ISSN 1364-0380 407

Geometry & Topology
Volume 4 (2000) 407–430
Published: 10 November 2000

Subexponential groups in 4–manifold topology

Vyacheslav S Krushkal
Frank Quinn

Department of Mathematics, Yale University
New Haven, CT 06520-8283, USA
and
Department of Mathematics, Virginia Tech
Blacksburg, VA 24061-0123, USA

Email: krushkal@math.yale.edu, quinn@math.vt.edu

Abstract

We present a new, more elementary proof of the Freedman–Teichner result that the geometric classification techniques (surgery, s–cobordism, and pseudoisotopy) hold for topological 4–manifolds with groups of subexponential growth. In an appendix Freedman and Teichner give a correction to their original proof, and reformulate the growth estimates in terms of coarse geometry.

AMS Classification numbers Primary: 57N13

Secondary: 57N37, 57N70, 57R65

Keywords: 4–manifolds, groups of subexponential growth, gropes

Proposed: Robion Kirby Received: 22 May 2000
Seconded: Wolfgang Metzler, Cameron Gordon Accepted: 3 November 2000

The disk embedding theorem for 4–manifolds with "good" fundamental group is the key ingredient of the classification theory: it is used in the proof of the 4–dimensional surgery theorem, and the 5–dimensional s–cobordism theorem and pseudoisotopy theorems. The homotopy hypotheses of the theorem always allow one to find a 2–stage immersed capped grope. If one can find such a grope so that loops in the image are nullhomotopic in the ambient manifold, then Freedman's theorem [1, 3] shows there is a topologically flat embedded disk. The current focus, therefore, is on obtaining this π_1–nullity condition. Freedman [2] showed this is possible if the fundamental group of the manifold is poly-(finite or cyclic). This was extended to groups of polynomial growth in [7]. The current best result is for groups of subexponential growth.

The disk theorem for subexponential groups was stated by Freedman and Teichner [4]. However the "key point" of [4] page 521, line 17, is incorrect. In the Appendix Freedman and Teichner show how to modify their construction to correct this. The present paper sidesteps the issue by using a different and more elementary construction developed by the first author (see [5]). It displays particularly clearly how the proof fails in the general (exponential growth) case, and suggests that an infinite construction may be necessary to make further progress.

The following result is the input needed for the disk embedding theorem for manifolds with subexponential fundamental groups. For a full statement of the disk embedding theorem and applications to surgery and s–cobordism, see [4]. For the application to pseudoisotopy see [6].

Theorem *Suppose $G \to M^4$ is a properly immersed (capped) grope of height ≥ 2, and ρ: $\pi_1 M \to \pi$ is a homomorphism with π of subexponential growth. Then the total contraction of G is regularly homotopic rel boundary to an immersion whose double point loops have trivial image in π.*

This slightly extends the usual immersion-improvement formulation in that we do not require the total contraction to be a disk, and the output immersion is regularly homotopic to the input. Neither extension has new consequences, but they come for free in the proof and they simplify applications. Capped gropes, contractions, and subexponential growth are all reviewed in the text.

In rough outline the proof goes as follows: The images of the double point loops of G give a finite subset of π. Subexponential growth implies that in a large collection of words of fixed length in the finite subset, somewhere there is a subword whose product is trivial. We organize the data so this subword can be realized geometrically as double point loops, by pushing intersections around in the grope. This eliminates a branch in the grope in exchange for π–trivial self-intersections of the base surface. Iterating this eliminates all branches (ie, gives a contraction) with π–trivial self-intersections. The key technique is splitting to dyadic branches.

Acknowledgements VK was supported at the Institute for Advanced Study by NSF grant DMS 97-29992. FQ is partially supported by NSF grant DMS 97-05168 and Harvard University.

1 Definitions

We briefly review the definition of gropes in order to fix terminology for caps, duals, immersions, contractions, etc.

1.1 Gropes Begin with a surface S. A *model grope* built from S, or an S–*like grope* is a 2–complex in $S \times I$ obtained by repeatedly replacing embedded disks by punctured tori with disks attached, see [3, Section 2.1]. The attached disks are the "caps" for the torus; there are two of them, they intersect in a point, and each is referred to as the *dual* of the other.

(1) The *caps* of the final grope are caps introduced at some stage, and which have not been modified later.

(2) A *subgrope* is a cap that has been modified, together with these modifications. More generally a subgrope is a disjoint union of these. Subgropes are disk-like gropes, or more generally (union of disk)-like gropes.

(3) *Dual* subgropes are subgropes obtained from caps that were dual at some stage of the construction. They are attached to the same lower surface, and the attaching circles intersect in a single point.

(4) The *base* of the grope is the surface obtained by modifying S.

(5) A *branch* is a dual pair of subgropes attached to the base.

(6) The grope has *height* $\geq k$ if any path from a cap to the base passes through at least k boundary curves of subgropes. In other words there are at least k levels of surfaces (counting the base) below each cap. Note each branch has height $\geq k-1$.

1.2 Contractions Suppose H_+, H_- are dual subgropes in a grope. The *contraction away from* H_-, or equivalently the *contraction across* H_+ is defined as follows: discard the subgrope H_-, cut open the surface to which H_+ is attached along the attaching circle, and glue in two parallel copies of H_+. See Figure 1. This process is sometimes referred to as surgery. The result is an $S-$like grope in a canonical way, ie, there is a canonical way to see this as obtained from S in $S \times I$.

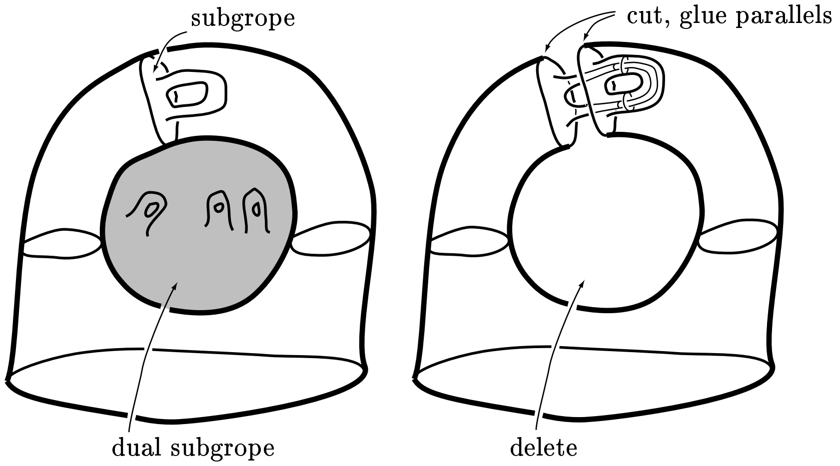

Figure 1: Contraction

There is also a "symmetric contraction" which cuts the surface along the attaching curves of both subgropes and glues in parallel copies of both, together with a small square near the intersection point of the curves. See [3, Section 2.3]. This operation won't be used here.

