Almost Sufficiently Large

Danny Calegari

1. The Engineer

Wolfgang Haken, one of the greatest and most original topologists of the 20th century, was born June 21, 1928, in Berlin, Germany. He obtained a PhD in mathematics at the University of Kiel in 1953, and then worked for almost a decade as an electrical engineer designing microwave devices at Siemens. While working there he solved what was perhaps the most fundamental unsolved problem in knot theory: the unknot recognition problem (hereafter URP) — i.e., the problem of giving an algorithm to determine whether a knot $K$ in $S^3$ is the unknot.

1.1. Knots and the URP. For our purposes a knot $K$ is a smooth embedded $S^1$ in $S^3$. Two knots $K,K'$ are equivalent if one may be moved to the other by a smooth isotopy. An unknot is any knot equivalent to a round circle in $S^3$ (e.g., a great circle in the round metric).

Some people find it difficult to understand why the URP is a difficult problem. Indeed, it is not easy to draw a diagram of an unknot that can’t be obviously simplified by moving one or two strands. But it is easy to be mislead if one thinks of knots that one can easily draw, that have maybe a dozen crossings or fewer. Maybe one gets a better sense of the complexity of the problem by contemplating the act of untangling a string of Christmas lights (or a long garden hose). Figure 1 is one of Haken’s examples of “complicated” unknots; the reader who can see easily how to begin to untangle it is a better topologist than I.
Unknots may be characterized in several ways:

1. $K$ is an unknot if and only if $\pi_1(S^3 - K) \cong \mathbb{Z}$; or
2. $K$ is an unknot if and only if it is the boundary of an embedded $D^2$ in $S^3$.

Any knot $K$ in $S^3$ is the boundary of some compact oriented embedded surface $\Sigma$ (a so-called Seifert surface) for $K$. The least genus of a Seifert surface is called the genus of the knot. Thus: a knot is the unknot if and only if its genus is zero.

1.2. Linear programming. Haken’s solution to the URP was revolutionary. It depends on the characterization of an unknot as one that bounds an embedded disk in its complement (a Seifert disk). Haken’s algorithm either finds such a disk, or certifies that none exists.

Let’s suppose we want to find a Seifert disk $D$. If it exists, it may be found among the set $S$ of properly embedded surfaces in the knot complement $S^3 - K$. The problem of course is that $S$ is an enormously complicated set with little apparent structure. The key is to replace $S$ by a more manageable collection of surfaces, as follows. Choose a triangulation $\tau$ of $S^3 - K$. A properly embedded surface $\Sigma$ in $S^3 - K$ is said to be normal with respect to $\tau$ if the intersection of $\Sigma$ with each tetrahedron $\Delta$ of $\tau$ is a finite disjoint collection of triangles and quadrilaterals, each of which has the combinatorial form depicted in Figure 2.

There are seven combinatorially distinct triangles and quadrilaterals in each tetrahedron; thus a normal surface $\Sigma$ determines a nonnegative integral vector $v(\Sigma)$ in $\mathbb{Z}^{7|\tau|}$, where $|\tau|$ is the number of tetrahedra in $\tau$. This enjoys the following rather miraculous properties:

1. any properly embedded surface $\Sigma$ may either be isotoped to be normal, or may be simplified by an operation (compression — we shall discuss this in § 2) that reduces its topological complexity;
2. the set $V$ of vectors $v$ of the form $v(\Sigma)$ for some normal surface $\Sigma$ is precisely the subset of $\mathbb{Z}^{7|\tau|}$ satisfying a finite constructible set of integer linear equalities and inequalities; and
3. normal surfaces may be recovered from their vectors; i.e., given $v \in V$ we may recover $\Sigma$ with $v(\Sigma) = v$ uniquely up to isotopy.

These properties, though easy to prove, have powerful consequences: they reduce topological questions about embedded surfaces in 3-manifolds — at least, the incompressible ones — to linear programming.

A Seifert disk $D$ for $K$, if one exists, may not be compressed (it is incompressible) so it may always be isotoped to be normal, and represented by a vector $v(D) \in V$. For any triangulation $\tau$ the set $V$ consists of the integer lattice points in a finite union of rational convex polyhedral cones in $\mathbb{R}^{7|\tau|}$; projectivizing this cone gives a finite set of rational projective polyhedra. If a Seifert disk $D$ exists, one may be found so that the projectivization of $v(D)$ is a vertex of one of these polyhedra. Thus the URP comes down to the (routine, though computationally expensive) problem of determining these polyhedra, enumerating their finite set of vertices, and checking whether any of these corresponds to the desired $D$.

