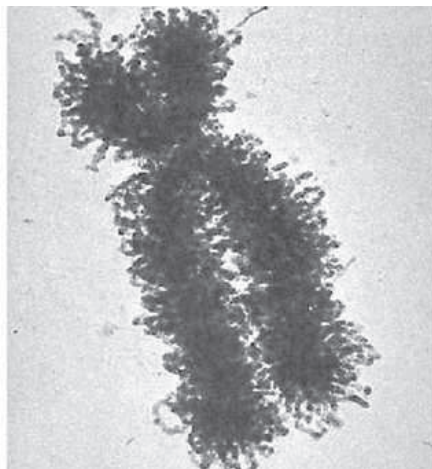


RESEARCH STATEMENT: KNOTS AND 3-DIMENSIONAL MANIFOLDS

ALEXANDER COWARD



CONTENTS

1. Introduction and motivation	1
2. Algorithms, moves and complexity	3
3. 3-manifolds and normal surface theory	6
4. Crossing changes	13
5. Physical knots and links	17
6. Epilogue	18
References	20

1. INTRODUCTION AND MOTIVATION

The rate of progress in 3-manifold theory over the past decade has been astounding: The Poincaré Conjecture, The Geometrization Conjecture, The Surface Subgroup Conjecture, The Virtual Haken and Virtual Fibered Conjectures, among other remarkable things, have all been solved in the last ten years.

And yet, despite this progress, it remains the case that many elementary questions about knots that were posed before the 1960s are still unresolved. Further, solutions to such questions, such as whether one can determine when two knots are related by a crossing change, whether crossing number is additive under connect sum, or if one can quickly determine if two knots are the same, seem some way off given the current state of mathematical technology.

This juxtaposition of astonishing progress with elementary unanswered questions is mirrored in the role of knots in the physical world, where they arise with ubiquity in microbiology. If one were to straighten out every DNA molecule in an adult human being and line them up end-to-end, one would create a strand longer than the orbit of Neptune. For something that long to fit into our physical bodies requires that the molecules of our DNA employ an astonishingly sophisticated process to package up the strands. If this was a completely static picture it would be remarkable enough, but it is not a static picture since DNA is constantly moving in order to perform its core functions of replication and expression. In order that DNA replicates the coils have to unwind and disentangle so that they can become physically separated. Further, in order for genes to be expressed, the relevant part of the DNA molecules must become exposed and untwisted.

We understand these processes of DNA replication and expression in broad principle, and recent decades have seen astonishing progress in our understanding of the underlying microbiology. However we would like to go further; just as today we understand the exact shapes of protein molecules in terms of the (x, y, z) -coordinates of every single atom in each molecule, we would like to understand the exact shape of DNA double helices as they move about, so that in the future we can control this motion with a similar level of precision to the way today we can control the positions of electrons in a microprocessor. We would like to be able to prevent genes for all manner of diseases from untwisting and expressing. Similarly we would like to be able to prevent DNA molecules that have undergone a specific, undesirable, mutation from untangling and replicating.

It is true that such ideas might seem a long way off, even fanciful, today. However there is an analogy between our understanding of microbiology today and our understanding of physics in the 19th and early 20th centuries. In that time we saw two things: First, experimental groundwork, like the 1887 Michelson-Morley Experiment that discredited the notion of a 'light medium', and the 1909 Gold Leaf Experiment from which we learned that atoms have nuclei; and second, mathematical groundwork, in particular geometry and analysis, enabling later physicists and mathematicians to interpret experimental observations and make predictions that would ultimately shape the course of history.

My research is about 3-dimensional topology and geometry, with a focus on knots, algorithms, and complexity. These are branches of mathematics that deal with shape and space, and should rightly be regarded as both fascinating areas of pure mathematics and also mathematical groundwork for understanding real world 3-dimensional phenomena. The following pages are intended to give a flavor of the kind of research that interests me. The intended audience is a non-specialist mathematician or scientist, although I hope it may be of interest to experts also. Because the focus is on the flavor of my research rather than technical details, I have allowed myself license to be a little imprecise with my language; for precise definitions and statements I refer the reader to my papers. I am by necessity highly selective, to the extent that any list of omitted key topics I might list here would itself be selective. My selectivity is

motivated by two factors. The first is my obvious desire to emphasize my own contributions, showing how they fit into the history of the subject and how they lead into my plans for the future. The second motivation is to counter the idea expressed in parts that knot theory and 3-manifold theory are, with all the remarkable progress of recent years, subjects where all the big and important problems have been solved, and to evangelize the view that knot theory and 3-manifold theory are young subjects whose full role in both the mathematical and real worlds are only just being uncovered.

2. ALGORITHMS, MOVES AND COMPLEXITY

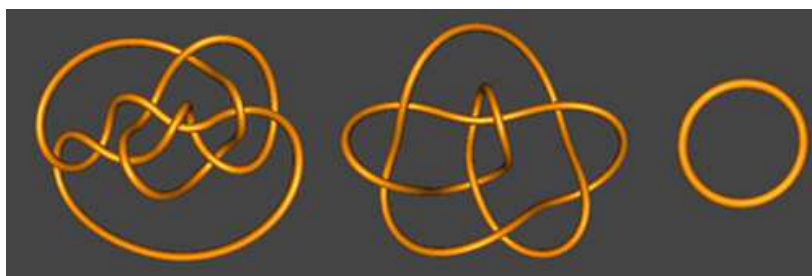


FIGURE 1. Three different knots.

Knot theory is motivated by understanding objects like those illustrated in Figure 1. They are called *knots*, and are defined to be embedded curves in 3-dimensional space. The first person to study knots with any semblance of mathematical rigor was Tait, a Scottish physicist motivated in large part by ideas of Kelvin at the time that knots were the building blocks of matter.

It is worth reflecting on why the theory that matter is made out of knots made some sense in the 19th century. First, by that time people were used to the idea that physical phenomena had mathematical explanations. Second, the idea that matter is composed of the elements combined in different ways can be thought of in mathematical terms as a ‘prime decomposition’ for matter. Third, the world we live in is 3-dimensional, so a mathematical explanation of the elements should be fundamentally 3-dimensional in nature. Once we realize that we are looking for 3-dimensional mathematical objects that admit a prime decomposition, knots seem like the most natural things to look at. Knots also have the advantage that they can be made very small, and they have the property that the same knot can be arranged in multiple ways, possibility explaining the way the same element can have different allotropes, like diamond and graphite. Finally, the idea of a ubiquitous aether, or ‘light medium’, fit with observations at the time that other mediums like air and water admit knotted vortices.

