AGT

Algebraic & Geometric Topology

Volume 25 (2025)

Shrinking without doing much at all

MICHAEL FREEDMAN
MICHAEL STARBIRD

msp

msp

Algebraic & Geometric Topology **25**:4 (2025) 2099–2114
DOI: 10.2140/agt.2025.25.2099
Published: 11 August 2025

Shrinking without doing much at all

MICHAEL FREEDMAN
MICHAEL STARBIRD

In 1952 Bing astonished the mathematical world with his wild involution on S^3. It has been among the most seminal examples in topology. The example depends on finding shrinking homeomorphisms of Bing's decomposition of S^3 into points and arcs. If Bing's original homeomorphisms are varied, Bing's original wild involution changes by conjugation, which preserves some analytic properties while altering others. In 1988, Bing published a second paper, *Shrinking without lengthening*, answering a question that one of the present authors posed to him in an effort to understand the geometry of the entire conjugacy class. Here we produce a counterintuitive construction, namely a method to shrink the Bing decomposition doing almost nothing at all: neither lengthening much nor rotating much.

57K30; 57M60

1 Introduction

In 1952, Bing [3] used decomposition space theory (DST) to produce a shrink of what is now called the Bing decomposition as the key step in building a *wild* involution of the three-sphere S^3. The involution is wild in that it cannot be made *smooth* in any system of coordinates. This single example invigorated several decades of research in DST, beginning with a rich theory in three dimensions, where manifold factors were first discovered by Andrews and Rubin [2]. Then the subject jumped into high dimensions, where J Cannon [6] and R Edwards [9] showed that the double suspensions of homology spheres are homeomorphic to a standard sphere, a project that culminated in fundamental results of F Quinn and R Edwards on manifold recognition (see Quinn [12] and Davis [7] for an introduction). The Bing decomposition and its close relatives were also fundamental to Freedman's proof [10] of the four-dimensional Poincaré conjecture. A Dranishnikov and collaborators constructed remarkable dimension-raising quotient maps; see Dranishnikov and Schchepin [8]. Combining Bing-style DST and Quinn-style surgery, J Bryant, S Ferry, W Mio and S Weinberger constructed the modern theory of ANR homology manifolds [5].

This paper is part of a reconsideration of DST with analytical aspects in mind. Our earlier paper [11] answered a long-standing question about the analytical properties of the Bing involution. In no coordinate system can it ever be made Lipschitz or even quasiconformal. In fact, any topological conjugate of the Bing involution was shown to have, up to a poly-log factor, an exponential modulus of continuity. The

estimate for this intrinsic modulus of continuity (imoc) requires thinking not just about *one* shrink of the Bing decomposition $\mathcal{D}$, but *all* possible shrinks (since the different shrinks can be thought of as conjugates of any single one).

Studying, in this sense, *all* shrinks of $\mathcal{D}$, we found a "surprise shrink", the subject of this paper. We present it for two reasons: First, because Bing's decomposition $\mathcal{D}$ is the ur-example of DST. Second, it is the hope of the authors that the method presented here might be combined with [11] to strengthen the main result of that paper and remove the annoying poly-log factor. At first this hope seems odd since this paper provides a novel *shrinking* method and [11] is, in a sense, a *nonshrinking* result: $\mathcal{D}$ can only be shrunk by doing great violence to its $\mathbb{Z}_2$-reflection symmetry. But, in the proof of [11], the poly-log originates from the possibility that the imagined adversarial shrinker at some point starts to "delay" by making only tiny motions. Our work here gives some insight into what classes of such "tiny motions" indeed result in shrinking — and, when they do, how to quantify the violations of symmetry.

What is the purpose of joining DST to analysis? Our answer is 4-manifolds. By a historical accident the topological theory of 4-manifolds arose simultaneously with Donaldson's theory of smooth 4-manifolds. Donaldson theory immediately implied that the infinite constructions of DST could not generally yield smoothable results. In a sense, the topologists were given an easy way out: a crisp no-go theorem. In 1982, there was no appetite to dig into shrinking arguments and determine exactly where, and how much, regularity was lost. With [11], the time seems ripe to study the regularity of DST constructions.

2 The Bing decomposition and his two shrinks

Bing's decomposition is made by intersecting finite stages called "Bing rings", nested solid tori (see Figure 2). The shrink amounts to figuring out a strategy for stretching, twisting, bending and/or rotating each finer pair of *daughter* solid tori within the previous *mother* stage. Bing found that the rings need be lengthened only infinitesimally during the shrink, whence his title. We find that not only can the lengths of the solid tori be nearly preserved, but the "rotations" of the daughters within the mother can also be made arbitrarily small and can be made to decay towards zero.

The Alexander horned sphere is the first known wild embedding of S^2 in S^3. One method of creating that wild embedding of S^2 is as follows: Start with a standard S^2 in S^3. For ease of visualization, think of S^2 as the yz plane in $\mathbb{R}^3$ with a point at infinity that makes $\mathbb{R}^3$ into S^3 and makes the yz plane a standard embedding of S^2 in S^3. On the positive x side of this S^2, construct a specific Cantor set's worth of arcs — one end of each arc will be on S^2, while the other ends of those arcs will be rather entangled among each other. Specifically, the Cantor set's worth of arcs are created by taking the components of an infinite intersection of families $\{C_i\}_{i=0}^{\infty}$ of U-shaped solid cylinders. Each family C_i has 2^i components. C_0 consists of a single U-shaped cylinder as pictured in Figure 1. Every component of C_i contains two U-shaped cylinders of C_{i+1} embedded as shown in Figure 1. With care, each component of $\bigcap_{i=0}^{\infty} C_i$ is an arc meeting S^2 in a single point, and the totality of those arc endpoints is a Cantor set on S^2.

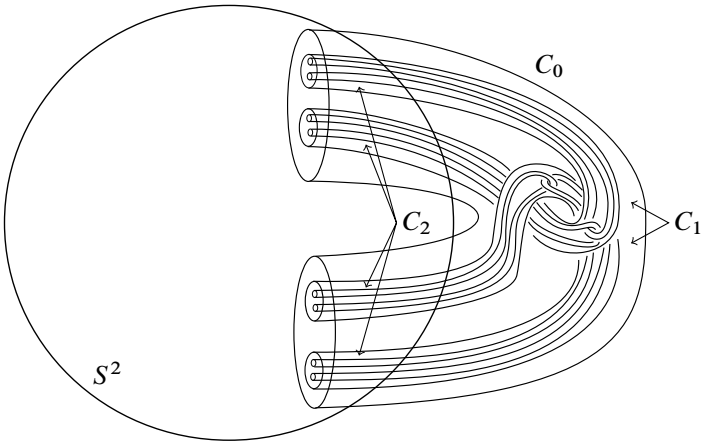

Figure 1: C_2 has four components, continuing the pattern as shown.

Shrinking that Cantor set's worth of arcs is key to creating the Alexander horned sphere embedding of S^2. The set consisting of those arcs together with the remaining points of S^3 is an upper-semicontinuous decomposition $\mathcal{D}_{\mathcal{A}}$ of S^3.

