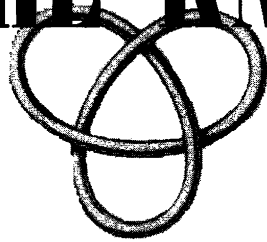
THE KNOT BOOK



An Elementary Introduction to the Mathematical Theory of Knots

Colin C. Adams

W. H. Freeman and Company
New York

Contents

Preface

Mathematics is an incredibly exciting and creative field of endeavor. Yet most people never see it that way. Nonmathematicians too often assume that we mathematicians sit around talking about what Newton did three hundred years ago or calculating a couple of extra million digits of π. They do not realize that more new mathematics is being created now than at any other time in the history of humankind.

Explaining the field of knot theory is a particularly effective way to dispel this misconception. Here is a field that is over one hundred years old, and yet some of the most exciting results have occurred in the last fifteen years. Easily stated open questions still abound, and one can get a taste for what it is like to do research very quickly. The other tremendous advantage that knot theory has over many other fields of mathematics is that much of the theory can be explained at an elementary level. One does not need to understand the complicated machinery of advanced areas of mathematics to prove interesting results.

My hope is that this book will excite people about mathematics—that it will motivate them to continue to explore other related areas of mathematics and to proceed to such topics as topology, algebra, differential geometry, and algebraic topology.

Unfortunately, mathematics is often taught as if the only goal were to pass a body of information from one person to the next. Although this is certainly an important goal, it is essential to teach an appreciation for the beauty of mathematics and a sense of the excitement of *doing* mathematics. Once readers are hooked, they will fill in the details themselves, and they will go a lot farther and learn a lot more.

Who, then, is this book for? It is aimed at anyone with a curiosity about mathematics. I hope people will pick up this book and start reading it on their own. I also hope that they will do the exercises: the only way to learn mathematics is to do it. Some of the exercises are straightforward; others take some thought. The very hardest are starred and can be a bit more challenging.

Scientists with primary interests in physics or biochemistry should find the applications of knot theory to these fields particularly fascinating. Although these applications have only been discovered recently, already they have had a huge impact.

This book can be and has been used effectively as a textbook in classes. With the exception of a few spots, the book assumes only a familiarity with high school algebra. I have also given talks on selected topics from this book to high school students and teachers, college students, and students as young as seventh graders.

The first six chapters of the book are designed to be read sequentially. With the exception that Section 8.3 depends on Section 7.4, the remaining four chapters are independent and can be read in any order. The topics chosen for this book are not the standard topics that one would see in a more advanced treatise on knot theory. Certainly the most glaring omission is any discussion of the fundamental group. My desire to make this book more interesting and accessible to an audience without advanced background has precluded such topics.

The choice of topics has been made by looking for areas that are easy to understand without much background, are exciting, and provide opportunities for new research. Some of the topics such as almost alternating knots are so new that little research has yet been done on them, leaving numerous open questions.

Although I drew on many sources while writing this book, I relied particularly heavily on the writings and approaches of Joan Birman, John Conway, Cameron Gordon, Vaughan Jones, Louis Kauffman, Raymond Lickorish, Ken Millett, Jozef Przytycki, Dale Rolfsen, Dewitt Sumners, Morwen Thistlethwaite, and William Thurston.

I would like to thank all the following colleagues, who contributed innumerable comments and suggestions during the writing of this book, and corrected many of the mistakes therein: Daniel Allcock, Thomas Banchoff, Timothy Bremer, Patrick Callahan, J. Scott Carter, Peter Cromwell, Alan Durfee, Dennis Garity, Jay Goldman, Cameron Gordon, Joel Hass, James Hoste, Vaughan Jones, Frank Morgan, Monica Nicolau, Jozef Przytycki, Joseph O'Rourke, Alan Reid, Yongwu Rong, Dewitt Sumners, Morwen Thistlethwaite, Abigail Thompson, and Jeffrey Weeks. The filmstrip format employed in the figures in Chapter 10 originated with J. Scott Carter.

I also would like to thank all the students who have contributed to this book. I was originally motivated to write this book through my participation in the SMALL Undergraduate Research Project at Williams College. Each summer since 1988, between fifteen and twenty-five students have come to Williams College to work on mathematical research with five to eight faculty over a ten-week period, through funds provided by the National Science Foundation, the New England Consortium for Undergraduate Science Education, Williams College, and other granting agencies. My group of students has usually worked on knot theory. Every summer, I would find myself teaching them the same material over again, without a reference at the right level. It was such beautiful material that I decided it would be worth writing a book. Although this list does not include all the students who have contributed to the book, I would like to thank the following: Aaron Abrams, Charene Arthur, Jeffrey Brock, Derek Bruneau, John Bugbee, Elizabeth Camp, Mark Chrisman, Tim Comar, Tara de Souza, Keith Faigin, Joseph Francis, Thomas Graber, Lisa Harrison, Daniel Heath, Martin Hildebrand, Hugh Howards, Amy Huston, Anne Joseph, Lisa Klein, Katherine Kollett, Joshua Kucera, John MacEachern, Lothar Mans, John Mynttinen, David Pesikoff, Jessica Polito, Dan Robb, William Sherman, John Terilla, Pinnarat Vongsinsirikul, and Edward Welsh.

Additional help came from Jeremiah Lyons, Marissa Barschdorf, and Pier Gustafson, Christine Hastings, and Mel Slugbate. Christine Heinitz and Thomas Banchoff deserve special credit for their work on the illustrations.

Colin C. Adams
February 1994

1 Introduction

1.1 Introduction

Take a piece of string. Tie a knot in it. Now glue the two ends of the string together to form a knotted loop. The result is a string that has no loose ends and that is truly knotted. Unless we use scissors, there is no way that we can untangle this string. (See Figure 1.1.)

Figure 1.1 Forming a knot from a piece of string.

A **knot** is just such a knotted loop of string, except that we think of the string as having no thickness, its cross-section being a single point. The knot is then a closed curve in space that does not intersect itself anywhere.

We will not distinguish between the original closed knotted curve and the deformations of that curve through space that do not allow the curve to pass through itself. All of these deformed curves will be considered to be the same knot. We think of the knot as if it were made of easily deformable rubber.

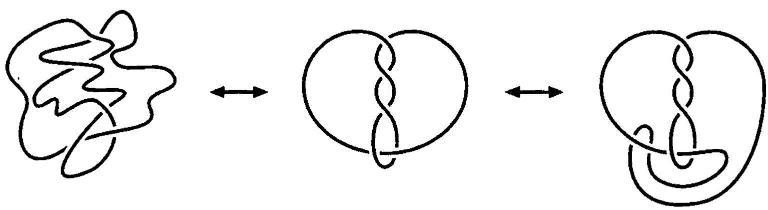


Figure 1.2 Deforming a knot doesn't change it.

In these pictures of knots (Figure 1.2) one section of the knot passes under another section at each crossing. The simplest knot of all is just the unknotted circle, which we call the **unknot** or the **trivial knot**. The next simplest knot is called a **trefoil knot**. (See Figure 1.3.) But how do we know these are *actually* different knots? How do we know that we couldn't untangle the trefoil knot into the unknot without using scissors and glue, if we played with it long enough?

a b

Figure 1.3 (a) The unknot. (b) A trefoil knot.

Certainly, if you make a trefoil knot out of string and try untangling it into the unknot, you will believe very quickly that it can't be done. But we won't be able to prove it until we introduce tricoloration of knots in Section 1.5.

In the table at the back of the book, there are numerous pictures of knots. All of these knots are known to be distinct. If we made any one of them out of string, we would not be able to deform it to look like any of the others. On the other hand, here is a picture (Figure 1.4) of a knot that is actually a trefoil knot, even though it looks completely different from the previous picture of a trefoil.

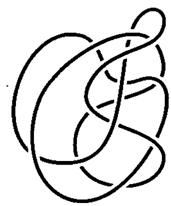


Figure 1.4 A nonstandard picture of the trefoil knot.

Exercise 1.1 Make this knot out of string and then rearrange it to show that it is the trefoil knot. (Actually, an electrical extension cord works better than string. You can tie a knot in it and then plug it into itself in order to form a knot. A third option is to draw a sequence of pictures that describe the deformation of the knot. This is particularly easy to do on a blackboard, with chalk and eraser. As you deform the knot, you can simply erase and redraw the appropriate sections of the picture.)

There are many different pictures of the same knot. In Figure 1.5, we see three different pictures of a new knot, called the **figure-eight knot.** We call such a picture of a knot a **projection** of the knot.

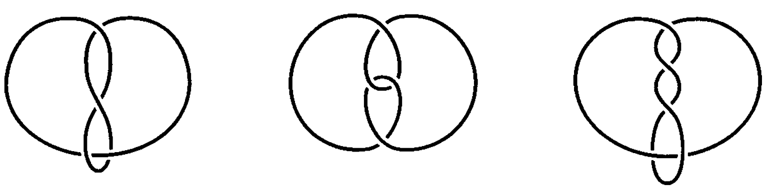


Figure 1.5 Three projections of the figure-eight knot.

The places where the knot crosses itself in the picture are called the **crossings** of the projection. We say that the figure-eight knot is a four-crossing knot because there is a projection of it with four crossings, and there are no projections of it with fewer than four crossings.

If a knot is to be nontrivial, then it had better have more than one crossing in a projection. For if it only has one crossing, then the four ends of the single crossing must be hooked up in pairs in one of the four ways shown in Figure 1.6. Any other projection with one crossing can be deformed to look like one of these without undoing the crossing. But each of these is clearly a trivial knot, as we can then untwist the single crossing.

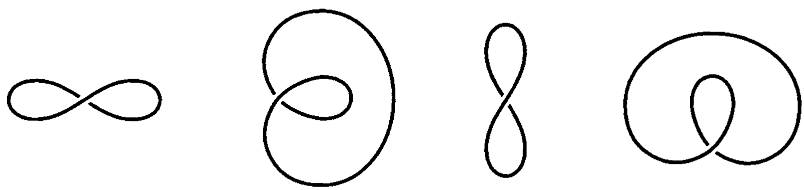


Figure 1.6 One-crossing projections.

Exercise 1.2 Show that there are no two-crossing nontrivial knots.

Much of knot theory is concerned with telling which knots are the same and which are different. One simplified version of this question is the following: "If we have a projection of a knot, can we tell whether it is the unknot?"

Certainly, if we play with a string model of the knot for a while and we do manage to untangle it completely, it is the unknot. But what if we play with it for two weeks and we still haven't untangled it? It still might be the unknot and for all we know, five more minutes of work might be enough to untangle it. So we can't quit.

But in fact, there is a way to decide if a given projection of a knot is the unknot. In 1961, Wolfgang Haken came up with a foolproof procedure for deciding whether or not a given knot is the unknot (see Haken, 1961). According to his theory, we should be able to give our projection of a knot to a computer (how to give a projection to computers is discussed in Chapter 2), and the computer would run the algorithm and tell us whether or not the given knot was the unknot. Unfortunately, even though Haken came up with his algorithm over 30 years ago, it is so complicated that no one has ever written a computer program to implement it.

Unsolved Problem

Write a computer program that can tell whether a knot that it is given is the unknot. (This is a difficult problem that requires a complete understanding of Haken's algorithm. But beware: His paper is 130 pages long!)

Aside: In 1974, Haken and Kenneth Appel solved one of the most famous problems in mathematics, the Four-Color Theorem. They proved that if you want to make a map, you only need to use four colors to make sure that no two countries of the same color touch each other along an edge. This was the first proof of a major theorem that used computers extensively to enumerate the thousands of cases that need to be examined. (See Haken, 1961).

Why should anyone be interested in knots? What's so important about being able to tell whether a tangled-up loop of string is truly tangled or can in fact be untangled without cutting and gluing?

Much of the early interest in knot theory was motivated by chemistry. In the 1880s, it was believed that a substance called ether pervaded all of space. In an attempt to explain the different types of matter, Lord Kelvin (William Thomson, 1824–1907) hypothesized that atoms were merely knots in the fabric of this ether. Different knots would then correspond to different elements (Figure 1.7).

This convinced the Scottish physicist Peter Guthrie Tait (1831–1901) that if he could list all of the possible knots, he would be creating a table of the elements. He spent many years tabulating knots. At the same time, an American mathematician named C. N. Little was working on his own tabulations for knots.

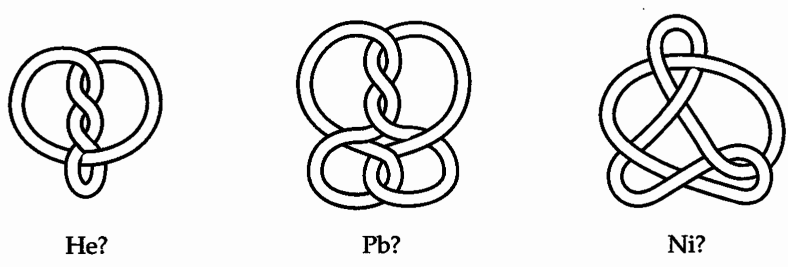


Figure 1.7 Atoms are knotted vortices?

Unfortunately, Kelvin was wrong. A more accurate model of atomic structure appeared at the end of the nineteenth century and chemists lost interest in knots for the next 100 years. But in the meantime, mathematicians had become intrigued with knots. A century of work on the mathematical theory of knots followed.

Interestingly enough, in the 1980s, biochemists discovered knotting in DNA molecules. Concurrently, synthetic chemists realized it might be possible to create knotted molecules, where the type of knot determined the properties of the molecule. A mathematical field that was born out of a misguided model for atoms has turned out to have several significant applications to chemistry and biology. We discuss these applications in Chapter 7.

Knot theory is a subfield of an area of mathematics known as **topology**. Topology is the study of the properties of geometric objects that are preserved under deformations. Just as we think of the knots as being made of deformable rubber, so we think of the more general geometric objects in topology as deformable. For instance, a topologist does not distinguish a cube from a sphere, since a cube can be deformed into a sphere by rounding off the eight corners and smoothing the twelve edges, as in Figure 1.8.

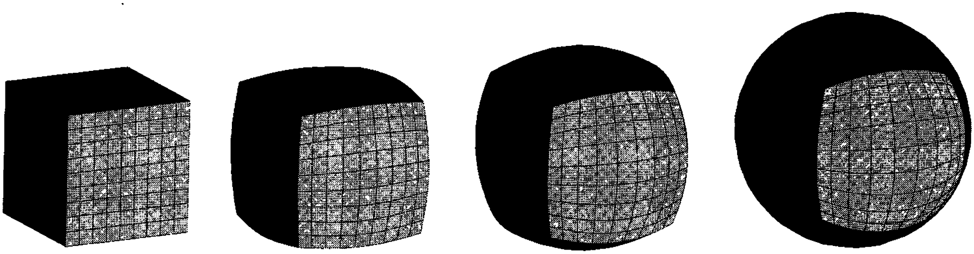


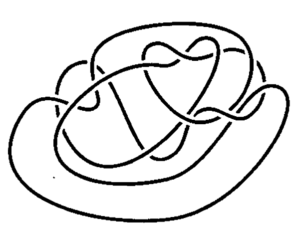
Figure 1.8 A cube and a sphere are the same in topology.

Topology is one of the major areas of research in mathematics today. Work in knot theory has led to many important advances in other areas of topology. We discuss some of these connections in Chapter 9.

In this book, we investigate the mathematical theory of knots. The emphasis is on current research in knot theory. Unlike the situation in some other fields of mathematics, many of the unsolved problems in knot theory are easily stated. Much of the theory is accessible to someone without any background in upper-level mathematics. There are open problems in the field that can be attacked and perhaps solved by nonexperts.

The best way to learn any kind of mathematics is by doing mathematics, not just by reading about what others have done. Therefore, throughout this book there are numerous open problems in knot theory. Try them! Think to yourself, "How would I solve this problem?" Maybe you can come up with the essential new idea and discover the solution.

Exercise 1.3 Use string (or an extension cord) to show that the following knot is the unknot.



Exercise 1.4 Show that any knot has a projection with over 1000 crossings.

Certain types of knots are particularly interesting. One such type is an alternating knot. An **alternating knot** is a knot with a projection that has crossings that alternate between over and under as one travels around the knot in a fixed direction. The trefoil knot in Figure 1.3 is alternating. So is the figure-eight knot in Figure 1.5, since the two projections of it on the left and middle are alternating.

Exercise 1.5 Choose crossings at each vertex in Figure 1.9 to make the resulting knot alternating.

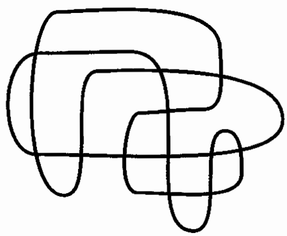


Figure 1.9 A projection without over- and undercrossings.

*Exercise 1.6** Show that by changing the crossings from over to under or vice versa, any projection of a knot can be made into the projection of an alternating knot. (This isn't as easy as it might seem. How do you know your procedure will always work?) In a projection with n crossings, what is the maximum number of crossings that would have to be changed in order to make the knot alternating?

*Exercise 1.7** Show that by changing some of the crossings from over to under or vice versa, any projection of a knot can be made into a projection of the unknot.

1.2 Composition of Knots

Given two projections of knots, we can define a new knot obtained by removing a small arc from each knot projection and then connecting the four endpoints by two new arcs as in Figure 1.10. We call the resulting knot the **composition** of the two knots. If we denote the two knots by the symbols J and K, then their composition is denoted by $J\#K$. We assume that the

* Exercise with asterisk denotes more difficult problem.

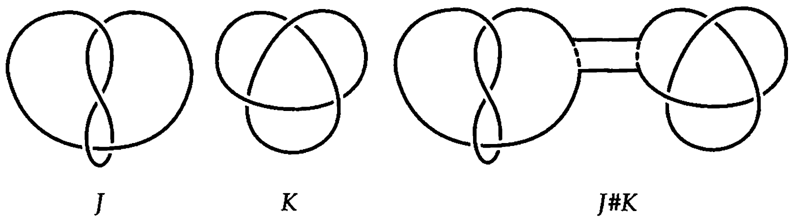


Figure 1.10 The composition $J\#K$ of two knots J and K.

two projections do not overlap, and we choose the two arcs that we remove to be on the outside of each projection and to avoid any crossings. We choose the two new arcs so they do not cross either the original knot projections or each other (Figure 1.11).

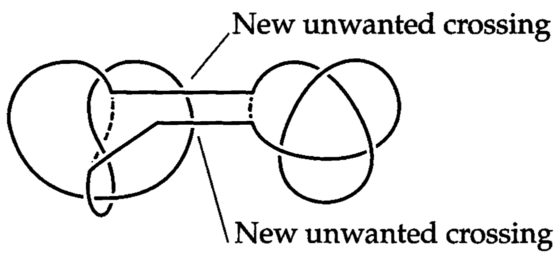


Figure 1.11 *Not* the composition of J and K.

We call a knot a **composite knot** if it can be expressed as the composition of two knots, neither of which is the trivial knot. This is in analogy to the positive integers, where we call an integer composite if it is the product of positive integers, neither of which is equal to 1. The knots that make up the composite knot are called **factor knots.**

Notice that if we take the composition of a knot K with the unknot, the result is again K, just as when we multiply an integer by 1, we get the same integer back again (Figure 1.12). If a knot is not the composition of

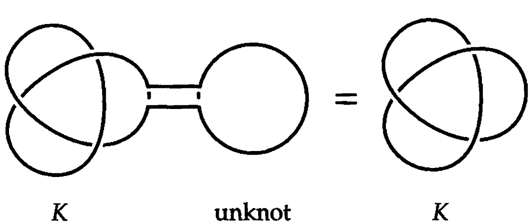


Figure 1.12 $K\#$(unknot) is just K.

any two nontrivial knots, we call it a **prime knot**. Both the trefoil knot and the figure-eight knot are prime knots, although this is not obvious.

For the knot *J*#*K* in Figure 1.10, it is clearly composite. We constructed it to be. But how about the knot in Figure 1.13? Is it composite? In fact, it is. If you make it out of string and play around with the knot, you can eventually get it into a projection that shows that it is composite.

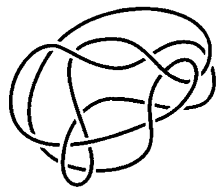


Figure 1.13 A potentially composite knot.

Here's a stranger question. Is the unknot composite? Obviously, from the picture in Figure 1.3a, it doesn't look composite. But maybe there is a way to tangle the unknot up so that we get a projection of it that makes it obviously a composite knot. That is, perhaps there is a picture of the unknot that has a nontrivial knot on the left, a nontrivial knot on the right, and two strands of the knot joining them (Figure 1.14). Maybe that part of the projection corresponding to the knot on the right somehow untangles that part of the projection corresponding to the knot on the left, resulting in the unknot.

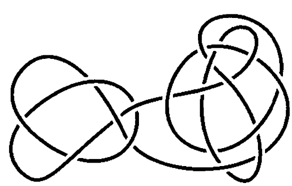


Figure 1.14 Could this untangle to be the unknot?

It's somewhat disconcerting to realize that if the unknot were a composite knot, then every knot would be a composite knot. Since every knot is the composition of itself with the unknot, every knot would be the composition of itself with the nontrivial factor knots that made up the unknot.

In fact, much to our relief, the unknot is not a composite knot. There is no way to take the composition of two nontrivial knots and get the unknot. We use surfaces to show this in Section 4.3. We can think of this result as analogous to the fact that the integer 1 is not the product of two positive integers, each greater than 1. Moreover, just as an integer factors into a unique set of prime numbers, a composite knot factors into a unique set of prime knots.

The appendix table, which contains projections of knots, and is located at the back of the book, lists only the prime knots and does not include any composite knots. It's like a table of prime numbers. Although all the positive integers aren't listed, any integer can be constructed by taking the appropriate product of the primes that are listed.