1.3. A “just so” story. The URP is not Haken’s best-known theorem in topology. In 1976, with the collaboration of Kenneth Appel, he proved the Four Color Theorem: the conjecture — first made in the 1850s — that one may label the vertices of a planar graph with at most four colors in such a way that no two adjacent vertices receive the
same label. Famously (or perhaps notoriously) the proof depended essentially on the use of a computer. The argument proceeds by first constructing and certifying a finite unavoidable set of graphs: a collection of graphs with the property that any minimal counterexample must contain one of these as a subgraph. Next one shows that each of the graphs $\Gamma$ in this collection is reducible: it may be simplified to a new graph $\Gamma'$ in such a way that if $\Delta$ contains $\Gamma$, and $\Delta'$ is the result of replacing $\Gamma$ in $\Delta$ with $\Gamma'$, then $\Delta$ may be four-colored if $\Delta'$ can. These two properties of a finite collection — unavoidability and reducibility — together show by induction that no minimal counterexample can exist, and therefore the Four Color Theorem is true.

The essential use of a computer for the verification of such a famous conjecture generated a storm of controversy, centered on the philosophical question of surveyability: whether one should accept the validity of a proof whose details, while humanly checkable one by one, were too numerous to be checkable in aggregate. If this strikes the contemporary reader as a rather frivolous point to get hung up on, it might be because the urgency of the controversy always had more to do with the sociology of mathematics than anything else.

Why was Haken able to solve such significant and long-standing problems when others failed? I do not know the answer, and probably it is unknowable.

Here is a guess. Haken appears to have been unusually open-minded about the use of tools (linear programming, computers) that were not part of the traditional mathematical repertoire. Linear Programming was largely invented by economists, and plays a significant role in many highly applied fields such as network flow problems and supply chain management. And although the first computers were invented by mathematicians, arguably as a natural development of the formalization program of Hilbert as pursued, e.g., by Russell–Whitehead, Gödel, and others, their use as a theoretical tool did not conform to Courant’s description of mathematics as “an expression of the human mind reflecting the active will, the contemplative reason, and the desire for aesthetic perfection.” Is it possible that Haken’s early career as an engineer made him more familiar with, and unprejudiced about tools such as these? What other tools, perspectives, metaphors, work-habits might prove of value to contemporary mathematicians if we were able to broaden our conception of our discipline enough to include them?

2. Hierarchies

Let’s take a closer look at incompressibility. If $\Sigma$ is an embedded two-sided surface in a closed 3-manifold $M$, a compressing disk for $\Sigma$ is a properly embedded disk $D$ in $M - \Sigma$ so that $\partial D \subset \Sigma$ is an essential simple loop in $\Sigma$ (one that does not bound a disk in $\Sigma$). One may cut out an annular neighborhood of $\partial D$ in $\Sigma$ and glue in two parallel copies of $D$ to produce a simpler surface, obtained from $\Sigma$ by compression. See Figure 3. If $\Sigma$ admits no compressing disk, it is incompressible (one also says $\Sigma$ is essential).

![Figure 3. A compressing disk in a surface and the result of compression.](image)

If $D$ is a compressing disk for $\Sigma$, the conjugacy class represented by $\partial D$ in $\pi_1(\Sigma)$ is in the kernel of the inclusion map $\pi_1(\Sigma) \to \pi_1(M)$. A remarkable theorem of Max Dehn gives the converse: if $\Sigma \subset M$ is incompressible, then $\pi_1(\Sigma) \to \pi_1(M)$ is injective.

It follows that in the universal cover $\tilde{M}$ the preimages of $\Sigma$ are properly embedded planes that cut up the universal cover into simply-connected complementary regions which each (universally) cover a component of $M - \Sigma$. A graph $\Gamma$ with one vertex for each complementary region and one edge for each plane in fact a tree (because $\tilde{M}$ is simply-connected), and by construction this tree comes with an action of $\pi_1(M)$ for which the edge stabilizers are the conjugates of $\pi_1(\Sigma)$ and the vertex stabilizers are the conjugates of $\pi_1$ of the components of $M - \Sigma$.

Friedhelm Waldhausen introduced the term sufficiently large for an irreducible 3-manifold $M$ (one in which every smoothly embedded 2-sphere bounds a 3-ball) that satisfies the first (and therefore all) of the following equivalent conditions:

1. $M$ contains an incompressible surface $\Sigma$;
2. $\pi_1(M)$ splits nontrivially as an amalgam $A \ast_C B$ or HNN extension $A \ast_{C'}$;
3. $\pi_1(M)$ acts nontrivially on a tree $\Gamma$.