The Alexander horned sphere can be created as the image of S^2 under a surjective map $g'\colon S^3 \to S^3$ whose point preimages are exactly the elements of $\mathcal{D}_{\mathcal{A}}$. The function g' can easily be constructed as the limit of homeomorphisms $\{g'_i\colon S^3 \to S^3\}_{i=0}^{\infty}$ that increasingly compress the arcs of $\mathcal{D}_{\mathcal{A}}$ toward their endpoints off S^2. Describing this process in terms of quotient spaces, the function g' shows that $S^3/\mathcal{D}_A \cong S^3$, so the 3-sphere has not been changed, but the 2-sphere $g'(S^2)$ is the *wild* Alexander horned sphere. Its wildness is reflected in the fact that the right side component of its complement is no longer simply connected but instead has an infinitely generated fundamental group. The simple closed curve around the center of the cylinder C_0 is an example of a nontrivial element of the fundamental group of the wild component of $S^3 - g'(S^2)$. The image under g' of the arcs in $\mathcal{D}_{\mathcal{A}}$ is a Cantor set that lies on the Alexander horned sphere.

Consider the two *closed* complementary regions of $S^3 - g'(S^2)$. The closed left side region is clearly homeomorphic to the ball B^3, whereas the closed right side region, the Alexander horned ball, is clearly not, having a nonsimply connected interior.

Now create a new decomposition $\mathcal{D}$ of S^3 by creating nontrivial elements that are symmetric across S^2. Let I_0 be the standard involution of S^3 with fixed-point set S^2; that is, $I_0(x, y, z) = (-x, y, z)$. The nondegenerate elements of $\mathcal{D}$ are the components of the infinite intersection $\bigcap_{i=0}^{\infty} \widetilde{\mathcal{T}}_i$, where $\widetilde{\mathcal{T}}_i = C_i \cup I_0(C_i)$. So each $\widetilde{\mathcal{T}}_i$ is the union of 2^i tori that each meet S^2 in two meridional disks. We will denote each torus component of $\widetilde{\mathcal{T}}_i$ by $\widetilde{T}_\sigma$, where σ is a binary string of length i and $\widetilde{T}_{\sigma 0}$ and $\widetilde{T}_{\sigma 1}$ are the two component tori of $\widetilde{\mathcal{T}}_{i+1}$ contained in $\widetilde{T}_\sigma$. Figure 2 shows $\widetilde{\mathcal{T}}_0$, $\widetilde{\mathcal{T}}_1$ and $\widetilde{\mathcal{T}}_2$. Notice that $\bigcap_{i=0}^{\infty} \widetilde{\mathcal{T}}_i$ is a Cantor set's worth of arcs, each one piercing S^2 at its center point. Those arcs comprise the nondegenerate elements of $\mathcal{D}$.

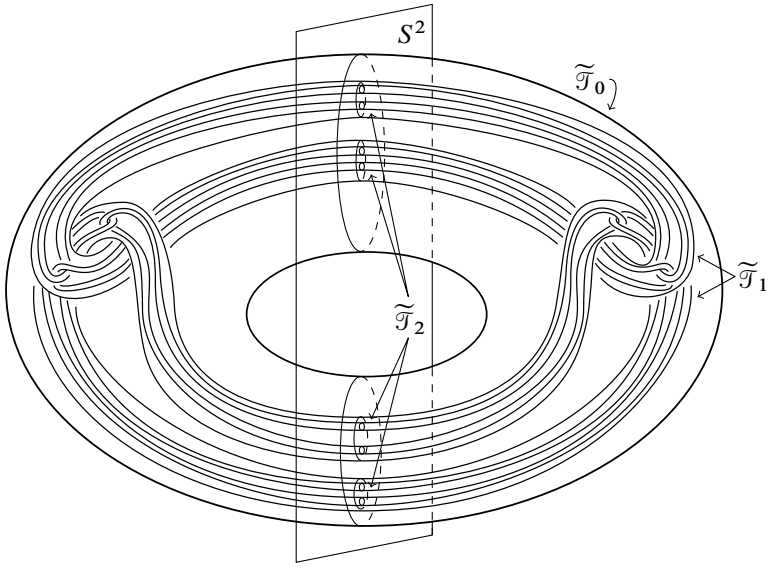

Figure 2

In his 1952 paper, Bing showed that there is a surjective map $g\colon S^3 \to S^3$ whose point preimages are precisely the sets in $\mathcal{D}$, thereby producing an involution $I : S^3 \to S^3$ defined by $I(x) = gI_0g^{-1}(x)$. The involution I is wild since its fixed-point set is the wild 2-sphere $g(S^2)$, and the closure of each component of the complement of $g(S^2)$ is an Alexander horned ball. The involution I swaps sides across a wild sphere.

