the our next task is to identify the pieces $\partial_+^k \widehat{\mathbf{C} \cdot \mathbf{H}}_I$ or, equivalently, $\partial_+^{\leq k} \widehat{\mathbf{C} \cdot \mathbf{H}}_I$. As in Section 7.2, where we studied the boundaries of Casson handles, one can show:

Exercise 9.16. Each $\partial_+^{\leq k} \widehat{\mathbf{C} \cdot \mathbf{H}}_I$ is diffeomorphic to the complement of a neighborhood of a mixed Bing-Whitehead link in $B^2 \times S^1$ with Whitehead doubles in its last stage.

We can thus choose embeddings

$$\psi_I: \partial_+^{\leq |I|} \widehat{\mathbf{C} \cdot \mathbf{H}}_I \hookrightarrow B^2 \times S^1 \quad (9.4)$$

with the following properties:

- (i) ψ_I maps the torus where $\partial_+^{\leq |I|} \widehat{\mathbf{C} \cdot \mathbf{H}}_I$ meets $\partial_- \widehat{\mathbf{C} \cdot \mathbf{H}}_I$ to $\partial B^2 \times S^1$.
- (ii) Whenever I contains J we have a commutative diagram

$$\begin{array}{ccc} \partial_+^{\leq |I|} \widehat{\mathbf{C} \cdot \mathbf{H}}_I & \xrightarrow{\psi_I} & B^2 \times S^1 \\ \downarrow & & \parallel \\ \partial_+^{\leq |J|} \widehat{\mathbf{C} \cdot \mathbf{H}}_J & \xrightarrow[\psi_J]{} & B^2 \times S^1 \end{array} \quad (9.5)$$

where the embeddings $\partial_+^{\leq |I|} \widehat{\mathbf{C} \cdot \mathbf{H}}_I \hookrightarrow \partial_+^{\leq |J|} \widehat{\mathbf{C} \cdot \mathbf{H}}_J$ are derived from (9.3).

As a final piece to write down an embedding of \mathbf{D}_I into $B^2 \times B^2$ we define a rescaling function $r_I: [0, 1] \rightarrow [0, 1]$ by the formula

$$r_I(t) = 1 - \sum_{j=1}^k \frac{i_j}{3^j} - \frac{1}{3^k} t$$

where $I = (i_1 \dots i_k)$. This maps the unit interval $[0, 1]$ affinely onto $[r_I(1), r_I(0)]$ while reversing the direction. Note that for $|I| = k$ these are precisely the intervals in $[\frac{2}{3}, 1]$ that are *not* thrown out in the k -th step of the middle third construction of the Cantor set.

Next we define embeddings $\Psi_I: D_I \hookrightarrow B^2 \times B^2$ as a composition

$$\begin{array}{ccc} D_I & \xrightarrow[\cong]{c_I^{-1}} & \partial_+^{[I]} \widehat{C \cdot H}_I \times [0, 1] \\ & & \downarrow \psi_I \times r_I \\ & & B^2 \times S^1 \times [r_I(1), r_I(0)] \hookrightarrow B^2 \times B^2 \end{array}$$

where the last arrow comes from using polar coordinates in the second factor of $B^2 \times B^2$. Our goal is to patch these embeddings together in order to eventually get an embedding of the whole Design.

Lemma 9.17. *The maps Ψ_I and Ψ_J coincide on $D_I \cap D_J$ and give rise to an embedding*

$$\Psi_I \cup \Psi_J: D_I \cup D_J \rightarrow B^2 \times B^2.$$

The proof is just a matter of going through the definitions and we leave it as an exercise. This is a good opportunity for the reader to get acquainted with our notation.

Exercise 9.18. Prove Lemma 9.17.

At this point we have found an embedding of the finite part of the Design

$$\Psi^{<\infty}: D^{<\infty} \hookrightarrow B^2 \times B^2.$$

It remains to treat the infinite part, that is, the endpoints of the CEQFAS handles with infinite dyadic expansions.