Exercise 1.8 Using the appendix table, identify the factor knots that make up the composite knot in Figure 1.15.

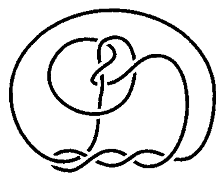


Figure 1.15 A composite knot.

Exercise 1.9 Show that the knot in Figure 1.16 is composite.

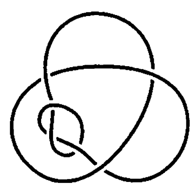


Figure 1.16 Another composite knot.

One way that composition of knots does differ from multiplication of integers is that there is more than one way to take the composition of two knots. We have a choice of where we remove the arc from the outside of each projection. Will this choice affect the outcome? Surprisingly, the answer is yes. It is often possible to construct two different composite knots from the same pair of knots J and K.

We first need to put an orientation on our knots. An **orientation** is defined by choosing a direction to travel around the knot. This direction is denoted by placing coherently directed arrows along the projection of the knot in the direction of our choice. We then say that the knot is **oriented**.

When we then form the composition of two oriented knots J and K, there are two possibilities. Either the orientation on J matches the orientation on K in $J\#K$, resulting in an orientation for $J\#K$, or the orientation on J and K do not match up in $J\#K$. All of the compositions of the two knots where the orientations do match up will yield the same composite knot. All of the compositions of the two knots where the orientations do not match up will also yield a single composite knot; however, it is possibly distinct from the composite knot generated when the orientations do match up (Figure 1.17).

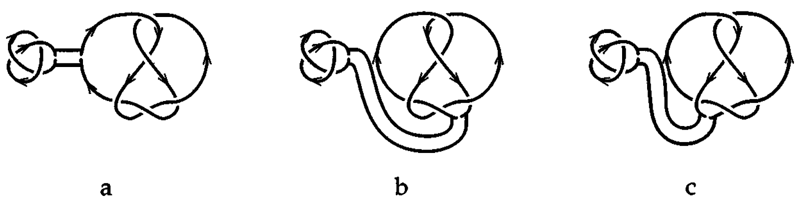


Figure 1.17 (a) Orientations match. (b) Orientations match. (c) Orientations differ.

To convince ourselves that the first two compositions in Figure 1.17 really do give us the same knot, we can shrink J down in the first picture and then slide it around K until we obtain the second picture (Figure 1.18). Although this will not be the case in general, in this particular example, the third composition in Figure 1.17 also gives the same knot as the two preceding compositions. This occurs because one of the factor knots is **invertible.** A knot is invertible if it can be deformed back to itself so that an orientation on it is sent to the opposite orientation. In the case that one of the two knots is invertible, say J, we can always deform the composite knot so that the orientation on K is reversed, and hence so that the orientations of J and K always match. Therefore, there is only one composite knot that we can construct from the two knots.

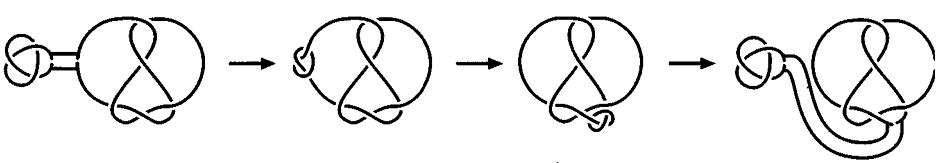


Figure 1.18 Two compositions that are the same.

The first knot that is not invertible in the table at the end of the book is the knot 8_{17}. Composing it with itself in the two different ways produces two distinct composite knots that are not equivalent (Figure 1.19). In order to determine the possible compositions of knots, we need to know which knots are invertible. So far, no one has come up with a general technique that will determine whether or not a given knot is invertible.

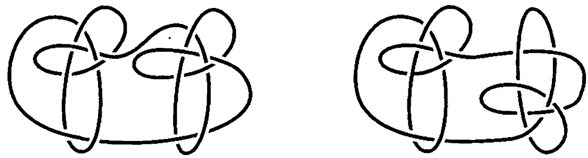


Figure 1.19 These two composite knots have the same factors, but they are distinct.

1.3 Reidemeister Moves

Suppose that we have two projections of the same knot. If we made a knot out of string that modeled the first of the two projections, then we should be able to rearrange the string to resemble the second projection. Knot theorists call the rearranging of the string, that is, the movement of the string through three-dimensional space without letting it pass through itself, an **ambient isotopy**. The word "isotopy" refers to the deformation of the string. The word "ambient" refers to the fact that the string is being deformed through the three-dimensional space that it sits in. Note that in an ambient isotopy, we are not allowed to shrink a part of the knot down to a point, as in Figure 1.20, in order to be rid of the knot. It's easiest to think of a knot made of string. Just as you can't get rid of a knot in a string by pulling it tighter and tighter, so an ambient isotopy doesn't allow us to get rid of a knot in this manner.

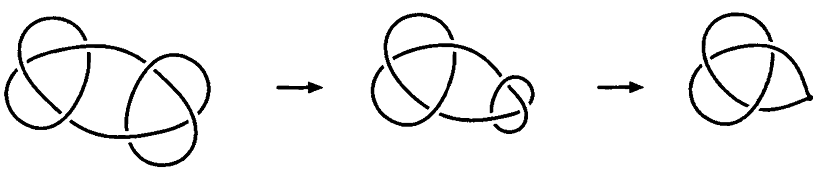


Figure 1.20 We are *not* allowed to shrink part of the knot to a point.

A deformation of a knot projection is called a **planar isotopy** if it deforms the projection plane as if it were made of rubber with the projection drawn upon it (Figure 1.21). The word "planar" is used here because we are only deforming the knot within the projection plane. Keep in mind that this is highly deformable rubber.

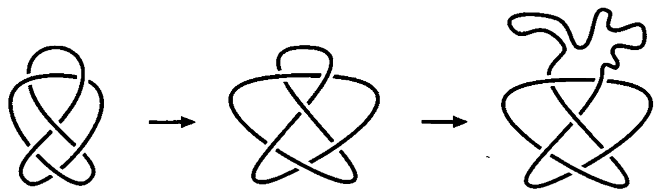


Figure 1.21 Planar isotopies.

A **Reidemeister move** is one of three ways to change a projection of the knot that *will* change the relation between the crossings. The **first Reidemeister move** allows us to put in or take out a twist in the knot, as in Figure 1.22. We assume that the projection remains unchanged except for the change depicted in the figure. The **second Reidemeister move** allows us to either add two crossings or remove two crossings as in Figure 1.23. The **third Reidemeister move** allows us to slide a strand of the knot from one side of a crossing to the other side of the crossing, as in Figure 1.24.

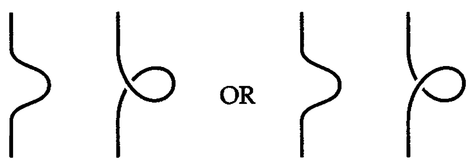


Figure 1.22 Type I Reidemeister move.

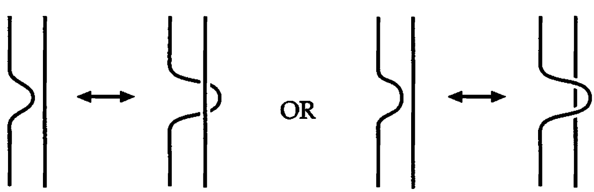


Figure 1.23 Type II Reidemeister move.

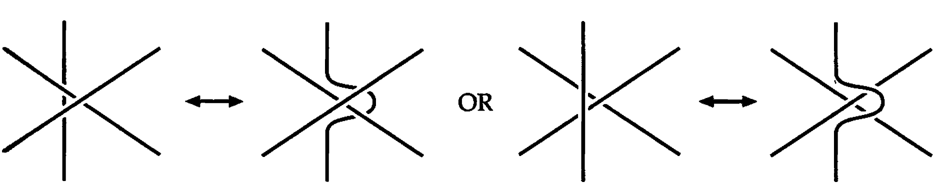


Figure 1.24 Type III Reidemeister move.

Notice that although each of these moves changes the projection of the knot, it does not change the knot represented by the projection. Each such move is an ambient isotopy.

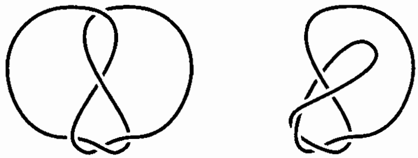


Figure 1.25 Two projections of the same knot.

In 1926,. the German mathematician Kurt Reidemeister (1893–1971) proved that if we have two distinct projections of the same knot, we can get from the one projection to the other by a series of Reidemeister moves and planar isotopies. For example, the two projections in Figure 1.25 correspond to the same knot. Therefore, according to Reidemeister, there is a series of Reidemeister moves that takes us from the first projection to the second. Figure 1.26 shows one series of moves that demonstrates this equivalence. As another example, the figure-eight knot is known to be

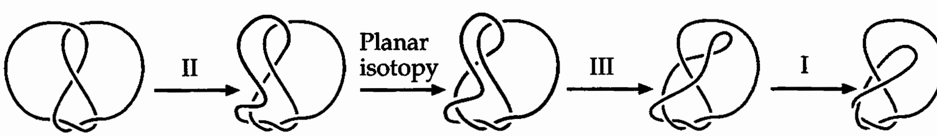


Figure 1.26 Reidemeister moves.

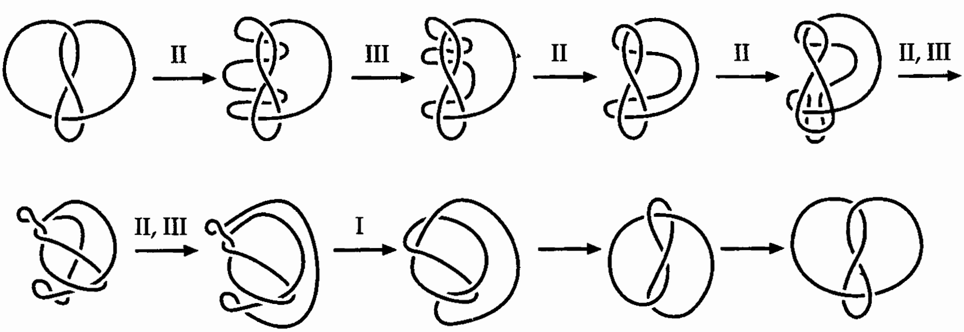


Figure 1.27 The figure-eight knot is equivalent to its mirror image.

equivalent to its mirror image, that is, the knot obtained by changing every crossing in the figure-eight knot to the opposite crossing. In Figure 1.27, we see the Reidemeister moves that show the equivalence. Incidentally, a knot that is equivalent to its mirror image is called **amphicheiral** by mathematicians and **achiral** by chemists. Although the knot tables do not list both a knot and its mirror image, we consider them to be distinct knots unless the knot is amphicheiral. More on amphicheirality in Chapter 7.

Exercise 1.10 Show that the two projections in Figure 1.28 represent the same knot by finding a series of Reidemeister moves from one to the other.

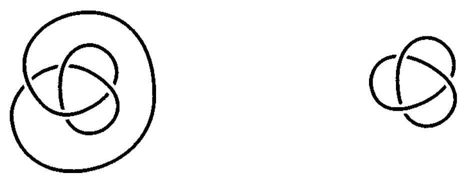


Figure 1.28 Find the Reidemeister moves.

The proof that Reidemeister moves and planar isotopy suffice to get us from any one projection of a knot to any other projection of that knot is not particularly difficult; however, it is technically involved, so we will not go into it here. A proof appears in Burde and Zeischang (1986). It might now seem that the problem of determining whether two projections represent the same knot would be easy. We just check whether or not there is a sequence of Reidemeister moves to get us from the one projection to the other. Unfortunately, there is no limit on the number of Reidemeister moves that it might take us to get from one projection to the other. If the two original projections have 10 crossings each, it is conceivable that in the process of performing the Reidemeister moves we will have to increase the number of crossings to 1000, before the moves simplify the projection back down to 10 crossings. For instance, the trefoil knot is not amphicheiral, but there is no known proof in terms of Reidemeister moves. Even if we could prove that we cannot get from the standard projection of the trefoil knot to its mirror image in 1,000,000,007 Reidemeister moves, maybe we could do it with 1,000,000,008 moves.

Here is an interesting example. Believe it or not, this is a projection of the unknot, so there has to be a series of Reidemeister moves that untangles it into an unknotted circle.

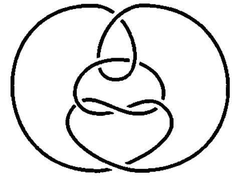


Figure 1.29 A nasty unknot.

*Exercise 1.11** Find a sequence of Reidemeister moves to untangle the unknot shown in Figure 1.29. (Note that this problem is asking a lot more than just showing that this knot can be untangled.)

Exercise 1.12 Prove that in Exercise 1.11, in any sequence of Reidemeister moves that unknot the projection with seven crossings in Figure 1.29, it is necessary to pass through a projection with more than seven crossings.

Unsolved Question

Could there be a constant c such that for any knot K and for any two projections P_1 and P_2 of K, each with no more than n crossings, one can get from one projection to the other by Reidemeister moves without ever having more than n + c crossings at any intermediate stage? It is highly unlikely such a constant exists; however, I know of no set of examples that demonstrate its nonexistence.

Note that even if such a c does not exist, it might be true that the increase in the number of crossings is never more than a simple function of n, say, the increase is never more than $2n+3$ or $3n^2-n+7$. Or perhaps you can find a sequence of examples that proves that the increase in the number of crossings is sometimes greater than any function of the form $ax+b$, where a and b are constants. (In mathematical parlance, you would have shown that there is no linear bound on the crossing increase.) Or perhaps there is a sequence of examples that shows that the crossing increase is sometimes greater than any polynomial in n. This would prove that the crossing increase is sometimes "exponential."

1.4 Links

So far, we have restricted our attention to knots; that is to say, single knotted loops. But there was no reason to say that there could only be one loop that we knotted.

A **link** is a set of knotted loops all tangled up together. Two links are considered to be the same if we can deform the one link to the other link without ever having any one of the loops intersect itself or any of the other loops in the process. Here are two projections of one of the simplest links, known as the **Whitehead link** (Figure 1.30).

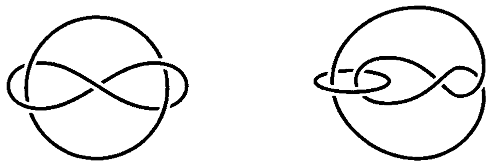


Figure 1.30 Two projections of the Whitehead link.

Exercise 1.13 Show that the two projections represent the same link.

Since it is made up of two loops knotted with each other, we say that it is a **link of two components**. Here is another well-known link with three components, called the **Borromean rings** (Figure 1.31). This link is named after the Borromeas, an Italian family from the Renaissance that used this pattern of interlocking rings on their family crest.

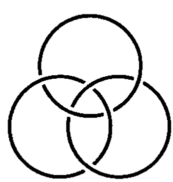


Figure 1.31 The Borromean rings.

A knot will be considered a link of one component. The table at the back of the book contains projections of some of the simpler links. Pretty much everything we have said about knots holds true for links. For instance, if two projections represent the same link, there must be a sequence of Reidemeister moves to get from the one projection to the other.

A link is called **splittable** if the components of the link can be deformed so that they lie on different sides of a plane in three-space. Sometimes it's obvious when a link is splittable, as in the first link in Figure 1.32. However, it's often the case that a link is splittable, but we can't easily tell that by looking at the projection, as in the second link in the figure.

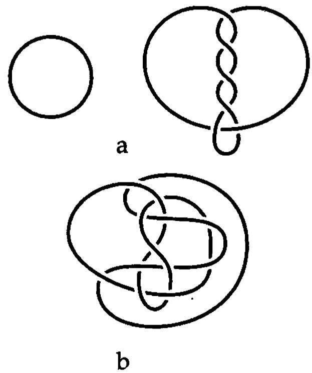


Figure 1.32 Two splittable links.

Exercise 1.14 Show that the second link is splittable.

Most of the links that we will be interested in are nonsplittable. There is one quick way for telling certain links apart: just count the number of components in the link. If the numbers are different, the two links have to be different. So obviously, the trefoil knot, the Whitehead link, and the Borromean rings all have to be distinct links.

If we have two projections of links, each with the same number of components, just as for knots, we would like to be able to tell if they represent the same link. In Figure 1.33, we show the two simplest links of two components. We call the first of these the **unlink** (or **trivial link**) **of two components** and the second the **Hopf link**. One difference between these two links is that the unlink is splittable, since its two components can be separated by a plane. But in the Hopf link, the two components do link each other once. We would like a method for measuring numerically how linked up two components are. We will define what's known as the **linking number**.

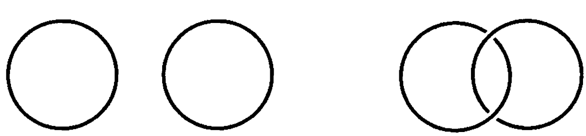


Figure 1.33 The unlink of two components and the Hopf link.

Let M and N be two components in a link, and choose an orientation on each of them. Then at each crossing between the two components, one of the pictures in Figure 1.34 will hold. We count a $+1$ for each crossing of the first type, and a -1 for each crossing of the second type. Sometimes it is hard to determine from the picture whether a crossing is of the first type or the second type. Note that if a crossing is of the first type, then rotating the understrand clockwise lines it up with the overstrand so that their arrows match. Similarly, if a crossing is of the second type, then rotating the understrand counterclockwise lines the understrand up with the overstrand so that their arrows match.

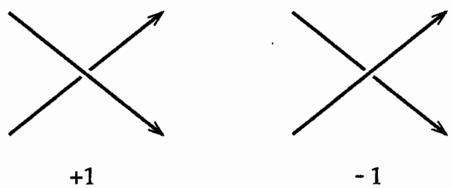


Figure 1.34 Computing linking number.

Now, we will take the sum of the $+1$s and -1s over all the crossings between M and N and divide this sum by 2. This will be the linking number. We do not count the crossings between a component and itself. For the unlink, the linking number of the two components is 0. For the Hopf link, the linking number will be 1 or -1, depending on the orientations on the two components. The two components in the oriented link pictured in Figure 1.35 have linking number 2. Notice that if we reverse the orientation on one of the two components, but not the other, the linking number of these two components is multiplied by -1. If we just look at the absolute value of the linking number, however, it is independent of the orientations on the two components.

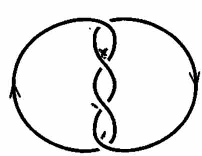


Figure 1.35 Linking number 2.

Exercise 1.15 Compute the linking number of the link pictured in Figure 1.36. Now reverse the direction on one of the components and recompute it.

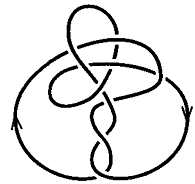


Figure 1.36 Compute the linking number.

Notice that we use a particular projection of the link in order to compute the linking number. In fact, we can show that the computed linking number will always be the same, no matter what projection of the link we use to compute it. We show this by proving that the Reidemeister moves do not change the linking number. Since we can get from any one projection of a link to any other via a sequence of Reidemeister moves, none of which will change the linking number, it must be that two different projections of the same link yield the same linking number.

Let's first look at the effect of the first Reidemeister move on the linking number. It can create or eliminate a self-crossing in one of the two components, but it will not affect the crossings that involve both of the components, so it leaves the linking number unchanged. Now, let's see what a Type II Reidemeister move does. In Figure 1.37 we have chosen a certain orientation on the strands of the link. We are assuming that the two strands correspond to the two different components, because otherwise the move has no effect on linking number. One of the new crossings contributes a +1 to the sum, and the other crossing contributes a −1, so the net contribution to the linking number is 0. Even if we change the orientation on one of the strands, we will still have one +1 and one −1 contribution, so Type II moves leave the linking number unchanged.

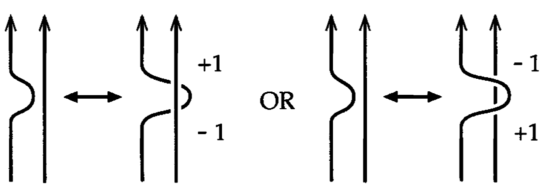


Figure 1.37 Type II Reidemeister moves don't affect linking number.

Finally, what about Type III moves? Once orientations are chosen for each of the three strands and +1s and −1s are assigned to each of the crossings, it is clear that sliding the strand over in the Type III move doesn't change the number of +1s or −1s, and so the linking number is preserved (Figure 1.38).

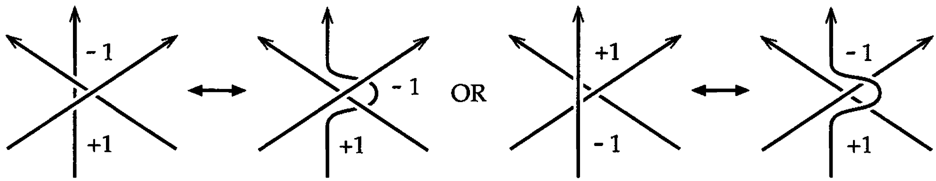


Figure 1.38 Type III Reidemeister moves don't affect linking number.

We say that the linking number is an **invariant** of the oriented link, that is, once the orientations are chosen on the two components of the link, the linking number is unchanged by ambient isotopy. It remains invariant when the projection of the link is altered. This is one of many invariants we will look at. Another invariant of links that we have already mentioned is simply the number of components in the link. It is unchanged by ambient isotopies of the link.

Exercise 1.16 Explain why the linking number of a splittable two-component link will always be 0, no matter what projection is used to compute it.

We can use linking number to distinguish links. Since we want to distinguish links that do not already have orientations on them, we will use the absolute value of the linking number. Any two links with two components that have distinct absolute values of their linking numbers have to be different links. For instance, the trivial link of two components has linking number 0. But the absolute value of the linking number of the Hopf link is 1, so the Hopf link cannot be the trivial link.