Lemma *Suppose H_+, H_- are dual caps in an S–like grope. Then the contractions across H_+, across H_-, and the symmetric contraction, give isotopic gropes.*

Proof Here "isotopic" means ambient isotopic in $S \times I$. Since H_+, H_- are both caps, all these contractions undo one disk-punctured torus replacement. The isotopy is clear in pictures; see [3, Section 2.3]. ◻

A *total contraction* of a grope is obtained by repeatedly contracting (in any order) until no caps are left. Iterating the lemma shows this always returns the original surface:

Corollary *All total contractions of S–like gropes are surfaces isotopic (rel boundary, in $S \times I$) to S.* □

When a grope is immersed the different contractions give regularly homotopic immersions of S. Their usefulness is that they may differ greatly in their intersection patterns.

1.3 Immersions A *proper immersion* of a grope $G \to M^4$ is an immersion of a regular neighborhood of the spine in $S \times I$ satisfying

(1) boundary goes to boundary, so the image intersects ∂M in $\partial S \times I$;

(2) all intersections come from transverse intersections among caps; and

(3) (if the base surface is noncompact) the immersion is proper in the topological sense.

Caps have I–bundle neighborhoods in $S \times I$. Condition (2) means these neighborhoods intersect in squares determined by transversality and the bundle structures, and that there are no other intersections.

In fact a regular neighborhood of the grope spine is isotopic to $S \times I$ itself. The definition is given in terms of regular neighborhoods so the standard D^2 bundle neighborhoods of caps will be easier to see. A consequence of this neighborhood uniqueness is that the original core copy of S is isotopic into any regular neighborhood of the grope spine. Composing this with a grope immersion gives an immersion of S, well-defined up to regular homotopy. As remarked above, total contractions give explicit descriptions of such immersions.

1.4 Transverse spheres and gropes A *transverse sphere* for a surface in a 4–manifold is a framed immersed sphere that intersects the surface in a single point, see [3, Section 1.9]. Similarly a transverse grope is an immersed sphere-like grope that intersects the surface in one point, and this point is in the base of the grope. Note that totally contracting a transverse grope gives a transverse sphere.

In a grope, every surface component not part of the base has a standard transverse grope. This is constructed in $S \times D^2$ (S is the original surface, or the total contraction), so there is a copy inside any neighborhood of the spine. This was a key part of Freedman's original constructions of convergent gropes. The *tight* transverse grope for a surface component A is obtained as follows. Let B be the next surface down, so A is attached to B along a circle. Let H be the subgrope dual to A. The base of the transverse grope is the torus obtained as the normal S^1 bundle to B, restricted to a parallel of the attaching curve of H. To this we add a parallel copy of H itself, and a D^2 fiber of the normal disk bundle of B. The result is a sphere-like grope, see Figure 2. Connected sum with this grope gives the "lollipop" move of [4].

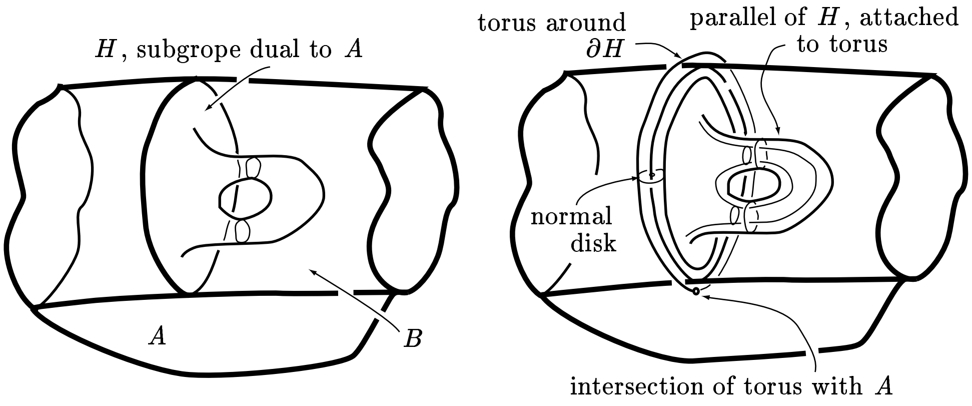

Figure 2: Tight transverse grope

The tight transverse grope has two points of intersection with the grope: one is the (desirable) intersection between the base torus and A. The other is the intersection of the normal-disk cap with B. Usually we avoid this one by contracting away from this cap. (See 1.2. Recall that this discards the normal-disk cap, cuts the torus along the boundary of the parallel of H, and fills in with two copies of H.) This contraction is the "loose" or standard transverse grope. It is described in [3, Section 2.6], and sums with it give the "double lollipop" move of [4].

Finally, transverse spheres for A are obtained by totally contracting the loose transverse grope.

We repeat here the warning of [3, Section 2.6]. If H_+, H_- are dual subgropes then the construction gives transverse gropes for the bottom-level surfaces in both. However these two gropes intersect each other, which in practice means they cannot be used simultaneously. (This is a mistake in [4].) This problem can be partially avoided by using them sequentially, see [3, Section 2.7].

Lemma *The spheres obtained by totally contracting loose transverse gropes bound embedded 3–disks in $S \times D^2$. Thus connected sum with these spheres changes a surface by regular homotopy.*

Proof To see this begin with the tight grope and totally contract the copy of H. This gives a torus with two caps. The torus bounds a solid torus (the D^2 bundle over the attaching circle of H) disjoint from the H cap. Attach to this solid torus a $D^2 \times I$ thickening of the H cap. The result is a 3–disk whose boundary is the transverse sphere. □

2 Dyadic branches and splitting

A disk-like grope is *dyadic* if all component surfaces are either disks or punctured tori. This means each non-cap surface has exactly one pair of dual subgropes attached to it. A grope has *dyadic branches* if all of the subgropes above the base level are dyadic. The benefits of dyadic branches are especially simple contractions and simpler tracking of their interactions. In 2.3 we describe the splitting construction, which converts any grope into one with dyadic branches. Later in the section it is used further to set up a (dyadic) situation in which product information in π can be exploited.

2.1 Contraction of dyadic gropes When a grope is contracted caps are discarded. In a total contraction most get discarded. Dyadic gropes are special in that there are total contractions in which *all but one* of the caps are discarded. Since intersections occur among caps, discarding them simplifies intersection data. Discarding all but one gives the greatest possible simplification of the data.