Lemma 9.19 (Automatic shrinking). *Let $r = (r_1 r_2 \dots) \in \{0, 2\}^\infty$ be an infinite dyadic expansion. Then the complement of the union $\bigcup_k \text{im } \psi_{(r_1 \dots r_k)} \subset B^2 \times S^1$ is a mixed Bing-Whitehead decomposition \mathcal{B}_r whose elements correspond bijectively to the endpoints of $\widehat{C \cdot H}_r$. Furthermore, the ψ -maps in (9.4) can be chosen such that all decomposition elements of all decompositions \mathcal{B}_r are points.*

Proof. According to Exercise 9.16 the complement of $\text{im } \psi_{(r_1 \dots r_k)}$ is a neighborhood of a mixed Bing-Whitehead link and we denote it by L_r^k . From the compatibility condition (9.5) we see that L_r^{k+1} is contained in L_r^k so that the intersection $\bigcap_k L_r^k$, which is just another way to write the complement of $\bigcup_k \text{im } \psi_{(r_1 \dots r_k)}$, is a mixed Bing-Whitehead decomposition \mathcal{B}_r of $B^2 \times S^1$.

In order to relate this decomposition to $\widehat{C \cdot H}_r$ we note that by construction $\widehat{C \cdot H}_r$ is contained in $\widehat{C \cdot H}_{(r_1 \dots r_k)}$ for each k and it is easy to see that the collar $c_{(r_1 \dots r_k)}$ gives rise to a diffeomorphism

$$\partial_+^{\leq k} \widehat{C \cdot H}_{(r_1 \dots r_k)} \longrightarrow \partial_+^{\leq k} \widehat{C \cdot H}_r$$

as in (9.3). From here on we can construct a diffeomorphism

$$\partial_+ \widehat{C \cdot H}_r \xrightarrow{\cong} \bigcup_k \text{im } \psi_{(r_1 \dots r_k)} \subset B^2 \times S^1$$

and we see that the ends of $\partial_+ \widehat{C \cdot H}_r$ correspond bijectively with those of $\bigcup_k \text{im } \psi_{(r_1 \dots r_k)}$ which, in turn, correspond to the elements of \mathcal{B} . On the other hand, the ends of $\partial_+ \widehat{C \cdot H}_r$ correspond to those of $\widehat{C \cdot H}_r$ as one readily checks.

Moreover, the arguments above show that we can identify the full frontier $\partial_+ \widehat{C \cdot H}_r$ with the decomposition space $B^2 \times S^1 / \mathcal{B}_r$ and, since $\widehat{C \cdot H}_r$ satisfies the Ancel-Starbird property, \mathcal{B}_r is shrinkable. A careful look at the Bing shrinking argument shows that the defining

sequence L_r^k can be repositioned by successively applying isotopies in deeper and deeper stages such that all decomposition elements of \mathcal{B}_r are points. Moreover, if two infinite dyadic expansions r, s agree in their first N digits, then for $k \leq N$ we can identify L_r^k and L_s^k according to (9.5) and the shrinking isotopies can be chosen compatibly. Finally, these modifications of the L_r^k can be translated into modifications of the ψ -maps. \square

If the ψ 's are chosen as in the lemma, then we say that the construction has *automatic shrinking*; we assume this property from now on. Automatic shrinking shows that the closure of $\Psi^{<\infty}(\mathbf{D}^{<\infty})$ inside $B^2 \times B^2$ can be identified with the endpoint compactification of $\mathbf{D}^{<\infty}$ and, according to Exercise 9.14 and its preceding remarks, it follows that $\Psi^{<\infty}$ extends to an embedding of the full Design

$$\Psi: \mathbf{D} \rightarrow B^2 \times B^2 \quad (9.6)$$

by sending the endpoints of $\widehat{C \cdot H}_r$, $r \in \{0, 2\}$, which make up \mathbf{D}_r to the complement of $\bigcup_k \text{im } \psi_{(i_1 \dots i_k)} \times \{r\} \subset B^2 \times B^2$.

9.5 Holes, gaps and the Endgame

Our strategy is to shrink the complements of the Design in both the standard handle and the Casson handle or, more precisely, the closures of the components of the complements. We will refer to these parts of the standard handle as *holes* and to those in the Casson handle as *gaps*. The holes and gaps are the pieces that are left unexplored and the philosophy is that whenever we can't explore something we try to crush it out and hope that there's enough technology in decomposition space theory so that we can get away with it. To make this a reasonable strategy we should be careful not to crush things that are not cellular.