Exercise 1.17 Compute the absolute values of the linking numbers of the two links shown in Figure 1.39 in order to show that they must be distinct links.

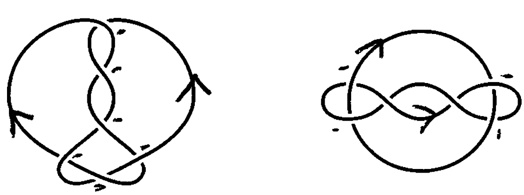


Figure 1.39 Compute the linking numbers.

So now you're thinking, "Well, at least we can tell all links apart." Unfortunately, life and links aren't that simple. Try computing the linking number for the Whitehead link in Figure 1.30. It has linking number

0, just like the trivial link of two components. So we can't even show that the Whitehead link is different from the trivial link of two components. We need some other ways to distinguish various knots and links. In the next section, we will see one such way. But first let's take another look at the Borromean rings (Figure 1.40). Note that if we removed any one of the three components of this link, the remaining two components would become two trivial unlinked circles. The fact that these three rings are locked together relies on the presence of all three components.

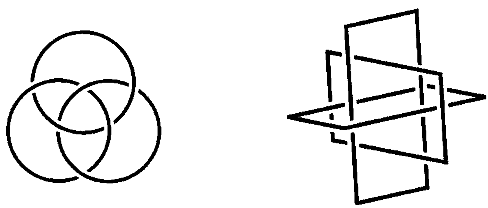


Figure 1.40 Two pictures of the Borromean rings.

A link is called **Brunnian** if the link itself is nontrivial, but the removal of any one of the components leaves us with a set of trivial unlinked circles. These links are named after Hermann Brunn, who drew pictures of such links back in 1892.

*Exercise 1.18** Find a Brunnian link of four components.

*Exercise 1.19** Find Brunnian links with arbitrarily many components.

Exercise 1.20 Make up your own conjecture about Brunnian links. Then see if you can prove it. (For example, can there be a Brunnian link such that each component is a round flat circle? What about ellipses? Think up your own.)

1.5 Tricolorability

We have talked a lot about telling knots and links apart, but actually we have not yet shown the most basic fact of knot theory. We have *not yet proved that there is any other knot besides the unknot.* For all we know right now, every projection of a knot in the table at the end of the book could simply be a messy projection of the unknot. Maybe every one of those projections can be turned into the projection of the unknot through a series of Reidemeister moves. The point is that of course they can't be, but we need some way to show this. So we will prove that there is *at least* one other knot besides the unknot. We will prove that the trefoil knot is not equivalent to the unknot. In order to do that, we need to introduce the idea of **tricolorability**.

We will say that a **strand** in a projection of a link is a piece of the link that goes from one undercrossing to another with only overcrossings in between. We will say that a projection of a knot or link is **tricolorable** if each of the strands in the projection can be colored one of three different colors, so that at each crossing, either three different colors come together or all the same color comes together. In order that a projection be tricolorable, we further require that at least two of the colors are used. Figure 1.41 shows that these two projections of the trefoil knot are tricolorable (using white, gray, and black as the colors).

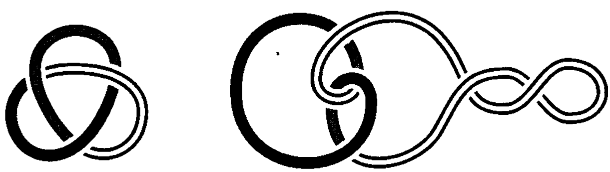


Figure 1.41 The trefoil is tricolorable.

In the first tricoloration, three different colors come together at each crossing, whereas in the second tricoloration, some of the crossings have only one color occurring. But none of the crossings in either picture have exactly two colors occurring, so these are valid tricolorations.

Exercise 1.21 Determine which of the projections of the three six-crossing knots 6_1, 6_2, and 6_3 in Figure 1.42 are tricolorable.

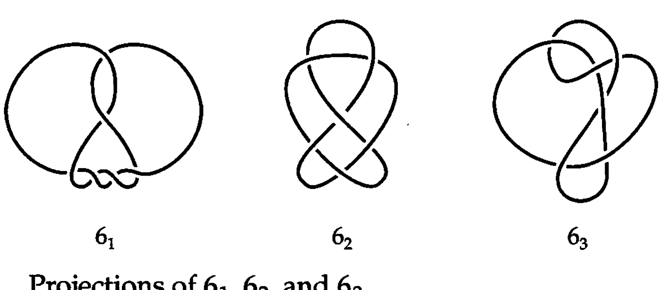


Figure 1.42 Projections of 6_1, 6_2, and 6_3.

Exercise 1.22 Show that the projection of the knot 7_4 in Figure 1.43 is tricolorable.

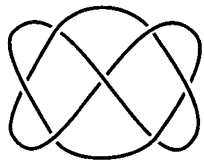


Figure 1.43 Show that this knot projection is tricolorable.

For our purposes, the most important fact is that if a projection of a knot is tricolorable, then the Reidemeister moves will preserve the tricolorability. If we do a Type I move and introduce a crossing, we can just leave all the strands involved the same color, and the new crossing will satisfy the requirements for tricolorability. Similarly, removing a crossing by a Type I move preserves tricolorability. If we do a Type II move to introduce two new crossings, and the two original strands are different colors, we can just change the color of the new strand to the third color and the resulting knot projection is tricolorable. If the two original strands are the same color, we can leave the new strand and the new crossings all that same color.

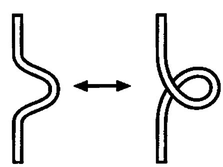


Figure 1.44 Type I moves preserve tricolorability.

Similarly, using a Type II move to reduce the number of crossings by two will also preserve tricolorability. Either all of the strands that appear in the diagram for the Reidemeister move are the same color, in which case we can color the strands that result from the Reidemeister move that same color, or three distinct colors come together at each of the two crossings, in which case we can color the two resulting strands as in Figure 1.45b. Note that in both these cases, since the original projection was colored with at least two distinct colors, the resulting projection will also be colored with at least two colors.

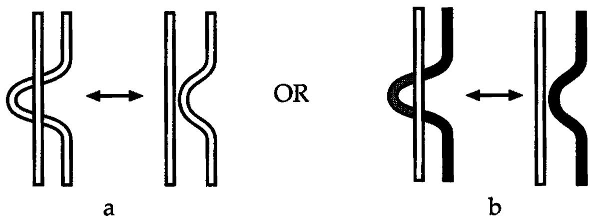


Figure 1.45 Type II moves preserve tricolorability.

Exercise 1.23 Show that the Type III Reidemeister move preserves tricolorability. (There are several cases to check.)

Therefore, since Reidemeister moves leave tricolorability unaffected, whether or not a projection is tricolorable depends only on the knot given by the projection. *Either every projection of a knot is tricolorable or no projection of that knot is tricolorable.* For instance, every projection of the trefoil knot is tricolorable. Since the usual projection of the unknot is not tricolorable (we certainly can't use at least two colors on it since it doesn't have distinct strands), it must be the case that the trefoil knot and the unknot are distinct.

We have just shown there is at least one other knot besides the unknot. In fact, any knot that is tricolorable must be distinct from the unknot.

Exercise 1.24 Determine which of the seven-crossing knots in the table at the end of the book are tricolorable.

Exercise 1.25 Show that the composition of any knot with a tricolorable knot yields a new tricolorable knot.

Exercise 1.26 Find an infinite set of tricolorable knots that are not obviously composite. (If a knot has a crossing in the tricoloration that has only one color, you can replace the crossing with a more complicated tangle. You needn't prove that the knots that you describe are actually different knots.)

Thus, many knots can be shown to be nontrivial using tricolorability. We can, in fact, conclude that any tricolorable knot must be distinct from any knot that is not tricolorable.

Exercise 1.27 Give an argument that shows that the figure-eight knot is not tricolorable. Conclude that the figure-eight knot and the trefoil knot are distinct knots.

Unfortunately, even if we can show that the figure-eight knot is not the same as the trefoil knot, tricolorability cannot be used to show that the figure-eight knot is nontrivial.

*Exercise 1.28** (a) Label the strands of the figure-eight knot with a selection of integers from the set 0, 1, 2, 3, and 4 so that they satisfy $x + y - 2z = 0 \pmod 5$ at each crossing, where z labels the overstrand. (That is to say, the remainder is 0 when $x + y - 2z$ is divided by 5.) Then show that such a labeling system on a knot projection is preserved under Reidemeister moves (Type III is the tricky one). Conclude that the figure-eight knot is not the trivial knot. (An argument is needed, even for this last step.)

(b) Reinterpret tricoloration in terms of a numerical scheme like the one we just applied to the figure-eight knot.

By Exercise 1.25, we know that the composition of the trefoil knot with any other knot is tricolorable. This proves that the unknot cannot be the composition of the trefoil knot with any other knot.

Unsolved Question

Is there a way to generalize tricolorability in order to show that the unknot is not a composition of any two nontrivial knots? Although we will see a proof of this fact later, the goal of this unsolved question is to find a simpler proof.

Tricolorability for links of two components is slightly different (Figure 1.46). Notice that the trivial link of two components *is tricolorable.* This is the reverse of what happened for tricolorability for knots. Now, if we have a link of two components that is *not* tricolorable, we know it can't be the unlink.

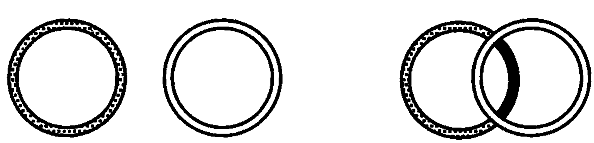


Figure 1.46 Two projections of the trivial link of two components.

Exercise 1.29 Prove that the Whitehead link in Figure 1.30 is not tricolorable and therefore is not the trivial link of two components. Remember, linking number wasn't enough to show this before.

Exercise 1.30 Determine which of the links of six or fewer crossings in Table 1.1 at the end of the book are and are not tricolorable.

Exercise 1.31 Show that the link in Figure 1.47 is tricolorable.

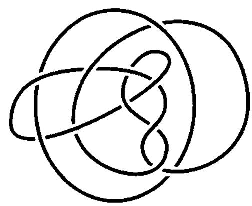


Figure 1.47 This link is tricolorable.

1.6 Knots and Sticks

Suppose we were given a bunch of straight sticks and we were told to glue them together end to end in order to make a nontrivial knot. The sticks can be any length that we want (Figure 1.48). How many sticks will it take to make a nontrivial knot? Try playing with some sticks to see what happens. Certainly, three sticks aren't enough, as they would just form a triangle that lies in a plane. If we looked down at the plane, we would see a projection of the knot with *no* crossings. So it would have to be the unknot.

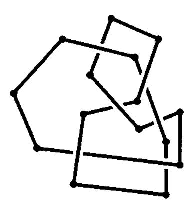


Figure 1.48 A knot made out of sticks.

How about four sticks? If we view the four sticks from any direction, we will see a projection of the corresponding knot. If two of the sticks are attached to each other at their ends, they cannot cross each other in the projection (since two straight lines can cross at most once, in this case at the point where they are attached to one another). So in the projection, each stick can only cross the one stick that is not attached to either one of its ends. Therefore, there can be at most two crossings in the projection.

But the only knot with a projection of two or fewer crossings is the unknot. (See Exercise 1.2.)

So four sticks aren't enough to make a nontrivial knot. How about five sticks? Let's view the knot so that we are looking straight down one of the sticks. In the projection of the knot that we see, we will only be able to see four of the sticks, since the fifth stick is vertical. For the same reason as in the previous paragraph, the four sticks that we see can have at most two crossings, and so the knot must the unknot.

Exercise 1.32 Prove that, in fact, a knot with four sticks in the projection can have at most one crossing.

Therefore, it must take at least six sticks to make a knot. In fact, it is possible to make a trefoil knot with six sticks, as shown in Figure 1.49. Although the picture looks believable, how do we know that we could really make a trefoil knot in space out of straight sticks like this? How do we know that the sticks needn't be bent or warped to fit together in this way, and that they only look straight when we see them from this view? We are only looking at a projection of the sticks in this picture.

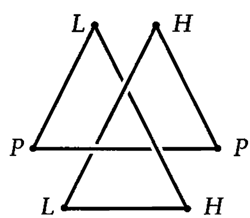


Figure 1.49 A trefoil knot from six sticks.

One solution is to actually build the knot with real sticks. But we can convince ourselves that this construction works without going to that much trouble. Let the vertices labeled P lie in the xy plane. The vertices labeled L lie low, underneath the plane. The vertices labeled H lie high, above the plane. Then it's clear that such a knot could actually be constructed from sticks.

If we want a hands-on demonstration that five sticks won't suffice to make a knot, we can try it with five "sticks" that we were born with. Namely, think of the first stick as being your left forearm, followed by a stick formed from your left upper arm, followed by a stick that goes from your left shoulder to your right shoulder, followed by a stick formed from your right upper arm, followed by a stick formed from your right lower arm. That's a total of five sticks that are attached end to end (Figure 1.50).

If you can tangle up your arms and then clasp your hands together so that the loop formed from these five sticks is knotted, you will have a knot made from five sticks. Don't hurt yourself, we have already demonstrated that you can't succeed.

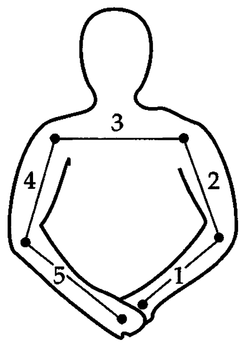


Figure 1.50 Making knots from your arms?

But supposedly, six sticks are enough to make a knot.

Exercise 1.33 Take a straight stick (say a yardstick or fireplace poker) as your sixth stick and demonstrate with your arms and this stick that a knot can be made out of six sticks.

What happens if we try to make knots using two people holding hands and their "ten sticks"? What knots can we make?

Exercise 1.34 How many sticks would it take to make a figure-eight knot?

*Exercise 1.35** Show that the only nontrivial knot you can make with six sticks is the trefoil knot.

Exercise 1.36 Show that you can make the knot 5_1 (see the table at the back of the book) or the Whitehead link using only 8 sticks (use P's L's, and H's to demonstrate that your constructions work).

Define the **stick number** $s(K)$ of a knot K to be the least number of straight sticks necessary to make K.

Exercise 1.37 Show that if J and K are knots, $s(J\#K) \leq s(J) + s(K) - 1$.

Unsolved Question

Can the inequality in Exercise 1.37 be improved to replace the -1 by -2 or -3? Amazingly, if J and K are trefoil knots (and hence each has stick number 6), then $s(J\#K)=8$, showing that in this very specific example, we have $s(J\#K) \leq s(J) + s(K) - 4$.

*Exercise 1.38** Let $c(K)$ be the least number of crossings in any projection of a knot K. Prove that if K is a nontrivial knot, then

$$\frac{5+\sqrt{(25+8(c(K)-2))}}{2} \leq s(K)$$

(*Hint*: Look straight down one edge and then count crossings to obtain a bound on $c(K)$ in terms of $s(K)$. Then invert the inequality.)

In fact, we also have an upper bound on the stick number of a knot in terms of the minimum crossing number $c(K)$ of the knot. In a paper that appeared in 1991, Seiya Negami, a professor at Yokohama National University in Japan, showed $s(K) \leq 2\,c(K)$. The proof is elementary; however, it depends on some results from graph theory, so we will not discuss it.

Unsolved Questions

1. By Exercise 1.37 and the preceding paragraph, we know that

$$\frac{5+\sqrt{(25+8(c(K)-2))}}{2} \leq s(K) \leq 2c(K)$$

Either show these are the best bounds we can obtain (by finding examples of knots of any crossing number that have $s(K)=5+\sqrt{(25+8(c(K)-2))}/2$ and other examples that have $s(K)=2c(K)$, or improve these bounds on $s(K)$ to narrow down the possibilities for $s(K)$ still further.

2. Does $s(K)$ change if we insist that the sticks that we make the knot out of must all be the same length? (It does not change in the case of the trefoil knot. But it seems unlikely that the same would be true for all knots.)

In Chapter 7, when we talk about applications of knot theory to synthetic chemistry, we will see why one might care about how many sticks it takes to make a knot.

Tabulating Knots

2

2.1 History of Knot Tabulation

Knot theory began in earnest around the end of the nineteenth century. Previously, several mathematicians had dabbled with knots, including Carl Friedrich Gauss (1777–1855), one of the greatest of all mathematicians. But it was Lord Kelvin's theory that atoms were knotted vortices in the ether that sparked serious interest in determining the possible knots.

The first work on tabulating knot projections was done in the 1880s by the Reverend Thomas P. Kirkman. These early explorations in knot theory suffered from Kirkman's opaque writing style. To quote:

> *By a knot of* n *crossings, I understand a reticulation of any number of meshes of two or more edges, whose summits, all tessaraces, are each a single crossing, as when you cross your forefingers straight or slightly curved, so as not to link them, and such meshes that every thread is either seen, when the projection of the knot with its* n *crossings and no more is drawn in double lines, or conceived by the reader of its course,*

> *when drawn in single line, to pass alternately under and over the threads to which it comes at successive crossings.*

In spite of Kirkman's obfuscation, his ideas were applied by a Scottish physicist named Peter Guthrie Tait in order to list all of the alternating knots up to 10 crossings. This was the first successful tabulation of knots.

A professor at the University of Nebraska named C. N. Little was the first to attack the problem of enumerating the nonalternating knots. In 1899, after six years of work, he published a table of 43 nonalternating knots of 10 crossings. His table was believed to be correct for 75 years. It wasn't until 1974 that it was discovered that two of the knots in Little's table were in fact the same knot and that there were only 42 distinct nonalternating knots of 10 crossings. The duplication was discovered by a part-time mathematician and New York lawyer named Kenneth A Perko. The two projections that actually correspond to the same knot are now called the Perko pair (Figure 2.1).

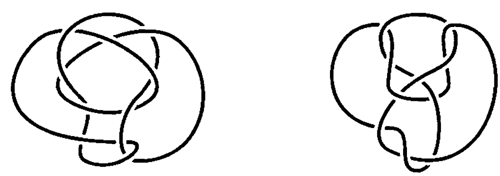


Figure 2.1 The Perko pair.

Exercise 2.1 Show that the Perko pair are the same knot.

Little went on to publish a census of 11-crossing alternating knots, eventually discovered to contain eleven omissions and one duplication. In 1917, Mary G. Haseman listed all amphicheiral knots (remember, that means knots that are equivalent to their mirror images) of 12 crossings in her doctoral thesis.

During this early period in the tabulation of knots, there were few attempts to rigorously prove that the knots claimed to be distinct in the tables were actually distinct. In fact, it wasn't until 1927 that two mathematicians named Alexander and Briggs provided the first proof that the knots of up to nine crossings in the tables were actually distinct, with only a few pairs of knots that they couldn't deal with. Their methods utilized the first polynomial applied to knots, now known as the Alexander polynomial. It remained the only polynomial for knots until 1984.

Kurt Reidemeister finished off the rigorous classification of knots up to nine crossings in 1932. There ensued a long period of inactivity in the tabulation of knots.

In 1969, an Englishman named John H. Conway invented a new notation for knots and used it to determine all of the prime knots of 11 or fewer crossings and all of the prime nonsplittable links of 10 or fewer crossings. His tabulation was all done by hand, without the aid of a computer. Conway had first become interested in knots while in high school and formulated many of his ideas then. But because of his wide-ranging mathematical interests, it wasn't until many years later that he applied these earlier ideas to the classification of knots.

In 1978, Alain Caudron of the University of Paris produced the first correct list of all prime knots through 11 crossings, repairing a few errors in Conway's table. In the meantime, a Canadian named Hugh Dowker invented a new notation for knots that was loosely based on Tait's ideas from the previous century. An algorithm for generating knots that utilized this notation was implemented on the computer by an Englishman named Morwen Thistlethwaite. This computer program resulted in a table of all prime knots through 12 crossings in 1981, and a table of all prime knots through 13 crossings in 1982. Thistlethwaite determined the following numbers:

Number of crossings	3	4	5	6	7	8	9	10	11	12	13	14
Number of prime knots	1	1	2	3	7	21	49	165	552	2176	9988	?

In this list of numbers, Thistlethwaite does not count both a knot and its mirror image. In the case that a knot is equivalent to its mirror image (that is, the knot is amphicheiral), no information is lost. In the case that the knot is not equivalent to its mirror image, however, a single knot in Thistlethwaite's list actually represents two distinct knots. Thistlethwaite also determined exactly which 12-crossing knots are amphicheiral, and he proved the previously conjectured fact that none of the knots of 13 crossings are amphicheiral. Thistlethwaite warned that his tabulation awaits independent verification. We will talk more about determining amphicheirality for knots in Chapter 6.

What does a 14-crossing knot look like? Let's draw one. Start drawing a curve on a piece of paper, allowing it to cross itself, but keeping track of how many times it does so as you go along. When you get near to 14 crossings, start heading for the point you started at. Try to close up the curve after exactly 14 crossings (Figure 2.2). You won't always be able to close it up after exactly 14 crossings, but after a few practice runs, you'll get better at it. Now, let's make the projection alternating. To do this, choose your favorite crossing and decide which string at the crossing goes under and which goes over. Then, follow one of the strings from that crossing to the next crossing, where you make the string do the opposite from what it did at the last crossing. Continue in this manner until you have a 14-crossing alternating knot. This gives you some feeling for how many 14-crossing knots there might be. Any 14-crossing scribbled curve corresponds to a 14-crossing alternating knot. Notice, that if the knot needn't be alternating, you have 2^{14} choices of how to put in the crossings on any one scribble, since at every one of the 14 crossings, there are two possibilities.