Cutting $M$ along $\Sigma$ produces a new (possibly disconnected) 3-manifold $M'$ with boundary. There is a notion of essential surface in a manifold with boundary. If $\partial M'$ contains a surface of positive genus, then $H_2(M', \partial M')$ is nontrivial, and any embedded surface $\Sigma'$ representing a nontrivial class may be repeatedly compressed or "boundary compressed" until it becomes essential. Thus $M'$ is also sufficiently large, and may be cut along $\Sigma'$ to produce $M''$ and so on, unless it happens at some point that every boundary component is a 2-sphere, which implies (because of the irreducibility of $M$) that the resulting 3-manifold is a union of 3-balls. Technically, after two or more cuts, one should think of the result as a manifold "with corners," and the 3-balls obtained at the end of the procedure have the structure of (combinatorial) polyhedra.
Haken proved that this sequence of cuts — called a hierarchy — must indeed necessarily terminate in a union of 3-balls after an effectively computable number of steps, and therefore one may understand and prove theorems about sufficiently large 3-manifolds by induction. This has turned out to be an extremely powerful perspective on 3-manifold topology.

Over the years the terminology changed, so that “sufficiently large” manifolds began to be referred to as “Haken manifolds”; as a benchmark, Hempel’s book from 1976 [8] uses the former term, whereas Jaco’s book from 1980 [9] uses the latter. Thus a Haken manifold is an irreducible 3-manifold that contains an incompressible two-sided embedded surface.

2.1. Topological rigidity. One of the first major applications of the theory of hierarchies, due to Waldhausen [19], was the proof that Haken manifolds are “topologically rigid”: a homotopy equivalence between Haken manifolds (satisfying certain further conditions if the manifolds have boundary) is homotopic to a homeomorphism. Examples of (necessarily non-Haken) 3-manifolds for which this fails are Lens spaces (quotients of $S^3$ by cyclic isometry groups): the Lens spaces $L(p; q_1)$ and $L(p; q_2)$ are homotopy equivalent if and only if one of $\pm q_1 q_2$ is congruent to a square mod $p$, but are homeomorphic if and only if $q_1$ is congruent to $\pm q_2^2$ mod $p$. Waldhausen straightens a homotopy equivalence surface by surface, reducing to the Alexander trick in the case of a ball.

In the same paper Waldhausen showed that the universal cover of a Haken manifold is homeomorphic to a 3-ball. He went on to remark that “of those irreducible manifolds, known to me, which have infinite fundamental group and are not sufficiently large, some (and possibly all) have a finite cover which is sufficiently large” ([19], page 87); the implication being that the universal covers of such manifolds are likewise homeomorphic to a 3-ball. If one interprets “(and possibly all)” as raising a question, this is the first appearance in print of what became known as the Virtual Haken Conjecture (hereafter VHC), the question of whether every irreducible 3-manifold with infinite fundamental group is finitely covered by a Haken manifold.

The first appearance in print of what became known as the Virtual Haken Conjecture, proving it for several classes of manifolds, including (most significantly) for Haken manifolds. Since every knot complement in $S^3$ is Haken, it follows from his proof that any knot $K$ falls into one of the following three classes:

1. **torus knots**: those that lie on an unknotted torus in $S^3$ (such knot complements are Seifert fibered);
2. **satellites**: those that lie in an essential way in a solid torus neighborhood of a nontrivial knot (the boundary of the solid torus is essential but not boundary parallel); and
3. **hyperbolic knots**: those whose complements admit a complete hyperbolic structure.

The hyperbolization theorem for Haken manifolds proceeds by induction. Even the base step of the argument is rather subtle: one needs to “hyperbolize” the polyhedra into which a Haken manifold is decomposed by a hierarchy, or exhibit a combinatorial obstruction why it cannot be done.

Each such polyhedron $P$ can be thought of as a 3-ball together with a trivalent graph $\Gamma$ embedded in the boundary, and the goal is to find a hyperbolic metric on the 3-ball for which the edges of $\Gamma$ are geodesics, and the complementary regions $\partial P^3 - \Gamma$ are totally geodesic, and meet at right angles along $\Gamma$. One necessary condition concerns the dual graph $\Gamma' \subset \partial P^3$ to $\Gamma$: it says that every loop $\gamma$ in $\Gamma'$ of length 3 should be in the link of a vertex of $\Gamma$, and every loop of length 4 should be in the link of an edge of $\Gamma$.