To be explicit, suppose C is a cap in a dyadic branch. The *total contraction across* C is defined as follows: C lies in one of the dual subgropes in the branch; form the contraction across this subgrope. The result has two dyadic branches parallel to the subgrope, so each contains a parallel copy of C. Repeat this construction, contracting each new branch across the subgrope containing the copy of C. Each iteration doubles the number of copies of C, and reduces their height above the base by 1. If the original height of C is k, then total contraction eliminates the branch and subsitutes 2^k parallel copies of C in the base surface.

2.2 Pushing down and back up Suppose C_1, C_2 are distinct caps in a dyadic branch in a grope. We construct a transverse sphere for C_1 that contains parallel copies of C_2, and otherwise lies in a small neighborhood of the body of the grope. Suppose a surface W intersects C_1. Then adding copies of this transverse sphere to remove the intersection points will be called *pushing* W *down off* C_1 *and back up across* C_2. Reasons for the terminology are:

(1) "pushing" because the new surface is regularly homotopic to W;

(2) "off C_1" because intersections with C_1 have been removed; and

(3) "across C_2" because the modified surface contains parallel copies of C_2, and will therefore intersect anything C_2 intersects.

We now construct the transverse sphere. Let H_1 and H_2 be dual subgropes in the branch, so that H_1 contains C_1 and H_2 contains C_2. Let $\hat{T}$ be the tight transverse grope for the bottom surface of H_1. This is a dyadic sphere-like grope with a cap parallel to C_2; let T be the sphere obtained by totally contracting $\hat{T}$ across this cap. T contains 2^j parallels of C_2, where j is the height of C_2 in H_2. If $H_1 = C_1$ then this is the desired sphere. Otherwise take a 2–sphere fiber of the normal sphere bundle of the attaching circle of C_1. This 2–sphere intersects C_1 in one point and intersects the surface to which it is attached in two points. Push these latter points down in the grope to the bottom surface of H_1. If C_1 has height k in H_1 then there will be 2^k intersection points there. Remove all these by connected sum with parallels of the sphere T. This gives the desired transverse sphere for C_1. Note that all together this sphere has 2^{j+k} copies of C_2.

This transverse sphere bounds a 3–disk in the model. To see this, first note that T does by Lemma 1.4. The fiber of a 2–sphere bundle bounds the fiber of the associated 3–disk bundle, and the isotopy pushing the boundary down in H_1 pushes the 3–disk too. Finally the boundary sum of a bunch of 3–disks gives a 3–disk whose boundary is the connected sum of spheres. It is easy to see that the 3–disk obtained this way is embedded.

Since the sphere bounds an embedded 3–disk, connected sum with the sphere changes surfaces by regular homotopy.

2.3 Splitting gropes The splitting operation splits a surface into two pieces, at the cost of doubling the dual subgrope. It can be used to decompose branches into dyadic branches, and can separate intersection points distinguished by properties unaffected by the dual doubling.

Suppose A is a component surface of a grope, not part of the base. Let B be the surface it is attached to, and H the dual subgrope. Now suppose α is an embedded arc on A, with endpoints on the boundary, and disjoint from attaching circles of higher stages. In the 3–dimensional model, sum B with itself by a tube about α (the normal S^1 bundle), and discard the part of A that lies inside the tube. This splits A into two components. H is a dual for one component; obtain a dual for the other by taking a parallel copy of H. See Figure 3.

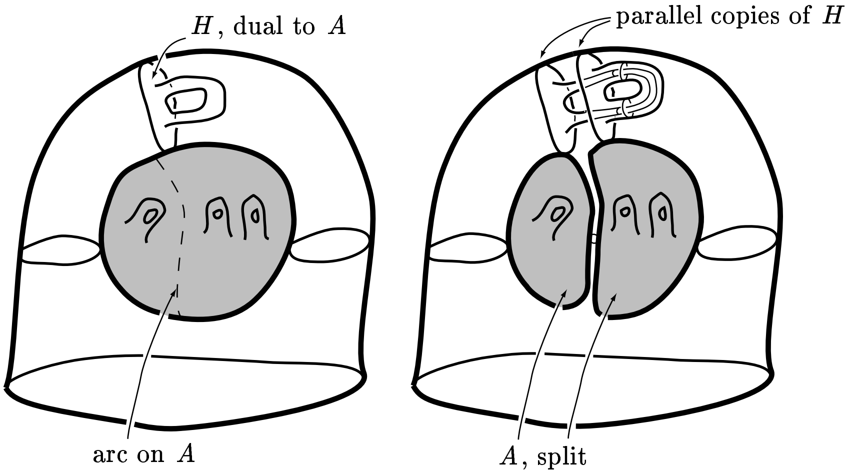

Figure 3: Splitting

Lemma *Any grope can be transformed by iterated splitting to one with dyadic branches. If no caps are split, and all caps in the result are parallel copies of the original caps, then the result is well-defined up to isotopy.*

Proof The splitting is done by induction downward from the caps. Suppose all subgropes of height $\leq k$ that do not include a component of the base are dyadic. To start the induction note that this is true for $k = 0$. For the induction step choose splitting arcs on base surfaces of subgropes of height $k+1$, provided this surface is not part of the global grope base. Choose the arcs so that each component of the complement has genus 1. Splitting one surface may double another, so we choose an order to ensure the process terminates. First split all surfaces whose dual subgrope has height $\leq k$. These duals are already dyadic, so doubling does not introduce nondyadic subgropes. Second, split one in each dual pair where both subgropes have height $k+1$. Nondyadic subgropes are introduced by doubling the dual. However these all now have dyadic duals. Thus splitting them does no further harm. After that we proceed by induction in m, splitting surfaces whose duals have height $k+m$. Previous steps ensure the height$-(k+1)$ subgropes of the duals are already dyadic, so the doubling of the dual does not introduce new nondyadic subgropes of that height. This proves the existence of dyadic splittings. □

This process can be seen as an expansion of a product of sums. Suppose H_1 and H_2 are dual subgropes, and each has a base surface of genus 2. This means each has two branches, say $H_{i,1}$ and $H_{i,2}$. Think of a grope as a sum of its branches, so $H_1 = H_{1,1} + H_{1,2}$. Think of a branch as a product of the two dual subgropes, so the branch formed by the H_i becomes

$$H_1 * H_2 = (H_{1,1} + H_{1,2}) * (H_{2,1} + H_{2,2}).$$

Splitting converts this into four branches:

$$H_{1,1} * H_{2,1} + H_{1,2} * H_{2,1} + H_{1,1} * H_{2,2} + H_{1,2} * H_{2,2}.$$

This formulation extends to represent an entire grope as an iterated composition of polynomials in its caps. Splitting corresponds to rewriting this composition as a sum of monomials. We caution, however, that the "algebra" of caps is not associative or commutative.