The theory that knots could be used to explain the different properties of the elements in Mendeleev’s periodic table ultimately did not stand up to scrutiny or experiment. However the proponents of this theory were serious physicists with good reasons behind their speculation, and their theory, though wrong, did provide strong motivation to understand knots from two points of view: First, if we were to explain properties of the elements with knots, then it would be important to have an accurate

and complete table of the different knots which one could compare to the elements in the periodic table. Second, if chemical reactions are really operations on knots, then we would need to understand these operations.

This was the motivation behind Tait's work. It set the direction and focus of knot theory for close to a hundred years, until the 1960s. In modern mathematics the questions arising from Tait's work are, as we shall see, expressed altogether differently: Seeking to understand how to construct a complete and accurate table of knots raises the question of whether properties of knots can be understood algorithmically, and understanding operations on knots is really about understanding different types of surgery.

In this section and the next we focus on the first of these themes, algorithms, deferring questions of surgery to Section 4.

Figure 2 shows what Tait is best remembered for. It purports to show a table of all the knots that can be flattened out into the plane with at most seven points where the knot passes over or under itself. Such an arrangement of a knot into the plane is called a *diagram*, and the number of crossing points, or *crossings*, is called the *crossing number* of the diagram. The minimum number of crossings in any diagram of a knot K is called the *crossing number of K* , is denoted $c(K)$, and this is the modern language for “*order of knottiness*”. So, in modern language, Tait's table in Figure 2 shows all knots K with $c(K) \leq 7$.¹

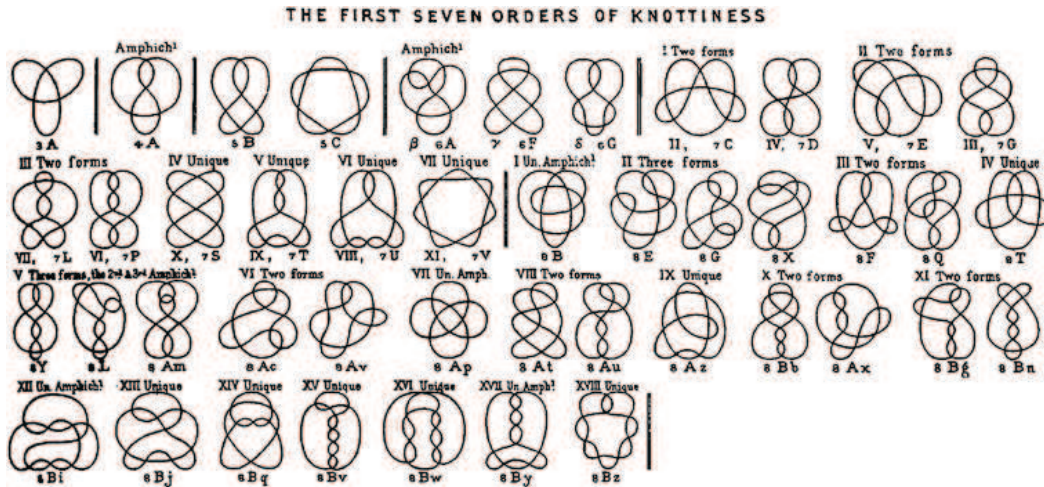


FIGURE 2. An extract from Tait's original 1885 table of knots.

The construction of the table in Figure 2 is a non-trivial task if one wishes to be sure that it is both complete and accurate. One needs to accomplish two tasks: First, one must make sure that no knot has been missed, and second, one must be sure that no knot has been repeated.

¹Note that the all the crossings in Tait's table are taken to run over, under, over, under, and so on, as one runs around the knot. In modern language we say that all the knots with crossing number at most 7 are *alternating*.

The first task, that of making sure that no knot has been missed, is simple. As Tait himself observed, “*We find that it becomes a mere question of skilled labour to draw all the possible knots having any assigned number of crossings.*” In modern language we might say that there are, up to continuous deformation in the plane, a finite number of knot diagrams, and that there is an algorithm to find them all.

To accomplish the second task, one must be able to decide if two arrangements of knots in \mathbb{R}^3 represent the same knot. Here, ‘the same knot’ means that there is a continuous motion of the knot in \mathbb{R}^3 that takes one arrangement to the other. Tait himself was well aware of this issue, although at the time, which predated Turing’s foundational work on algorithms by 50 years, there was no mathematical language with which to phrase, let alone answer, the key question: *Is there an algorithm that will determine whether two knot diagrams specify the same knot?*

This is called the *recognition problem* for knots. It is an extremely hard problem, and was ultimately not solved (for the most part) until the early 1970s with groundbreaking work of Haken, and even then there was a mistake that remained unidentified and unfilled until 2003, when Matveev published a book [11] containing a full and complete proof.

It is worth reflecting on why the problem is so hard. The reason is that given two knot diagrams, in order to be sure that they do not specify the same knot, one must rule out the possibility that there is a continuous motion taking one to the other. However there is an infinite space of candidate continuous motions, with no clear way of restricting to a finite list of possible motions to try.

It is with this idea in mind, of restricting the possible motions from one diagram to another, that Kurt Reidemeister [13] in 1927 proved the following theorem.

Theorem 2.1. *Any two diagrams for the same knot or link may be joined by a sequence of Reidemeister moves.*

The Reidemeister moves referred to in Theorem 2.1 are shown in Figure 3. They provide a simple combinatorial way of describing a continuous motion of a knot or link² from an arrangement as one diagram to an arrangement as another.

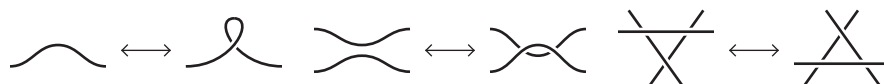


FIGURE 3. Reidemeister moves.

Reidemeister’s theorem was certainly progress towards a solution to the recognition problem because it provides a simple way of declaring ‘Yes, these two diagrams specify the same knot’ in a finite amount of time. One simply starts with one diagram, applies every possible Reidemeister move to it, and then every possible Reidemeister move to the resulting diagrams, and so on. If the two diagrams specify the same knot, then eventually a sequence of Reidemeister moves joining them will be found.