To see why these conditions are necessary, let $D$ be a disk in $P$ bounding $\gamma$, and suppose $Q$ is the closed 3-manifold obtained from $P$ by reflection in the sides. This means the following: if $P$ has $n$ faces $F_1, \ldots, F_n$, we take $2^n$ copies of $P$ indexed by maps from $\{1, \ldots, n\}$ to $\{0, 1\}$ and glue two copies whose associated maps differ only at $i$ by identifying their respective copies of the face $F_i$. If $3D$ has length 3 then 8 copies of $D$ fit together in $Q$ to form a sphere; similarly, if $3D$ has length 4 then 4 copies of $D$ fit together in $Q$ to form a torus. Such spheres and hyperbolic 3-space. A closed hyperbolic 3-manifold is necessarily irreducible, and furthermore its fundamental group is infinite but does not contain a $\mathbb{Z}^2$ subgroup; the geometrization conjecture implies that these necessary conditions are also sufficient. The fundamental group of a closed Haken manifold contains $\mathbb{Z}^2$ if and only if it contains an essential embedded torus; thus the conjecture says that a closed Haken manifold is hyperbolizable if and only if it contains no incompressible tori. Likewise, the conjecture says that a Haken manifold with boundary is hyperbolizable if and only if every incompressible torus is isotopic to a boundary component, and it is not Seifert fibered (roughly, a circle bundle over a surface orbifold).

Thurston gave an enormous amount of evidence for this conjecture, proving it for several classes of manifolds, including (most significantly) for Haken manifolds. Since
tori will be essential unless the necessary condition holds; thus they are an obstruction to hyperbolization. If $P$ is not a tetrahedron, these necessary conditions turn out to be sufficient, and $P$ may be hyperbolized.

The induction step is rather complicated, depending on the quasiconformal deformation theory of hyperbolic structures on 3-manifolds with incompressible boundary, and we do not discuss it here. For details see, e.g., Misha Kapovich’s book [11].

2.3. On proof and progress in mathematics. Thurston’s proof of the hyperbolization theorem for Haken manifolds was written up and distributed as a series of preprints, only one of which was ever formally published. Far more influential than the details of the proof was the vision of 3-manifold theory implied by the geometrization conjecture, and a suite of associated conjectures and questions laid out in a famous Bulletin article [18].

This article contains few proofs and many examples. The geometrization conjecture and its proof for Haken manifolds is not the end but a starting point. One can take the conclusion of the conjecture as a hypothesis, and hyperbolic 3-manifolds as objects of interest in their own right. Apparently disparate fields — complex and quasiconformal analysis, bounded cohomology, arithmetic lattices, geometric group theory — are revealed as parts of a deeper and more unified whole.

The article ends with a list of 24 questions/projects, of which the geometrization conjecture was only the first, and not in fact to learn my proof of the geometrization conjecture for Haken manifolds. It is unlikely that the proof of the general geometrization conjecture will consist of pushing the same proof further.

Number 16 on the problem list is the Virtual Haken Conjecture (posed as a question). One may have viewed — I and many colleagues I knew did view — this conjecture as an intermediate step towards a possible approach to geometrization: given $M$ closed and irreducible with infinite $\pi_1$, if one knew $M$ had a finite (regular) cover $\tilde{M}$ which was Haken, one could geometrize $\tilde{M}$ and then analyze the action of the deck group on $\tilde{M}$ to show that $M$ was geometric too (the details of how this argument might go emerged as a corollary of work of Dave Gabai in 1997 [7]).

But how to get started? Suppose $M$ is a closed irreducible 3-manifold. If the only thing we know about its fundamental group is that it is infinite, how do we construct any nontrivial finite cover of $M$ at all?

A curious chicken-and-egg situation emerges. Suppose one already knew $M$ to be hyperbolic (say). Because the group of isometries of hyperbolic 3-space is a matrix group, we may deduce $\pi_1(M)$ is linear and therefore (by a well-known lemma of Atle Selberg) residually finite; this means that for every nontrivial element $\gamma \in \pi_1(M)$ there is a finite index subgroup of $\pi_1(M)$ that does not contain $\gamma$. In particular, any hyperbolic 3-manifold admits many finite coverings, and one can start to explore whether any Haken manifolds may be found among them. Turning the conventional view on its head, one could think of the geometrization conjecture itself as a stepping stone to the VHC.