The problem is that, while there is one "central" hole in the standard handle that is homeomorphic to the 4-ball and is easy to shrink out, we will see that all other holes are homeomorphic to $S^1 \times B^3$ and it would be unproductive to crush them directly into points. To remedy this we try to add spanning disks to these holes to turn them into cellular and potentially more shrinkable decomposition elements. This means that we have to take a step back and give up some pieces that we've already explored but it turns out that what we get in return is worth it.

9.5.1 From holes and gaps to holes⁺ and gaps⁺

We first take a closer look at the holes. For that purpose it is useful to understand how the Design in the standard handle intersects the *sleeves* $B^2 \times S^1 \times \{r\}$ where we identify $S^1 \times \{r\}$ with the sphere of radius r in the cocore factor. The first thing to note is that the Design lives in the collar $B^2 \times S^1 \times [\frac{2}{3}, 1]$ of the belt region $B^2 \times S^1$ so that $\mathbf{D} \cap B^2 \times S^1 \times \{r\}$ is empty for radii $r < \frac{2}{3}$. For $r \geq \frac{2}{3}$ two things can happen. For a Cantor set radius we get the full sleeve as we've already noted in Remark 9.13. If we're not in a Cantor set radius, then we are in some "middle-third" interval given by some $I \in \{0, 2\}^k$ and the lower and upper endpoints are given by $r_I(0)$ and $r_I(1)$ respectively. In this case the design part in $B^2 \times S^1 \times \{r\}$ is given as the image of ψ_I . Therefore, the holes are given by the components of $L_I \times [r_I(0), r_I(1)]$ for all I , where L_I is the closure of the complement of $\text{im } \psi_I$. As already mentioned L_I is a neighborhood of a Bing-Whitehead link. The last Whitehead link components, which we denote by h_I^j where j runs from one to the number of link components, naturally span immersed disks Δ_I^j as shown in Figure 87.

We want to add the disks $\Delta_I^j \times \{r_I(\frac{1}{2})\}$ to the holes to make them birdlike equivalent. But we first have to perturb them in the second factor, to make them embedded disjointly



Figure 87: A spanning disk of a Whitehead double.

from all other disks and the holes. The disks intersect the associated link component in two disjoint sub disks and themselves in two intervals connecting these sub disks with the boundary. The way we will make them embedded is by pushing a neighborhood of one sub disk to a cantor set level $d_I > r_I(1)$ and a neighborhood of the other to a cantor set level $\hat{d}_I < r_I(0)$. More precise, we will do it in the following inductive way.

We isotope the disk such that they intersect the torus in the sub disks $B_{\frac{1}{4}}(\{\pm\frac{1}{2}\} \times \{0\})$ and themselves in the intervals $[-1, -\frac{3}{4}] \times \{0\}$ and $[\frac{3}{4}, 1] \times \{0\}$. Let $C := B_{\frac{9}{10}}(\{\frac{1}{2}\} \times \{0\}) \cup B_{\frac{9}{10}}(\{-\frac{1}{2}\} \times \{0\})$ and let $b : B^2 \rightarrow [0, 1]$ be a smooth function with $b \cong 1$ on $B_{\frac{1}{4}}(\{\pm\frac{1}{2}\} \times \{0\})$, $b > 0$ on C and $b \cong 0$ on $B^2 \setminus C$. Now we can start perturbing the disks. Define $C = \{\frac{2}{3}\} \cup \{1\}$. In every step choose one $I \in \{0, 2\}^k$ of shortest length where we not yet have spanned disks, then choose points d_I, \hat{d}_I in the cantor set which are no endpoints and such that $r_I(1) < d_I < \min C \cap [r_I(1), 1]$ and $r_I(0) > \hat{d}_I > \max C \cap [\frac{2}{3}, r_I(0)]$. Define $\phi_I : B^2 \rightarrow [0, 1]$ by

$$\phi_I(s) = \begin{cases} r_I(\frac{1}{2}) + b(s)(d_I - r_I(\frac{1}{2})) & \text{if } s \in B_{\frac{9}{10}}(\{\frac{1}{2}\} \times \{0\}) \\ r_I(\frac{1}{2}) + b(s)(\hat{d}_I - r_I(\frac{1}{2})) & \text{if } s \in B_{\frac{9}{10}}(\{-\frac{1}{2}\} \times \{0\}) \\ r_I(\frac{1}{2}) & \text{else} \end{cases}$$

For every component of the corresponding hole take the parametrized disks $\Delta_I^j : B^2 \rightarrow B^2 \times S^1$ as above and embed disks by

$$\tilde{\Delta}_I^j : B^2 \xrightarrow{\Delta_I^j \times \phi_I} B^2 \times S^1 \times [0, 1]$$

At last add the points $\{d_I, \hat{d}_I, r_I(0), r_I(1)\}$ to C . Now we can do the same for the next I .