²A *link* is a generalization of a knot where more than one closed curve component is allowed. Thus, a knot is a one component link.

The limitation of Reidemeister's theorem is that it provides little help in the case where the two diagrams do not specify the same knot. The problem is that without information about how many moves are needed, there is no way of telling the computer when to stop looking for possible sequences. Marc Lackenby and I [7] resolved this matter with the following theorem.

Theorem 2.2. *Let D_1 and D_2 be connected diagrams for some knot or link in \mathbb{R}^3 , and let n be the total number of crossings in both diagrams. Then D_2 may be obtained from D_1 by a sequence of at most $\exp^{(c^n)}(n)$ Reidemeister moves, where $c = 10^{1,000,000}$.*

Theorem 2.2 provided the first conceptually simple solution to the recognition problem for knots and links; one simply calculates the upper bound on the number of Reidemeister moves and tells the computer to stop looking for sequences of Reidemeister moves once that limit is reached.

Now, while this is excellent news, there is a substantial problem if one actually wishes to implement the algorithm arising from Theorem 2.2. That is that the bound one obtains is unimaginably large, meaning the computer program, though theoretically finite, would never stop in any reasonable timeframe; the function $\exp(x)$ is the exponential function 2^x , and $\exp^{(r)}(x)$ means iterate this function r times. Thus, $\exp^{(c^n)}(n)$ is shorthand for a tower of 2s with an n at the top, the height of the tower being c^n .

We are now faced with a new problem; we not only wish to show the existence of a theoretical bound, yielding a theoretical algorithm, but rather find bounds and algorithms that could be implemented in practice, on a real computer considering real world knots. As we shall see, such goals are both difficult and within reach, but to understand how one may approach this we need to look at why the bound in Theorem 2.2 is so big and the technology with which this bound is obtained. It is this technology that we look at in the next section.

3. 3-MANIFOLDS AND NORMAL SURFACE THEORY

A three dimensional manifold, or *3-manifold*, is a mathematical object that locally looks like \mathbb{R}^3 . The modern theory of knots is now a branch of 3-manifold theory because knots are best studied via their complements. If we compactly \mathbb{R}^3 with a point at infinity, we obtain a 3-manifold called the 3-sphere, denoted by S^3 , and if a small open neighborhood of a knot K is removed from S^3 we obtain a 3-manifold M_K with toral boundary called the *exterior of K* . For many years it was conjectured that the topology of the M_K determines the knot K , although it took until 1989 for this statement to finally be proved by Gordon and Luecke [9].

Theorem 3.1. *The exteriors of two knots K_1 and K_2 are homeomorphic if and only if K_1 and K_2 are the same knot.*

Theorem 3.1 focuses the mind on understanding 3-manifolds if one wishes to understand knots. In particular, the question of how to determine algorithmically whether two knots are the same may be translated with Theorem 3.1 into the question of how

to determine algorithmically whether their two exteriors are homeomorphic.³ The goal of this section is to provide an overview of how this question this may be answered.

The first step, when seeking to answer any question algorithmically, is to have a scheme to specify the input with a finite amount of data. If the input is a knot or link, then a diagram suffices and these may be stipulated with a finite amount of data. If, on the other hand, the input is a 3-manifold then another approach is needed. There are several schemes that one can use. In this overview we will focus of one of the most common, a *triangulation*.

A triangulation is a way of building a 3-manifold out of a collection of tetrahedra by gluing their faces together in pairs, as shown in Figure 4.

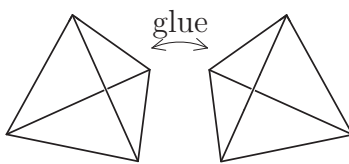


FIGURE 4. A 3-manifold may be specified by a triangulation.

A triangulation may easily be specified with a finite amount of data; one simply declares a list of tetrahedra, together with information about how their faces are identified.

We now have the setup. We have a 3-manifold, M say, specified by a triangulation, and we wish to answer some question about M . So far we have focused on the question of whether M is homeomorphic to some other triangulated 3-manifold M' , but there are many other questions one could ask. Some examples of other interesting questions include:

- Is every embedded 2-sphere in M the boundary of an embedded 3-ball?
- Is there an essential curve on the boundary of M^4 that bounds an embedded disk in M ?
- Can M be constructed by gluing the boundaries of two genus- n handlebodies?⁵

There are a wide variety of questions one can ask. However the list above illustrates a theme that runs through 3-manifold topology, namely that all these questions may be encoded by the existence, or not, of a surface in M with a specific property. The first question asks if there is an embedded 2-sphere in M not bounding a 3-ball. The second question asks if M contains a properly embedded disk whose boundary is essential. Finally, the third question asks if M contains an embedded genus- n surface that bounds a handlebody on each side. *Normal surface theory* is a collection of tools that enables us to construct a finite list of candidate surfaces in a given triangulated

³Haken's solution to the recognition problem for knots and links predated Theorem 3.1 by two decades. For this reason Haken's original work considered 3-manifolds decorated with a little extra structure to specify how they arise as knot or link exteriors.

⁴An *essential curve* on the boundary of M is a properly embedded curve that does not bound a disk in the boundary of M .

⁵A *genus- n handlebody* is a 3-manifold with boundary that looks like a solid pretzel with n holes.

3-manifold, with the view of testing each one in turn to see if any of them have the property in question.

A *normal surface* in a triangulated 3-manifold M is a properly embedded surface which intersects each tetrahedron of the triangulation in a disjoint collection of *triangles* and *quadrilaterals*, as shown in Figure 3. These triangles and quadrilaterals are called *elementary disks*.

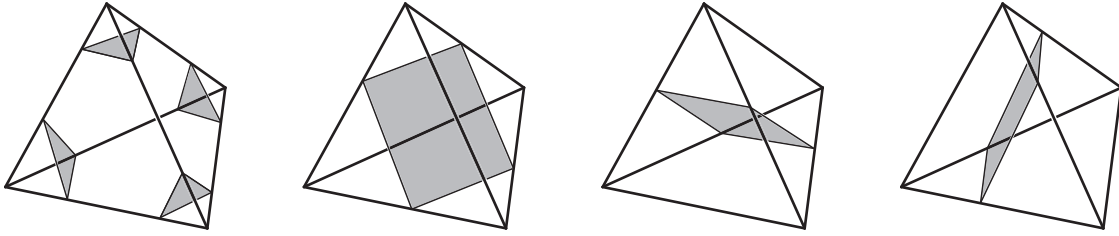


FIGURE 5. The seven types of elementary disc in a tetrahedron.