As Thurston’s quote anticipates, the geometrization conjecture was proved round the turn of the millenium using tools from completely different fields (geometric PDE methods — specifically Ricci flow — and a deep analysis of its singularity formation) by Grisha Perelman [12, 13] (a short-cut to the Poincaré Conjecture spun off in a separate paper in [14]). Thurston’s theorem for Haken manifolds (and for that matter, most of 3-manifold topology) plays almost no logical role in the argument.

3. Of LERF and RAAGs

After the work of Perelman the Virtual Haken Conjecture narrowed to the question of whether every closed hyperbolic 3-manifold $M$ has a finite cover that contains an incompressible embedded surface. This question factorizes naturally into two subquestions:

1. does $\pi_1(M)$ contain a subgroup isomorphic to $\pi_1(S)$ for $S$ a closed surface (of genus at least 2)? and
2. does every immersed surface $S \to M$ injective on $\pi_1$ lift to an embedding in a finite cover?

Immersed $\pi_1$-injective surfaces in hyperbolic 3-manifolds come in two distinct kinds: those that are (virtual) fibers of a fibration of the manifold over the circle; and all the rest. This distinction exactly captures a key geometric property
of $\pi_1(S)$ as a subgroup of $\pi_1(M)$. If $S$ is a (virtual) fiber, then $\pi_1(S)$ is exponentially distorted in $\pi_1(M)$, in the sense that one may write certain elements of $\pi_1(S)$ in terms of a (fixed) generating set for $\pi_1(M)$ far more efficiently than they may be written in terms of a (fixed) generating set for $\pi_1(S)$. On the other hand, if $S$ is not a virtual fiber, $\pi_1(S)$ is undistorted in $\pi_1(M)$: translating elements of $\pi_1(S)$ back and forth between generating sets for $\pi_1(S)$ and for $\pi_1(M)$ does no more than multiply word length by a bounded multiplicative factor; one says in this situation that $\pi_1(S)$ is a \textit{quasiconvex} subgroup of $\pi_1(M)$.

Topologically, an incompressible immersed surface $S$ lifts to an embedded plane $\hat{S}$ in the universal cover $\hat{M}$, which is homeomorphic to $\mathbb{R}^3$. But geometrically, if one thinks of $\hat{M}$ conformally as the interior of the round unit ball in $\mathbb{R}^3$, when $S$ is a virtual fiber the plane $\hat{S}$ becomes wilder and wilder near infinity, and limits to a sphere-filling (Peano) curve; whereas when $S$ is not a virtual fiber, the plane $\hat{S}$ may be compactified to a closed disk in the closed unit ball whose boundary is a topological circle (technically, it enjoys the analytic regularity condition of being a \textit{quasicircle}).

Work of Peter Scott from 1978 [16] shows that the question of lifting an immersed $\pi_1$-injective surface to an embedding can be re-expressed in purely algebraic terms. A group $G$ is said to be \textit{LERF} — an acronym for "locally extended residually finite" — if every finitely generated subgroup $H$ is the intersection of some family of finite index subgroups of $G$; equivalently, if, for every $g \in G - H$ there is a finite index subgroup $H'$ of $G$ containing $H$ but not $g$. Weaker than LERF is \textit{QCERF}, or "quasiconvex extended residually finite," which asks for the LERF property only for quasiconvex subgroups $H$ of $G$ (the analogous property for hyperbolic 3-manifolds with (torus) boundary is \textit{GFERF}).

3.1. Nearly geodesics. How can one find (build?) injective surfaces in a hyperbolic 3-manifold? There are two problems: to find (build) the surface, and to show that it is $\pi_1$-injective. An \textit{embedded} 2-sided surface that is not $\pi_1$-injective has a simple essential loop in the kernel. For an immersed surface this is unknown (this is the so-called \textit{Simple Loop Conjecture}), and in general it seems very hard to give a purely topological criterion guaranteeing that an immersion $S \to M$ is $\pi_1$-injective. However it is possible to give a \textit{geometric} criterion.

In Euclidean space, submanifolds that are nearly flat on a small scale may be badly distorted on a large scale. But in hyperbolic space, quasiconvexity is a local condition. Because hyperbolic space is negatively curved, geodesics diverge from each other exponentially quickly. This divergence dominates the behavior of normal geodesics to a submanifold whose extrinsic principal curvatures have absolute value bounded away from 1, and (by comparison with the endpoint map at infinity) certifies that such submanifolds are $\pi_1$-injective.