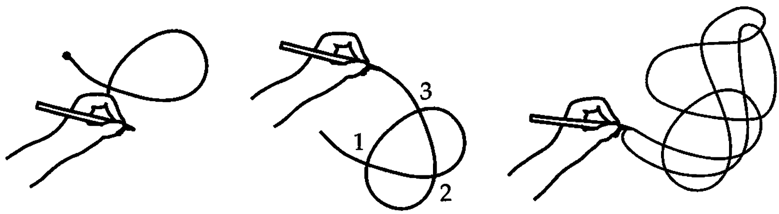


Figure 2.2 Scribbling to make a 14-crossing projection.

There are also millions of different scribbled curves that we could draw. It seems like there are many more 14-crossing knots than we could ever catalog. But many of the scribbles and choices of crossings actually correspond to the same knots.

Unsolved Questions

1. Find all of the 14-crossing prime knots. This is probably hard and requires some new ideas. See the next section for why this could be difficult.

2. Classify the alternating knots of 14 crossings. When you restrict yourself to alternating knots, the number of cases you have to look at is reduced by a factor of 2^{14}.

3. Determine the sequence of integers that begins 1, 1, 2, 3, 7, 21, 49, 165, 552, 2176, 9988, ... Perhaps this sequence of numbers giving the number of prime knots with a given crossing number is in fact a reasonable function, like $f(n) =$ greatest integer less than e^{n-4}. This is a hard open question. Perhaps no elementary function gives this sequence.

4. Show that the number of distinct prime $(n+1)$-crossing knots is greater than the number of distinct prime n-crossing knots, for each positive integer n. It is remarkable that we cannot yet show this.

We do know that the number of prime knots of n crossings grows at an exponential rate. In 1987, Claus Ernst and Dewitt Sumners, then both at Florida State University, used recent results on alternating knots due to Kauffman, Murasugi, and Thistlethwaite in order to prove that the number of distinct prime knots of n crossings is at least $(2^{n-2}-1)/3$ for $n \geq 4$.

(See Ernst and Sumners, 1987.) Note that in this lower bound, both a given knot and its mirror image are counted if they are not equivalent. Hence, this number can exceed the number of prime knots of n crossings given by Thistlethwaite for $n \leq 13$. We talk more about this result in Section 3.2.

Very recently, Dominic Welsh of Oxford University has proved that the number of distinct prime n-crossing knots is bounded above by an exponential in n.

2.2 The Dowker Notation for Knots

The Dowker notation is an extremely simple way to describe a projection of a knot. First, let's start with an alternating knot. Suppose we have a projection of an alternating knot that we want to describe, like the one in Figure 2.3. Choose an orientation on the knot, given by placing coherently directed arrows along the knot. Pick any crossing and label it 1. Leaving that crossing along the understrand in the direction of the orientation, label the next crossing that you come to with a 2. Continue through that crossing on the same strand of the knot, and label the next crossing with a 3. Continue to label the crossings with the integers in sequence until you have gone all the way around the knot once. When you are done, each crossing will have two labels on it, as the knot passes through each crossing twice (Figure 2.4). Notice that, in fact, each crossing has one even number and one odd number labeling it.

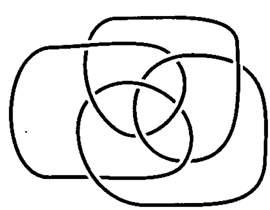


Figure 2.3 An alternating knot.

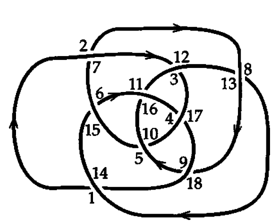


Figure 2.4 Label each crossing of the knot with two numbers.

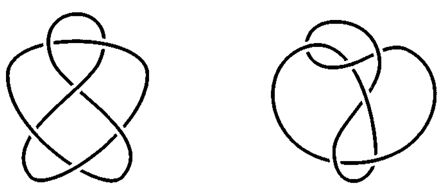
Exercise 2.2 Why does every crossing get one even numbered label and one odd numbered label?

Thus, we can think of this labeling as giving us a pairing between the odd numbers from 1 to 18 and the even numbers from 1 to 18. In this case, we get

1	3	5	7	9	11	13	15	17
14	12	10	2	18	16	8	6	4

As a shorthand, we could just write 14 12 10 2 18 16 8 6 4, and keep in mind that this means 1 is paired with 14, 3 with 12, 5 with 10, and so forth. Thus, from a projection of a knot, we obtain a sequence of even integers, where the number of even integers is exactly the number of crossings in the knot.

Exercise 2.3 Find a sequence of even integers that represents the projection of the knots 6_2 and 6_3 (Figure 2.5). How about a second sequence of even integers that also represents the same projection of 6_3?



6_2 6_3

Figure 2.5 The knots 6_2 and 6_3.

Now, suppose we want to go the other way. Given a sequence of even integers that represents a projection of an alternating knot, how do we draw the projection? Say the sequence is 8 10 12 2 14 6 4. This is shorthand for

1	3	5	7	9	11	13
8	10	12	2	14	6	4

So let's begin drawing the knot. Start by drawing just the first crossing, labeling it with a 1 and an 8. We extend the understrand of the knot and then draw in the next crossing, which corresponds to 2. Since 2 is paired with 7, we label this crossing with a 2 and a 7. Because the knot is alternating, we know that the strand that we are on goes over this crossing. We continue the overstrand through this crossing to the next crossing where it becomes the understrand, labeling the new crossing with a 3 and the integer that is paired with 3, namely 10 (Figure 2.6). We continue this process until the next integer that should be placed on a crossing already labels an existing crossing. We then know that the knot must now circle around to pass through that crossing. Note that we have two choices as to how to circle around: either circling to the right or to the left in order to pass back through the previously drawn crossing. For the time being, let's ignore this ambiguity and just choose either direction for circling around.

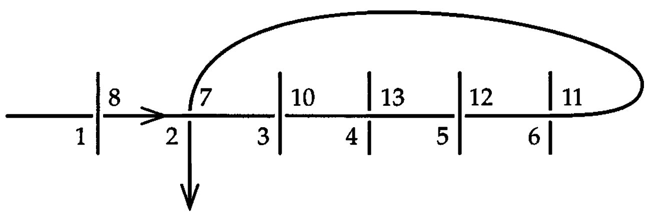


Figure 2.6 Constructing a knot projection from the Dowker notation.

We continue in this manner. If neither of the labels on the next crossing has occurred before, then we make a new crossing. But if one of the labels has occurred before, we circle the knot through that crossing. All the way along, we will be sure that the crossings alternate as we progress along the knot. Finally, we end up with a picture of our knot (Figure 2.7).

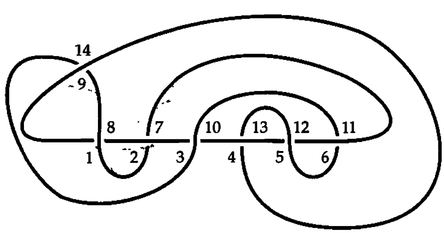


Figure 2.7 The knot that comes from 8 10 12 2 14 6 4.

Exercise 2.4 Which seven-crossing knot from the table at the end of the book is this knot?

Exercise 2.5 Draw a picture of the projection of an alternating knot corresponding to the sequence 10 12 8 14 16 4 2 6.

Now what about that ambiguity in our choice of how the knot circles around? Our choice *can* change the resulting knot. For instance, the sequence 4 6 2 10 12 8 represents two distinct knots, as shown in Figure 2.8. Note that the two knots are composite knots, and that this is reflected in the fact that the sequence 4 6 2 10 12 8 is actually a shuffling of the three numbers 2, 4, 6 and then a shuffling of the three numbers 8, 10, and 12. When the permutation of the even numbers can be broken into two separate subpermutations, the resulting knots are composite (assuming each of the factor knots is nontrivial) and the knot is not completely determined by the Dowker notation. However, if we restrict ourselves to sequences of even numbers that cannot be split into subpermutations, either a particular knot or its mirror image results (Figure 2.9). When the knot is amphicheiral, only one knot can be the result.

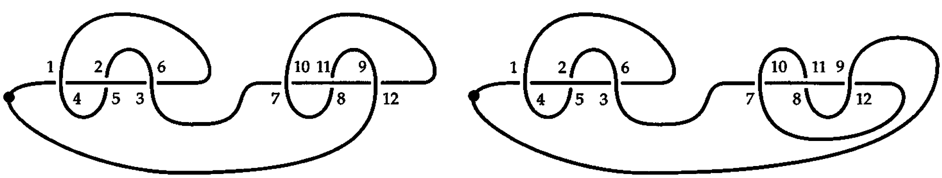


Figure 2.8 Two knots with the same Dowker notation.

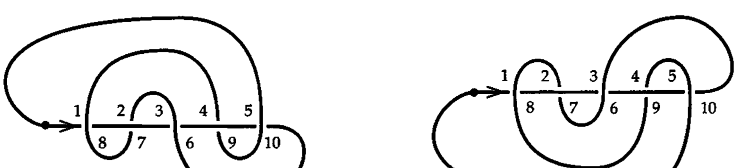


Figure 2.9 A knot and its mirror image are both given by 8 6 10 2 4.

Although the possible projections look different, they will all correspond to the same pair of knots. The best way to see this is to think of projecting the knot onto a sphere (Figure 2.10) rather than onto a plane. (Just as the earth looks planar until you get far enough away from it, so does any sphere.) The advantage to projecting onto a sphere is that there is no special outer region with infinite area as there is in a projection onto the plane. Figure 2.11 contains two projections described by 8 6 10 2 4 that are distinct as projections on the plane but that are equivalent projections on the sphere.

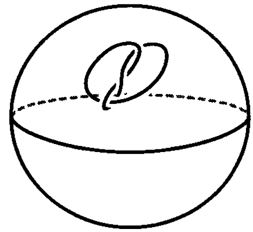


Figure 2.10 Projecting a knot onto a sphere.

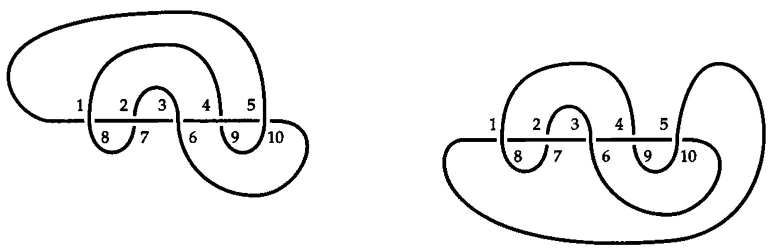


Figure 2.11 Two projections of 8 6 10 2 4.

Exercise 2.6 Draw two projections given by 10 12 2 14 6 4 8, which are inequivalent as projections in the plane but which are equivalent as projections on the sphere.

Exercise 2.7 How many different sequences of the integers 2 4 6 8 10 12 14 are there? (This exercise gives us an upper bound on the number of possible alternating knot projections with seven crossings; however, it's far from accurate.)

The system that we have explained works very well for describing the projection of an alternating knot, but how can we extend it to knots that aren't alternating? We add in + and − signs to our sequence of even numbers. Our rule is as follows: When traversing the knot using the labeling system that we have described, we assign an even integer and an odd integer to each crossing. If the even integer is assigned to the crossing while we are on the overstrand at that crossing, we leave the even integer positive. But if the even integer is assigned to the crossing while we are on the understrand of that crossing, we make the corresponding even number negative. So, for example, in the knot in Figure 2.12, the numbers 14, 12, 4, and 8 become negative.

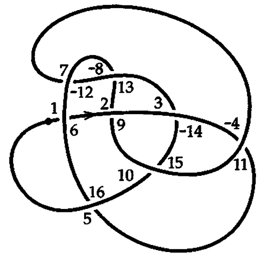
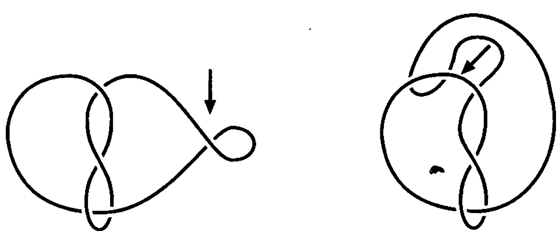


Figure 2.12 A nonalternating knot with sequence 6 −14 16 −12 2 −4 −8 10.

Exercise 2.8 Draw a projection of the knot corresponding to the sequence 14 12 −16 2 18 6 8 10 −4.

Exercise 2.9 How do you recognize from the sequence of numbers that a projection has a trivial crossing in it like this? How about recognizing a Type II Reidemeister move that will reduce the number of crossings by two? (See Figure 2.13.)



a b

Figure 2.13 (a) Trivial crossing. (b) Type II Reidemeister move.

Dowker's notation allows us to feed projections of knots into the computer simply as a sequence of numbers. In particular, suppose we wanted to attempt a classification of 14-crossing knots. The number of sequences of the 14 numbers 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 is 14!, which is about 87 billion. Then we can put a +1 or −1 in front of each of the even numbers, giving us another factor of 2^{14}. Of course, there aren't this many different knots with 14 crossings. Lots of the sequences represent the same knot. In fact, lots of the sequences represent the same projection of the same knot.

Morwen Thistlethwaite used the Dowker notation to list all of the prime knots of 13 or fewer crossings. Perhaps it will turn out to be the best way to list knots of 14 or fewer crossings.

2.3 Conway's Notation

In this section, we introduce a notation for knots due to John H. Conway. This was the notation he used in order to tabulate the prime knots through 11 crossings and prime links through 10 crossings in 1969. (Although he did not use a computer, he missed only four knots.) The Conway notation has been utilized in order to prove numerous results and recently has been applied to knotting in DNA (see Section 7.2 and Sumners, 1992). It is particularly suited to calculations involving what are called tangles.

A **tangle** in a knot or link projection is a region in the projection plane surrounded by a circle such that the knot or link crosses the circle exactly four times (Figure 2.14). We will always think of the four points where the knot or link crosses the circle as occurring in the four compass directions NW, NE, SW, and SE.

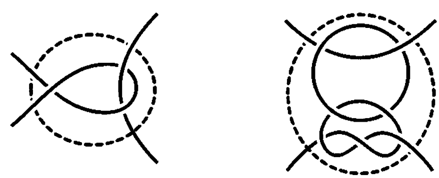


Figure 2.14 Tangles.

We can use tangles as the building blocks of knot and link projections (Figure 2.15). Therefore, understanding tangles will be very helpful in understanding knots. We will say two tangles are **equivalent** if we can get from one to the other by a sequence of Reidemeister moves while the four endpoints of the strings in the tangle remain fixed and while the strings of the tangle never journey outside the circle defining the tangle. So, for instance, the two tangles in Figure 2.16a and e are equivalent by the sequence of Reidemeister moves in Figure 2.16b, c and d, e.

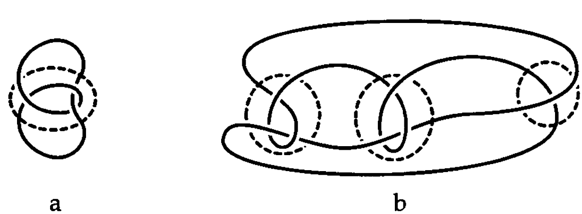


Figure 2.15 Knot projections formed from tangles.

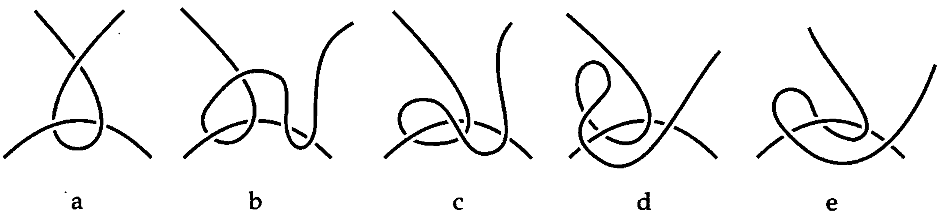


Figure 2.16 These tangles are equivalent.

Notice that if we form a knot from a single tangle by gluing together the ends in pairs as we did in Figure 2.15a, then two such knots are equivalent whenever the corresponding tangles are equivalent. Let's look at some particular tangles that are easy to form. One of the simplest tangles is two vertical strings, as in Figure 2.17a. We denote this tangle as the ∞ tangle. We denote the tangle consisting of two horizontal strings as the 0 tangle. We could wind two horizontal strings around each other to get Figure 2.17c. We denote this tangle by the number of left-handed twists we put in. In this case, the number is 3. If we had twisted the other way around, we would have denoted the resulting tangle by -3. Note that for a positive-integer twist, the overstrand always has a positive slope, if we think of it as a small segment of a line.

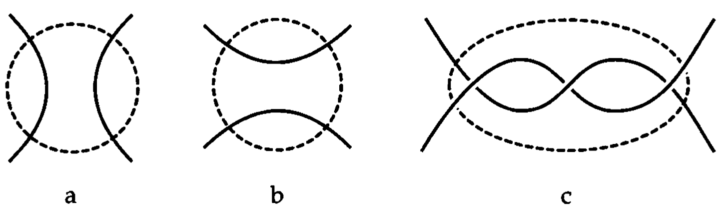
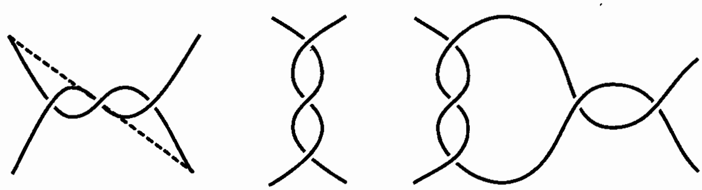


Figure 2.17 (a) The ∞ tangle. (b) The 0 tangle. (c) The 3 tangle.

We are going to form a more complicated tangle, starting from the 3 tangle. First, we reflect the tangle through the NW and SE diagonal line in Figure 2.18a to obtain Figure 2.18b. Think of this reflection as if we reflected in a mirror that was perpendicular to the projection plane and intersected the projection plane along the NW and SE diagonal line. Note that the two ends of the tangle along the diagonal are fixed when we perform the reflection, while the two ends of the tangle that are not on the diagonal are switched. It is sometimes difficult to picture what happens to the crossings under the reflection. Usually, we can figure out what happens to one crossing and then we can infer what must happen to the other crossings. Note that for a positive-integer twist, it is still true that after reflection the overstrand has positive slope.

Now we wind the two right-hand ends of the tangle around each other to get Figure 2.18c. We denote this tangle by 3 2, as the original tangle had three twists of the horizontal strings followed by a reflection and then two twists of the horizontal strings.

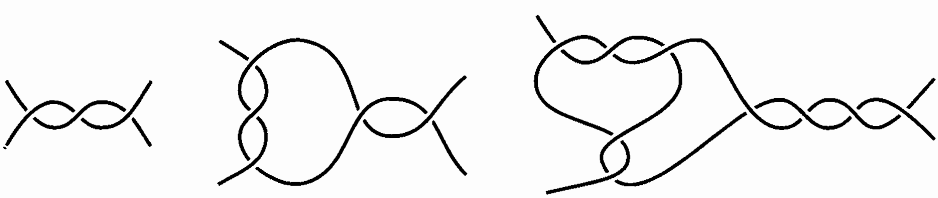


3 2

a b c

Figure 2.18 Constructing a tangle.

Let's complicate this tangle still further. First, we take the tangle 3 2, as in Figure 2.19b, and again reflect about the NW to SE diagonal. Then we add −4 twists to the right-hand strings, as in Figure 2.19c. We denote this tangle 3 2 − 4. Figure 2.20 gives some additional examples.



3 3 2 3 2 -4

a b c

Figure 2.19 Constructing the 3 2 −4 tangles.

2 1 -2 1 1 1 1

a b

Figure 2.20 More tangles.

We call any tangle that we could construct in this manner a **rational tangle.** Notice that if the rational tangle is represented by an even number of integers, we can think of constructing it by simply starting with two vertical strings (that is, the ∞ tangle) and then twisting the two bottom endpoints around each other some number of times, while holding the top two endpoints fixed. Then we could twist the two right-hand endpoints around each other while keeping the left-hand endpoints fixed. We could then alternately twist the bottom two endpoints and the right two endpoints to create the tangle. A positive-integer twist always gives the overstrand a positive slope, regardless of whether the twist is occurring in two vertical strands or two horizontal strands (Figure 2.21).

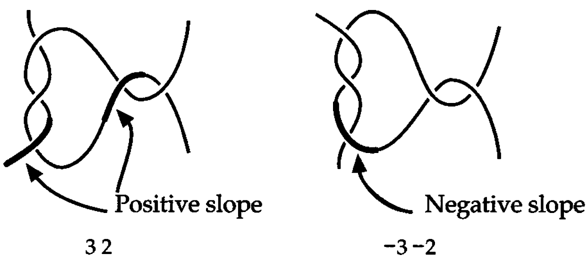


Figure 2.21 Positive integer twists give the overstrand positive slope.

Similarly, if the rational tangle is represented by an odd number of integers, we can construct it by starting with two horizontal strings (the 0 tangle) and alternately twisting the two right-hand endpoints appropriately, followed by twisting the two bottom endpoints appropriately.

Exercise 2.10 Draw the rational tangles corresponding to 2 −3 4 5 and 3 −1 3 −3 2.

Exercise 2.11 Show that the two tangles 2 1 1 and −1 −2 2 are equivalent.

Amazingly enough, there is an extremely simple way to tell if two rational tangles are equivalent. Suppose the two tangles are given by the sequences of integers −2 3 2 and 3 −2 3. We compute the so-called **continued fractions** corresponding to these integers. The continued fraction corresponding to −2 3 2 is

$$2 + \frac{1}{3 + (1/-2)}$$

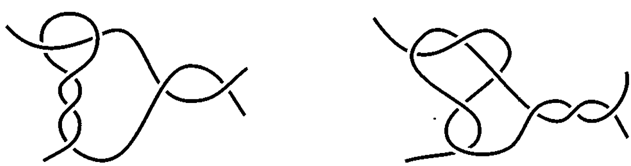
We put the first -2 in the denominator of a fraction with numerator 1. We add to the $-1/2$ the next number 3, and then put the result in the denominator of a fraction with numerator 1. We then added the last number to the result. Notice that we can clean up this fraction:

$$2 + \frac{1}{3+(1/-2)} = 2 + \frac{1}{5/2} = 2 + \frac{2}{5} = \frac{12}{5}$$

The continued fraction corresponding to $3\ -2\ 3$ is

$$3 + \frac{1}{-2+(1/3)}$$

which also equals $12/5$. It is in fact the case that, since their continued fractions are equal, these two rational tangles *are* equivalent (Figure 2.22).