It turns out that many interesting classes of surface in a 3-manifold can be arranged to be normal. That is, if there is a surface S in M with some property, then it is often the case that this implies the existence of a normal surface S' with this same property. The philosophy of normal surface theory is to enumerate finitely many possibilities for S' .

Consider a face of the 2-skeleton of the triangulation of M . Any normal surface in M must intersect this triangle in *normal arcs*, that is arcs which start and end on different edges.

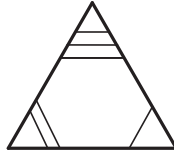


FIGURE 6. The intersection of a normal surface with a typical face of the 2-skeleton.

There are three types of normal arc in any face of the 2-skeleton, as shown in Figure 6. Each of these represents an edge of a normal disc on each side. These normal discs must match up, as shown in Figure 7. Note that on each side of a triangle of the 2-skeleton, there are only two possible types of normal disc, one triangle and one quadrilateral, which can give rise to each type of normal arc.

Let x_1, \dots, x_n be the number of normal discs of each type in a particular normal surface. In a manifold with t tetrahedra, n will equal $7t$. The observations above lead us to conclude that x_1, \dots, x_n satisfy a system of linear equations of the form

$$x_i + x_j = x_k + x_l$$

with three such equations for each face of the triangulation where two tetrahedra are glued.

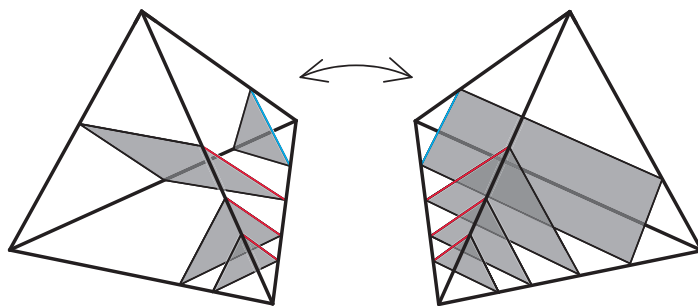


FIGURE 7. The normal discs must match up on the common face of two adjacent tetrahedra.

This system of equations is known as the *matching equations* for the triangulation. There is a natural correspondence between normal surfaces and non-negative integer solutions to the matching equations which satisfy the additional condition that they only give rise to at most one type of quadrilateral in each tetrahedron. This is excellent news, because it means that we have a method to translate questions about topological 3-manifolds into questions that can be understood using *integer linear programming*, and this is an area that is both widely studied and well understood.

Now, the matching equations for a particular triangulated 3-manifold have infinitely many non-negative integer solutions in general, and so there can be no hope of computing a finite list of all of them. This may be readily seen by considering many parallel copies of any normal surface.

Despite the non-existence of a finite list of solutions, the matching equations always have a finite set of *fundamental solutions* which gives rise to all solutions as linear sums. Furthermore, the fundamental solutions may be found algorithmically. The fundamental solutions are also known as a *Hilbert basis* for the equations. Note that not all the fundamental solutions necessarily correspond to normal surfaces because they may give rise to conflicting quadrilateral types. However, for those fundamental solutions that correspond to normal surfaces we refer to the surfaces as *fundamental surfaces*.

Now, it can often be shown that if there is a normal surface with a particular property, then there is a fundamental surface with that property. For example, any triangulation of the exterior of the unknot⁶ always contains a spanning disk for the unknot in the set of fundamental surfaces. This is how the unknot recognition problem is solved: One starts with any knot diagram and builds from it a triangulation for the knot exterior. Then one finds all the fundamental surfaces coming from the matching equations for this triangulation. Finally one inspects all these surfaces to see if any is a disk whose boundary runs once around the knot. The initial knot diagram specifies the unknot if and only if we find such a disk.

This overview of normal surface theory presents a very clean picture, and at its best the theory is very clean. In practice however there can be many different difficulties that arise when seeking to apply the general scheme. From a theoretical point of view,

⁶The *unknot* is the simplest of all knots, shown on the right of Figure 1.

many interesting surfaces cannot be arranged to be normal, and even when they can, showing that one can be taken to lie in the set of fundamental surfaces is usually difficult and often impossible. From a practical point of view, the fundamental solutions to the matching equations have, as a general rule, coordinates that are exponential in the number of tetrahedra in the triangulation, making the practical implementation of normal surface theory a challenging exercise for complexity time reasons.

These theoretical and practical difficulties in implementing the philosophy of normal surface theory are a major area of current research, both of myself and the wider 3-manifold community. On the theoretical side, I developed a theory of *almost interiorly-normal surfaces* to calculate an invariant called the *bridge number* for hyperbolic knots. The precise theorem I proved in [2] is the following.

Theorem 3.2. *Let K be a hyperbolic knot in S^3 . Let M_K be the exterior of K in S^3 . Then, up to ambient isotopy, there are only finitely many bridge punctured 2-spheres for M_K of given Euler characteristic. Furthermore there is an algorithm to find all of these surfaces.*

Corollary 3.3. *There exists an algorithm to determine the bridge number of a hyperbolic knot.*

A *bridge punctured 2-sphere* for the figure-eight knot is illustrated as the horizontal plane in Figure 8. It has the property that it cuts the knot into collections of simple unknotted arcs above and below called *bridges*. Thus Theorem 3.2 says that given any hyperbolic knot⁷ and an integer n there are only finitely many essentially different ways of arranging it as a union of n simple unknotted arcs above and below the plane, and furthermore there is an algorithm to find all these configurations. The *bridge number* of a knot is the minimum number of bridges required in any such arrangement.

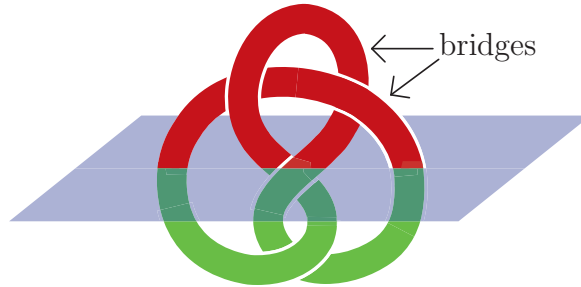


FIGURE 8. A two bridge arrangement of the figure-eight knot.