In 2009 Jeremy Kahn and Vladimir Markovic announced their proof [10] that every hyperbolic 3-manifold contains (many!) $\pi_1$-injective immersed surfaces. In fact if $M$ is a closed hyperbolic 3-manifold, if $p \in M$ is arbitrary, and if $v \in T_pM$ is an arbitrary vector of length 1, they showed one may find for any $\epsilon$ an immersed surface $S \to M$ with principal curvatures bounded in absolute value by $\epsilon$, so that $p \in S$ and the normal to $S$ at $p$ is $v$. The argument has two parts. Let’s fix a 3-manifold $M$ and a real number $\ell > 0$. Say that an immersed pair of pants $P \to M$ is \textit{good} if the cuff lengths are all closed geodesics of length very close to $\ell$, and if the extrinsic curvature of $P$ is very close to 0 (i.e., it is very nearly totally geodesic). The first part of the argument uses ergodic theory (technically: exponential mixing of the frame flow) to demonstrate the existence of a proliferation of good pants: if $\ell$ is large enough (depending on $M$) then every geodesic $\gamma$ of length close to $\ell$ is the boundary of some good pair of pants in many ways. The second part of the argument explains how to glue up these good pants in such a way that the resulting closed surface has extrinsic curvatures very close to 0 and is therefore $\pi_1$-injective.

To get good enough estimates to control the distribution of good pants one needs to know a lot about the frame flow on a hyperbolic manifold. Negative curvature once more plays a key role. So this step of the argument depends essentially on knowing that the 3-manifold is hyperbolic (or, at least, negatively curved) and therefore builds on Perelman’s proof of geometrization.

3.2. Right-angled pentagons. Suppose we have a hyperbolic 3-manifold $M$ and an immersed, $\pi_1$-injective quasi-geodesic surface $S \to M$. How may we find a finite cover in which a lift of $S$ embeds (up to homotopy)? One dimension lower the analogous question is: given a hyperbolic surface $S$ and an immersed essential loop $\gamma \to S$ (let it be an immersed simple geodesic for simplicity) how may we find a finite cover in which a lift of $\gamma$ embeds (up to homotopy)?

Peter Scott solved this problem in 1978, in the paper [16] cited above. His argument, on the face of it, has a rather ad hoc flavor, and depends on the fact that any surface with negative Euler characteristic admits a hyperbolic metric in which it may be tiled by regular right-angled pentagons. How does this help? Since $\gamma \to S$ is injective, there is an (infinite index) cover $\hat{S}$ of $S$ with $\pi_1(\hat{S})$ equal to the image of $\pi_1(\gamma)$ (which is infinite cyclic) in $\pi_1(S)$. This cover is topologically an annulus, and $\gamma$ lifts to a curve $\hat{\gamma} \subset \hat{S}$ which is homotopic to the core curve of this annulus, so that after a homotopy it may be taken to be \textit{embedded}.

The tiling of $\hat{S}$ by right-angled pentagons lifts to a tiling of $\hat{S}$, and we may find a compact convex region $P \subset \hat{S}$,
containing \( \hat{\gamma} \), entirely tiled by right-angled pentagons. To see this, let's first lift all the way to the universal cover \( \hat{S} \), and let \( \hat{\gamma} \) be the preimage of \( \gamma \) under \( S \to \hat{S} \), so that \( \hat{\gamma} \) is a bi-infinite curve in \( \hat{S} \) stabilized by the deck group \( \mathbb{Z} \). The convex hull \( \hat{C} \) of \( \hat{\gamma} \) is the intersection of all the (closed) half-spaces of \( \hat{S} \) that contain it. One may obtain a bigger (but still functorial) convex subset \( \hat{P} \) by taking the intersection of all the (closed) half-spaces of \( \hat{S} \) containing \( \hat{\gamma} \) that are tiled perfectly by right-angled pentagons. There are “enough” such half-spaces that \( \hat{P} \) and \( \hat{C} \) are a finite Hausdorff distance apart; since both \( \hat{P} \) and \( \hat{C} \) are evidently invariant under the deck group, they project to convex subsets \( P \) and \( C \) in \( \hat{S} \) respectively, each of which may be seen to be compact.

Now, \( P \) is not a closed surface, but (just as we did in § 2.2 in the base step for the inductive proof of hyperbolization for Haken manifolds) we may obtain a closed surface \( Q \) by reflecting in the sides of \( P \). The surface \( Q \), like \( S \), is tiled by right-angled regular pentagons, so their fundamental groups are commensurable in the group of isometries of the hyperbolic plane. Thus, one may find a common finite cover of both \( S \) and \( Q \) in which some lift of \( \gamma \) covers \( \hat{\gamma} \) in \( Q \) and is therefore embedded.