Exercise 9.20. Check that the disks constructed above are embedded and disjoint from each other and the holes.

As mentioned earlier we now add the disks to the holes, that is, we define

$$h_I^{j+} = h_I^j \cup \Delta_I^j \subset B^2 \times B^2$$

and call the resulting objects *holes*⁺. Also, since the disks are contained in the design we can map them over to the CEQFAS handle $\widehat{C \cdot H}$ and add them to the gaps resulting in the *gaps*⁺.

9.5.2 The Endgame

To set the stage for the Endgame we introduce three decompositions:

- \mathcal{H}^+ is the collection of holes⁺ in $B^2 \times B^2$.
- \mathcal{G}^+ be the collection of gaps⁺ in $\widehat{\mathbf{C} \cdot \mathbf{H}}$.
- \mathcal{D} is the decomposition of the Design \mathbf{D} whose elements are the boundary components of the holes union the corresponding disks.

As it turns out, all these decompositions have a “common quotient” in the following sense.

Lemma 9.21. *The quotients $B^2 \times B^2 / \mathcal{H}^+$ and $\widehat{\mathbf{C} \cdot \mathbf{H}} / \mathcal{G}^+$ are both canonically homeomorphic to \mathbf{D} / \mathcal{D}*

Proof. This would follow if we knew that the collection of holes and the collection of gaps were both in one-to-one correspondence with the boundary components of the Design. While this is obviously true for the holes, it could happen that a single gap spans between different boundary components of the Design and we have to exclude this possibility. In other words, we have to show that any two boundary components ∂h_I^j and ∂h_J^k can be separated in $\widehat{\mathbf{C} \cdot \mathbf{H}}$ by something that lives inside the Design. We first note that for any finite or infinite dyadic expansion I the frontier $\partial_+ \widehat{\mathbf{C} \cdot \mathbf{H}}_I$ separates $\widehat{\mathbf{C} \cdot \mathbf{H}}$. Now, by construction the boundary component ∂h_I^j and any gap that meets it are contained in $\widehat{\mathbf{C} \cdot \mathbf{H}}_I$. But if we look more closely, then we actually see that

- they reside above the first $|I|$ stories of $\widehat{\mathbf{C} \cdot \mathbf{H}}_I$,
- they are contained in $\widehat{\mathbf{C} \cdot \mathbf{H}}_{I(0222\dots)} \subset \widehat{\mathbf{C} \cdot \mathbf{H}}_I$, and
- they don’t meet the interior of $\widehat{\mathbf{C} \cdot \mathbf{H}}_{I(2)} \subset \widehat{\mathbf{C} \cdot \mathbf{H}}_I$.

Moreover, any boundary component contained in this region must also be of the form ∂h_I^k , that is, it must have the same lower index.

This immediately shows that ∂h_I^j and ∂h_J^k are separated in $\widehat{\mathbf{C} \cdot \mathbf{H}}$ whenever $I \neq J$. So it remains to separate ∂h_I^j and ∂h_I^k for $j \neq k$. For that purpose we look at $\widehat{\mathbf{C} \cdot \mathbf{H}}_I$ and remove its first $|I|$ stories. What we’re left with is a disjoint union of towers whose attaching regions in the boundary of the first $|I|$ stories is in one-to-one correspondence with the boundary components of the Design. Moreover, each tower contains a gap which shows that no gap can meet two boundary components. \square

To obtain an actual common quotient of $B^2 \times B^2$ and $\widehat{\mathbf{C} \cdot \mathbf{H}}$ we let $Q = \mathbf{D} / \mathcal{D}$ and consider the following diagram:

$$\begin{array}{ccc}
 B^2 \times B^2 & & \widehat{\mathbf{C} \cdot \mathbf{H}} \\
 \downarrow / \mathcal{H}^+ & \searrow \alpha & \swarrow \beta \downarrow / \mathcal{G}^+ \\
 B^2 \times B^2 / \mathcal{H}^+ & \xrightarrow{\cong} Q & \xleftarrow{\cong} \widehat{\mathbf{C} \cdot \mathbf{H}} / \mathcal{G}^+
 \end{array} \tag{9.7}$$

Now the Endgame boils down to approximating α and β by homeomorphisms. This is where we finally get to use our toolbox from decomposition space theory.