-2 3 2 3 -2 3

Figure 2.22 Two equivalent rational tangles.

Exercise 2.12 Draw a sequence of pictures to show that these two tangles are equivalent.

On the other hand, the tangle $3\ 2\ -4$ in Figure 2.19 has continued fraction

$$-4 + \frac{1}{2+(1/3)}$$

which equals $-25/7$. Thus, this tangle is distinct from the two equivalent tangles $-2\ 3\ 2$ and $3\ -2\ 3$.

In general, suppose we have two rational tangles given by the sequences of integers *ijk . . . lm* and *npq . . . rs*. We can compute the corresponding continued fractions

$m+1/(+1/(l\ldots 1/(k+1/(j+1/i))))$ and
$s+1/(r+1/(\ldots 1/(q+1/(p+1/n))))$

These fractions are both rational numbers. The two tangles are equivalent if and only if these two rational numbers are the same.

Exercise 2.13 Determine which of the four rational tangles in Figure 2.23 are equivalent.

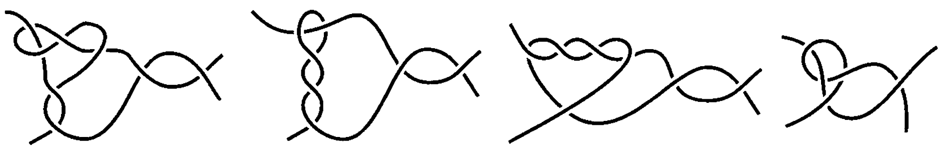


Figure 2.23 Which of these tangles are equivalent?

The proof that two rational tangles are equivalent if and only if their continued fractions yield the same rational number is difficult. If you are interested, a proof appears in Burde and Zieschang (1986) (see Suggested Readings for Chapter 1).

Exercise 2.14 Show that the rational tangle $2\ 1\ a_1 a_2 \ldots a_n$ is equivalent to the rational tangle $-2\ 2\ a_1 a_2 \ldots a_n$ both by using continued fractions and by drawing a picture.

If we close off the ends of a rational tangle as in Figure 2.24, we call the resulting link a **rational link**. So for instance, the figure-eight knot is a rational knot, with rational tangle 22 (Figure 2.24). We can use our notation for rational tangles to denote the corresponding rational knot. In the table at the end of the book, you can see this notation applied to the knots. We call this notation **Conway's notation**.

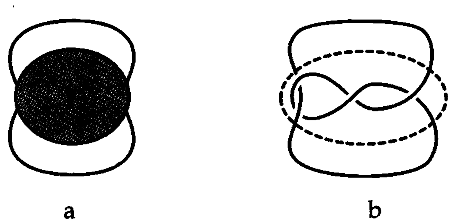


Figure 2.24 (a) A rational link. (b) The figure-eight knot.

*Exercise 2.15** Determine a Conway's notation for each of the knots in Figure 2.25. (You do not need to use the given projections.)

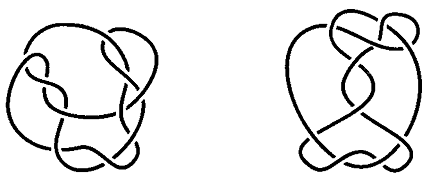


Figure 2.25 What is the Conway notation for these knots?

Exercise 2.16 (a) Show that a rational link has either one or two components.

(b) For which sets of Conway notations do the corresponding rational links have two components?

*Exercise 2.17** Show that any rational link is alternating (by showing that it has an alternating projection).

We can use the rational tangles to construct more complicated tangles. For instance, we will define a way to "multiply" two tangles to obtain a new tangle, as in Figure 2.26. We reflect the first tangle across its NW to SE diagonal line, and then we glue it to the second tangle. Note that this definition of multiplication fits in nicely with our definition of a rational tangle. We can think of the rational tangle 32 as coming from multiplying together the two tangles 3 and 2. Note also that multiplying together two rational tangles will always generate a rational tangle. Moreover, if we ever want to reflect a tangle across its NW to SE diagonal line, we can simply multiply it by the tangle 0.

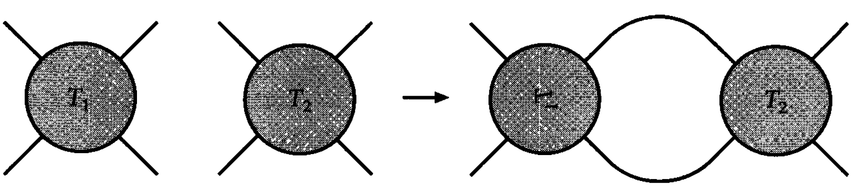
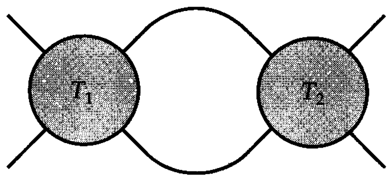


Figure 2.26 Multiplying tangles.

We can also "add" together two tangles, as in Figure 2.27. As an example, note that the knot 8_5 can be written as the knot corresponding to the tangle 30 + 30 + 20, as it is simply the sum of these three rational tangles.

If we multiply each tangle in a sequence of tangles by 0, and then add them all together, we denote the resultant tangle by the sequence of numbers that stand for the original tangles, only now separated by commas. So we would denote the tangle for 8_5 by 3, 3, 2 (Figure 2.28). (A knot obtained from a tangle represented by a finite number of integers separated by commas is often called a **pretzel knot**.)



$T_1 + T_2$

Figure 2.27 Adding tangles.

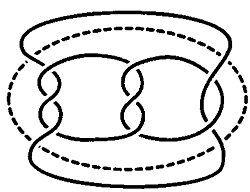


Figure 2.28 The knot 8_5 has Conway notation 3, 3, 2.

Exercise 2.18 Draw the tangles 2, −32, 41 and −23, 1, 42 and the corresponding knots obtained by connecting the NW string to the NE string and the SW string to the SE string.

Numerous additional examples appear in the appendix table. So we have the operations of addition of tangles and multiplication of tangles. We will call any tangle obtained by the operations of addition and multiplication on rational tangles an **algebraic tangle**.

Exercise 2.19 Draw the algebraic tangle $(3, 2, 1) \cdot (1, 2, 2)$.

An **algebraic link** is simply a link formed when we connect the NW string to the NE string and the SW string to the SE string on an algebraic tangle. We denote the link the same way we denote the corresponding tangle. (Such a link is also sometimes called an arborescent link.)

Exercise 2.20 Show that an algebraic knot with Conway notation containing no negative signs must be an alternating knot.

These algebraic tangles are behaving a lot like the real numbers. We can add two of them or multiply two of them. But the real numbers have an element 0 so that adding 0 to a number doesn't change the number. We call 0 an **additive identity** for the real numbers.

Exercise 2.21 Is there an additive identity for tangles?

The real numbers also have the number 1 so that multiplying any number by 1 doesn't change it. We call 1 a **multiplicative identity**.

Exercise 2.22 Is there a multiplicative identity for tangles? Is it the same if you multiply a tangle by it on the right side or the left side?

There are differences between the structure of the real numbers and the structure of algebraic tangles. For instance, multiplication on tangles is not commutative. It's not true that $ab = ba$ for all tangles. Multiplication on tangles is also not associative. Usually, it's not true that $(ab)c = a(bc)$. Moreover, we don't have inverses. In the real numbers, there is always an additive inverse, so if c is a real number, $-c$ is its additive inverse, that is, $c + -c = 0$. But for a tangle T, in general, there is no inverse tangle, no tangle that when added to T gives back the trivial tangle 0.

Although many tangles are algebraic, there are tangles that are not algebraic. For instance, the tangle in Figure 2.29 is not algebraic.

While we are discussing tangles, let's mention another way to obtain new knots, called **mutation**. Suppose we have a knot K that we think of as being formed from two tangles, as in Figure 2.30. We form a new knot by cutting the knot open along four points on each of the four strings coming out of T_2, flipping T_2 over, and gluing the four strings back together. The resulting knot looks like Figure 2.31a. We could also cut the four strings coming out of T_2, flip T_2 left to right, and then glue the strings back together as in Figure 2.31b. If we did both operations in turn, it's as if we rotated the tangle 180 degrees and then reglued it as in Figure 2.31c. Any of these three operations is called a mutation, and the three resultant knots together with the original knot are call **mutants** of one another. Figure 2.32 shows two famous mutants called the **Kinoshita–Terasaka mutants**.

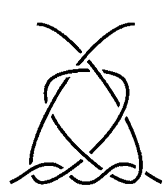


Figure 2.29 This tangle is *not* algebraic.

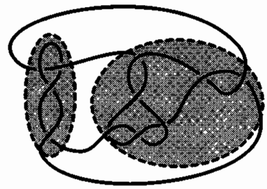
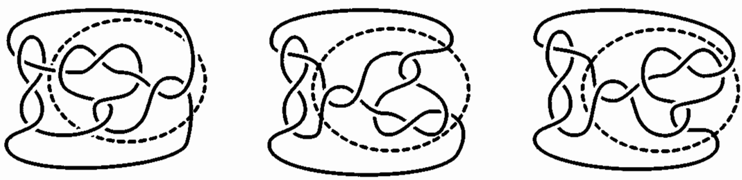


Figure 2.30 A knot formed from two tangles.



a b c

Figure 2.31 Mutant knots.

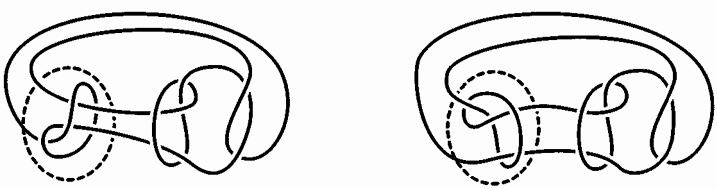


Figure 2.32 The Kinoshita–Terasaka mutants.

Exercise 2.23 Show that mutation applied to an alternating projection of a knot always yields an alternating knot.

Exercise 2.24 Show that the mutation of a knot is always another knot, rather than a link.

Exercise 2.25 Show that if we have three tangles as in Figure 2.33a, we can mutate several times in order to permute the tangles. Note that we can then permute n tangles in a row.

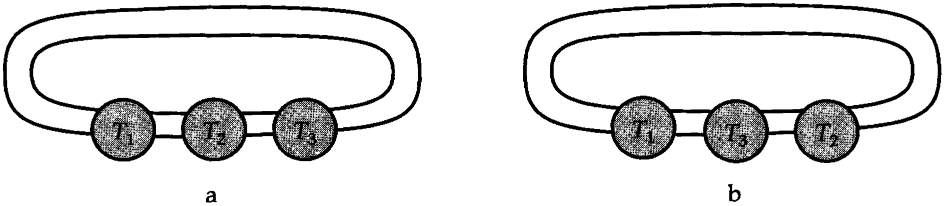


Figure 2.33 Show that these knots are related through a sequence of mutations.

Exercise 2.26 Show that the two knots in Figure 2.34 are related through a sequence of mutations.

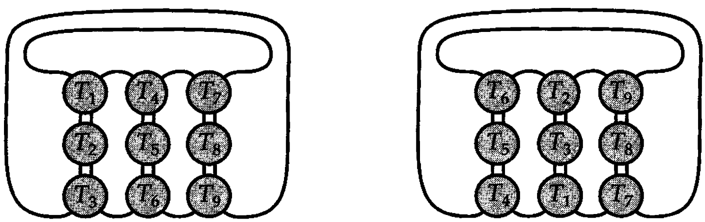


Figure 2.34 Two nasty mutants.

Although mutation can turn one knot into another, it cannot turn a nontrivial knot into the trivial knot. At least, we don't have to worry about that possibility. Still, mutant knots are some of the most difficult knots to tell apart. We discuss them again in Chapter 6. In Chapter 7, we use tangles to help us understand knotting in DNA.

2.4 Knots and Planar Graphs

In this section, we introduce a notation for knot projections that has been useful in the past for knot tabulation. It provides a bridge between knot theory and graph theory, with the potential for commerce in both directions.

What is a graph? It consists of a set of points called vertices and a set of edges that connect them. Here we are interested in planar graphs, that is, graphs that lie in the plane, as in the first two examples in Figure 2.35. From a projection of a knot or link, we create a corresponding planar graph in the following way. First shade every other region of the link projection so that the infinite outermost region is not shaded (Figure 2.36).

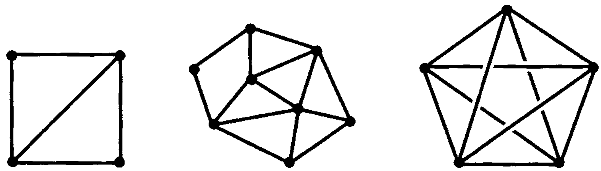


Figure 2.35 Some graphs.

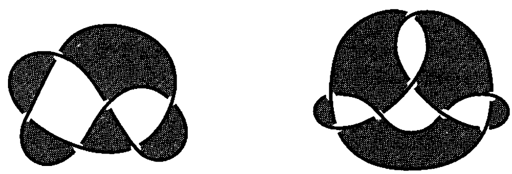


Figure 2.36 Shaded link projections.

Exercise 2.27 Prove that any link projection can be shaded in the checkerboard manner portrayed in Figure 2.36.

Put a vertex at the center of each shaded region and then connect with an edge any two vertices that are in regions that share a crossing (Figure 2.37). This is the graph corresponding to our projection. There is only one problem. It doesn't depend in any way on whether a crossing is an overcrossing or an undercrossing. So we define crossings to be positive or negative depending on which way they cross as in Figure 2.38. Now we label each edge in the planar graph with a + or a −, depending on whether the edge passes through a + crossing or a − crossing. We call the result a signed planar graph (Figure 2.39). (Note that this sign convention is dependent from the way that we labeled crossings with ± 1s when we were computing linking number in Section 1.4.) We now have a way to turn any link projection into a signed planar graph.

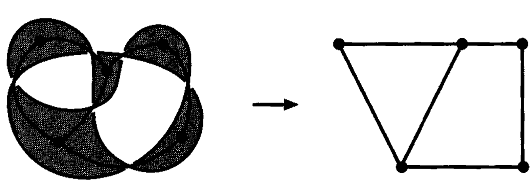


Figure 2.37 A graph from a knot projection.

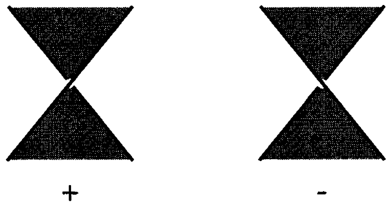


Figure 2.38 Signs on crossings.

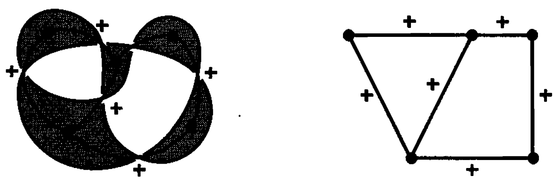


Figure 2.39 A signed planar graph from a knot projection.

Exercise 2.28 Turn the knot projection in Figure 2.40 into a signed planar graph.

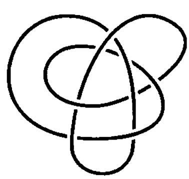


Figure 2.40 Find the corresponding signed planar graph.

What if we want to go in the other direction? Can we turn any signed planar graph into a knot projection? Certainly. Starting with the signed planar graph, put an x across each edge as in Figure 2.41b. Connect the edges inside each region of the graph as in Figure 2.41c. Shade those areas that contain a vertex. Then, at each of the x's, put in a crossing corresponding to whether the edge is a + or a − edge. The result is a link (Figure 2.42).

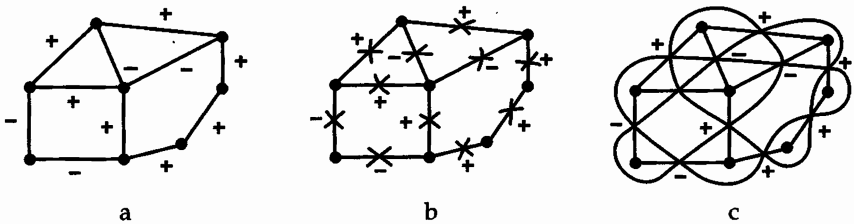


Figure 2.41 Turning a signed planar graph into a link.

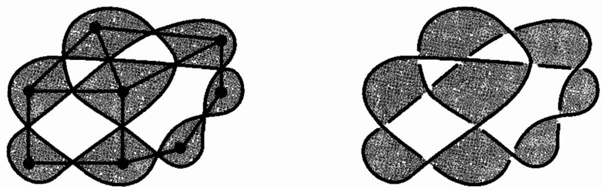


Figure 2.42 A link generated from a signed planar graph.

Exercise 2.29 Determine the link projection corresponding to the signed planar graph in Figure 2.43.

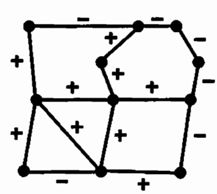


Figure 2.43 What link projection does this signed planar graph represent?

Exercise 2.30 Show that a link projection is alternating if and only if all the edges in the corresponding signed planar graph have the same sign.

Thus, we now have a way to go from knot projections to signed planar graphs and back again. In particular, we can turn questions about knots into questions about graphs. For example, one of the open problems in knot theory is to find a practical algorithm for determining if a projection is a projection of the unknot (see Section 1.1). This is equivalent to asking whether or not there is a sequence of Reidemeister moves that takes us from the given projection to the projection of the unknot.

But we can turn knot and link projections into signed planar graphs. We can turn Reidemeister moves into operations on signed planar graphs. The question of whether knot projections are equivalent under Reidemeister moves becomes one of whether signed planar graphs are equivalent under operations that the Reidemeister moves become.

Exercise 2.31 What do the Reidemeister moves become when translated into operations on signed planar graphs? (Make sure you consider what happens when different regions are shaded.)

We will come back to signed planar graphs when we look at the relationship between knot theory and statistical mechanics in Section 7.4.

3

Invariants of Knots

3.1 Unknotting Number

In this chapter, we introduce several new invariants for knots. We begin with a very intuitive invariant, known as the unknotting number. Notice first of all that if we changed the crossing circled in Figure 3.1, the knot 7_2 would become the unknot. The one change of crossing completely unknots the knot. We say that 7_2 has unknotting number 1. More generally, we say that a knot K has **unknotting number** n if there exists a projection of the knot such that changing n crossings in the projection turns the knot into the unknot and there is no projection such that fewer changes would have turned it into the unknot. We denote the unknotting number of a knot by $u(K)$.

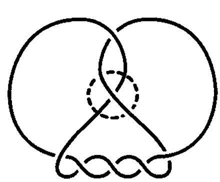


Figure 3.1 The knot 7_2 becomes the unknot.

Exercise 3.1 Find the unknotting number of the figure-eight knot.

Exercise 3.2 Find an infinite family of knots, all of which have unknotting number 1. (You need not prove that the knots in the family are distinct.)

Aside (for people who know the traditional definition of unknotting number): In our definition of the unknotting number, we performed all the crossings changes in a single projection of the knot. Traditionally, the unknotting number is defined to be the least number of crossing changes necessary to change a knot into an unknot, where we can perform the first crossing change in one projection of the knot, then do an ambient isotopy of the resulting projection to a new projection and change the second crossing in that projection. We can then do another ambient isotopy to a new projection before we change our third crossing, and continue in this manner until we have done all n crossing changes. That these two definitions are equivalent follows from the fact that we can keep track of each crossing change in the second definition with an arc that runs to and from the two points on the knot where the crossing change occurs. As we do our ambient isotopy to another projection, we carry along these arcs, stretching and deforming them as necessary. By the time we are finished with our n crossing changes, we have n such arcs. However, we can then shrink each of these arcs down to a tiny arc, pulling the knot along, and make a single projection of the knot so that each arc appears as a vertical arc running from the top of a crossing to the bottom. Then, changing these crossing in this single projection is equivalent to changing the crossings one by one and allowing ambient isotopy to occur between the crossing changes.

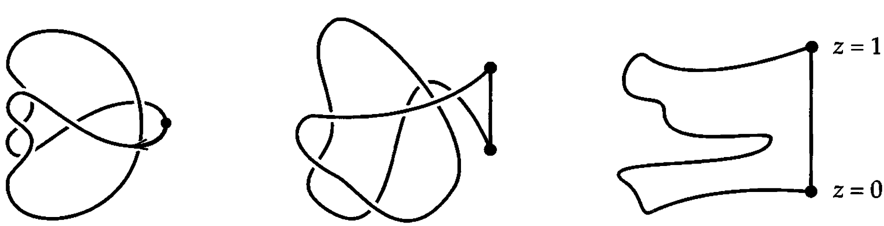
The fact that every knot has a finite unknotting number follows from the fact that every projection of a knot can be changed into a projection of the unknot by changing some subset of the crossings in the projection. Although this fact appeared as Exercise 1.7, let's verify it, since we are depending on it here.

Given a projection of a knot, let's pick a starting point on the knot that for convenience is not at a crossing, and let's pick a direction to traverse the knot. Now, beginning at that point, we head along the knot in our chosen direction. The first time that we arrive at a particular crossing, we change the crossing if necessary so that the strand that we are on is the overstrand. Then we continue through that crossing on our merry way along the knot. If we come to a crossing that we have already been through once, we do not change that crossing, but rather continue through it on what must necessarily be the understrand. Once we have returned to our initial starting point, we have a projection of a knot that was obtained from our original knot by changing crossings and that will in fact be the trivial knot, as we will demonstrate (Figure 3.2).

a b

Figure 3.2 (a) Original projection. (b) Altered projection.