The uniqueness for minimal dyadic splittings can easily be proved by formally following the proof of uniqueness of monomial expansions of iterated polynomials. We omit this since we have no application for it.

2.4 Splitting and labeling caps Suppose there are several different "types" of intersection points on a cap. We can separate the types by arcs, then split along the arcs to get caps containing only one type of point. Other caps get doubled during this process, so for it to succeed "type" must be defined so that new intersections with parallel copies of a cap have the same "type." We apply this principle to get uniform local patterns of intersection invariants.

DATA To facilitate the definition of invariants we now require every grope branch to have *dyadic labels*, and *a path to the basepoint.* (This last is not necessary if the grope is modeled on a simply-connected surface, ie, D^2–like, or S^2–like).

Dyadic labels are obtained as follows: for each dual pair of surfaces, label one by 0 and the other by 1. A cap gets a label by reading off the sequence of 0 or 1s encountered in a path going from the base to the cap. In a dyadic branch the labels uniquely specify the caps, since at each level there are only two ways to go up and these are distinguished by the labels. When branches are split the fragments inherit labelings and paths to the basepoint in the evident ways, so this data is preserved.

The following is the first application of splitting to simplify cap types. It will be iterated in later constructions.

Lemma *A grope whose branches have dyadic labels and paths to the basepoint can be split so that it has dyadic branches and each cap satisfies:*

(1) *there are no self-intersections;*

(2) *all caps intersecting the given one have the same label; and*

(3) *the fundamental group classes of the loops through the intersection points are the same.*

Further, the subset of $\pi_1 M$ *occuring as double point loops is the same as that of the original grope.*

We clarify that condition (2) requires all the intersecting caps to have the same label, but this may be different from the label of the cap being intersected. The loops in (3) go from the fixed cap to the others. In detail they go along the path from the basepoint and up through the branch to the fixed cap, through the intersection point to the other cap, then back down and back to the basepoint. Note that viewing an intersection point from the other cap reverses this path, so gives the inverse element in $\pi_1 M$.

Proof In each cap choose arcs to separate intersection points into sets, so that self-intersection points are separated, and all points in a set have the same group element and intersect caps with the same label. Initially these sets may each contain only one point. Now split to separate all these sets, and continue splitting to obtain dyadic branches. If parallel copies of a cap are used before it is split then use parallels of the chosen splitting arcs in these copies. An argument similar to the proof of Lemma 2.3 shows that this process terminates.

The doubling involved in splitting means that each original intersection point bifurcates to many intersections with parallels of pieces of the other cap. However by the way labels are chosen in splittings all these pieces will have the same label as the original. Similarly they have the same group element. Therefore the end result has the properties specified in the lemma. □

2.5 Intersection types We iterate the construction of 2.4 to arrange all intersections with a given cap to have the same "type" in more elaborate senses. The first definition formalizes the situation of 2.4, the next inductively extends it.

Definition *1–types*

(1) A *1–type* is a function from a set of dyadic labels to pairs (α, x) with $\alpha \in \pi_1 M$ and x a dyadic label;

(2) a dyadic branch *has 1–type* ρ if

(a) the domain of ρ is the set of labels on caps in the branch; and

(b) if y is a cap label and $\rho(y) = (\alpha, x)$, then all caps intersecting y have label x, and all intersection points have π_1 element α.

The output of Lemma 2.4 is a grope each of whose branches has a 1–type.

Definition *n–types* If $n > 1$ then

(1) An *n–type* is a function from a set of dyadic labels to pairs (α, x) with α an $(n-1)$–type and x a dyadic label;

(2) a dyadic branch *has n–type ρ* if

  (a) the domain of ρ is the set of labels on caps in the branch; and

  (b) if y is a cap label and $\rho(y) = (\alpha, x)$, then all caps intersecting y have label x, and the branches containing these caps have $(n-1)$–type α.

Branches with 1–types have uniform intersections with adjacent branches. Branches with n–types have uniform patterns of intersection going out through chains of n branches, see Figure 4.

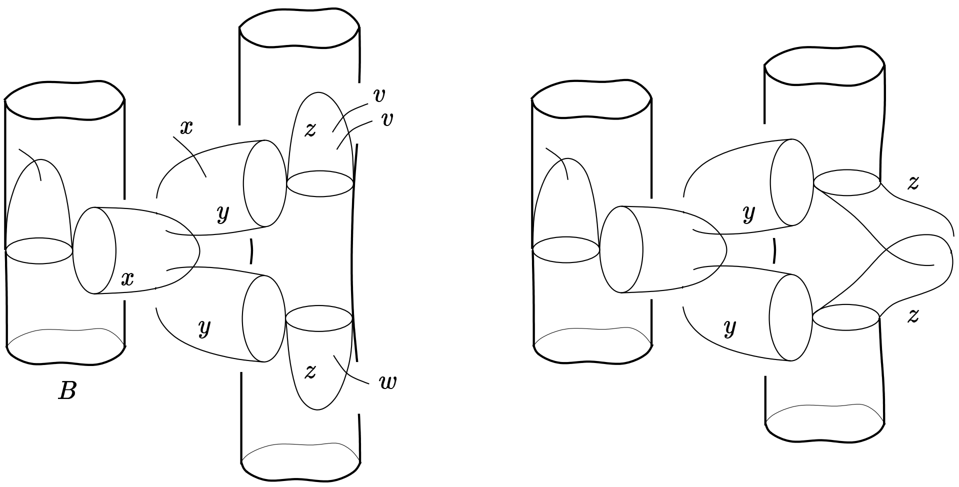

Figure 4: Example: the branch B of a capped surface has a 2–type if (1) all "$x-y$" intersections determine the same element in $\pi_1 M$, (2) the dyadic cap labels v and w coincide, and (3) all "$z-v$" and "$z-w$" intersections determine the same element in $\pi_1 M$. The second figure shows a collision at distance 2.

Lemma

(1) *If a branch with an n–type is split into dyadic branches, then each of these has the same n–type.*

(2) *Any grope can be split to one in which every branch has an n–type.*

Proof The first statement is straightforward. Statement (2) proceeds by induction, with Lemma 2.4 starting the induction with $n = 1$.