Proposition 9.22 (The α shrink). *The decomposition \mathcal{H}^+ of $B^2 \times B^2$ given by the holes⁺ is shrinkable. In other words, the map $\alpha: B^2 \times B^2 \rightarrow Q$ is ABH.*

Proof. This will follow from Theorem 4.5 or rather its “birdlike” generalization indicated in Remark 4.8. Each hole⁺ is obviously birdlike equivalent, so in order to show that α is ABH we only have to show that \mathcal{H}^+ is null. Clearly, it is enough to show that the holes as well as their spanning disks are null independently.

We first take care of the holes arguing by contradiction. Let’s assume that the collection of holes is not null, that is, there is an $\epsilon > 0$ such that infinitely many holes have diameter at least ϵ (measured with respect to some fixed metric on $B^2 \times B^2$) – to save letters later we call things *large* if their diameter is at least ϵ . Recall that each hole h_I^j is the product of a solid torus L_I^j – which is a component of a thickened Bing-Whitehead link L_I in $B^2 \times S^1$ – and some interval depending on I . A close look at the construction reveals that the collection of these intervals is null so that there also have to be infinitely many large solid tori L_I^j . Moreover, according to (9.5) the links L_I are nested in the sense that if $I \subset J$, then $L_I \subset L_J$ so that each component of L_I is contained in some component of L_J . In particular, if L_J has a large component, then so does L_I . It follows that among the finitely many tori L_I^j with fixed length of I , we can always find a large one which contains infinitely many other large ones. We call such tori *huge*.

To obtain a contradiction we will construct an infinite expansion $r = (0r_2r_3\ldots) \in \{0, 2\}^\infty$ that violates the automatic shrinking property of Lemma 9.19. We construct r inductively along with a sequence of huge solid tori $L_k = L_{(0r_2\ldots r_k)}^{j_k}$ such that $L_{k+1} \subset L_k$. Starting with $k = 2$ we know that there must be some huge torus of the form $L_{(0r_2)}^{j_2}$ which we take as L_2 and, of course, we also get an r_2 . Assuming that we have already chosen r_2, \ldots, r_k and L_1, \ldots, L_k for some $k \geq 2$, then one of the finitely many tori of the form $L_{(0r_2\ldots r_k i)}^j \subset L_k$ must be huge which gives r_{k+1} and L_{k+1} . Now, since all the L_k are huge, their intersection $\bigcap_k L_k$ must be large. But according to automatic shrinking $\bigcap_k L_k$ must be a point since it is one of the decomposition elements in Lemma 9.19 as we saw during its proof. So the collection of holes must be null.

Turning our attention to the disks $\tilde{\Delta}_I^j$, we first observe that the diameter of each unperturbed disk (which was the spanning disk of some Whitehead curve) is bounded by the diameter of the corresponding hole. So we only have to check that the perturbations don’t increase the size of the disks too much. But this is essentially built-in in our construction. Indeed, recall that the perturbations were also constructed inductively and in each step we were only allowed to move in a subinterval of $[0, 1]$. Another look at the construction shows that the collection of these intervals is null so that, altogether, the disks must be null, too. \square

Proposition 9.23. *The decomposition \mathcal{G}^+ of $B^2 \times B^2$ given by the gaps⁺ is shrinkable. In other words, the map $\beta: \widehat{C \cdot H} \rightarrow Q$ is ABH.*

Proof. Thanks to Proposition 9.22 we now know that Q is a ball and we want to apply the ball to ball theorem to β . But there’s the tiny problem that we don’t know yet that $\widehat{C \cdot H}$ is a ball. So we need to find a clever way to reduce the problem to a ball-to-ball situation. The idea is to add an exterior collar to the frontier $\partial_+ \widehat{C \cdot H}$ and to make use of the set E that appears in the statement of the ball-to-ball theorem.