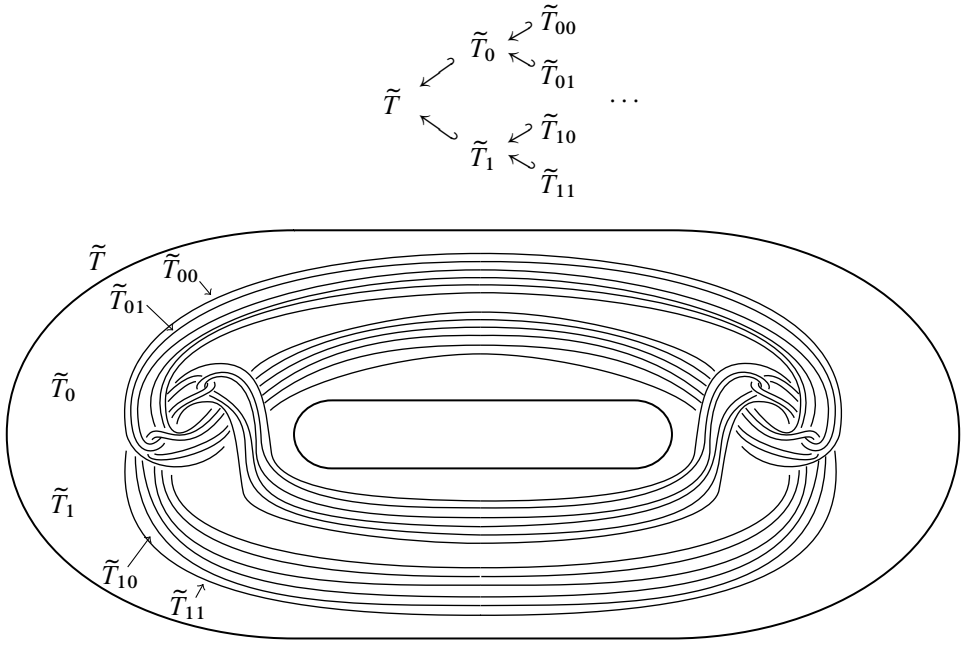

Figure 3

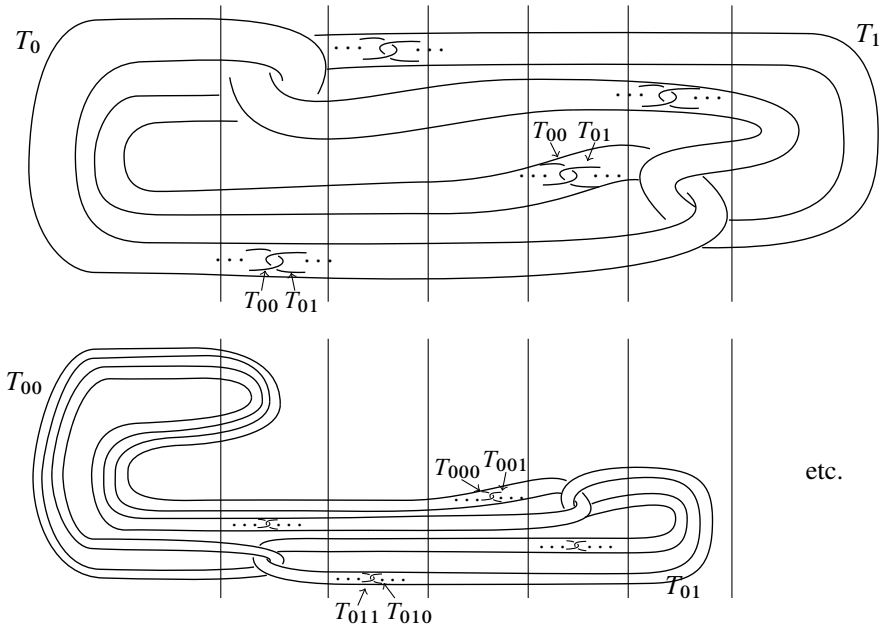

Figure 4

Bing's map $g\colon S^3 \to S^3$ is produced as the limit of shrinking homeomorphisms $\{g_i : S^3 \to S^3\}_{i=0}^{\infty}$. The g_i homeomorphisms eventually shrink the nondegenerate elements of $\mathcal{D}$ to increasingly smaller diameters as i increases in such a way that the limit g has the property that each nondegenerate element of $\mathcal{D}$ shrinks to a point and $\{g^{-1}(x) \mid x \in S^3\} = \mathcal{D}$. So the challenge is to produce the homeomorphisms g_i that shrink the arcs of $\mathcal{D}$.

Bing's shrinks The astonishing conclusion of [3] is that such shrinking homeomorphisms exist. Let us review two shrinks of $\mathcal{D}$ that Bing published in [3; 4]. In both shrinks, Bing reduces the shrinks to an essentially 1D model where the only important measure of diameter is displacement along the x-axis. Denoting the "Bing tori" dyadically, he lays them out along the x-axis and measures their diameters discretely by choosing a large integer n and erecting parallel planes in intervals of $\frac{1}{n}$. So the original tori are positioned as in Figure 3.

Bing's original shrink [3] was accomplished by describing a sequence of homeomorphisms $\{g_i\}_{i=0}^{\infty}$. Each homeomorphism g_{i+1} agrees with g_i on $S^3 - \widetilde{\mathcal{T}}_i$. The homeomorphisms g_i are defined in sets, meaning we first define the first n_1 g_i's and then sort of pause while we celebrate a certain amount of shrinking; for example, we could choose our first collection of homeomorphisms so that, for every component torus $\widetilde{T}_\sigma$ of $\widetilde{\mathcal{T}}_{n_1}$, $\operatorname{diam}(g_{n_1}(\widetilde{T}_\sigma))$ meets some diameter goal, say $< \frac{2}{10^1}$. Then we start with a new diameter goal — say $< \frac{2}{10^2}$ — and create the next set of g_i's, say $i = n_1 + 1, \dots, n_2$. It is only towards the end of each set of homeomorphisms that the diameters shrink. In other words, when we look at the diameters of the components of $\{g_j(\widetilde{\mathcal{T}}_j)\}_{j=n_1+1}^{n_2-1}$, generally those images of tori do not have increasingly smaller diameters; however, finally, at stage n_2, for each torus $\widetilde{T}_\sigma \subset \widetilde{\mathcal{T}}_{n_2}$, $\operatorname{diam}(g_{n_2}(\widetilde{T}_\sigma)) < \frac{2}{10^2}$.

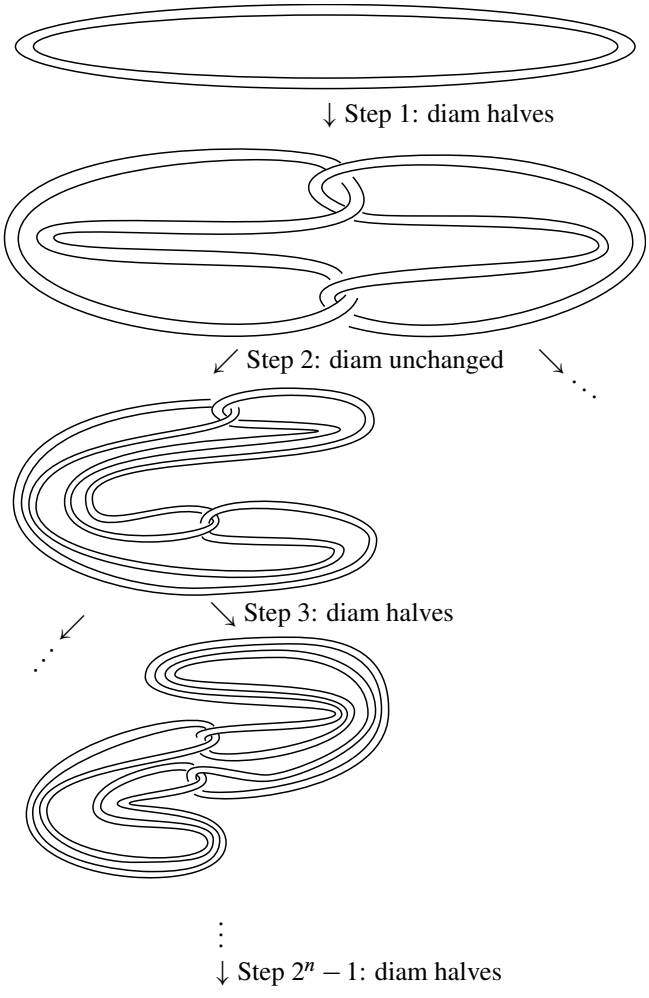

Figure 5

The homeomorphisms $\{g_i\}_{i=1}^{n_1}$ are defined as follows: Let's choose our first goal to be to shrink each torus at some stage to diameter less than, say, $\frac{2}{10^1}$. Imagine the first stage torus as long, thin and flat. Then position sufficiently many parallel planes — say n_1 planes — so that the distance between consecutive parallel planes is less than $\frac{1}{10^1}$ and the torus $\widetilde{T}_0$ intersects each of the n_1 planes in a pair of meridional disks. The first homeomorphism g_1 leaves $S^3 - \widetilde{\mathcal{T}}_0$ fixed and rotates $\widetilde{T}_0 \cup \widetilde{T}_1$ in $\widetilde{\mathcal{T}}_0$ so that each intersects one less plane (see Figure 4). For simplicity, for each component torus $\widetilde{T}_\sigma \subset \widetilde{\mathcal{T}}_i$, we will denote $g_i(\widetilde{T}_\sigma)$ by T_σ. So after the rotation and using our new notation, T_0 and T_1 each intersects only $n_1 - 1$ planes. In general, each subsequent pair of daughter tori are rotated in their already moved mother so that each of the k-stage daughters meet only $n_1 - k$ planes (Figure 4).

After n_1 generations, no T_σ, where σ is a binary word of length $|\sigma| = n_1$, meets more than one of the planes, so its x-axis extent is $< \frac{2}{10^1}$. Normal to the x-axis, we are free to have chosen a strong

compression, so this procedure produces a homeomorphism g_{n_1} that shrinks each n_1-stage torus to diameter less than $\frac{2}{10}$. Shrinking the tori of course shrinks the decomposition elements therein.

Next we choose a new diameter goal, say, $\frac{2}{10^2}$. Make many tick marks (actually, meridional disks on parallel planes) along the partially shrunk T_σ's for $|\sigma| = n_1$ such that the distance between consecutive (around T_σ) meridional disks chosen is less than $\frac{1}{10^2}$. Now start our process over. That is, let $g_{n_1}+1$ rotate the two daughters in each T_σ in such a way that those daughters each intersect one less meridional disk. Continue defining the g_i's, each reducing the number of meridional disks intersected by each stage torus, until we reach a number n_2 such that every T_σ, where $|\sigma| = n_2$, meets at most one meridional disk. Again, by compressing dimensions other than the x-extent means every such T_σ has diameter less than $\frac{2}{10^2}$, as desired. Notice that, because of the folded nature of the T_σ's for $|\sigma| = n_1$, the images under the g_i's starting with $i = n_1 + 1$ do not decrease the diameters of the T_σ's for a long time, but, when we reach n_2, we can again pause to celebrate successful shrinking.