The strategy for the proof of Theorem 3.2 is to try to search for bridge punctured 2-spheres using normal surface theory. However classical normal surface theory breaks down here for many reasons that are beyond the scope of this statement. For a flavor, Figure 9 shows an example of an almost interiorly-normal disk, which is a building

⁷A *hyperbolic knot* is one whose complement admits a complete Riemannian metric with constant curvature -1 . William Thurston famously showed that every knot, with the exception of certain specific types of knot called torus knots and satellite knots, is hyperbolic.

block for almost interiorly-normal surfaces. It is out of pieces like this that we build bridge punctured 2-spheres like the horizontal plane in Figure 8.

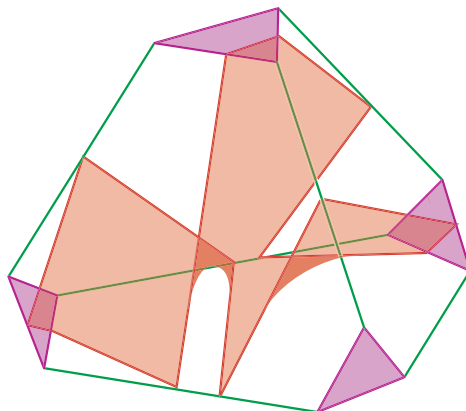


FIGURE 9. An almost interiorly-normal disk.

The practical application of normal surface theory is also an area I am interested in. Recall that finding the fundamental solutions to the matching equations is an exponential-time problem in general. Perhaps surprisingly in the light of this, a software suite that applies normal surface theory, called Regina, is available. It solves the exponential-time problem of finding the fundamental solutions to the matching equations, and it achieves this with carefully constructed algorithms that, though theoretically running in exponential-time, run fast enough to provide a useful tool for researchers using nothing more than a home computer. For an excellent survey paper about Regina's many features the reader is referred to [1] by Ben Burton, the main developer of Regina.

So far we have seen that normal surface theory is a powerful and highly adaptable tool that one can use to answer many questions about knots and 3-manifolds, and we have also seen that, despite theoretically exponential run times, finding exponentially complicated fundamental normal surfaces is feasible in practice. However there is no amount of care in reducing the constants that can make an algorithm running in multiply iterated exponential-time, such as the knot recognition algorithm arising from Theorem 2.2, feasible in practice. I am currently involved in several projects with (separately) Joel Hass and Marc Lackenby that seek to bring multiply iterated exponential-time problems into the realm of single exponential-time problems, and also bring iterated exponential bounds, such as that in Theorem 2.2, down to single exponential bounds.

To understand how these projects work, we need to understand why the bound in Theorem 2.2 contains the iterated exponential. The reason is that rather than relying on a single surface, as in the above description of normal surface theory, Theorem 2.2 requires an analysis of something called a *hierarchy of surfaces*. This is a finite sequence of surfaces with the properties that the first surface is properly embedded in the ambient 3-manifold M , the second surface is properly embedded in the manifold obtained by cutting M along the first surface, the third surface is properly embedded

in the manifold obtained by cutting M along both the first and the second surfaces, and so on. We also require that the manifold obtained by cutting M along all the surfaces is a collection of 3-balls.

Hierarchies are extremely useful tools with which to understand 3-manifolds, and they have a rich and highly developed theory. Furthermore, they interact well with normal surface theory from the point of view of constructing theoretical algorithms. This is because it is often the case the the kinds of hierarchies one studies have the property that each surface in the hierarchy can be taken to be a fundamental surface in a triangulation for the cut-up 3-manifold in question. This is the key construction for Haken's solution to the recognition problem for knots and links. The idea is to use a hierarchy, found algorithmically using normal surface theory, to cut up the knot or link exteriors into balls, and then use the combinatorics of how the balls glue together as a way of encoding whether the two knots or links were the same.

We are now in a position to see where iterated exponentials like that in Theorem 2.2 come from. One arranges the first surface in the hierarchy to be normal in M , and then one cuts M along this surface. We now wish to make the second surface normal, and for this we require a triangulation for the cut 3-manifold. Traditionally this is achieved by coning every piece of every tetrahedron that is obtained from cutting. This is illustrated a dimension down in Figure 10.

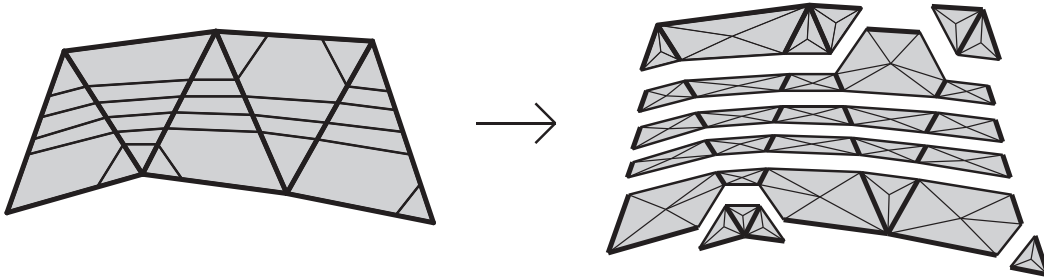


FIGURE 10. Traditional cutting and re-triangulating.

Since the first surface has exponentially many normal disks, the cut 3-manifold generally has exponentially many tetrahedra. Now the number of normal disks in the second surface is exponential in the number of tetrahedra in the cut manifold, i.e. double exponentially many in terms of the number of tetrahedra in the original 3-manifold M . Continuing in this fashion, we obtain bounds involving towers of exponentials in terms of the number of tetrahedra in M , where the height of the tower is the number of surfaces in the hierarchy.

The iterated exponential in Theorem 2.2 arises in exactly this way. There we use a normal hierarchy to describe an explicit movie taking one diagram to the other. We then keep track of the number of Reidemeister moves that take place throughout this movie. The iterated exponential in the normal hierarchy is the exact source of the iterated exponential in the bound we obtain.