To see that this is the trivial knot, we view it in three-space. Think of the z axis as coming straight out of the projection plane toward us. Starting at the initial point again, we place that point in three-space with z-coordinate $z = 1$. Now, as we traverse the knot, we decrease the z-coordinates of each of the points on the knot until we get almost back to where we started. That last point will have z-coordinate $z = 0$. But, since we gave the initial point and the last point z-coordinates $z = 0$ and $z = 1$, and these are supposed to be the same point, we had better put in a vertical bar from one to the other to complete the knot (Figure 3.3). Note then that when we look straight down the z axis at our knot, we see the projection that we had changed the crossings to create. But when we look at our projection from the side, we see a projection *with no crossings*. Hence this knot is a trivial knot.



a b c

Figure 3.3 (a) Altered projection. (b) Partial side view. (c) Side view. The altered projection is the trivial knot.

Exercise 3.3 Find an inequality that relates $u(K)$ and the minimum crossing number $c(K)$ of the knot.

In general, it's very hard to find the unknotting number of a knot. For instance, if we change crossings in the projection of 7_4 in the table at the back of the book, it looks like the unknotting number is 2 (which it is). But how do we know that there isn't some other projection of 7_4 that can be unknotted by only one crossing change? In order to prove that the unknotting number is 2, quite a bit more work would have to be done. For example, it wasn't until 1986 that Taizo Kanenobu of Kyushu University and Hitoshi Murakami of Osaka City University, both in Japan, proved that the unknotting number of the knot 8_3 is 2 (Figure 3.4). It's not hard to find two crossing changes that make this projection into the unknot. (Find them.) But how do we know there isn't some other projection of this knot that can be made into the unknot with one crossing change? Kanenobu and Murakami applied the powerful Cyclic Surgery Theorem, due to Marc Culler and Peter Shalen (both of the University of Illinois at Chicago), and Cameron Gordon and John Luecke (both of the University of Texas at Austin), to prove that no such projection exists.

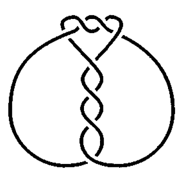


Figure 3.4 The unknotting number of 8_3 is 2.

Here's an interesting question: Can a composite knot have unknotting number 1 (Figure 3.5)?

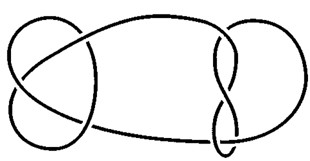


Figure 3.5 Can a composite knot be unknotted with one crossing change?

We might expect the answer to be no, for if we have a composite knot, changing one crossing might allow us to untangle one of the two factor knots that make up the composite knot, although it seems unlikely that the one crossing change would allow us to untangle both factor knots. In fact, the answer *is* no, but it took 100 years for someone to find the proof. In 1985, Martin Scharlemann at the University of California–Santa

Barbara (Scharlemann, 1985) proved that a knot with unknotting number 1 is prime. His proof is very technical.

Unsolved Questions

1. Find a simple proof that a knot with unknotting number 1 is prime.

2. Is it true that a knot with unknotting number 2 cannot be a composite knot made from three factor knots? Is it true that a knot with unknotting number n cannot be a composite knot with $n + 1$ factor knots?

3. If K is a knot with unknotting number 1, is there always a crossing in any minimal crossing projection that we can change to make it the unknot?

4. If K is an alternating knot with unknotting number 1, is there always a crossing in each alternating projection that we can change to make it the unknot? (Unsolved Question 3 in fact implies Unsolved Question 4.)

5. Old conjecture: $u(K_1 \# K_2) = u(K_1) + u(K_2)$. Note that it is certainly always true that $u(K_1 \# K_2) \leq u(K_1) + u(K_2)$. (Show this.) Scharlemann's result says the conjecture is true in the case $u(K_1 \# K_2) = 1$. Perhaps you could prove it when $u(K_1 \# K_2) = 2$.

Exercise 3.4 Show that a knot like the one in Figure 3.6 is alternating by finding an alternating projection. Then show that it has unknotting number 1 by showing that there is a crossing in this projection that can be changed to yield the trivial knot.

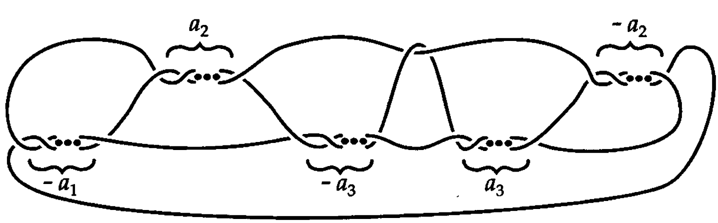


Figure 3.6 Knots of this type are alternating and have unknotting number 1.

One would expect that the unknotting number of a knot is realized in a projection of the knot with a minimal number of crossings. Amazingly enough, this is *not* always the case. In 1983, Steve Bleiler and Y. Nakanishi independently discovered the following example. Here is a knot with

Conway notation 514 (Figure 3.7). It is known that this knot cannot be drawn with fewer crossings, so its minimal crossing number is 10. It is also known that this is the only projection (up to planar isotopy and mirror reflection) of this knot with 10 crossings.

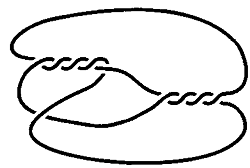


Figure 3.7 The knot 514.

Exercise 3.5 Check that it takes at least three crossing changes in the projection in Figure 3.7 to unknot this knot.

Here is another projection of the same knot (Figure 3.8). It has Conway notation 2 −2 2 −2 2 4. (Check for yourself that the two continued fractions give the same rational number.)

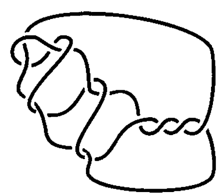


Figure 3.8 The knot 2 −2 2 −2 2 4.

Exercise 3.6 Show that the projection of the knot in Figure 3.8 can be unknotted by changing only two crossings.

In fact, one can prove that the unknotting number of the knot in Figure 3.8 is in fact 2. Thus, the unknotting number of this knot is realized by a projection that is *not* minimal! That's surprising.

While we are at it, let's discuss a concept related to unknotting number. Given a projection of a knot, define a **k-move** to be a local change in the projection that replaces two untwisted strings with two strings that twist around each other with k crossings in a right-handed manner. A $-k$-move will be the same, except that the twist is a left-handed twist (Figure 3.9). (Again, if k is positive, we make the overstrand in the new crossings have positive slope.)

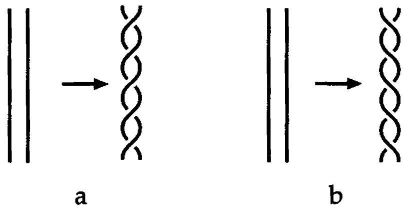


Figure 3.9 (a) A 5-move. (b) A −5-move.

We say that two knots or links are ***k*-equivalent** if we can get from a projection of one to a projection of the other through a series of *k*-moves and −*k*-moves. We allow ourselves to change the projections via ambient isotopies between the various moves that we perform.

Exercise 3.7 Show that every link is two-equivalent to the trivial link with the same number of components.

Unsolved Conjecture 1

Show that every link is three-equivalent to a trivial link. This conjecture is due to Y. Nakanishi of Kobe University. It's surprising that no one has succeeded in proving or disproving the conjecture yet.

Exercise 3.8 Show that the knots in Figure 3.10 are each three-equivalent to a trivial link.

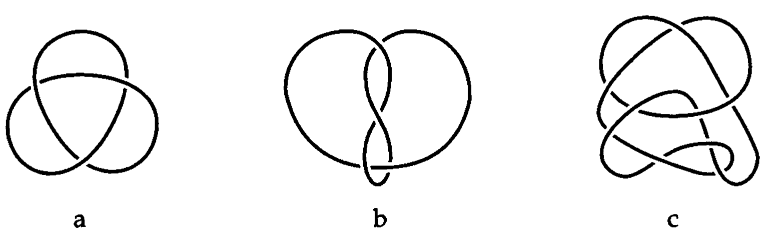


Figure 3.10 Knots that are three-equivalent to trivial links. (a) 3_1. (b) 4_1. (c) 9_{42}.

Unsolved Conjecture 2

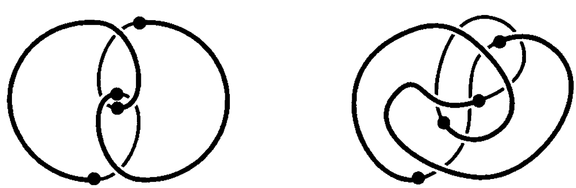
Show that every knot is four-equivalent to the trivial knot. This is known to be true for rational knots, pretzel knots, and closed three-string braids (which we discuss in Section 5.4).

Exercise 3.9 Show that if a link is tricolorable, then any link that is three-equivalent to it must also be tricolorable.

Note that if the first unsolved conjecture is proved to be true, then the links with tricoloration are exactly the links that are three-equivalent to a trivial link with more than one component. The links without tricoloration would be exactly those links three-equivalent to the trivial link of one component, namely the unknot.

3.2 Bridge Number

In Figure 3.11, we show a pair of unusual projections of the trefoil and figure-eight knots, respectively. In these pictures, think of the knots as cutting through the projection plane, rather than lying in it. Think of the darkened portions of the knots as lying above the plane and the rest of the knots as lying below the plane. Each knot intersects the plane in four vertices. In both of the pictures, there are two unknotted arcs from each knot lying above the plane. This is the least number of such unknotted arcs in any projection of these knots. Hence, we say these knots both have bridge number 2.



a b

Figure 3.11 The (a) trefoil and (b) figure-eight knots.

In general, given a projection of a knot to a plane, define an **overpass** (Figure 3.12a) to be a subarc of the knot that goes over at least one crossing but never goes under a crossing. A **maximal overpass** is an overpass that could not be made any longer (Figure 3.12b). Both of its endpoints occur just before we go under a crossing. The bridge number of the projection is then the number of maximal overpasses in the projection (those maximal overpasses forming the *bridges* over the rest of the knot). Note that each crossing in the projection must have some maximal overpass that crosses over it. The **bridge number of *K*,** denoted ***b(K)*,** is the least bridge number of all of the projections of the knot *K*.

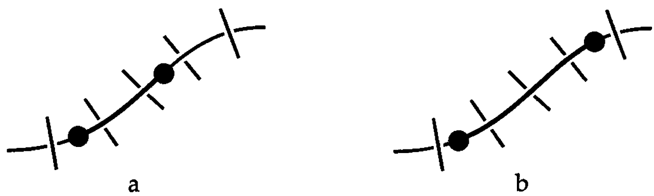


Figure 3.12 (a) An overpass. (b) A maximal overpass.

Exercise 3.10 Show that if a knot has bridge number 1, it must be the unknot.

Exercise 3.11 Show that the knot 5_2 has bridge number 2.

Exercise 3.12 (a) Show that the bridge number $b(K)$ of a nontrivial knot K is always less than or equal to the least number of crossings in any projection of the knot. (*Hint:* It may help to think about the cases where the projection is alternating or nonalternating separately.)

(b)* Show that the bridge number $b(K)$ of a nontrivial knot K is strictly less than the least number of crossings in any projection of the knot.

Knots that have bridge number 2 are a special class of knots, known as **two-bridge knots**. Suppose we cut a two-bridge knot open along the projection plane. We would be left with two unknotted untangled arcs from the knot above the plane, corresponding to the two maximal overpasses, and two unknotted untangled arcs from the knot below the plane. Note that they are unknotted and untangled, since they can have no crossings with each other. All of the crossings came from a maximal overpass and one of these arcs. So, if we want to construct all possible two-bridge knots, we just glue the endpoints of two unknotted untangled strings above the plane to the endpoints of two unknotted untangled strings below the plane. The tricky part is that although the strings to each side of the plane are individually unknotted, they can twist around each other and themselves. So from the side view, we see something like Figure 3.13.

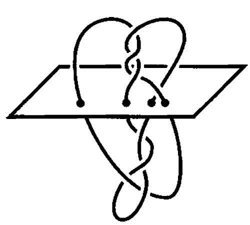


Figure 3.13 A two-bridge knot (side view).

Figure 3.14 shows the side view of the two-bridge representations of the trefoil and figure-eight knots from Figure 3.11. Given a picture of a two-bridge knot as in Figure 3.13, we can always free one of the strings and redraw our projection as in Figure 3.15. Now we can see that this two-bridge knot is in fact simply a rational knot, by turning every other integer tangle horizontal, starting with the bottom one (Figure 3.16). *In fact, the two-bridge knots are exactly the rational knots.*

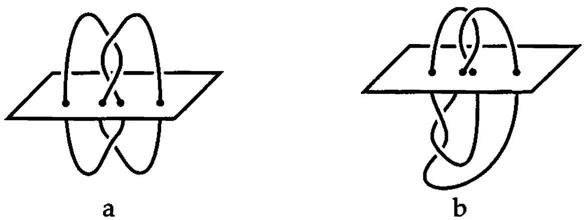


Figure 3.14 Another view of the (a) trefoil and (b) figure-eight knots.

Figure 3.15 Two-bridge knot redrawn.

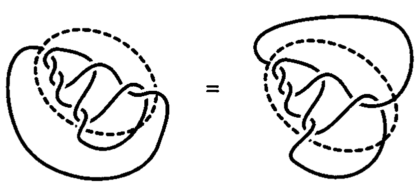


Figure 3.16 A two-bridge knot is a rational knot.

The two-bridge knots are a very well understood class of knots. Often, a property that is suspected to hold for all knots is first proved to hold for this particular class of knots. For instance, Claus Ernst and Dewitt Sumners proved that the number of distinct two-bridge knots of n crossings is at least $(2^{n-2} - 1)/3$. Since two-bridge knots are known to be prime, this implies that the number of distinct prime knots of n crossings is at least $(2^{n-2} - 1)/3$. Note that we are counting a knot and its mirror image as distinct knots if they are not equivalent.

The first three-bridge knot in the appendix table is 8_{10} (Figure 3.17).

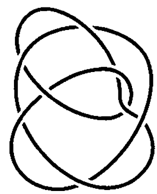


Figure 3.17 The knot 8_{10} is a three-bridge knot.

Exercise 3.13 Find a picture of 8_{10} that shows that it is at most a three-bridge knot.

Unsolved Question

Classify the three-bridge knots. The two-bridge knots are well understood, simply corresponding to the rational knots. No one has yet understood all of the three-bridge knots.

In 1954, a German mathematician named H. Schubert proved that $b(K_1 \# K_2) = b(K_1) + b(K_2) - 1$.

Exercise 3.14 Explain how Schubert's result implies that rational knots are all prime.

3.3 Crossing Number

We have discussed this invariant before. The **crossing number** of a knot K, denoted $c(K)$, is the least number of crossings that occur in any projection of the knot.

How do we determine the crossing number of a knot K? First, we find a projection of the knot K with some number of crossings n. Then we know the crossing number is n or smaller. If all of the knots with fewer crossings than n are known, and if K does not appear in the list of knots of fewer than n crossings, then K must have crossing number n. So, for instance, the knot 7_3 has crossing number 7 since it has a projection with 7 crossings and it is distinct from all the knots of fewer than 7 crossings (Figure 3.18). (This last fact is very difficult and gets at the essence of knot theory. How do you prove that 7_3 does not equal 3_1, 4_1, 5_1, 5_2, 6_1, 6_2, 6_3 or $3_1 \# 3_1$? The answer will have to wait until Chapter 7, when we utilize polynomials to distinguish these knots.)

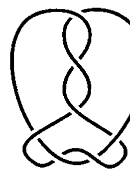


Figure 3.18 The knot 7_3 has crossing number 7.

In general, it is very difficult to determine the crossing number of a given knot. If we have a knot in a projection with 15 crossings, how can we hope to show that it can't be drawn with fewer than 15 crossings? Nobody yet knows what all the knots of 14 crossings are.

Sometimes, we can still determine the crossing number. In 1986, Louis Kauffman (From the University of Illinois at Chicago), Kunio Murasugi (from the University of Toronto), and Morwen Thistlethwaite (from the University of Tennessee) independently proved the first major result concerning crossing number. Call a projection of a knot **reduced** if there are no easily removed crossings, as in Figure 3.19. Kauffman, Murasugi, and Thistlethwaite proved that an alternating knot in a reduced alternating projection of n crossings has crossing number n. There cannot be a projection of such a knot with fewer crossings. They utilized the Jones polynomial for knots in order to prove this. We discuss the Jones polynomial in Chapter 6.

Figure 3.19 These crossings are easily removed, lowering the crossing number.

Since we can tell by just looking at an alternating projection whether or not it is reduced, and since we can lower the number of crossings if it is not reduced, we can tell the crossing number of *any* alternating knot. For instance, the crossing number of the knot 7_3 is in fact 7 since it appears in Figure 3.18 in a reduced alternating projection of 7 crossings. Here is an alternating knot in a reduced alternating projection of 23 crossings (Figure 3.20). Hence its crossing number is 23. There cannot be a projection of this knot with fewer than 23 crossings.

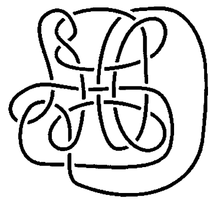


Figure 3.20 This knot has crossing number 23.

The question of determining the crossing number for a nonalternating knot is still very much open. In fact, we can't yet say anything about the crossing number of a composite knot.

Big Unsolved Question

Show that the crossing number of a composite knot is the sum of the crossing numbers of the factor knots, that is, $c(K_1\#K_2) = c(K_1) + c(K_2)$ (Figure 3.21).

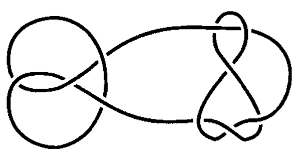


Figure 3.21 Is $c(K_1\#K_2) = c(K_1) + c(K_2)$ for a composite knot?

This problem has been open for 100 years. Note that Kauffman, Murasugi, and Thistlethwaite's result implies that the conjecture does hold when $K_1\#K_2$ is an alternating knot (see Kauffman, 1988).

Exercise 3.15 Show that if K_1 and K_2 are alternating, then so is $K_1\#K_2$. Hence $c(K_1\#K_2) = c(K_1) + c(K_2)$ holds when K_1 and K_2 are alternating, even if $K_1\#K_2$ does not appear alternating (Figure 3.22).

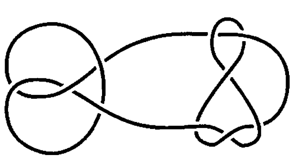


Figure 3.22 $K_1\#K_2$ appears nonalternating.

We will come back to crossing number when we discuss particular categories of knots.

4 Surfaces and Knots

4.1 Surfaces without Boundary

In this chapter, our goal is to use surfaces to help understand and distinguish knots. But first of all, what do we mean by a surface? Certainly, all of the objects in Figure 4.1 qualify as surfaces. Note that these are not solid objects. They are just the surface of the object. For instance, an example of a surface is the glaze on a doughnut, not the doughnut itself. (Keep in mind that we think of the glaze as being infinitely thin.) We call the surface of a doughnut a **torus**.

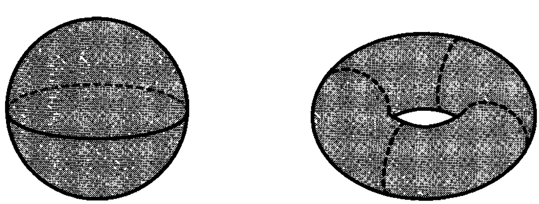


Figure 4.1 Some surfaces.

What is the property that surfaces have in common? At any point on a surface, there is a small region on the surface surrounding and containing the point that looks like a disk (Figure 4.2). The disk doesn't have to be flat, it can be deformed, but it still must be a disk. (If you ironed it, it would be a flat disk.) For example, in Figure 4.3 we see some objects that are *not* surfaces. They fail to be surfaces because each of them has at least one point such that the region on the object surrounding that point does not form a disk on the object, no matter how small a region we take. In each of the three examples, there exist points with "neighborhoods" around them, appearing as in Figure 4.4.

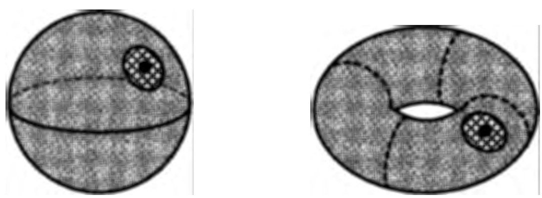
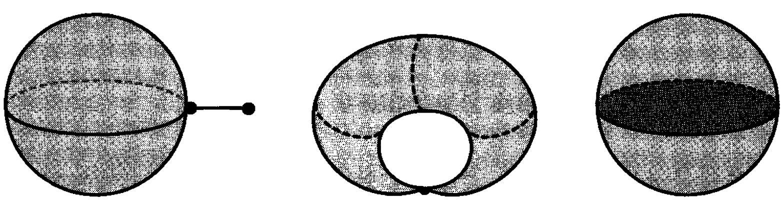
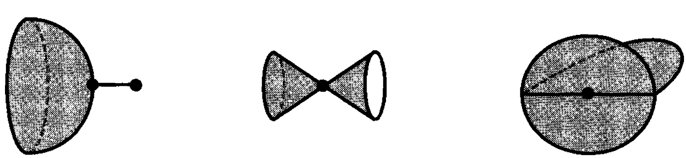


Figure 4.2 Each point on a surface is surrounded by a disk.



a b c

Figure 4.3 These are *not* surfaces.



a b c

Figure 4.4 Nondisk neighborhoods of points.

Another name for a surface is a **two-manifold**. A two-manifold is defined to be any object such that every point in that object has a neighborhood in the object that is a (possibly nonflat) disk.