So, after sufficient celebration, we start again and repeat the process with an even more ambitiously small diameter goal. Continue producing such g_i's. In the limit, the g_i's converge to a surjective function $g\colon S^3 \to S^3$ whose nondegenerate point preimages are precisely the nondegenerate elements of $\mathcal{D}$. This then was Bing's original method of shrinking the decomposition $\mathcal{D}$.

Next we summarize Bing's 1988 shrink, which he produced in answer to questions we asked him at that time. In his 1988 shrink, every other rotation of tori is *greedy*, as it tries (usually in vain) to cut diameters in half by rotating the daughters maximally, that is, rotating the clasp points of the daughters to the midpoints of their mother. The alternate rotations are *patient* — just rotating by $\frac{1}{2^n}$ when going from diameter $\frac{1}{2^{n-1}}$to diameter $\frac{1}{2^n}$. It turns out greed does not speed the shrinking; it is only at steps indexed by $2^n - 1$ that diameters *actually* are halved. So, again, many steps are taken during which no diameter shrinking is accomplished. Figure 5 shows the idea of Bing's 1988 shrink [4].

3 A small displacement Bing shrink

As described in the introduction, every known shrink of the Bing decomposition consists in starting with a standard torus and describing at each stage how to displace the clasp points of the two daughters relative to the two clasp points of the mother. For known shrinks, the clasps can be imagined as arbitrarily tight, and the reduction of diameter can be studied in a strictly 1D model where the starting torus is configured to tightly surround the unit interval. In this way, the diameter of a torus at a future stage is measured by the x-axis width of the part of $[0,1]$ covered by that folded solid torus.

Describing the positions of the folded tori at each stage can therefore be captured as a binary tree of functions into $[0,1]$ as follows. We start with an example.

Example The first, straight torus $T_\varnothing$ is modeled by the identity function $f_\varnothing\colon [0,1] \to [0,1]$. In the example in Figure 6, the clasp positions of the daughters T_0 and T_1 are displaced by distance $\frac{1}{3}$ at the left

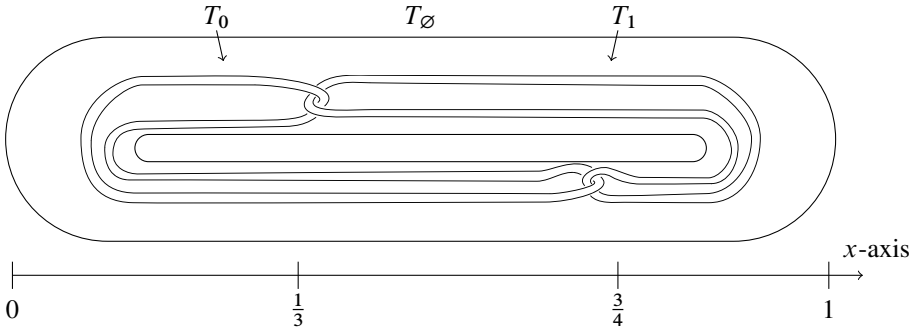

Figure 6

and distance $\frac{1}{4}$ at the right end. So the folded length of T_0 is $\frac{1}{3}+\frac{3}{4}=\frac{13}{12}$ and the folded length of T_1 is $\left(1-\frac{1}{3}\right)+\left(1-\frac{3}{4}\right)=\frac{11}{12}$.

The following two piecewise linear functions model the configurations of T_0 and T_1. The function $f_0\colon\left[-\frac{1}{3},\frac{3}{4}\right]\to[0,1]$ takes the interval $\left[-\frac{1}{3},0\right]$ backwards from $\frac{1}{3}$ to 0 and then proceeds forwards from 0 to $\frac{3}{4}$. Specifically,

$$f_0(x):=\begin{cases} f_\varnothing(-x) & \text{if } -\frac{1}{3}\le x\le 0,\\ f_\varnothing(x) & \text{if } 0<x\le\frac{3}{4}.\end{cases}$$

Thus, the image of f_0 models the shape of T_0 projected into the interval $[0,1]$.

Likewise, $f_1\colon\left[\frac{1}{3},\frac{5}{4}\right]\to[0,1]$ is defined to model the shape of T_1. Specifically,

$$f_1(x):=\begin{cases} f_\varnothing(x) & \text{if } \frac{1}{3}\le x\le 1,\\ f_\varnothing(1-(x-1))=f_\varnothing(2-x) & \text{if } 1<x\le\frac{5}{4}.\end{cases}$$

Notice that the lengths of the domains of f_0 and f_1, respectively, roughly equal the lengths of the folded tori T_0 and T_1, although, since T_0 and T_1 are clasped, they must be slightly longer than the domains of f_0 and f_1.

Following the pattern of this example, we next describe how to inductively produce a binary tree of functions $\{f_\sigma\}$ indexed by finite binary strings σ into $[0,1]$ whose images trace the patterns of the folded tori created by displacing the clasping positions of daughter pairs relative to clasping positions of their mothers when shrinking the Bing decomposition.

Definition 3.1 We define $f_\varnothing$ as the identity function on $[0,1]$ and define $[c_\varnothing,d_\varnothing]:=[0,1]$. Suppose σ is a finite binary string and the function $f_\sigma\colon[c_\sigma,d_\sigma]\to[0,1]$ has been produced. The string σ has two daughters, $\sigma0$ and $\sigma1$, and f_σ will have two daughters $f_{\sigma0}$ and $f_{\sigma1}$. The definitions of the daughter functions of f_σ are determined by the choice of a pair of points a_σ,b_σ in $[c_\sigma,d_\sigma]$ with $c_\sigma\le a_\sigma\le b_\sigma\le d_\sigma$. We call a_σ and b_σ the daughter *clasp points*. As in the example, the domain of $f_{\sigma0}$ extends the domain of f_σ to the left so that $f_{\sigma0}$ begins by retracing f_σ backwards from a_σ to c_σ, then forwards from c_σ to b_σ. So the domain of $f_{\sigma0}$ is $[c_\sigma-(a_\sigma-c_\sigma),b_\sigma]=[2c_\sigma-a_\sigma,b_\sigma]$. Likewise, $f_{\sigma1}$ retraces f_σ forward from a_σ to d_σ and then backwards from d_σ to b_σ (see Figure 7).