### 3.3. RAAGs and cube complexes

Already in Scott’s paper he observed that the argument could be generalized to a quasiconvex \( \pi_1(S) \) in a 3-manifold group \( \pi_1(M) \) if one could find a 3-dimensional right-angled hyperbolic polyhedron \( X \) for which \( \pi_1(M) \) is commensurable with the group \( \Gamma_X \) generated by reflections in the sides of \( X \).

Unfortunately, 3-manifolds with this property are rare. One can get further by finding a suitable injection \( \pi_1(M) \to \Gamma_X \) where \( X \) is a higher-dimensional right-angled hyperbolic polyhedron, and in fact Ian Agol, Darren Long, and Alan Reid [3] used this idea to prove GFERF for the fundamental groups of the Bianchi manifolds — certain non-compact finite volume hyperbolic 3-manifolds important for number theory. But this trick has its limits: finite volume right-angled hyperbolic polyhedra do not exist in dimensions above 12 [4] (compact ones do not exist above dimension 4).

Having learned since Thurston the profound importance of hyperbolic geometry for 3-manifold topology, mathematicians now had to unlearn it. What was ultimately most important in Scott’s argument about the right-angled hyperbolic pentagons was not their hyperbolicity perse, but their right-angledness.

A right-angled Artin group — or RAAG for short — is a group \( \Gamma_\Delta \) associated to a finite graph \( \Delta \) by taking one free generator for each vertex, and imposing the relations that two generators commute if and only if the associated vertices share an edge (and no other relations that do not follow from these). Starting around 2000, Dani Wise and his collaborators developed a program to prove that certain classes of groups \( G \) are QFERF by showing that they contain a finite index subgroup \( G' \) that embeds suitably in a RAAG.

If \( \Gamma_\Delta \) is a RAAG then it is the fundamental group of a cell complex \( K_\Delta \) which is the union (in a natural way) of a torus \( T^k \) for each \( k \)-tuple of vertices of \( \Delta \) that span a clique. This cell complex may be built in a natural way from Euclidean cubes, all of side length 1, which are glued isometrically along their faces. With this metric the universal cover \( \hat{K}_\Delta \) enjoys a form of nonpositive curvature expressed by saying that it is CAT(0) (these letters stand for Cartan, Alexandrov, and Topologov, who all proved significant theorems in comparison geometry). Distance between parameterized geodesics in a CAT(0) space is a convex function, and this convexity, together with a proliferation of totally geodesic separating subspaces (obtained by gluing together codimension one cubes in the dual complex) lets one simulate Scott’s argument and separate quasiconvex subgroups.

Suppose \( G \) may be exhibited as \( \pi_1(K) \) for some Euclidean cube complex whose universal cover is CAT(0). In 2008 Wise and Frédéric Haglund [5] gave a combinatorial criterion (called special) for a cube complex \( K \) to immerse isometrically in \( K_\Delta \) for some RAAG \( \Gamma_\Delta \), thereby realizing \( G \) as a quasiconvex subgroup of \( \Gamma_\Delta \). If a finite cover of \( K \) is special, one says that \( K \) (and by abuse of notation \( G \)) is “virtually special”; this is evidently good enough to show that \( G \) is QFERF.

### 3.4. Cubulating 3-manifold groups

All well and good. But how on Earth to exhibit \( \pi_1(M) \) as \( \pi_1(K) \) for some locally CAT(0) cube complex?! Rather astonishingly, a cube complex of exactly the sort we need falls into our laps by feeding the Kahn–Markovic surfaces into a beautiful construction due to Micah Sageev [15].

First let’s work one dimension lower and consider a proper configuration \( L \) of lines in the plane. If the lines are disjoint, \( L \) is dual to a tree with one edge for each line in \( L \) and one vertex for each complementary region. If the lines cross in general position, \( L \) is dual to a 2-dimensional complex whose cells are squares. This complex is not typically CAT(0); if there are three lines \( \ell_1, \ell_2, \ell_3 \) that intersect in pairs, these intersections give rise to three squares that share a common vertex with an “atom” of positive curvature, violating the CAT(0) condition. This can beameliorated by going up a dimension: we may fill in the three squares with a 3-dimensional cube in the obvious way; and likewise, for every configuration of \( n \) lines that pairwise intersect we may help ourselves to an \( n \)-cube. The resulting complex \( K \) will be CAT(0), and if there is an upper bound on the size of the cliques in the incidence graph of \( L \), it will be finite dimensional. If the configuration \( L \) is invariant under the action of a group \( G \), then \( G \) acts (combinatorially) on \( K \) and under the right circumstances,
some finite index subgroup of $G$ will act freely with quotient $K$.