Exercise 4.1 Based on the definition for two-manifolds, decide what the definition should be for a one-manifold. Find two different one-manifolds.

We will eventually generalize to three-manifolds in Chapter 9. (Any thoughts on what the definition of a three-manifold should be? Our spatial universe appears to be a three-manifold.)

In order to apply surfaces to the study of knots, we first have to determine the possibilities for surfaces. In what follows, we think of all surfaces as being made of rubber, and hence deformable. Thus, we consider a sphere and a cube to be equivalent surfaces, since we could pull out eight points on a rubber sphere to make it look like a cube, without having to do any cutting and pasting (see Figure 4.5). This idea of treating objects as if they were made of rubber is the fundamental concept behind topology. Topologists are interested in the properties of objects that remain unchanged, even as the object is deformed.

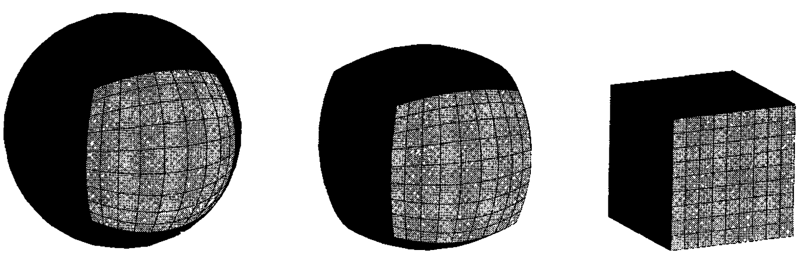


Figure 4.5 A rubber sphere is equivalent to a cube.

Similarly, we consider each of the surfaces shown in Figure 4.6 to be equivalently placed in space because we could get from any one to any other by a rubber deformation. Mathematicians call such a rubber deformation an **isotopy**. (Isotopy is a generic name for a rubber deformation, whether it's a deformation of a knot or a surface.) Two surfaces in space that are equivalent under a rubber deformation are called **isotopic surfaces.**

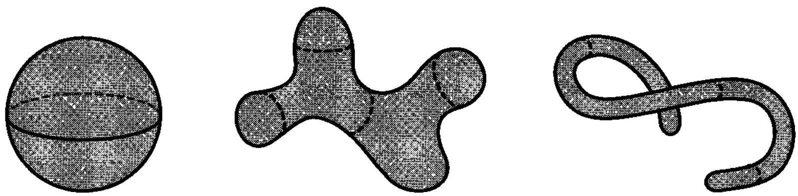


Figure 4.6 These are isotopic surfaces.

Exercise 4.2 By drawing a sequence of pictures to depict the rubber deformations, show that the three surfaces in space in Figure 4.7 are all isotopic to one another.

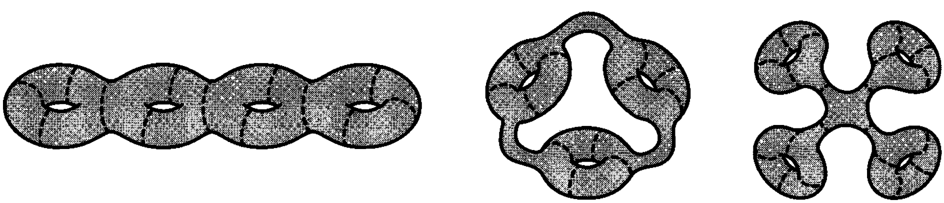


Figure 4.7 These three surfaces are all isotopic.

The two surfaces in Figure 4.8 are not isotopic, however, because we could not deform the first to look like the second without doing some cutting and pasting. (A proof that they are not isotopic would require a lot more work.)

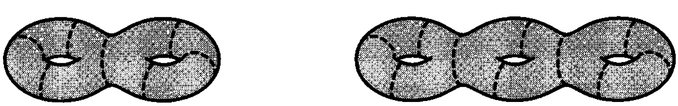


Figure 4.8 These are nonisotopic surfaces.

In order to better work with surfaces, we cut them into triangles. The triangles have to fit together nicely along their edges so that they cover the entire surface. They cannot intersect each other in any of the ways pictured in Figure 4.9. The triangles needn't be flat with straight edges. Just like all the other objects in topology, they are deformable. We can think of them as disks with a boundary made up of three edges connecting three vertices. We call such a division of a surface into triangles a **triangulation**. Examples of triangulations of the sphere and the torus are given in Figure 4.10.

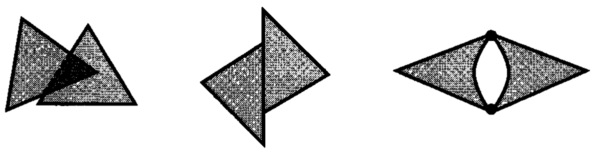


Figure 4.9 Triangles cannot intersect like this.

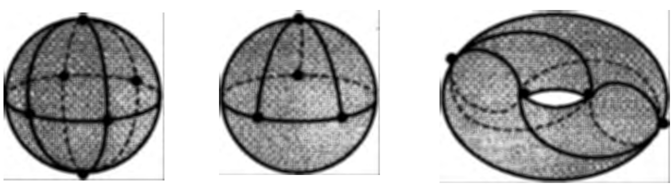


Figure 4.10 Triangulations of the sphere and the torus.

Given a surface with a triangulation, we can cut it into the individual triangles, keeping track of the original surface by labeling the edges that should be glued back together, and placing matching arrows on the pairs of edges that are to be glued (Figure 4.11).

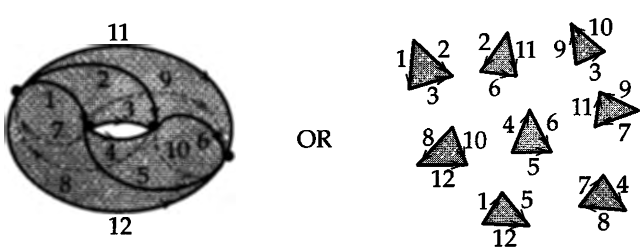


Figure 4.11 Two representations of a torus.

We say that two surfaces are **homeomorphic** if one of them can be triangulated, then cut along a subset of the edges into pieces, and then glued back together along the edges according to the instructions given by the orientations and labels on the edges, in order to obtain the second surface. For example, here are two homeomorphic copies of the torus. We simply cut along a subset of the edges of a triangulation that form a circle, knot the resulting cylinder, and then glue the two circles back together (Figure 4.12). Notice that we didn't even draw the rest of the triangulation, since it is clear we can find a triangulation of the torus such that the circle that we just cut along is contained within the edges of the triangulation.

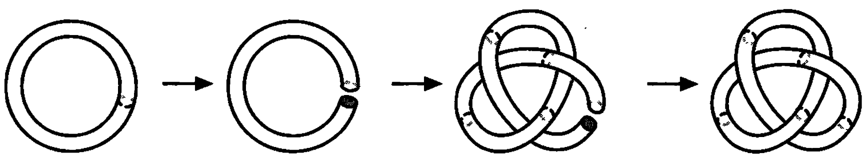


Figure 4.12 These two surfaces are homeomorphic.

Figure 4.13 shows another example of two surfaces that are not isotopic but that are homeomorphic. We can see the chain of cutting and gluing that takes us from the one surface to the other. Again, we don't actually need a complete triangulation, but rather a set of circles and edges that we cut the surface open along. We could always find a triangulation that contained this set of circles and edges as part of the union of the set of edges. (The fact that these two surfaces are not isotopic is not obvious and would take some work to prove.)

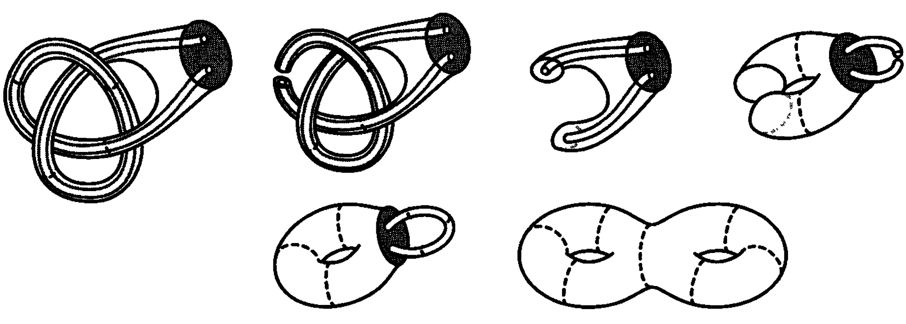


Figure 4.13 Two homeomorphic surfaces.

Exercise 4.3 Show that the two surfaces in Figure 4.14 are homeomorphic by drawing a sequence of pictures that show how to cut and paste the first in order to get the second.

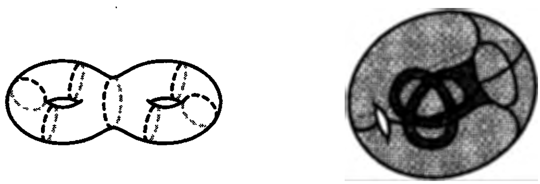


Figure 4.14 These two surfaces are homeomorphic.

A sphere and a torus are not homeomorphic (Figure 4.15). There is no triangulation of either one that can be rearranged and repasted to create the other surface. (There is clearly an inherent difference between a sphere and a torus. Every closed loop on a sphere cuts it into two pieces. However, there exist loops on a torus that do not cut it into two pieces. Unfortunately, we don't have time to prove that they are not homeomorphic.)

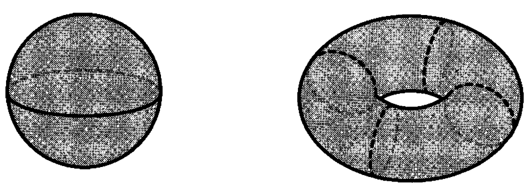


Figure 4.15 A sphere and a torus.

We could also have the surface of a two-hole doughnut or a three-hole doughnut. These possibilities are pictured in Figure 4.16. None of these four examples are homeomorphic to one another.

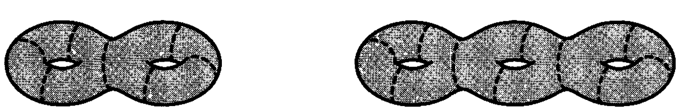


Figure 4.16 More surfaces.

Since we could just keep increasing the number of holes in our doughnuts, there are an infinite number of distinct (nonhomeomorphic) surfaces. We call the number of holes in the doughnut the **genus** of the surface. So the sphere has genus 0 and the torus has genus 1. The surfaces in Figure 4.16 have genera 2 and 3. Each of these surfaces can be placed in space in different ways. For instance, we saw two ways to put a torus in space in Figure 4.12. Even though both of those surfaces were tori (plural for torus), they were not isotopic, since there was no rubber deformation that would take us from the one to the other. However, they were still homeomorphic surfaces, just placed in space in two different ways. We call a choice of how to place a surface in space an **embedding** of the surface. Figure 4.12 depicts two distinct embeddings of the torus in three-space.

Figure 4.17 shows three distinct embeddings of a genus 3 surface in space. Although they are all homeomorphic to one another, only two of the three are isotopic to one another. Which two? You might think that it is the second and third surfaces. In fact, it is the first and third surfaces. Remember, the surfaces are made of highly deformable rubber. On the third surface, we can slide the end of one of the tubes along another tube to unknot the knotting. We call the third surface the surface of a **cube-with-holes**, as it is the surface of the solid object obtained by drilling three wormholes out of a cube.

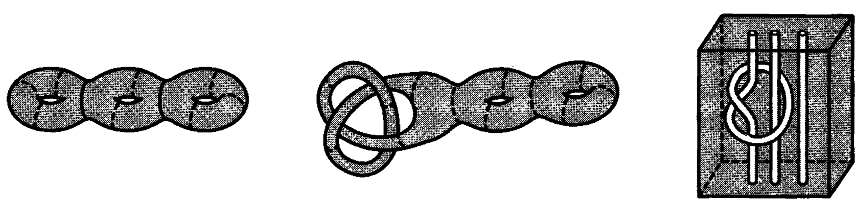


Figure 4.17 Three genus 3 surfaces.

Exercise 4.4 Draw a series of pictures that show the isotopy between the first and third surfaces.

Given a random surface in space, how do we tell what surface it is? (In the language of topology, what is its **homeomorphism type**?) It might be a sphere or torus, but it is so mangled that we don't recognize it. One option is to cut and paste to simplify the appearance of our surface until we can identify it. But this technique requires us to make a clever choice of how to cut up the surface and rearrange the pieces before regluing. It would be better if there were a method for recognizing surfaces that didn't require the cut-and-paste technique.

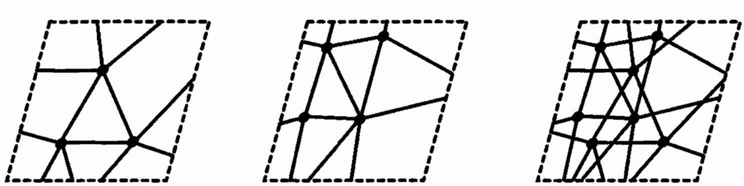
Let's take a triangulation of the surface. Let V be the number of vertices in the triangulation. Let E be the number of edges and let F be the number of triangles. (F stands for faces. It turns out that the formula can also be applied when the faces are not just triangles, but are polygons with more than three edges.) We define the **Euler characteristic** of the triangulation to be $\chi = V - E + F$. So, for example, in the case of the first triangulation of the sphere in Figure 4.10, $V = 6$, $E = 12$, and $F = 8$, so $\chi = 6 - 12 + 8 = 2$.

Exercise 4.5 Compute the Euler characteristic of the second triangulation of the sphere in Figure 4.10.

Exercise 4.6 Compute the Euler characteristic of the triangulation of the torus in Figure 4.10.

Notice that in Exercise 4.5 you obtained the same answer that we had already obtained for the sphere using a different triangulation. In fact, this will always be the case. The Euler characteristic depends only on the surface, not on the particular triangulation of the surface that we use. Although the rigorous proof is a bit technical, let's take a look at the idea behind the proof.

Suppose that we have two different triangulations of the same surface S, call them T_1 and T_2. Let's place them both on the surface at the same time, so that they are overlapping (Figure 4.18). We will build a new triangulation T_3 of S that "contains" each of T_1 and T_2 within it. As we build it up, we will show that it has the same Euler characteristic as T_1. Since the same argument can be used to show that it has the same Euler characteristic as T_2, we will have shown that T_1 and T_2 have the same Euler characteristic.



a b c

Figure 4.18 (a) T_1 (b) T_2 (c) $T_1 \cup T_2$

We will assume that each edge of T_1 intersects each of the edges of T_2 a finite number of times. There is a technical proof that the edges of T_1 can be moved just slightly, to make sure that this is the case, but we will not go into it as it is too time-consuming and would take us too far afield. We will also assume that the vertices of T_2 do not lie on top of a vertex or edge from T_1, which can be made true by moving T_1 slightly.

We begin to build the new triangulation T_3 by starting with T_1 (as in Figure 4.19a). One at a time, we add to the vertices of T_1 a new set of vertices corresponding to where the edges of T_2 cross the edges of T_1. Each new vertex also cuts an edge into two edges. Since when computing the Euler characteristic, the number of vertices is added and the number of edges is subtracted, the Euler characteristic is unchanged by this operation. (See Figure 4.19b.) We also add each vertex in the second triangulation T_2 to T_3, together with one edge that runs from that vertex to one of the vertices that is already in T_3, as in Figure 4.19c. We choose each of these new edges to be a subset of one of the edges from T_2. Note also that the addition of each new vertex and edge doesn't change the Euler characteristic, since the number of faces (admittedly, funny-looking faces) hasn't changed, while the number of vertices and the number of edges has each gone up by one. Sometimes we will need to add a chain of edges to connect a vertex of T_2 and T_3; however, the Euler characteristic remains unchanged.

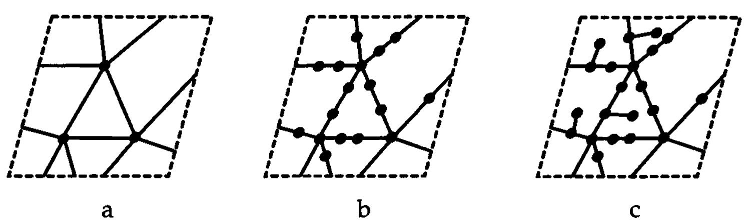


Figure 4.19 (a) T_1. (b) Add vertices. (c) Add pairs of vertices and edges.

Now we add all of the pieces of edges from T_2 that have not been added yet, each of which becomes a separate edge in T_3. Note that as we add one of these edges, as in Figure 4.20a, we cut a face in two. Hence, the number of edges and the number of faces each goes up by one, leaving the Euler characteristic unchanged. We now have a picture as in Figure 4.20b. Of course, at this point, as is the case with our picture, we may not have a triangulation. Some of the faces may not be triangles. So now we just add edges to cut the faces into triangles, as in Figure 4.20c. When we add such an edge, it cuts an existing face into two pieces, so both the number of edges and the number of faces goes up by one, again leaving the Euler characteristic unchanged. Thus, we have shown that there exists a third triangulation, T_3, with the same Euler characteristic as T_1, such that it "contains" both T_1 and T_2. Since we could have built it by starting with T_2, it also has the same Euler characteristic as T_2. Hence, we have shown that T_1 and T_2 must have the same Euler characteristic.

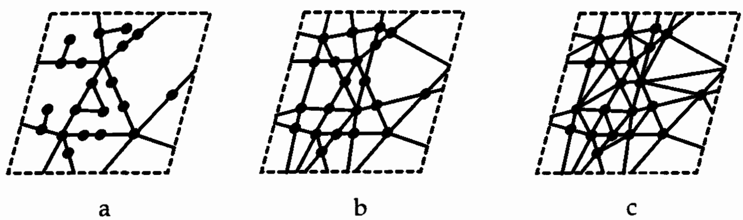


Figure 4.20 (a) Adding one more edge. (b) Adding the rest of T_2. (c) Triangulating the result.

Exercise 4.7 Find two triangulations of the sphere. Overlap them and find a third triangulation that "contains" both of them. Check that they all yield the same Euler characteristic.

Great, so Euler characteristic only depends on the type of surface that we have, and not on the particular triangulation. Any triangulation of the sphere has Euler characteristic 2, and any triangulation of the torus has Euler characteristic 0. But what about the Euler characteristic of a genus 2 surface? We could just take a triangulation of the surface and then compute its Euler characteristic. But instead, let's be a little bit more clever. One way to obtain a genus 2 surface is to remove a disk from each of two tori and then to glue the tori together along the resulting circle boundaries (Figure 4.21). This is called taking the **connected sum** of the tori.

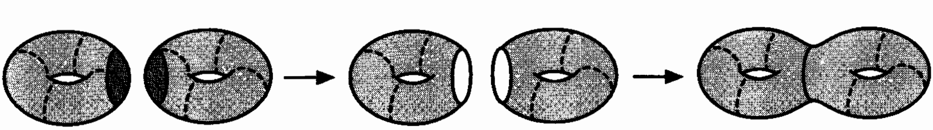


Figure 4.21 The connected sum of two tori is a genus 2 surface.

Suppose that we already have triangulations of the two tori. Then we can think of taking their connected sum as removing the interior of a triangle from each torus and then gluing together the boundaries of the two missing triangles by pairing up the vertices and edges (Figure 4.22). The result is a triangulated genus 2 surface. Since we have a triangulation for it, we can figure out what the Euler characteristic will be.

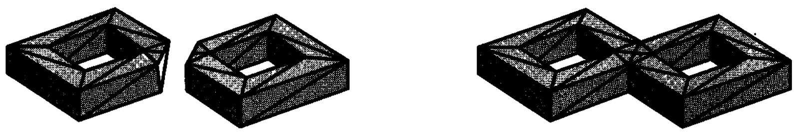


Figure 4.22 The connected sum of two triangulated tori.

The total number of vertices, edges, and faces in the triangulation of S is just the total number of vertices, edges, and faces in T_1 and T_2, with three fewer vertices, since we identified three vertices in T_1 with three vertices in T_2, three fewer edges, since we identified three edges in T_1 with three edges in T_2, and two fewer faces, since we threw away the interiors of two triangles in order to construct the connected sum. But since V is added into the formula and E is subtracted from the formula, the loss of three vertices and three edges has no net effect on the Euler characteristic. Hence the only effect is the loss of two faces. Therefore we obtain

$$\chi(S) = \chi(T_1) + \chi(T_2) - 2$$

Since we know that the Euler characteristic of a torus is 0, this says $\chi(S) = -2$.

Exercise 4.8 Use connected sums to show that the Euler characteristic of a genus 3 surface is -4.

Exercise 4.9 Use induction to show that the Euler characteristic of a surface of genus g is $2 - 2g$.

Let's make the computation of Euler characteristic even easier. We no longer insist that the faces be triangles. Instead, we can subdivide the surface into vertices, edges, and faces, where a face is simply a disk with its boundary made up of a sequence of edges connecting the vertices (better known as a polygon). Our only restriction on a face is that it be a single piece that has no holes in it. For example, we could subdivide the torus into a single face, with one vertex and two edges, obtaining $\chi = 0$. Or we could cut the genus 2 surface up into 4 faces, with 6 vertices and 12 edges, yielding the expected $\chi = -2$ (Figure 4.23).

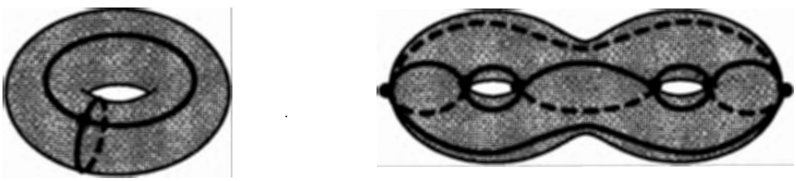


Figure 4.23 Subdividing the torus and genus 2 surface.