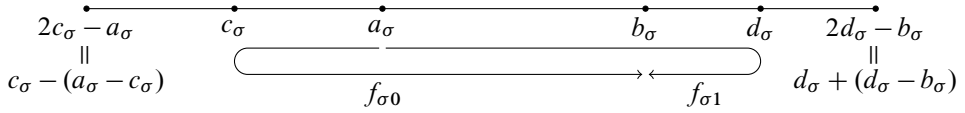

Figure 7

Here are the formal definitions of $f_{\sigma 0}$, $f_{\sigma 1}$, $c_{\sigma 0}$, $d_{\sigma 0}$, $c_{\sigma 1}$ and $d_{\sigma 1}$:

$$f_{\sigma 0}\colon [2c_\sigma - a_\sigma, b_\sigma] \to [0,1], \qquad f_{\sigma 0}(x) := \begin{cases} f_\sigma(-x + 2c_\sigma) & \text{if } 2c_\sigma - a_\sigma \le x \le c_\sigma, \\ f_\sigma(x) & \text{if } c_\sigma \le x \le b_\sigma, \end{cases}$$
$$[c_{\sigma 0}, d_{\sigma 0}] := [2c_\sigma - a_\sigma, b_\sigma],$$
$$f_{\sigma 1}\colon [a_\sigma, 2d_\sigma - b_\sigma] \to [0,1], \qquad f_{\sigma 1}(x) := \begin{cases} f_\sigma(x) & \text{if } a_\sigma \le x \le d_\sigma, \\ f_\sigma(-x + 2d_\sigma) & \text{if } d_\sigma \le x \le 2d_\sigma - b_\sigma, \end{cases}$$
$$[c_{\sigma 1}, d_{\sigma 1}] := [a_\sigma, 2d_\sigma - b_\sigma].$$

□

Notice that each f_σ defined above is piecewise linear with each piece having slope ± 1. There are many interesting choices for the tree of clasp points $\{a_\sigma, b_\sigma\}$ and, hence, the tree of functions f_σ associated with potential shrinks of Bing's decomposition.

Definition 3.2 (1) A binary tree of functions $\{f_\sigma\}$ starting with $f_\varnothing$ being the identity on $[0,1]$ and defined as above will be called a *Bing tree of functions*.

(2) A Bing tree of functions $\{f_\sigma\}$ *shrinks* if and only if $\operatorname{length}\big(f_\sigma([c_\sigma, d_\sigma])\big) \to 0$ wherever the string length $|\sigma| \to \infty$.

(3) The values $|a_\sigma - c_\sigma|$ and $|d_\sigma - b_\sigma|$ are *displacements*. □

A Bing tree of functions that shrinks indicates a method of shrinking the Bing decomposition; however, such a tree of functions is just a model because the clasping positions of daughter tori cannot occur at single points. So the values a_σ and b_σ just indicate where the clasps occur, although in reality both $T_{\sigma 0}$ and $T_{\sigma 1}$ will be slightly longer than the domains $[c_{\sigma 0}, d_{\sigma 0}]$ and $[c_{\sigma 1}, d_{\sigma 1}]$ of $f_{\sigma 0}$ and $f_{\sigma 1}$, respectively. Nevertheless, a Bing tree of functions that shrinks can guide a shrink of the Bing decomposition simply by making the clasps sufficiently tight.

Previously known shrinks of the Bing decomposition included some large displacements, that is, instances where $|a_\sigma - c_\sigma|$ and $|d_\sigma - b_\sigma|$ were relatively large compared to $|d_\sigma - c_\sigma|$. However, we show in this paper that it is possible to construct a Bing shrink — or, actually, a Bing tree of functions $\{f_\sigma\}$ that shrinks — even though all the displacements are small. The shrinks we will produce have the additional property that the length of each torus, or equivalently the domain of each f_σ, grows by less than any desired quantity.

Before constructing our small displacement shrink, let's make some observations about the functions f_σ in a Bing tree of functions. When visualizing the following, it might be useful to imagine the displacements — that is, the $|a_\sigma - c_\sigma|$'s and $|d_\sigma - b_\sigma|$'s — as very small compared to the length of $|d_\sigma - c_\sigma|$.

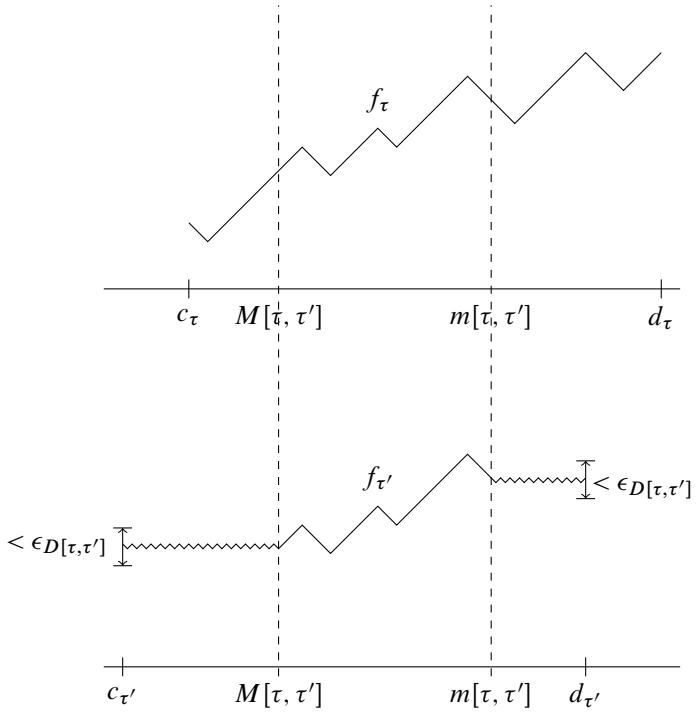

Figure 8: Note that $c_{\tau'}$ may be greater than, or less than, or equal to c_τ, and similarly for $d_{\tau'}$. The salient observation is that f_τ and $f_{\tau'}$ are identical on the interval $[M[\tau,\tau'], m[\tau,\tau']]$.

Consider the daughter $f_{\sigma 0}$ of f_σ. The domain of f_σ, namely $[c_\sigma, d_\sigma]$, shifts downward to create the domain of $f_{\sigma 0}$, namely $[c_\sigma - (a_\sigma - c_\sigma), b_\sigma] = [2c_\sigma - a_\sigma, b_\sigma]$. Let's think about the relationship between the images of f_σ and $f_{\sigma 0}$.

First notice that the domains of f_σ and $f_{\sigma 0}$ share the interval $[c_\sigma, b_\sigma]$, and by definition agree there. The function $f_{\sigma 0}$ is defined on the additional interval $[2c_\sigma - a_\sigma, c_\sigma]$, but $f_{\sigma 0}([2c_\sigma - a_\sigma, c_\sigma]) = f_\sigma([c_\sigma, a_\sigma])$ as sets, so no new points are added to the image of $f_{\sigma 0}$ compared to the image of f_σ; that is, $f_{\sigma 0}([c_{\sigma 0}, d_{\sigma 0}]) \subset f_\sigma([c_{\sigma,d_\sigma}])$. Similarly, the domain of $f_{\sigma 1}$ shifts up compared to the domain of f_σ, namely $[c_{\sigma 1}, d_{\sigma 1}] = [a_\sigma, d_\sigma + (d_\sigma - b_\sigma)] = [a_\sigma, 2d_\sigma - b_\sigma]$. These observations allow us to record some relationships between an f_τ and its descendants in the lemma and corollary below.

Definition 3.3 Let τ' be a binary string descendant of τ (that is, τ is the beginning of the string τ') and let $\{f_\sigma\}$ be a Bing tree of functions. Then $M[\tau,\tau'] := \max\{a_\mu \mid \tau \le \mu \le \tau'\}$, $m[\tau,\tau'] := \min\{b_\mu \mid \tau \le \mu \le \tau'\}$, and $\epsilon_{D[\tau,\tau']} := \max\{|a_\mu - c_\mu|, |d_\mu - b_\mu| \mid \tau \le \mu \le \tau'\}$. □

Lemma 3.4 *Let $\{f_\sigma\}$ be a Bing tree of functions and let τ' be a descendant of the binary string τ. Then*:

(1) *$f_{\tau'}$ restricted to $[M[\tau,\tau'], m[\tau,\tau']]$ equals f_τ restricted to $[M[\tau,\tau'], m[\tau,\tau']]$.*

(2) $\text{length}\big(f_{\tau'}([c_{\tau'}, M[\tau,\tau']])\big) \le \epsilon_{D[\tau,\tau']}$.