Lemma 9.24. *$\widehat{C \cdot H} \cup (\partial_+ \widehat{C \cdot H} \times [0, 1])$ is homeomorphic to a ball.*

Proof (sketch). As we saw in Proposition 9.7 we can consider $\widehat{C \cdot H}$ as a decomposition space $B^2 \times B^2 / \tilde{\mathcal{D}}$ where the elements of \mathcal{D} are cones on elements of the mixed Bing-Whitehead decomposition \mathcal{D} that appears in the belt region $B^2 \times S^1$ and whose quotient is $\partial_+ \widehat{C \cdot H}$. Since we can arrange the elements of \mathcal{D} to be points by the Ancel-Starbird property, it follows that the elements of $\tilde{\mathcal{D}}$ are arcs in $B^2 \times B^2$.

Note that this situation is very similar to our decomposition space description of the Alexander gored ball. There we had a decomposition of B^3 whose elements were cones on a Cantor set in S^2 , that is, arcs the emanate from the boundary into the interior. Although this decomposition doesn't shrink, we observed in Remark 3.16 that it does shrink when considered as a decomposition of \mathbb{R}^3 ; in fact, we don't have to go to the whole \mathbb{R}^3 but any bigger ball containing B^3 is enough. In other words, it is enough to add an exterior collar to B^3 . Now, the same trick works for $\widehat{C \cdot H}$. We extend $B^2 \times B^2$ by adding an exterior collar $E = (B^2 \times S^1) \times [0, 1]$ of the $+$ -part of the boundary and the quotient map

$$B^2 \times B^2 \cup E \longrightarrow (B^2 \times B^2 \cup E)/\widehat{\mathcal{D}}$$

will become ABH. Moreover, the quotient map restricts to a homeomorphism on E so that the quotient is homeomorphic to $\widehat{C \cdot H} \cup (\partial_+ \widehat{C \cdot H} \times [0, 1])$ \square

In order to reduce to a ball-to-ball situation observe that α and β are homeomorphisms on the boundary so that we can glue $E = (B^2 \times S^1) \times [0, 1]$ onto Q using α and then further onto $\widehat{C \cdot H}$ using β . Then we extend β to a map

$$\tilde{\beta}: \widehat{C \cdot H} \cup_{\beta} E \longrightarrow Q \cup_{\alpha} E$$

in the obvious way. This is now a map between balls and, if applicable, the ball-to-ball theorem will tell us that $\tilde{\beta}$ can be approximated by homeomorphisms relative to E where $\tilde{\beta}$ already is a homeomorphism. Thus it follows that β must be ABH.

It remains to verify the assumptions of the ball-to-ball theorem and this means checking two things. First, that the collection of gaps⁺ is null and second, that their image in Q is nowhere dense. For the nullity, observe that we don't have to worry about the disks but only about the gaps. Indeed, the disks live inside the design which, as an abstract space, is compact and metrizable and it is easy to see that whether or not something is null doesn't depend on a specific metric. As for the gaps, we have already observed that the gap g_I^j is contained in $\widehat{C \cdot H}_I$ above the first $|I|$ stages. But the control part in the construction forces these stages to have diameter less than $1/|I|$, so as the length of I increases, the diameter of the corresponding gaps tends to zero.

To see that the set of singular points of β is nowhere dense in Q we observe that we can consider the singular points of α instead. This means that we don't have to worry about gaps, but only about holes.

Exercise 9.25. Let \mathcal{D} be a decomposition of a space X such that for any element $D \in \mathcal{D}$ and any open neighborhood U of D there is an open subset $V \subset U$ whose closure does not meet any decomposition element. Then the set of singular points in X/\mathcal{D} is nowhere dense.

Now, the condition in this exercise is obviously satisfied for the holes⁺. \square

10 Epilogue: Edwards' original shrink

Coming soon...