In several projects, some joint with Joel Hass, and others joint with Marc Lackenby, I am working on ways to overcome this iterated exponential blow-up associated with

normal hierarchies. The basic strategy for all of these projects involves the same central idea, which is that when cutting a triangulated 3-manifold along a normal surface, it is very inefficient to re-triangulate every piece of the cut-up tetrahedra. Instead it makes more sense to consider parts of the cut 3-manifold where the surface we cut along runs parallel to itself on their own. Figure 11 shows these parts of the cut 3-manifold in blue.

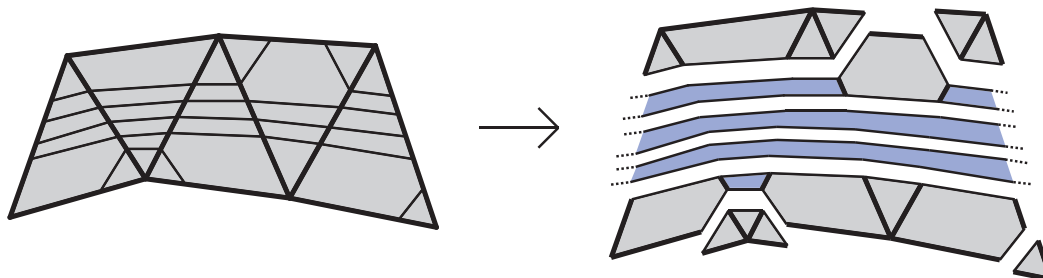


FIGURE 11. A better way to cut along a normal surface.

With this approach, the number of pieces obtained when cutting along a normal surface only goes up by a linear amount instead of exponentially. This means that the overall complexity of the normal hierarchy is still exponential, but no worse than exponential.

The projects I am working on with Joel Hass are focused on developing an improved theory of normal hierarchies that could be implemented in practice, for example in a future version of Regina. The run times for these future algorithms would still be exponential, but as Burton has shown, this is not prohibitive if care is taken in the implementation. We also hope to use this theory to show that certain problems in 3-manifold theory can be certified in polynomial time.

The projects I am working on with Marc Lackenby are of a slightly different flavor. We are focused on obtaining improved upper bounds like that in Theorem 2.2. For example, we have a plan for showing that any diagram of an alternating knot can be turned into an alternating diagram with a number of Riedemeister moves that is exponential in the number of crossings in the original diagram. We also have a plan to bound the number of Pachner moves, which are a natural set of moves for passing from one triangulation to another, required to pass from a triangulation of S^3 to a standard triangulation with a function that is polynomial in the number of tetrahedra in the initial triangulation.

4. CROSSING CHANGES

The previous two sections have been focused on problems, such as the recognition problem for knots, that are known to have an algorithmic solution, and obtaining improved bounds on the complexity of such problems. However there are many problems in knot and 3-manifold theory which are not known to have an algorithmic solution. In this section we examine some of the most important problems in this class, namely those that arise in the study of crossing changes.

One of the most natural ways that a knot can change is via a *crossing change*, which is the operation shown in Figure 12. It is a very simple operation whereby one passes the curve out of which the knot is made through itself exactly once, much like the way metal rings appear to pass through each other in the conjuror's ring trick.

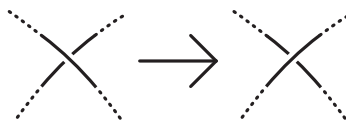


FIGURE 12. A crossing change.

Despite the wealth of knowledge that we have about knots and 3-manifolds, crossing changes are extraordinarily poorly understood and there are many elementary questions that we do not know the answer to. For example, the minimum number of crossing changes required to turn a knot K into the unknot is a measure of the knot's complexity called its *unknotting number*, denoted $u(K)$. Figure 13 shows a knot whose unknotting number is not known.

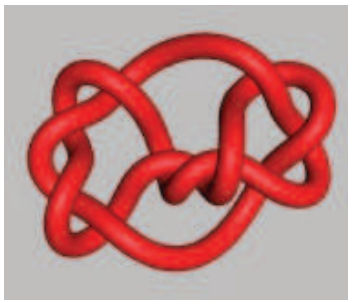


FIGURE 13. A knot with unknown unknotting number.

The reason that crossing changes and unknotting numbers are hard to understand is similar to the reason the recognition problem for knots is hard. If one wishes to determine whether two knots are the same, one is really asking if there is a continuous motion of \mathbb{R}^3 from one to the other. On the other hand the question of whether two knots are related by a single crossing change, for example, is also a question of whether there is a continuous motion taking one to the other, but where the continuous motion is allowed to pass the knot through itself exactly once.

The solution to the recognition problem for knots and links came about because Haken found a way of encoding whether two knots are the same using certain hierarchies of surfaces, and showed how to enumerate possibilities for those hierarchies using normal surface theory. This raises the question of how may one encode when two knots are related by a crossing change.

A *crossing circle* for a knot K is a simple closed curve C in the complement of K which bounds a disc that intersects K transversely in two points where K passes through with opposite orientation, as illustrated in Figure 14.

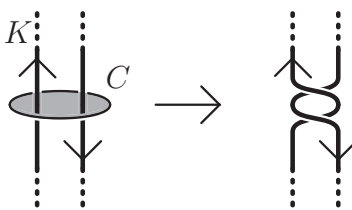


FIGURE 14. A crossing circle encodes a crossing change.

A crossing circle describes a crossing change specified by inserting a full turn to the two strands of K near the disk bounded by C , as shown in Figure 14. We call this the *crossing change specified by C* .⁸ The question of deciding whether two knots K and K' are related by a crossing change is best understood as whether there is a crossing circle for K specifying a crossing change that turns it into K' . This is hard to do because the crossing circle is a curve and not a surface. This means that one cannot use normal surface theory, or any of its adaptations, to search for crossing circles directly.

Crossing circles provide a natural way of saying when two crossing changes on a knot are essentially the same. We say two crossing changes for a knot K are *equivalent* if they are specified by crossing circles that are related by a continuous motion of the knot complement, and the handedness associated with the two crossing circles is the same. Marc Lackenby and I used this notion of equivalence in [6] where we proved the following theorem.