(3) $\text{length}(f_{\tau'}[m_{[\tau,\sigma]}, d_{\tau'}]) \le \epsilon_{D[\tau,\tau']}$.

Corollary 3.5 *Let $\{f_\sigma\}$ be a Bing tree of functions and let τ' be a descendant of the binary string τ. Then*

$$\operatorname{length}\big(f_{\tau'}([c_{\tau'}, d_{\tau'}])\big) \le |m[\tau, \tau'] - M[\tau, \tau']| + 2\epsilon_{D[\tau,\tau']}$$

We are now ready to construct a shrink that uses only small displacements.

Theorem 3.6 *For any $\epsilon_L > 0$ and $\{\epsilon_i \ge 0 \mid \sum_{i=0}^{\infty} \epsilon_i^2 = \infty\}$, there exists a Bing tree of functions $\{f_\sigma\}$ that shrinks such that, for every σ,*

(1) $|d_\sigma - c_\sigma| < 1 + \epsilon_L$, *and*

(2) *for $|\sigma| = i$, $|a_\sigma - c_\sigma| \le \epsilon_i$ and $|d_\sigma - b_\sigma| \le \epsilon_i$.*

Proof We will describe the functions f_σ by describing the domain intervals $[c_\sigma, d_\sigma]$. Notice that knowing the domain intervals of f_σ and of $f_{\sigma 0}$ automatically implies what a_σ and b_σ are and what $f_{\sigma 1}$ is. Specifically, if the domain of f_σ is $[c_\sigma, d_\sigma]$ and the domain of $f_{\sigma 0}$ is $[c_{\sigma 0}, d_{\sigma 0}]$, then $a_\sigma = c_\sigma + (c_\sigma - c_{\sigma 0}) = 2c_\sigma - c_{\sigma 0}$ and $b_\sigma = d_{\sigma 0}$. So the domain and definition of $f_{\sigma 1}$ are also determined.

It will be convenient to associate each domain interval $[c_\sigma, d_\sigma]$ with the point in the plane (c_σ, d_σ). Notice the following:

(1) Since during our construction c_σ will always be less than d_σ, each point (c_σ, d_σ) lies above the main diagonal Δ; specifically, Δ is the graph of $y = x$.

(2) The horizontal (or vertical) distance from Δ to (c_σ, d_σ) equals the length of $[c_\sigma, d_\sigma]$.

(3) Interval inclusion $[u, v] \subset [c_\sigma, d_\sigma]$ corresponds to the point (c_σ, d_σ) lying in the NW quadrant with respect to the point (u, v).

(4) The vector from the point in the plane (c_σ, d_σ) to $(c_{\sigma 0}, d_{\sigma 0})$, namely $(c_\sigma - a_\sigma, b_\sigma - d_\sigma)$, is the negative of the vector from (c_σ, d_σ) to $(c_{\sigma 1}, d_{\sigma 1})$.

These observations allow us to construct the desired Bing tree of functions $\{f_\sigma\}$ that shrinks by describing an associated binary tree of points in the plane (c_σ, d_σ) with $\operatorname{length}([c_\sigma, d_\sigma]) < 1 + \epsilon_L$ and $\operatorname{length}\big(f_\sigma([c_\sigma, d_\sigma])\big)$ eventually getting short.

Definition 3.7 Let τ be a finite binary string. Then a finite set of finite binary strings $\{\tau_i\}_{i=1,\dots,n}$ is a *complete* set of descendants of τ if and only if

(1) for each i, τ_i is a descendant of τ, and

(2) for every infinite binary string $\hat{\tau}$ that starts with τ, there exists exactly one τ_i such that $\hat{\tau}$ starts with τ_i.

Similarly, we will refer to finite, complete sets of descendants of f_τ and (c_τ, d_τ) if their subscripts are a finite, complete set of descendants of τ. □

The proof of the following lemma contains the heart of the construction.

Lemma 3.8 *Let $\epsilon_L > 0$, $\{\epsilon_i \geq 0 \mid \sum_{i=0}^{\infty} \epsilon_i^2 = \infty\}$, $\delta > 0$, and τ' be a finite binary string descendant of τ. Suppose the beginning of a Bing tree of functions has been defined starting at $f_\varnothing$ and including all descendants of $f_\varnothing$ that are ancestors of $f_{\tau'}$. Suppose*:

(1) $|d_{\tau'} - c_{\tau'}| < 1 + \epsilon_L$.

(2) $\text{length}\big(f_{\tau'}([c_{\tau'}, M[\tau, \tau']])\big) < \delta$.

(3) $\text{length}(f_{\tau'}([m[\tau, \tau'], d_{\tau'})) < \delta$.

(4) $|m[\tau, \tau'] - M[\tau, \tau']| \geq 3\delta$.

Then there exists a finite complete set A of descendants of τ' such that, for each $f_{\tau''}$ in A,

(1) $\max\{|a_\mu - c_\mu|, |d_\mu - b_\mu|\} \leq \min\{\epsilon_{|\mu|}, \delta\}$ *for each μ with $\tau' \leq \mu \leq \tau''$, and*

(2) $|m[\tau, \tau''] - M[\tau, \tau'']| \leq \frac{2}{3}|m[\tau, \tau'] - M[\tau, \tau']|$.

Corollary 3.9 *Given the hypotheses of Lemma 3.8, there exists a finite complete set A of descendants of τ' such that, for each $f_{\tau''} \in A$,*

$$\text{length}\big(f_{\tau''}([c_{\tau''}, d_{\tau''}])\big) \leq \tfrac{2}{3}|m[\tau, \tau'] - M[\tau, \tau']| + 2\delta.$$

Proof of Lemma 3.8 We will describe descendants of $f_{\tau'}$ by looking at points in the plane. The interval $[M[\tau, \tau'], m[\tau, \tau']]$ is contained in $[c_{\tau'}, d_{\tau'}]$. And, of course, the middle third of $[M[\tau, \tau'], m[\tau, \tau']]$, namely $I := \big[M[\tau, \tau'] + \frac{1}{3}(m[\tau, \tau'] - M[\tau, \tau']), M[\tau, \tau'] + \frac{2}{3}(m[\tau, \tau'] - M[\tau, \tau'])\big]$, is also contained in $[c_{\tau'}, d_{\tau'}]$. So the point $(c_{\tau'}, d_{\tau'})$ lies in the NW quadrant above I. Call that NW quadrant Q. Figure 9 shows that relationship.

Figure 9 also suggests a large number of concentric circles centered at some distant point p in that NW quadrant Q on the slope -1 ray from I heading up and left. Recall that $|d_{\tau'} - c_{\tau'}| < 1 + \epsilon_L$. Let $C_{\tau'}$ be the circle centered at p and containing the point $(c_{\tau'}, d_{\tau'})$. The point p is so distant that every point $(x, y) \in C_{\tau'} \cap Q$ has $y - x < 1 + \epsilon_L$. This choice of circle center will guarantee that no point (c_μ, d_μ) that is created during our process has corresponding interval $[c_\mu, d_\mu]$ with length greater than $1 + \epsilon_L$.