Suppose that each branch has an $(n-1)$–type. Associate to each intersection point on a cap the $(n-1)$–type of the branch intersected. Choose arcs in the cap to separate these points into sets, so that all the points in each set have the same type. Do this for every cap, then split along these arcs and continue splitting to get dyadic branches. Note that each intersection point in the original may bifurcate into a large number of intersections with fragments of splittings. However since (by (1)) all these fragments have the same $(n-1)$–type, we end up with all branches intersecting a cap having the same $(n-1)$–type.In other words the branch has an n–type, and the conclusion of the lemma holds. $\square$

2.6 Collisions The construction of 2.4 separates self-intersections and makes π_1 classes of intersections uniform. The construction of 2.5 extends the uniformity out to distance n from each branch. Here we similarly extend the self-intersection condition out distance n, see Figure 4.

Definition A *collision at distance* n from a branch B is a sequence of dyadic labels (x_i, y_i) and two sequences of branches $B_{1,i}$, $B_{2,i}$ for $1 \leq i \leq n$ so that

(1) $B = B_{1,1} = B_{2,1}$

(2) for each $i < n$ and $j = 1$ or 2, there is an intersection between the x_i cap of $B_{j,i}$ and the y_i cap of $B_{j,i+1}$; and

(3) the x_n cap of $B_{1,n}$ transversally intersects the x_n cap of $B_{2,n}$.

We will see that if the branches have n–types then the only "unexpected" intersections that can occur must satisfy these conditions.

Lemma *Given n and a grope, there is a splitting so that there are no collisions at distance$\leq n$.*

Proof This proceeds by induction in n. A collision at distance 1 is just a self-intersection in a cap, so Lemma 2.4 starts the induction with $n = 1$. Suppose there are no collisions at distance $n-1$. If there is a collision at distance n then there are sequences of branches and labels as specified in the definition. If $B_{2,1} = B_{2,2}$ then there would be a collision at distance $n-1$, so the induction hypothesis implies these are distinct. Their intersection points with the cap in B must also be distinct. If we split the cap to separate these points we avoid the collision, or more precisely, we postpone it to distance $n+1$. Moreover splitting does not introduce new collisions at shorter distances. The induction step thus proceeds by choosing arcs to separate any points that lead to a collision at distance n, and splitting along these. $\square$

3 Groups with subexponential growth

A group π is said to have *subexponential growth* if given any finite subset $S \subset \pi$ there is n so that the set of all products of length n of elements of S determine fewer than 2^n elements of the group. The formulation we actually use is a variation on this.

Lemma *Suppose π has subexponential growth and S is a finite subset of π. Then the n specified in the definition also has the property:*

(1) *suppose T is a rooted tree so that all leaves have distance n from the root,*
(2) *each vertex of T, other than the leaves and the root, has valence ≥ 3, and*
(3) *each edge of T is labeled with an element of S.*

Then there is a path in T with distinct endpoints, so that the product of elements along the path gives $1 \in \pi$.

Proof First a few clarifications. The distance of a leaf from the root is the number of vertices with valence at least 3 (so branching actually occurs) between the leaf and root. Orient each edge to go toward the leaves. Then the "product of elements along a path" is the product either of the label or its inverse, depending on whether the path has the same or opposite direction as the edge.

For each leaf we get an element of π by taking the product of elements along the path from the root to the leaf. By the distance condition there are at least 2^n leaves. The choice of n implies that there are two leaves with the same associated product. Since reversing the direction of the path inverts the product, the path from one of these leaves to the root, then back out to the other leaf, has total product 1. The geodesic (unique embedded path) between these leaves also has product 1. Note that the maximum length of this path is $2n$. □

4 Proof of the theorem

The initial data is a properly immersed grope $G \to M^4$ of height at least 2, and $\rho\colon \pi_1 M \to \pi$ with π of subexponential growth. We also allow self-intersections of the base of the grope, provided their loops have trivial image in π.

Choose dyadic labelings for the branches of G, and paths from branches to the basepoint of M. Let S be the set of images in π of π_1 classes of double point loops in G, as in 2.4. Let n be the integer associated to S by the subexponentiality of π.

Use Lemmas 2.5 and 2.6 to split G so that it has dyadic branches, each branch has a $2n$–type, and there are no collisions at distance $2n$ or less. Note that having $2n$–types and having no collisions $\leq 2n$ are both preserved under further dyadic splitting, so doing one construction and then the other gives a grope with both properties.

Note that if a branch has a cap with no intersections, then we can totally contract across that cap. This eliminates the branch without introducing any new intersections. Repeating this reduces to the case where all caps have an intersection point.

4.1 Branches with trivial product We now claim that if there are any branches at all then there is a path going from branch to branch through at most $2n$ intersection points, so that the image in π of the corresponding loop is trivial.

Pick some branch, and define a tree as follows: At the first stage attach an edge to the root corresponding to each cap of the chosen branch. Label the edge by the image in π of the π_1 class of the intersection points on the cap. Associate to the vertex at the outer end of the edge the $(2n-1)$–type of the branches intersecting the cap. Recall that according to the definition of $2n$–type, all of these branches have the same $(2n-1)$–type.

We now do the induction step. Suppose we have gone out distance k, with edges associated to intersecting caps and labeled by corresponding elements of π, and with vertices at this distance labeled by $(2n-k)$–types. Since the edge coming in to such a vertex comes from intersecting caps, it corresponds to one of the dyadic cap labels in the vertex type. Attach outgoing edges corresponding to the other cap labels, label them with images in π of the $\pi_1 M$ elements provided by the 1–type, and associate to the new vertices the $(2n-k-1)$–type provided by the $(2n-k)$–type and the cap label. The number of edges at a vertex is the number of caps in the branch. If the grope has height ≥ 2 then there are at least 4 caps in each branch. Since the Lemma of section 3 only requires 3 edges we see that the proof works for gropes with height "3/2". By this we mean each branch consists of a cap on one side and a grope of height 1 on the other.

This inductive construction continues until we reach $k = 2n$, but in fact we only use the tree out to distance n.

For each branch we now have a tree of radius n, with root corresponding to this branch, and with edges labeled by elements of π. According to the Lemma of Section 3 there is a path in this tree with trivial product. The geodesic with the same endpoints has length $\leq 2n$ and still has trivial product.

This verifies the claim of 4.1 in general. There is a minor sublety in that this path was constructed from abstract patterns in the type, so we must check that it is realized by a sequence of actual intersections. (Recall that a type requires that if there are any intersections then they satisfy certain conditions, but does not require that there be any). The types were originally constructed to correspond to actual intersections, so the only way this can fail is that previous steps have contracted away all the branches that used to lie at some point along the path. However in this case there must be branches with free caps somewhere along the path between this point and the root. This contradicts the standing assumption that all branches with free caps have already been contracted. The conclusion is that if there are any branches at all remaining then paths in these trees are realized by actual intersections.