References

- [Anc84] F. D. Ancel, *Approximating cell-like maps of S^4 by homeomorphisms*, Four-manifold theory (Durham, N.H., 1982), Contemp. Math., vol. 35, Amer. Math. Soc., Providence, RI, 1984, pp. 143–164.
- [AS89] F. D. Ancel and M. P. Starbird, *The shrinkability of Bing-Whitehead decompositions*, Topology **28** (1989), no. 3, 291–304.
- [AR65] J. J. Andrews and L. Rubin, *Some spaces whose product with E^1 is E^4* , Bull. Amer. Math. Soc. **71** (1965), 675–677.
- [Bea67] R. J. Bean, *Decompositions of E^3 with a null sequence of starlike equivalent non-degenerate elements are E^3* , Illinois J. Math. **11** (1967), 21–23. MR0208581 (34 #8390)
- [Bin52] R. H. Bing, *A homeomorphism between the 3-sphere and the sum of two solid horned spheres*, Ann. of Math. (2) **56** (1952), 354–362.
- [Bro60] M. Brown, *A proof of the generalized Schoenflies theorem*, Bull. Amer. Math. Soc. **66** (1960), 74–76.
- [Cas86] A. J. Casson, *Three lectures on new-infinite constructions in 4-dimensional manifolds*, À la recherche de la topologie perdue, Progr. Math., vol. 62, Birkhäuser Boston, Boston, MA, 1986, pp. 201–244. With an appendix by L. Siebenmann.
- [CG78] A. J. Casson and C. McA. Gordon, *On slice knots in dimension three*, Algebraic and geometric topology (Proc. Sympos. Pure Math., Stanford Univ., Stanford, Calif., 1976), Part 2, Proc. Sympos. Pure Math., XXXII, Amer. Math. Soc., Providence, R.I., 1978, pp. 39–53.
- [COT03] T. D. Cochran, K. E. Orr, and P. Teichner, *Knot concordance, Whitney towers and L^2 -signatures*, Ann. of Math. (2) **157** (2003), no. 2, 433–519.
- [COT04] ———, *Structure in the classical knot concordance group*, Comment. Math. Helv. **79** (2004), no. 1, 105–123.
- [Dav86] R. J. Daverman, *Decompositions of manifolds*, Pure and Applied Mathematics, vol. 124, Academic Press Inc., Orlando, FL, 1986.
- [DS83] R. Denman and M. Starbird, *Shrinking countable decompositions of S^3* , Trans. Amer. Math. Soc. **276** (1983), no. 2, 743–756.
- [Edw80] R. D. Edwards, *The topology of manifolds and cell-like maps*, Proceedings of the International Congress of Mathematicians (Helsinki, 1978), Acad. Sci. Fennica, Helsinki, 1980, pp. 111–127, available at <http://www.mathunion.org/ICM/ICM1978.1/>.
- [Edw84] ———, *The solution of the 4-dimensional annulus conjecture (after Frank Quinn)*, Four-manifold theory (Durham, N.H., 1982), Contemp. Math., vol. 35, Amer. Math. Soc., Providence, RI, 1984, pp. 211–264.
- [Fre82] M. H. Freedman, *The topology of four-dimensional manifolds*, J. Differential Geom. **17** (1982), no. 3, 357–453.
- [FQ90] M. H. Freedman and F. Quinn, *Topology of 4-manifolds*, Princeton Mathematical Series, vol. 39, Princeton University Press, Princeton, NJ, 1990.
- [GS84] R. E. Gompf and S. Singh, *On Freedman reimbedding theorems*, Four-manifold theory (Durham, N.H., 1982), Contemp. Math., vol. 35, Amer. Math. Soc., Providence, RI, 1984, pp. 277–309.
- [Kir89] R. C. Kirby, *The topology of 4-manifolds*, Lecture Notes in Mathematics, vol. 1374, Springer-Verlag, Berlin, 1989.
- [Maz59] B. Mazur, *On embeddings of spheres*, Bull. Amer. Math. Soc. **65** (1959), 59–65.
- [Mor60] M. Morse, *A reduction of the Schoenflies extension problem*, Bull. Amer. Math. Soc. **66** (1960), 113–115.
- [Qui82] F. Quinn, *Ends of maps. III. Dimensions 4 and 5*, J. Differential Geom. **17** (1982), no. 3, 503–521.
- [Whi44] H. Whitney, *The self-intersections of a smooth n -manifold in $2n$ -space*, Ann. of Math. (2) **45** (1944), 220–246.