Now we are ready to construct our binary tree of points (c_μ, d_μ). We begin at $(c_{\tau'}, d_{\tau'})$ and proceed along the tangent of the circle $C_{\tau'}$ in both directions until we hit the next circle. We choose the size of the next circle so that the distance along the tangent of $C_{\tau'}$ to the next concentric circle equals $\min\{\epsilon_{|\tau'|}, \delta\}$. Those two points will be $(c_{\tau'0}, d_{\tau'0})$ and $(c_{\tau'1}, d_{\tau'1})$. From each of those points we do the same thing — that is, from $(c_{\tau'0}, d_{\tau'0})$ we move along the tangent of its circle $C_{\tau'0}$ to find the points $(c_{\tau'00}, d_{\tau'00})$ and $(c_{\tau'01}, d_{\tau'01})$ on the next circle $C_{\tau'00}$. At each step, when going from (c_μ, d_μ) to points $(c_{\mu0}, d_{\mu0})$ and $(c_{\mu1}, d_{\mu1})$, we choose the size of circle $C_{\mu0}$ so that the distance along the tangent of circle C_μ to a point on $C_{\mu0}$ equals $\min\{\epsilon_{|\mu|}, \delta\}$.

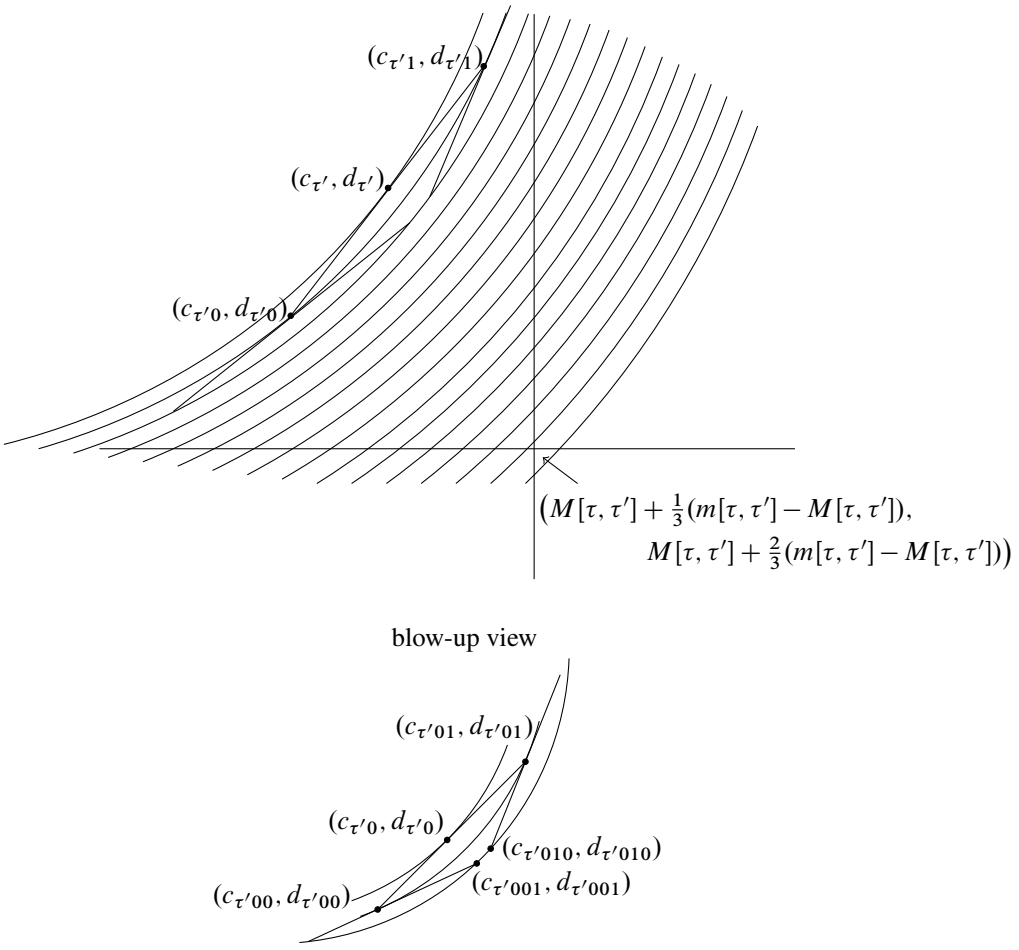

Figure 9

We continue creating this binary tree of points until we arrive at a point $(c_{\tau''}, d_{\tau''})$ where either

(1) $c_{\tau''} > M[\tau,\tau'] + \frac{1}{3}(m[\tau,\tau'] - M[\tau,\tau'])$, or

(2) $d_{\tau''} < M[\tau,\tau'] + \frac{2}{3}(m[\tau,\tau'] - M[\tau,\tau'])$,

that is, $(c_{\tau''}, d_{\tau''}) \notin Q$.

The following lemma implies that every branch of descendants does eventually leave Q.

Lemma 3.10 *Let $\{\epsilon_i \geq 0 \mid \sum_{i=0}^{\infty} \epsilon_i^2 = \infty\}$, $\delta > 0$, and let $\{C_i\}_{i=1}^{\infty}$ be a nested sequence of concentric circles in $\mathbb{R}^2$ centered at point p such that, for every i, the radius of C_i is r_i and the distance from a point on C_i along the tangent to C_i to a point on C_{i+1} is $\min\{\epsilon_i, \delta\}$. Then the sequence of radii $(r_i)_{i=1}^{\infty}$ is unbounded.*

Proof Let $M \in \mathbb{R}^+$. Given the hypotheses, we will show that there is a k such that $r_k \geq M$. If not, then, for every i, $r_i < M$ (see Figure 10).

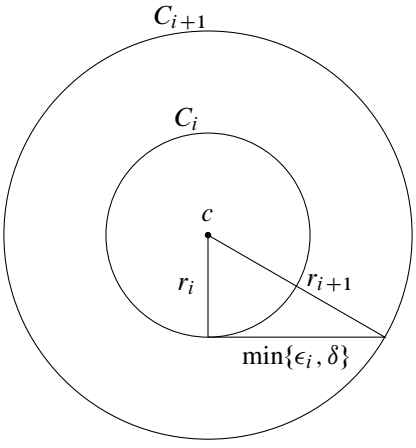

Figure 10

By the Pythagorean theorem,

$$r_{i+1}^2 - r_i^2 = \epsilon_i^2, \text{ so } r_{i+1} - r_i = \frac{\min\{\epsilon_i, \delta\}^2}{r_{i+1} + r_i} > \frac{\min\{\epsilon_i, \delta\}^2}{2M}.$$

But $\sum_{i=0}^{\infty} \min\{\epsilon_i, \delta\}^2/2M$ diverges, meaning some $r_k > M$. □

Every path branch of the binary tree of points that we construct stops by a predictable stage since our hypothesis about the divergence of the sum of the squares of the ϵ_i's guarantees that there are only a finite number of circles before the point $\big(M[\tau,\tau'] + \frac{1}{3}(m[\tau,\tau'] - M[\tau,\tau']), M[\tau,\tau'] + \frac{2}{3}(m[\tau,\tau'] - M[\tau,\tau'])\big)$ is itself inside a circle C_μ. Any point constructed with our procedure on such a circle C_μ must be outside Q. When we have arrived at a point $(c_{\tau''}, d_{\tau''})$ that is not in Q, let us define $a_{\tau''} := c_{\tau''}$ and $b_{\tau''} := d_{\tau''}$. The function $f_{\tau''}$ defined using these values $c_{\tau''}$, $a_{\tau''}$, $b_{\tau''}$ and $d_{\tau''}$ then satisfies the conditions for being an element of the finite, complete set of descendants of $f_{\tau'}$ that we seek, thereby concluding the proof of Lemma 3.8. □

Note Our proof of Lemma 3.8 has actually proved a stronger result. Since we always stop at the first time that a $(c_{\tau''}, d_{\tau''})$ is not in Q — that is, either $c_{\tau''} > M[\tau,\tau'] + \frac{1}{3}(m[\tau,\tau'] - M[\tau,\tau'])$ or $d_{\tau''} < M[\tau,\tau'] + \frac{2}{3}(m[\tau,\tau'] - M[\tau,\tau'])$ — then, in the former case, $c_{\tau''} = M[\tau,\tau'']$, while, in the latter case, $d_{\tau''} = m[\tau,\tau'']$. So Corollary 3.9 could actually be improved by replacing the 2δ with δ. However, this subtlety is not necessary when applying Lemma 3.8 and Corollary 3.9 to prove Theorem 3.6.