4.2 Eliminating branches The next step is to eliminate the branch at the beginning of a path of the type found in 4.1, without changing any of the global data ($2n$–types, etc.) The path gives us a sequence of branches B_i and dyadic labels (x_i, y_i) so the x_i cap in B_i intersects the y_i cap in B_{i+1}. Further, if α_i is the image in π of the loop through this intersection then the product $\Pi_i \alpha_i$ is trivial. This is true for one path of intersections starting at B_1. However since B_1 has a $2n$–type it will also be true for *all* paths starting at B_1 and following the same pattern of dyadic labels.

Totally contract B_1 through the cap x_1. This eliminates B_1, but introduces intersections between the base and the y_1 caps of adjacent branches, and these have π images α_1. Push these intersections down and back up through the

x_2 caps (see 2.2 for the definition of this operation). This eliminates these intersections, but introduces intersections between the base and the y_2 caps of branches of distance 2 from B_1. These new intersections all have the same associated π elements because B_1 had a $2n$–type, and by construction this element is $\alpha_1\alpha_2$. Note that the branch B_2 (and all other branches that used to intersect the cap x_1) now has a free cap y_1, which does not intersect anything else. Totally contract these branches along y_2, thus reducing the genus of the base without introducing any new intersections.

Continue in this fashion,pushing y_i intersections down and back up across x_{i+1}, thereby introducing intersections with y_{i+1} caps at distance $i+1$. As soon as the cap y_i frees up, totally contract the branch B_{i+1} along it. At the end of the path all the base–cap intersections will have the same image in π, which by choice of the path is 1. Push these intersections down off branches to the base. They now are base–base intersections with loop image 1.

We must check that this construction does not introduce any new base–base intersections until the base–cap intersections are pushed down at the end. The first step, total contraction of B_1 across a cap, gives an embedded disk because caps have no self-intersections. Inductively suppose no new base–base intersections have occurred in step k. In each individual branch pushing down and back up does not introduce intersections. This is because the transverse sphere used for the operation is embedded. This in turn follows from the requirement that the path be a geodesic in the tree, so it enters and exits each branch through different caps, and there are no self-intersections in the cap it exits through. Thus the only way intersections can arise is if exit caps on different branches intersect. This, however, gives a collision in the sense of 2.6. All collisions of distance up to the length of the path were eliminated by splitting, so they stay disjoint.

Eliminating a branch in this way eliminates all intersections with its caps. After doing this we check to see if this leaves branches with caps without intersections. If so contract these repeatedly until none are left. If there are any branches left after this then repeat the main step (4.1, 4.2). Eventually all branches are eliminated, leaving a surface satisfying the conclusions of the theorem. □

Appendix: Clarification of linear grope height raising

M Freedman
P Teichner

Slava Krushkal and Frank Quinn recently brought to our attention misstatements in the proof of our linear grope height raising procedure which we published in 1995 [4]. This appendix replaces pages 518–522 of that paper with a proof along the same lines but with correct details. The main difference is that we are more careful in which order we add surface stages. This resolves in particular the problem of how to deal with intersections that involve a dual pair of circles on a surface stage: Even though the "key point" in the middle of page 521 is not true as stated (the Borromean rings are not slice after all), the intersections that arise can be dealt with by picking an order and correspondingly decreasing the scale of the relevant lollipops.

We also reformulate the final word length count in terms of coarse geometry, mainly for clarity but also for possible future use.

Since the "warm up" and "warm down" parts of the proof of Theorem 2.1 in [4] are correct, it suffices to explain the core construction and show that the word length grows linearly. More precisely, we prove the asserted estimate for the word length

$$(*) \qquad \ell(g^\bullet_{k+r}) \leq 2r+1$$

in terms of the double point loops of G_k. In the last paragraph on page 522 this assertion is correctly used to finish the proof of Theorem 2.1. We now begin the revision on the top of page 518:

As we start the core construction we have a Capped Grope $G^c := G^c_k$ of height $k \geq 3$. The inductive set up is a Grope G_{h-1} of height $h-1 \geq k$ and an embedding $(G_{h-1}, \gamma) \hookrightarrow (G^c, \gamma)$. One works with the spines, proceeding from g_{h-1} to g_h by adding a finite number of connected surfaces $\Sigma(t)$ to g_{h-1}. To underline the importance of the order in which the surfaces $\Sigma(t)$ are attached, we write

$$g_{h-1} =: g(0) \subset g(1) \subset g(2) \subset \ldots \subset g(n) = g_h$$

where $g(t) := g(t-1) \cup \Sigma(t)$. Even though technically the $g(t)$ are not gropes (since they have heights in between $h-1$ and h), we will still consider them as such. In particular, each $g(t)$ will be thickened to a "Grope" $G(t)$. The surfaces $\Sigma(t)$ are obtained in two steps:

- Step 1 finds surfaces $\Sigma'(t)$ which have (illegal) self-intersections and intersections with grope stages at various heights, but only above

  $$Y := \text{base stage} \cup \text{second stage surfaces } \Sigma_1 \cup \{\Sigma_2\} \text{ of } G.$$

  The subspace Y is protected in the construction so that the dual spheres $\{S\}$ will remain geometrically dual to $\{\Sigma_2\}$, the second stages of G, and disjoint from everything else.
- Step 2 only changes the surface $\Sigma'(t)$ to $\Sigma(t)$, removing double points with itself and with earlier stages (and in the process increases the genus of the surface).

Every application of Step 1 involves choosing some obvious surface (often a disk) so, formally, the presence of these obvious surfaces is an inductive hypothesis which must be propagated in passing from g_{h-1} to g_h. The surfaces $\Sigma'(t)$ for Step 1 are of three types:

(1) "parallel" copies of the initial caps $g^c \smallsetminus g$,

(2) meridional disks to some surface stages of $g(t-1)$, and

(3) "parallel" copies of stages of the original Grope G.

Every application of Step 2 is accomplished by a finite number of moves called a *lollipop move* or a *double lollipop move*. The Step 2 algorithm removes all self-intersections and intersections of $\Sigma'(t)$ (in a particular order) to produce the surface $\Sigma(t)$. The caps $g_h^c \smallsetminus g_h$, necessary to define $\ell(g_h)$, are constructed last and in two steps. The preliminary caps cross all grope stages above Y (stages ≥ 3); these are refined to caps disjoint from the grope using the dual spheres $\{S\}$.