Theorem 4.1. *Suppose that K is a knot with genus one⁹ and unknotting number one. Then, if K is not the figure-eight knot, there is precisely one crossing change that turns K into the unknot, up to equivalence. If K is the figure-eight knot then there are precisely two crossing changes that turn K into the unknot, up to equivalence.*

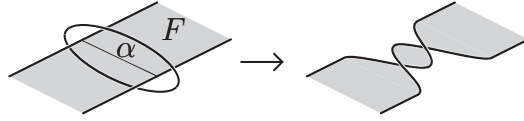
The proof of Theorem 4.1 fits into a major area of focus for my research, which is to find ways to encode crossing circles with surfaces, thereby making them amenable to using normal surface theory and other techniques for understanding surfaces. In a pair of papers [3, 4] I prove that crossing changes can very often be understood in terms of a closely related operation on surfaces called *twisting*.

A *twist* on a surface with boundary F is the operation shown in Figure 15. It involves drawing a properly embedded arc α on F , cutting F along α and then re-gluing F with a single twist inserted. Note that a crossing change is applied to the boundary of F , specified by the curve obtained by pushing α off F a little in all directions.

The problem of determining when two surfaces with boundary are related by a twist seems easier than determining when two knots are related by a crossing change. This is because the later amounts to searching for a curve in a 3-manifold, namely the knot complement, and the former amounts to searching for an arc on a surface. I

⁸Strictly speaking there are two crossing changes specified by C , depending on whether one applies a left-handed or a right-handed turn. Thus a crossing circle is endowed with a little extra information to make this choice.

⁹The *genus* of a knot is the minimum genus of any embedded orientable spanning surface for the knot. Thus, genus one knots bound embedded punctured tori.

FIGURE 15. Performing a twist along α

proved that very often a solution to this apparently easier problem implies that one can determine when two knots are related by a crossing change. The precise theorem I prove is the following:

Theorem 4.2. *Let K and K' be oriented knots in S^3 , both either hyperbolic or fibered, and with the genus of K being bigger than the genus of K' . Then there are finite lists of oriented spanning surfaces $\{S_1, \dots, S_n\}$ for K and $\{S'_1, \dots, S'_{n'}\}$ for K' such that if K and K' are related by a single crossing change, then some $S_i \in \{S_1, \dots, S_n\}$ and some $S'_i \in \{S'_1, \dots, S'_{n'}\}$ are related by a single twist, up to ambient isotopy of S^3 . Furthermore, there is an algorithm that will take diagrams for K and K' as input, and output such finite lists of spanning surfaces.*

Corollary 4.3. *Suppose there is an algorithm to determine whether two compact oriented surfaces in S^3 , each with a single boundary component, are related by a single twist, up to ambient isotopy of S^3 . Then there is an algorithm to determine if two given knots in S^3 of different genus and both either hyperbolic or fibered are related by a single crossing change. In particular there is an algorithm to determine whether a given hyperbolic or fibered knot in S^3 has unknotting number one.*

Finding an algorithm that will determine whether a given knot has unknotting number one has been a central ambition for most of my adult life. I hope that Corollary 4.3 is a stepping stone in this direction.

There are many ingredients for the proof of Theorem 4.2. These include objects called sutured manifolds, angle structures for ideal triangulations, circular Heegaard splittings and almost interiorly-normal surfaces. The details of the proof are beyond the scope of this statement, but there are three key points that one should take from this discussion. First, crossing changes are about as natural an operation as one could possibly wish to perform on knots, and yet they are very poorly understood, despite all we know about knots and 3-dimensional manifolds. Second, there is now a wealth of technology in knot and 3-manifold theory that one can try to apply to problems like determining when two knots are related by a crossing change, but doing so is often technically difficult because curves such as crossing circles naturally interact poorly with much of this technology, which is generally better at controlling surfaces rather than curves. Third, understanding crossing changes is extremely important because when DNA replicates it untangles itself with the help of enzymes that apply crossing changes to the double helix. So if we want to understand these most important of real world knots we need to vastly improve our understanding of crossing changes.

5. PHYSICAL KNOTS AND LINKS

So far in this statement we have regarded knots as 1-dimensional curves that are infinitely thin and infinitely stretchy. Of course, real world knots like DNA are not like this. They have a certain thickness, length and geometric shape. Such knots and links are called *physical knots and links*.

The two most fundamental theoretical problems concerning physical knots and links are to show the existence of a *Gordian Unknot* and a *Gordian Split Link*. A Gordian Unknot is a loop of fixed thickness and length whose core is unknotted, but which cannot be deformed to a round circle by a continuous motion fixing its length and thickness. A candidate for a Gordian Unknot appeared in work of Freedman, He and Wang [8], who studied certain energies associated to curves in \mathbb{R}^3 . This curve was studied by Pierański [12], who developed a computer program called SONO (Shrink On No Overlaps) to numerically shorten a curve of fixed thickness while avoiding overlaps. The program unexpectedly succeeded in unraveling the Freedman-He-Wang example. However there are more complicated examples that do fail to unravel under SONO, and hence give numerical evidence for the existence of Gordian Unknots. Figure 16 shows such an example of a conjectural Gordian Unknot.



FIGURE 16. A conjectural Gordian Unknot.

A Gordian Split Link is a pair of loops of fixed thickness whose core curves have the property that there is a continuous motion that takes the two components to opposite sides of a plane, but where this is not possible if one wishes to fix each component's length and thickness. Joel Hass and I [5] proved that such a link exists.

Theorem 5.1. *A Gordian Split Link exists.*

The proof of Theorem 5.1 is by a construction of a link, illustrated in Figure 17, whose components can be topologically but not physically split.

The proof of Theorem 5.1 returns us to a theme in this statement. To prove that there is no physical motion separating the components one must rule out any motion from an infinite space of possible motions. We have seen this before; when studying the recognition problem for knots and links one must be able to rule out any continuous motion joining two knots that are not the same, and when studying crossing changes one is interested in continuous motions that pass the knot or link through itself a certain number of times. In both these previous examples the general strategy is to try to encode possible continuous motions with surfaces.

Our proof of Theorem 5.1 continues this theme. We wish to show that there is simply not enough room inside the blue component to pass the orange component through. This is achieved by constructing a continuously moving family of spanning

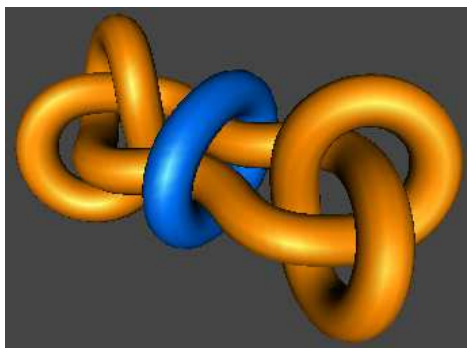


FIGURE 17. A Gordian Split Link.

disks for the blue component together with an area estimate for each disk in the family. We are then able to analyze the way the orange component intersects the disk at each time and show that if the two components become separated we contradict the area bound for the disk. I am currently trying to find a way to apply similar techniques to prove the existence of a Gordian Unknot such as that in Figure 16.