The Kahn–Markovic construction gives rise to a large but finite collection of almost totally geodesic immersed surfaces $S$ in a hyperbolic 3-manifold $M$. Lifting to the universal cover one obtains a proper collection $\tilde{S}$ of almost totally geodesic planes in hyperbolic 3-space. Roughly speaking, one may obtain a cube complex $\tilde{K}$ with one $n$-cube for each $n$-tuple of planes in $\tilde{S}$ that mutually intersect. Elementary coarse properties of hyperbolic geometry imply that $\tilde{K}$ is finite dimensional, and that $\pi_1(M)$ acts on it effectively and properly.

3.5. The long goodbye. Thus things stood in March 2012: the VHC was “reduced” to Wise’s conjecture, that every locally CAT(0) cube complex $K$ with $\pi_1(K)$ (word-)hyperbolic was virtually special. In a blog post [20] written March 6 2012, Henry Wilton wrote that

(Wise’s conjecture is) such a remarkable conjecture that it’s difficult to believe it’s true, but it’s also a win–win in the sense that either a positive or a negative answer would be a huge advance in geometric group theory ... implausible or not, I think this conjecture is already a major open problem.

In a remarkable development, only six days later Ian Agol announced a proof of Wise’s conjecture (and, consequently, of the VHC) in a talk at the Institut Henri Poincaré in Paris. Two weeks later Agol gave lectures outlining the details of his argument in a workshop at the IHP, and his preprint (containing an appendix written jointly with Daniel Groves and Jason Manning) was posted to the arXiv on April 12.

It is beyond the scope of this article to explain the main ideas of the argument, except to say that one first constructs an infinite regular cover $\tilde{K}$ of $K$ in which the “hyperplanes” (which correspond in a suitable way to the quasi-convex subgroups to be separated) are compact, two-sided and embedded, and then one builds a finite cover of $K$ “modeled” (in some sense) on $\tilde{K}$; and that the construction of $\tilde{K}$ rests on a certain technical tool in geometric group theory (hyperbolic Dehn surgery) with origins in hyperbolic 3-manifold topology and the work of Thurston.

Combined with work of Wise and previous work of Agol and others, Agol’s work resolved most of the remaining problems (numbers 15–18) on Thurston’s list, marking the end of an era in 3-manifold topology.

What determines which questions become central in a field? What constitutes progress, and how can we recognize it when we see it? Academic research, whatever else it is, is a social activity, and the forces that shape social activities are complicated and have an enormous psychological component. Fashion, value-judgments, familiarity, taste all play a role.

Agol has been unusually thoughtful about the social psychology of mathematical practice, sometimes describing conjectures with which he wrestled over many years as “friends” whose company he was pleased to spend so much time with. He writes [2]:

Mathematicians use part of the social wiring of the brain to engage with mathematical ideas and objects. I certainly feel like this for myself, where the figure 8 knot complement plays the role in my mind that a celebrity (like Bob Dylan) might in others. I think this helps us engage with abstract mathematics, which can be socially isolating because there’s very few that we can communicate with about these ideas on a regular basis. I don’t think this is special to mathematics, for example astronomers might have a social connection to the Andromeda Galaxy.

The long and meandering route that led from the resolution of one conjecture (the URP) to another (the VHC) winds in and out of topology, geometry, combinatorics, group theory, PDE, and many other fields. Along the way objects and theories emerged, grew, evolved; seedlings transplanted to foreign soils became established and bore brave new fruit. If one could plot progress over time (if one could even agree on an axis along which to measure it) the graph would be the Devil’s Staircase. Headway, when it came, came suddenly, like a fire alarm or the finale of Beethoven’s Fifth Symphony, and caught most of us naked and napping.

A partial entity-relationship graph of the situation might look like Figure 4. Is the situation as irreducibly complex and muddled as the graph suggests, or might there be some hitherto undreamed of principle that would untangle the skeins of this picture like one of Haken’s unknots and let us really see clear to the bottom of the well?

Wolfgang Haken died in his home in Champaign, Illinois, on October 2, 2022, at the age of 94.

ACKNOWLEDGMENT. I would like to thank Christine Magnotta and the anonymous referees for encouragement and valuable feedback on earlier drafts.

References


Figure 4. An entity-relationship graph connecting URP to VHC.


Danny Calegari

Credits

The opening image is courtesy of Olivier Vandeginste via Getty.

Figure 1 is courtesy of Cameron Gordon.

Figures 2, 3, and 4 are courtesy of the author.

Photo of the author is courtesy of the author.