We next explain the central move in our grope height raising procedure. Every surface stage Σ in the Grope $G(t-1)$ has a symplectic basis of circles $\alpha_1, \beta_1, ..\alpha_g, \beta_g$ where g is the genus of Σ, along which higher surface stages or caps have been attached. We consider tori $T_{\alpha_i}, i = 1, \dots, g$ which are ϵ normal circle bundles to Σ in $G(t-1)$ restricted to α_i where ϵ is a small positive number depending on Σ. Notice that all these tori are disjoint. Suppose x is a double point with local sheets $S \subset \Sigma'(t)$ and $S_\beta \subset \Sigma_\beta$, and that the surface stage or cap Σ_β is attached to Σ along β. Symmetrically, if the surface $\Sigma'(t)$ intersects Σ_α then interchange α and β in the next paragraphs.

The *lollipop move* replaces a disk neighborhood S of x with a slightly displaced copy of T_α, made by taking normal ϵ–bundles over a parallel displacement (depending on x) of α in Σ, boundary connected summed to S along a tube which is the normal $\epsilon/10$–bundle of Σ_β in $G(t-1)$ restricted to an arc $\lambda \subset \Sigma_\beta$ from $(T_{\alpha(\text{displaced})}) \cap \Sigma_\beta$ to x. Denote the lollipop by L_α. It is the punctured torus made by attaching the tube (or *stem*) to $T_{\alpha(\text{displaced})}$, see Figure 2.1 in [4].

We are now ready to describe the core construction in detail. Let $h-1=k$. The very first application of Step 1 simply attaches one cap of g^c to g. When regarded as a grope stage the self-intersections in the cap are impermissible and thus the cap only gives $\Sigma'(1)$.

We specify that the initial application of Step 2 removes (in some order) all intersections of $\Sigma'(1)$ using lollipop moves. This gives $\Sigma(1)$ and hence $g(1)$. To obtain $\Sigma(2)$ one just repeats Step 1 and Step 2 by starting with the next cap. Note that now the self-intersections of the second cap as well as the intersections with the first cap have to be removed (in some order) by lollipop moves. In the same manner, one constructs all surfaces $\Sigma(t)$ and hence the grope g_{k+1}. Here the scale ϵ of the lollipops is getting rapidly smaller so that they do not intersect the previously constructed surface stages. This is where the order of things is relevant.

In subsequent applications of Step 1 we must specify which surfaces we choose and what the intersections are. Each L_α contains a meridional circle to which we attach the meridional disk (type (2) above) and a longitude ℓ_α (picked out by the standard framing used to thicken g to G) to which we attach a surface of type (1)–(3) above. Type (2) arises if ℓ_α is the meridional circle of a previously constructed lollipop. Types (1) or (3) arise if ℓ_α is parallel to a circle on the original grope g. In this case, the new surface or cap is only crudely parallel in the sense that we need to glue an annulus A to get from the longitude ℓ_α to $\partial\Sigma_{\alpha(\text{displaced})}$, the attaching circle of a slightly displaced copy of one of the surfaces or caps of G^c. The surface $\Sigma'(t)$ is then defined to be $A \cup \Sigma_{\alpha(\text{displaced})}$. The framing assumption of G implies that for type (3) the surface stage $\Sigma_{\alpha(\text{displaced})}$ will be disjoint from everything constructed previously, ie from $g(t-1)$. However, for both types (1) and (3), the annulus A may intersect many $\Sigma(s), s<t$, so that $\Sigma'(t)$ has many intersections with $g(t-1)$. For type (2), $\Sigma'(t)$ is a meridional disk and it will intersect $g(t-1)$ in a single point.

The reader may expect that the next application of Step 2 will use lollipop moves on $\Sigma'(t)$ to remove these intersection points. This is part of the picture, but there is a difficulty. The lollipop moves, if repeated, produce a branch heading inexorably down G: namely resolving (meridian disk) $\cap\Sigma_i$ with a lollipop capped by a (meridian disk) meeting a Σ_{i-1} lead toward the base of G which is Σ_1. There is no way of using a lollipop to remove a point of (meridian disk) $\cap\Sigma_1$. The solution is to use the *double lollipop move* to resolve any intersection of a current top stage meridional disk with a third stage surface Σ_3. This move turns the branch of the growing grope back "upward" to avoid the bottom part Y.

The *double lollipop move* removes an intersection x between a surface $\Sigma'(t)$ and a third story surface Σ_3. This move replaces a small disk neighborhood $S \subset \Sigma'(t)$ of x with L_α/Σ_α. The notation assumes Σ_3 attaches to β (otherwise reverse the labels α and β), L_α is the lollipop made from T_α as describe above, Σ_α is the third story surface attached to α and finally L_α/Σ_α denotes the embedded surface that results by surgering L_α along a parallel copy $\Sigma_{\alpha(\text{displaced})}$ of Σ_α, ie, $L_\alpha/\Sigma_\alpha = (L_\alpha \setminus \text{nbh. of } \alpha(\text{displaced})) \cup$ two copies of $\Sigma_{\alpha(\text{displaced})}$. Because we have assumed G^c is an untwisted thickening the two copies of $\Sigma_{\alpha(\text{displaced})}$ are disjoint from each other and from the original Σ_α.

Now suppose that we have constructed the grope g_{h-1}. Then the top layer of surfaces has a natural symplectic basis coming from the original grope g and the (meridian, longitude) pair on each lollipop. These bound obvious surfaces $\Sigma'(t)$ of types (1)–(3) as explained above. Applying Step 2 to these surfaces in some chosen order, we remove intersection points by a lollipop move except in the case of intersection with a third stage surface Σ_3 in which case a double lollipop is used. This gives the embedded surfaces $\Sigma(t)$ and hence an embedded grope $(g_h, \gamma) \hookrightarrow (G^c, \gamma)$.

We next check the normal framing. If we assume that each cap has algebraically zero many self-intersections then all surfaces $\Sigma'(t)$ are 0–framed. A lollipop move on a $\pm$–self-intersection changes the relative Euler class by ± 2 (this is best checked in the closed case, $S^2 \times S^2$, where adding the framed dual $0 \times S^2$ to $S^2 \times 0$ gives the diagonal). All other lollipop moves leave the 0–framing unchanged. Thus the passage to $\Sigma(t)$ leaves the relative Euler class trivial so the neighborhood of $g(t)$ agrees with the standard thickening $G(t)$.

To obtain caps $\{\delta\}$ for g_h, we examine the symplectic basis for the top stage of g_h. Some of the curves bound meridian disks to earlier stages of the construction. Some bound "parallel" copies of sub capped gropes of G^c. Contracting, the latter also yield disks. We set $h = k + r$ and

$$g^\bullet_{k+r} := g_{k+r} \cup \{\delta\}$$

The superscript $\bullet$ warns the reader that $g^\bullet_{k+r}$ does not satisfy the definition of a capped grope owing to the cap–grope intersections. These will be removed in the last step, see the last paragraph of page 522 in [4].