We can now conclude the proof of how to construct the desired Bing tree of functions that shrinks as follows. We begin with $f_\varnothing\colon [0,1] = [c_\varnothing, d_\varnothing] \to [0,1]$ being the identity and $a_\varnothing := 0$ and $b_\varnothing := 1$.

We apply Lemma 3.8, starting with $\tau = \varnothing$ and $\tau' = \varnothing$, and with $\delta = \frac{1}{10}$ repeatedly until we have created a finite, complete set of descendants $A_\varnothing$ where the small displacement conditions are satisfied at each stage and such that, for each $\tau \in A_\varnothing$, $f_\tau\colon [c_\tau, d_\tau] \to [0,1]$ has the property that $\text{length}\big(f_\tau([c_\tau, d_\tau])\big) < 3\delta + 2\delta = \frac{5}{10} = \frac{1}{2}$.

We then start with each such f_τ and choose $a_\tau := c_\tau$ and $b_\tau := d_\tau$, and use it and itself as its descendants, with $\delta = \frac{1}{10^2}$ as the hypotheses of Lemma 3.8. We use Lemma 3.8 repeatedly to create a finite, complete set A_τ of descendants of f_τ such that the construction respects the small displacement and nonlengthening conditions and, for each $\sigma \in A_\tau$, $\operatorname{length}\big(f_\sigma([c_\sigma, d_\sigma])\big) < \frac{3}{10^2} + \frac{2}{10^2} = \frac{5}{10^2}$.

Continuing in this fashion produces the desired Bing tree of functions, that shrinks while respecting the small displacement conditions, thus proving Theorem 3.6. □

4 Conclusion and questions

Bing's 1952 construction of a wild involution of S^3 opened the door to many further insights — and many further questions. In [11], we proved that all conjugates of Bing's involution *must* share certain analytic features with the involution derived from Bing's original shrink. Understanding such shared features turned up a surprise, the subject of this paper, whose exact relation to the previous paper is still to be worked out. Bing always told us not to form fixed beliefs about what you do not know. Following his advice, let us state some questions without presuming to guess their answers.

Question The first question involves the rate at which small displacements can shrink. Bing's original method of shrinking and his shrinking without lengthening method shrink stage tori to size about $\frac{1}{n}$ their original diameter at stage order n. The analysis of [1] shows that there is no faster way to shrink the Bing decomposition $\mathcal{D}$. It appears that if one imposes constraints on displacement and lengthening, as we have here, shrinking must be even slower. For example, if lengthening and displacement are restricted to 0.1%, our algorithm takes about 10^{12} stages to get from diameter $= 1$ to diameter $= 0.001$. What is the actual functional form for shrinking using our algorithm? And are there more efficient algorithms respecting the same constraints? In both cases, what are the analytical properties, the modulus of continuity (moc), of the corresponding involutions?

Question Suppose we use the 1D model for shrinking as in this paper. We showed that it is possible to shrink the Bing decomposition with small displacement shrinks. Bing's shrinks and the shrink in this paper seem to require some insight or even cleverness, but might that apparent cleverness be an illusion? Suppose the a_σ's and b_σ's were simply chosen randomly in the intervals $[c_\sigma, d_\sigma]$. Would such a random selection lead to a shrink of the Bing decomposition with probability 1? If so, wouldn't we feel silly. Shrinking (or not) is a *tail event*, meaning independent of any initial segment of choices, so Kolmogorov's 0-1 law tells us that, within a probabilistic model, shrinking will occur with probability either 0 or 1, which depends on the model. If the model is artificially concentrated near our explicit shrink, the probability will be 1, but if $\{(a_\sigma, b_\sigma)\}$ are independent and uniformly distributed, we do not know.

We were slow to accept that Bing's decomposition could be shrunk using only tiny jiggles. Bing understood that the unknown is actually *unknown*. He told us that he would alternately work toward

proving a conjecture was true and toward producing a counterexample with the idea that both attempts would help to expose the truth, whatever that truth turns out to be.

Acknowledgement

We would like to thank the referee for many helpful suggestions that have greatly improved the exposition of the construction.

References

[1] **F D Ancel**, **M P Starbird**, *The shrinkability of Bing–Whitehead decompositions*, Topology 28 (1989) 291–304 MR

[2] **J J Andrews**, **L Rubin**, *Some spaces whose product with E^1 is E^4*, Bull. Amer. Math. Soc. 71 (1965) 675–677 MR

[3] **R H Bing**, *A homeomorphism between the 3-sphere and the sum of two solid horned spheres*, Ann. of Math. 56 (1952) 354–362 MR

[4] **R H Bing**, *Shrinking without lengthening*, Topology 27 (1988) 487–493 MR

[5] **J Bryant**, **S Ferry**, **W Mio**, **S Weinberger**, *Topology of homology manifolds*, Ann. of Math. 143 (1996) 435–467 MR

[6] **J W Cannon**, $\Sigma^2 H^3 = S^5/G$, Rocky Mountain J. Math. 8 (1978) 527–532 MR

[7] **R J Daverman**, *Decompositions of manifolds*, Pure and Applied Math. 124, Academic, Orlando, FL (1986) MR Zbl

[8] **A N Dranishnikov**, **E V Shchepin**, *Cell-like mappings: the problem of the increase of dimension*, Uspekhi Mat. Nauk 41 (1986) 49–90 MR In Russian; translated in Russ. Math. Surv. 41 (1986) 59–111

[9] **R D Edwards**, *The topology of manifolds and cell-like maps*, from "Proceedings of the International Congress of Mathematicians" (O Lehto, editor), Acad. Sci. Fennica, Helsinki (1980) 111–127 MR

[10] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geometry 17 (1982) 357–453 MR

[11] **M Freedman**, **M Starbird**, *The geometry of the Bing involution*, preprint (2022) arXiv 2209.07597

[12] **F Quinn**, *Ends of maps, II*, Invent. Math. 68 (1982) 353–424 MR

MF: *Station Q, Microsoft*
Santa Barbara, CA, United States

Department of Mathematics, University of California, Santa Barbara
Santa Barbara, CA, United States

MS: *Department of Mathematics, The University of Texas at Austin*
Austin, TX, United States

mikehartleyfreedman@outlook.com, starbird@math.utexas.edu

Received: 24 January 2023 Revised: 4 October 2023

ALGEBRAIC & GEOMETRIC TOPOLOGY

msp.org/agt

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

The subscription price for 2025 is US $760/year for the electronic version, and $1110/year (+$75, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Algebraic & Geometric Topology (ISSN 1472-2747 printed, 1472-2739 electronic) is published 9 times per year and continuously online, by Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840. Periodical rate postage paid at Oakland, CA 94615-9651, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 25 Issue 4 (pages 1917–2526) 2025