6. EPILOGUE

The figure shown on the front page of this statement shows a human chromosome. It consists of two DNA molecules, each about 4cm in length, wound up together. The extent to which the two molecules are tied up is absolutely breathtaking; if one were to straighten out all the DNA in a single adult human being one would obtain a strand longer than the orbit of Neptune. These words are hardly believable: Neptune travels around the Sun at about five and a half kilometers per second and it takes about 165 Earth years for it to complete a single orbit. That is the length of DNA that is wrapped up in the confines of our physical bodies, not to mention that only about one part in a thousand of our mass is made of DNA.

The remarkable nature of DNA doesn't stop at the sheer length. It also has two amazing properties. First, it expresses itself, giving living things their many unique and wonderful properties. Second, it replicates itself, allowing one generation of organism to pass on characteristics to the next generation. The basic chemical mechanisms that facilitate these two processes are the subject of an undergraduate course in microbiology, but these descriptions mask the full complexity of the mechanisms that are still shrouded in mystery to a large extent. These mechanisms are hard to understand because we do not have the technology to make a live, detailed movie of whole DNA molecules as they express and replicate, either in vivo or in vitro.

This situation is very similar to our understanding of electricity and the atom one hundred years ago. During the 19th century there was considerable progress in developing electrical technology like transformers and AC motors, both of which existed in a form that is recognizable today by the late 1800s. Further, Maxwell's theory of electromagnetism provided physicists with a sense that they understood much of the basic science. However, on a deeper level their understanding was limited because

they did not understand the atom; the Gold Leaf Experiment that established the existence of atomic nuclei, for example, did not take place until 1909. That observation, among others such as energy existing in discrete ‘quanta’, set the scene for quantum mechanics because it raised more questions than it answered. The ideas that electrons move around nuclei and that energy can be understood discretely did not transform the world overnight, but they did say, loud and clear: *This is important and we should try to understand this better.*

Returning to our understanding of microbiology today, we have enormously sophisticated drugs being developed by researchers at universities and drug companies to control all kinds of physiological and disease processes. Furthermore, techniques like X-ray crystallography mean that we have a very detailed understanding of the way many important molecules like proteins and nucleic acids interact with each other. However, just as physicists in the late 19th century were hampered by a lack of understanding of the atom, today microbiologists are hampered by a lack of understanding of many underlying mechanisms, in particular in relation to malignancy and the role of DNA topology and geometry. For example, as Koster *et al* observed in their 2007 letter to *Nature* [10], ‘Increasing the ability of chemotherapeutic drugs to kill cancer cells is often hampered by a limited understanding of their mechanism of action.’ In that paper they showed that a certain chemotherapeutic drug inhibits a special type of crossing change on DNA where one strand of the double helix passes over the other strand, a key process for DNA replication that is facilitated by enzymes called I topoisomerases. This work, which is just one example of extensive research in the microbiology community showing the importance of DNA topology and geometry, again does not answer every question we are interested in overnight, but the message is the same in respect of understanding DNA topology and geometry as the message from the Gold Leaf Experiment in respect of electron movement: *This is important and we should try to understand this better.*

With the benefit of hindsight we now know how important the mathematics of the 19th century was for the development of quantum mechanics. It provided the mathematical groundwork for building a theory that could be used to explain experimental observations governed by processes that could not be observed directly. Today we do not know exactly what the right mathematical theory is for understanding DNA topology and geometry, however it is almost certainly related, at least in some measure, to the central theme of this statement: building technology to understand motions of knots.

This technology, being developed by researchers in the mathematical, scientific and computing communities is not going to revolutionize the way we live today or tomorrow, but history suggests that it will, soon. The full extent to which my research fits into this broader picture remains to be seen. However even if my research turns out to have no practical use, I hope to have persuaded the reader that it is worthwhile mathematics nonetheless.

REFERENCES

- [1] Benjamin A. Burton, *Computational topology with Regina: Algorithms, heuristics and implementations*, Geometry and Topology Down Under, Proceedings of the American Mathematical Society, to appear.
- [2] Alexander Coward, *Algorithmically detecting bridge numbers of hyperbolic knots*, arxiv.org/abs/0710.1262.
- [3] ———, *Crossing changes and circular Heegaard splittings*, arxiv.org/abs/1210.5812.
- [4] ———, *Algorithms relating to circular Heegaard splittings*, in preparation.
- [5] Alexander Coward and Joel Hass, *Topological and physical link theory are distinct*, arxiv.org/abs/1203.4019.
- [6] Alexander Coward and Marc Lackenby, *Unknotting genus one knots*, Commentarii Mathematici Helvetici **86** (2011), 383–399.
- [7] ———, *An upper bound for Reidemeister moves*, arxiv.org/abs/1104.1882.
- [8] Michael H. Freedman, Zheng-Xu He, and Zhenghan Wang, *Möbius energy of knots and unknots*, Ann. of Math. (2) **139** (1994), no. 1, 1–50.
- [9] C. McA. Gordon and J. Luecke, *Knots are determined by their complements*, J. Amer. Math. Soc. **2** (1989), no. 2, 371–415.
- [10] Daniel A. Koster, Komaraiah Palle, Elisa S. M. Bot, Mary-Ann Bjornsti, and Nynke H. Dekker, *Antitumour drugs impede DNA uncoiling by topoisomerase I*, Nature **448** (2007), 213–217.
- [11] Sergei Matveev, *Algorithmic topology and classification of 3-manifolds*, Algorithms and Computation in Mathematics, vol. 9, Springer, 2003.
- [12] Piotr Pierański, *In search of ideal knots*, Ideal knots, Ser. Knots Everything, vol. 19, World Sci. Publ., River Edge, NJ, 1998, pp. 20–41.
- [13] Kurt Reidemeister, *Knotentheorie*, Springer, Berlin, 1932.