Let us next bound the word length $\ell(g^{\bullet}_{k+r})$ in terms of the original generators (= double point loops) of the free group $F := \pi_1 G^c$. Recall that we need to prove

$$(*) \qquad \ell(g^{\bullet}_{k+r}) \leq 2r + 1.$$

For this purpose, we put a *pseudo metric* on the universal covering X of G^c. This is a distance function which still satisfies the triangle inequality but distinct points may have distance zero. Note that pseudo-metrics can be pulled back by arbitrary maps which we will use in the construction as follows. First project X onto the Cayley graph of F such that lifts of the Grope body G map bijectively onto the vertices and lifts of the plumbed squares in the Caps map bijectively onto the centers of the edges. Then take a coarse or pseudo version of the usual path metric on the Cayley graph (in which all edges have length 1) by saying that edge centers have distance $1/2$ from all the vertices the edge meets and that all path components of the Cayley graph minus the edge centers have diameter zero. Finally, use the above map to pull this pseudo metric back to X.

For any map $f\colon Y \to G^c$ which is trivial on π_1, we may then measure the *diameter* of a lift $\tilde{f}(Y)$ in X. For example, if Y is a model capped grope (ie with unplumbed caps) such that $f(Y) = g^{\bullet}_{k+r}$ then the diameter of $\tilde{f}(Y)$ is just the word length $\ell(g^{\bullet}_{k+r})$.

If Y happens to be a disk, surface or (capped) grope such that ∂Y maps to G, it is very useful to consider the *radius* of $\tilde{f}(Y)$ around the "point" $\tilde{f}(\partial Y)$. This uses the fact that each lift of G projects onto a vertex in the Cayley graph of F and thus has radius zero itself. For example, if Y is a disk mapping onto a cap of G^c which has one self-intersection, then the radius of $\tilde{f}(Y)$ is $1/2$ whereas the diameter is 1.

Let X_r be a lift of g_{k+r} to X and let $X^c_r := \tilde{f}(Y)$ where $f(Y) = g^{\bullet}_{k+r}$ as above. Then the triangle inequality shows that $radius(X^c_r) \leq radius(X_r) + 1/2$ and hence

$$\ell(g^{\bullet}_{k+r}) = diam(X^c_r) \leq 2 \cdot radius(X^c_r) \leq 2 \cdot radius(X_r) + 1.$$

It thus suffices to check that $radius(X_r) \leq r$. This in turn follows by the triangle inequality (applied to the usual tree structure of the grope g_{k+r}) from knowing that the radii of all $S(t)$ are ≤ 1. Here $S(t)$ are lifts to X of the surfaces $\Sigma(t)$ used in the construction of g_{k+r} and the radii are again measured w.r.t. $\partial S(t)$.

We prove that radius $S(t) \leq 1$ by induction on t: Recall that the first surface $\Sigma(1)$ was obtained by applying lollipop moves to the first cap of G^c. Before the lollipop moves, we can lift the (unplumbed) cap to X and as explained above it has radius $1/2$ (if the cap is embedded then the radius is zero but we won't consider this easy case). The lollipops then increase this radius to at most 1, independently of how many are used. This follows from the triangle inequality applied to the decomposition of each lollipop into its stem and body (or toral piece). The body has diameter zero since it lies in G whose lift projects to a vertex. The stem has by definition diameter $1/2$ since it leads from a plumbed square to the base of the cap.

Now assume by induction that radius $S(s) \leq 1$ for all $s < t$. Let $S'(t)$ be a lift to X of $\Sigma'(t)$. If $\Sigma'(t)$ is of type (2) or (3) then the radius of $S'(t)$ is zero since it lies in a lift of G. For every intersection point of $\Sigma'(t)$ with $g(t-1)$ we add a lollipop or a double lollipop to obtain $\Sigma(t)$. Only the stems of these (double) lollipops will contribute to the radius of $S(t)$ since the bodies lie in G. The induction hypothesis implies that all these stems have diameter ≤ 1 and thus we are done in this case.

Finally, consider the case where $\Sigma'(t)$ has type (1), ie is a "parallel" cap. Then its radius is $1/2$ as explained above. For every self-intersection of $\Sigma'(t)$ and every intersection point of $\Sigma'(t)$ with $g(t-1)$ we add a lollipop to obtain $\Sigma(t)$ (note that double lollipops don't occur for caps). Again, only the stems of these lollipops will contribute to the radius of $S(t)$. There are two types of lollipops: One type removes self-intersections and intersections with surface stages of $g(t-1)$ that come from the caps of g^c. As for $\Sigma(1)$ the corresponding lollipop stems have diameter $1/2$ and thus can only increase the radius to 1. The other type of lollipops remove intersections of the annulus $A =$ (collar of $\partial\Sigma'(t)$). This means that, as far as our pseudo metric can measure, the stems of the lollipops start essentially on $\partial\Sigma'(t)$ which is the base point with respect to which we measure the radius. By the induction hypothesis these stems can only bring the radius up to 1. □

Note added in proof Slava Krushkal has pointed out that in the above proof, the "warm-up" and "warm-down" steps can be replaced by the following easier and shorter argument:

Do the core construction on the originally given Capped Grope of height $k \geq 2$, preserving only the bottom surface Σ_1 instead of the first two stages Y as done above. (No dual spheres need to be constructed.) After the core construction, we have a Capped Grope of height $k+r$ and word length $\leq 2r+1$, with many cap–body intersections but caps are disjoint from the bottom surface Σ_1. Now do symmetric contraction of the bottom surface. This requires taking parallel copies of whatever is attached to it, and reduces the height of the entire Capped Grope by 1. Then push all cap–body intersections down and off the contraction. This at most doubles the estimate on the double point loop length and thus leads to a clean Capped Grope of height $k+(r-1)$ and word length

$$\leq 2(2r+1) = 4(r-1)+6.$$

Thus linear grope height raising is established.

References

[1] **M H Freedman**, *The topology of four–dimensional manifods*, J. Diff. Geom. 17 (1982) 357–453

[2] **M H Freedman**, *The disk theorem for four-dimensional manifods*, Proc. ICM Warsaw (1983) 357–453

[3] **M H Freedman, F Quinn**, *Topology of 4-manifolds*, Princeton University Press (1990)

[4] **M H Freedman, P Teichner**, *4–manifold topology I: subexponential groups*, Invent.Math. 122 (1995) 509–529.

[5] **V Krushkal**, *Exponential separation in 4-manifolds*, Geom. Topol. 4 (2000) 397–405

[6] **F Quinn**, *Pseudoisotopies of 4–manifolds*, preprint

[7] **R Stong**, *Four-manifold topology and groups of polynomial growth*, Pacific J. Math. 157 (1993) 145–150.