Exercise 4.10 Use Euler characteristic to determine the genus of the surface in Figure 4.24.

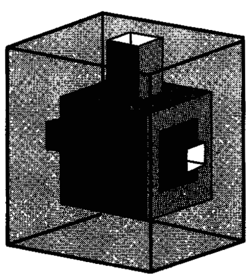


Figure 4.24 What is the genus of this surface?

One question remains: How do we know that every surface has a triangulation? This turns out to be a hard technical fact that was proved in the 1930s. However, even though every surface does have a triangulation, not every surface has one with a finite number of triangles.

We say that a surface is **compact** if it has a triangulation with a finite number of triangles. So the sphere and torus are certainly compact surfaces. But neither the plane nor a torus minus a disk (Figure 4.25) is compact, as neither can be triangulated with finitely many triangles. In the case of the plane, this is obvious. In the case of the torus minus a disk, we would have to use infinitely many triangles, getting smaller and smaller as we approached the boundary of the missing disk. Note that both the plane and the torus minus a disk do satisfy the definition of a surface.

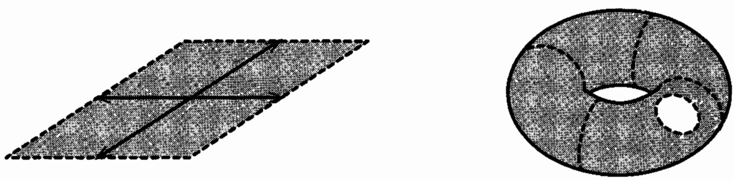


Figure 4.25 A plane and a torus minus a disk.

We are primarily interested in compact surfaces. They have the advantage that we can compute their Euler characteristic.

Where do surfaces appear in knot theory? In the space around the knot. Let R^3 be the three-dimensional space that the knot K sits in. The space around the knot is everything but the knot, which we denote $M = R^3 - K$. We call M the **complement of the knot**. It is what is left over if we drill the knot out of space (Figure 4.26). All of the surfaces that we look at live in the complement of the knot.

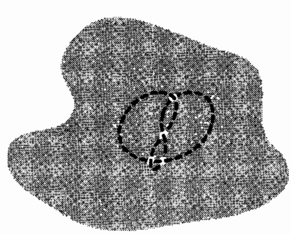


Figure 4.26 The complement of the knot is everything *but* the knot.

Figure 4.27 shows an example of a surface in the complement of a link when the link is splittable. Since we can pull the components of the link apart, we can think of there being a sphere that separates the components from one another. In fact, an alternative way to define a splittable link is simply to say that it is a link such that there is a sphere in the link complement that has components of the link on either side of it.

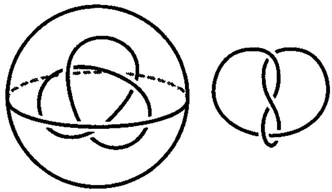


Figure 4.27 A sphere in the complement of a splittable link.

Note that every knot is contained in a torus like the one in Figure 4.28. But Figure 4.29 contains a torus that surrounds a knot in a more unusual way. We will see more examples of this in Section 5.2 when we discuss satellite knots. And Figure 4.30 is an example of a genus 2 surface in the complement of a knot.

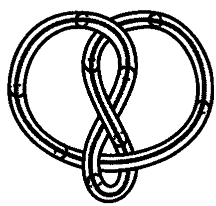


Figure 4.28 Every knot is contained in a torus.

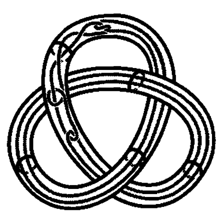


Figure 4.29 A torus surrounding a knot.

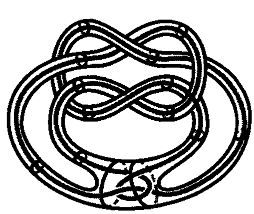


Figure 4.30 A genus 2 surface around a knot.

We are particularly interested in surfaces in knot and link complements that cannot be simplified. In particular, let L be a link in R^3 and let F be a surface in the complement $R^3 - L$. We say that F is **compressible** if there is a disk D in $R^3 - L$ such that D intersects F exactly in its boundary *and* its boundary does not bound another disk on F. Note that D is not allowed to intersect L.

For instance, the surface F in Figure 4.31 is compressible since the disk D is a disk in R^3 that does not intersect the link L, intersects F exactly in its boundary, and its boundary does not bound a disk on F. A compressible surface can be simplified, by cutting it open along the boundary of the disk and then gluing two copies of the disk to the two curves that result. We obtain a simpler surface (or sometimes a pair of surfaces) that still lies in the link complement. This simplifying operation is called a **compression** of the original surface.

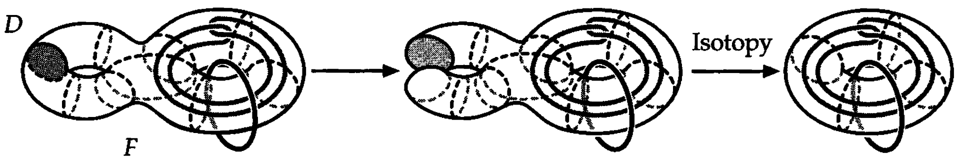


Figure 4.31 A compressible surface in $R^3 - L$ and the simpler surface we can construct from it.

Exercise 4.11 Show that a compression always increases Euler characteristic. Use this to show that the genus of the resulting surface or surfaces is always less than the genus of the original surface.

If a surface is not compressible, we say that it is **incompressible**. For instance, the torus in Figure 4.32 is incompressible, although proving it is somewhat difficult. But notice that any disk that intersects the torus in its boundary looks like it either must intersect the link L or its boundary must cut a disk off of the torus.

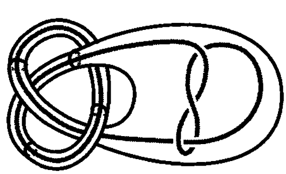


Figure 4.32 This torus is incompressible since no compressing disks exist.

An incompressible torus like the one in Figure 4.32 exists any time that we have a composite knot. It is called a **swallow-follow torus** because it swallows one of the two factor knots and follows the other one. Surprisingly, the genus 2 surface in Figure 4.30 is compressible.

*Exercise 4.12** Find a disk in Figure 4.30 that demonstrates that the surface in the figure is compressible. If we simplify the surface using this disk, what surface do we get?

All of the surfaces we have looked at so far are surfaces that do not have boundaries. We now want to look at surfaces with boundaries.

4.2 Surfaces with Boundary

In order to obtain surfaces **with boundary**, we can just remove the interiors of disks from the surfaces that we already have (Figure 4.33). We leave the boundaries of the disks in the surfaces. These become the boundaries of the surfaces. All of the resulting boundaries are circles, which we will call **boundary components**. Since all of our surfaces are made of rubber, they can look very different when we deform them. For instance, Figure 4.34 shows two different pictures of a torus with one boundary component and the deformation for getting from the one picture to the other.

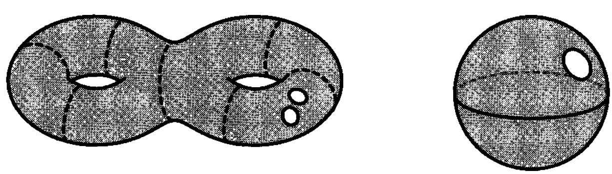


Figure 4.33 Surfaces with boundary.

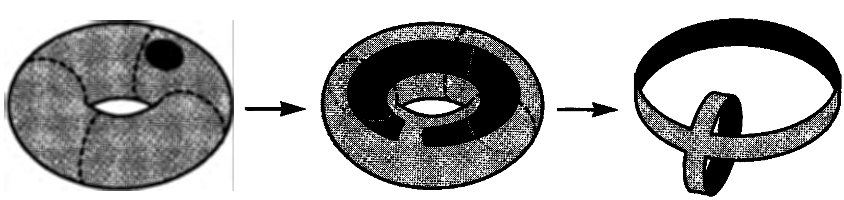


Figure 4.34 Pictures of a torus with one boundary.

How does the Euler characteristic apply to surfaces with boundary? When we remove a disk from a surface without boundary, we can think of it as removing the interior of one triangle in a triangulation of the surface. Hence the Euler characteristic goes down by one. Thus, a surface with three boundary components has an Euler characteristic three less than the Euler characteristic of the surface obtained by filling in the three boundaries with disks. Filling in boundary components by attaching disks is called **capping off** a surface with boundary.

Exercise 4.13 Find the Euler characteristics of each of the surfaces in Figure 4.35 without triangulating them.

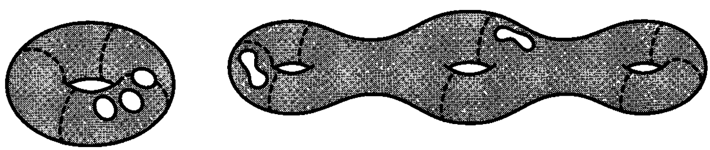


Figure 4.35 Find the Euler characteristics of these surfaces.

Unlike surfaces without boundary in three-space, surfaces with boundary cannot all be distinguished by Euler characteristic. For instance, Figure 4.36 contains two surfaces with boundary that have the same Euler characteristic, but that are not homeomorphic. It might help to picture these surfaces by thinking of their boundaries as wire frames and the surfaces as soap films spanning the wires.

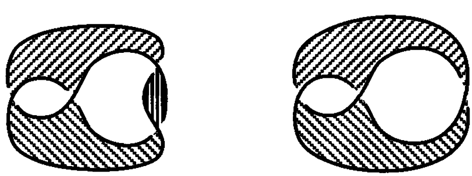


Figure 4.36 Both surfaces have the same Euler characteristic.

We can calculate Euler characteristic for surfaces with boundary just as we did for surfaces without boundary, by adding vertices and edges to cut the surface into a finite set of faces. Note that when we add vertices to the boundary of the surface, the resulting pieces of the boundary count as edges in our calculation of Euler characteristic.

Exercise 4.14 Verify that the two surfaces with boundary in Figure 4.36 have the same Euler characteristic.

We can actually construct these surfaces out of paper. For instance, in order to construct the first surface in Figure 4.36, cut out two larger disks and three thin strips of paper. At one end of each of the disks, tape two of the strips running from one disk to the other, each with a half twist in it, then tape the last strip from the one disk to the other with a full twist in it, making sure that the direction of the twist matches the direction of the twist in Figure 4.36 (see Figure 4.37).

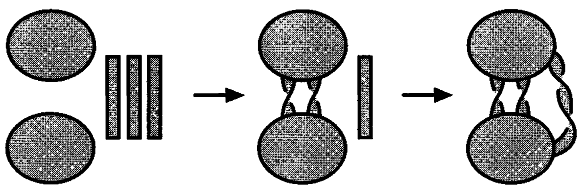


Figure 4.37 Constructing a paper surface.

There must be some trait other than Euler characteristic that distinguishes between these two surfaces. Suppose we start painting one side of the first surface gray. If we continue to paint that side, eventually we will end up painting the entire surface gray on both sides (Figure 4.38a). In essence, the surface doesn't have two distinct sides. Rather, the two sides are connected. On the other hand, we could paint the two sides of the second surface black and white, and nowhere would any black paint touch any white paint (Figure 4.38b). There really are two distinct sides of the surface.

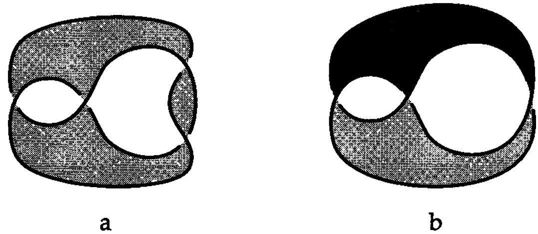
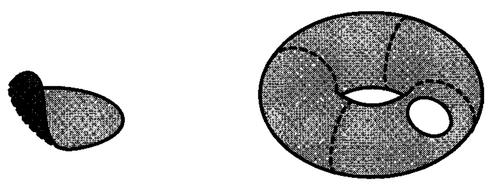


Figure 4.38 (a) One side. (b) Two sides.

We say that the second surface is **orientable**. A surface sitting in three-dimensional space is orientable if it has two sides that can be painted different colors, say black and white, so that the black paint never meets the white paint except along the boundary of the surface. So, for example, a torus is an orientable surface, because we could always paint the outer side black and the inner side white. Also, a disk and a torus with one boundary component are both orientable (Figure 4.39). In fact, any of the surfaces in Figures 4.15 and 4.16 with any number of disks removed to create boundary components will be orientable (Figure 4.40).



a b

Figure 4.39 A disk (a) and a torus with boundary (b) are both orientable.

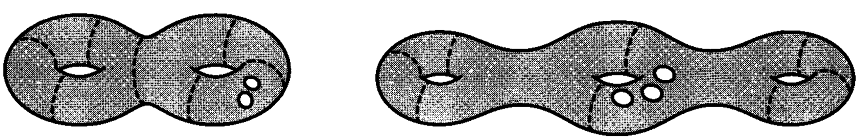


Figure 4.40 These surfaces are all orientable.

So what's another example of a surface that is not orientable? One of the simplest such surfaces is the **Möbius band** (Figure 4.41). This surface is not orientable because if we started painting one side of it black and continued working on that side, we would find that when we were done, we had painted all of it black. Because of the twist in the Möbius band, it only has one side. We call such a surface **nonorientable**. (Why do we use the word "nonorientable"? Print the letter S on a Möbius band so that the ink bleeds through to the other side. Now slide the letter S once around the Möbius band. We will now see Ƨ. The orientation of the letter has been reversed.)

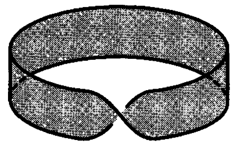


Figure 4.41 A Möbius band.

Figure 4.42 shows a stranger nonorientable surface. It also has only one side. In fact, a surface is nonorientable if and only if it contains a Möbius band within it. (The Möbius band may have an odd number of half-twists in it rather than just one half-twist, since it would be homeomorphic to the usual Möbius band.) In Figure 4.42, we have shaded such a Möbius band.

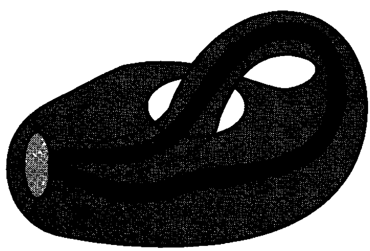


Figure 4.42 A Klein bottle with one boundary component.

Exercise 4.15 Decide which of the two surfaces in Figure 4.43 is orientable and which is nonorientable.

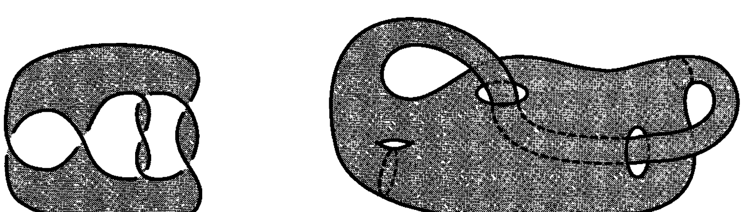


Figure 4.43 One surface is orientable, and one is not.

Now suppose that we have a messy surface with boundary, say the one in Figure 4.44, and we want to figure out what surface it is. To do so, we need to know three facts.

1. Is it orientable or nonorientable?
2. How many boundary components does it have?
3. What is its Euler characteristic?

These three pieces of information will completely determine the homeomorphic type of the surface. [See (Massey, 1967) for a proof.]

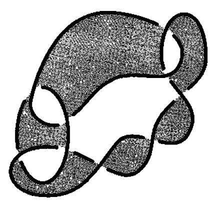


Figure 4.44 What surface is this?

In the case of Figure 4.44, the surface is orientable, it has three boundary components, and we can subdivide it, as in Figure 4.45, in order to determine that its Euler characteristic is -3. Therefore, if we cap off its boundary components with three disks, the resulting surface without boundary will have $\chi = 0$. So the resulting surface without boundary is a torus. Hence, our surface is simply a torus with three disks removed.

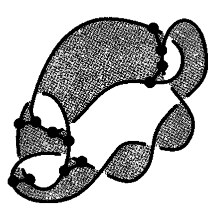


Figure 4.45 A subdivision.

Exercise 4.16 Use the three criteria to identify the surfaces with boundary in Figure 4.46 .

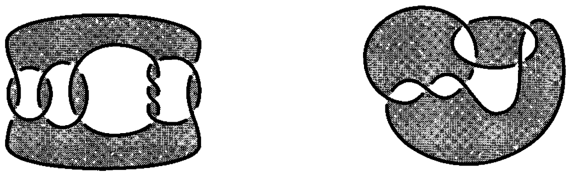


Figure 4.46 Identify these surfaces with boundary.

If a surface has boundary, we define its **genus** to be the genus of the corresponding surface without boundary obtained by capping off each of its boundary components with a disk. Thus, the genus of the surface with boundary shown in Figure 4.44 must be 1.

We would now like to apply surfaces with boundary to knot theory. As a first example, let's look at the unknot. One way to define the unknot is to say that it is the only knot that forms the boundary of a disk (Figure 4.47). In some projections of the unknot, the disk is not at all obvious, but it is always there.

Figure 4.47 The unknot always bounds a disk.

Another example of a surface with boundary in knot theory comes from composite knots. As in Figure 4.48, if we have a composite knot, there is a sphere with two boundary components that lies outside the knot. This surface is also called an **annulus**. Note that we thickened the knot up a little in this picture. Otherwise, if we had left the knot infinitely thin, we would have said the surface was a sphere with two punctures, the punctures occurring where the knot passed through the sphere. Thus, an alternative definition of a composite knot is a knot such that there is a sphere in space punctured twice by the knot such that the knot is nontrivial both inside and outside the sphere.

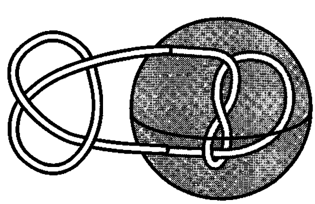


Figure 4.48 An annulus outside a composite knot.

Another place that surfaces snuck by was when we discussed tangles in Section 2.3. There we thought of a tangle as a region in the projection plane with four outgoing strands. We can also think of it as a portion of the knot surrounded by a sphere with four punctures, the punctures occurring where the knot passes through the sphere. Such a sphere is aptly called a **Conway sphere** (Figure 4.49). If we thicken up the knot, the punctures become holes and we have a sphere with four boundary components.

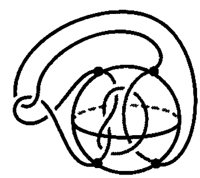


Figure 4.49 A Conway sphere.

A third example of a surface in knot theory appears in Figure 4.50. We see a Möbius band with boundary the trefoil knot. Even though the band has three twists instead of one, it is still a Möbius band. (This band and the Möbius band are homeomorphic since we can cut this band open along an arc, untwist one full twist, and then reidentify the points we first cut along, obtaining the Möbius band.)

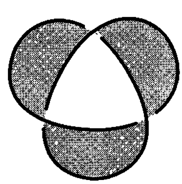


Figure 4.50 A Möbius band with boundary the trefoil knot.

We will be particularly interested in orientable surfaces with one boundary component such that the boundary component is a knot. For example, here is a torus with one boundary component where that boundary component is a trefoil knot (Figure 4.51). Admittedly, the surface pictured doesn't look much like a torus with one boundary, but as we saw in Figure 4.34, these surfaces can look kind of strange.

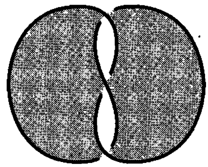


Figure 4.51 A torus with one boundary component, that boundary component being a trefoil knot.

Exercise 4.17 Use Euler characteristic to show that the surface in Figure 4.51 is indeed a torus with one boundary component.

4.3 Genus and Seifert Surfaces

We have seen that surfaces appear in knot theory in many ways. Particular types of knots have particular types of surfaces in their complements. However, surprisingly enough, there is one type of surface that appears in the complement of any knot. In 1934, the German mathematician Herbert Seifert came up with an algorithm so that, given any knot, one can create an orientable surface with one boundary component such that the boundary circle is that knot. This is pretty amazing. On first thought, it's hard to imagine how to get any orientable surface with one boundary component such that the boundary component is knotted at all. We are supposed to take a surface like a torus with one boundary component and embed it in space so that the boundary circle is knotted (Figure 4.52)? But we did see one example in the previous section. There, we saw a torus with one boundary component where that boundary component was knotted into a trefoil knot.

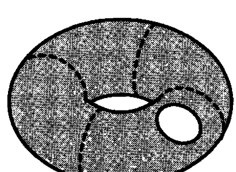


Figure 4.52 Embed this in space so that the boundary circle is knotted?

Seifert's algorithm tells us that not only can we embed surfaces in space with a knotted boundary component but we can do so to get *any* knot whatsoever. Suppose we want to construct such a surface for a particular knot *K*. Starting with a projection of the knot, choose an orientation on *K*. At each crossing of the projection, two strands come in and two strands go out. Eliminate the crossing by connecting each of the strands coming into the crossing to the adjacent strand leaving the crossing (Figure 4.53). Now all of the resultant strands will no longer cross. The result will be a set of circles in the plane. (They are not round circles in the usual sense, but rather, they can be deformed to round circles. So, to us topologists, they are circles.) These circles are called **Seifert circles**. Each circle will bound a disk in the plane. Since we do not want the disks to intersect one another, we will choose them to be at different heights rather than having them all in the same plane (Figure 4.54).

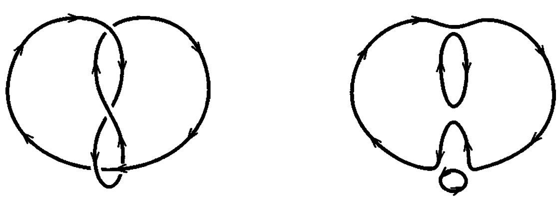


Figure 4.53 Eliminate all crossings.

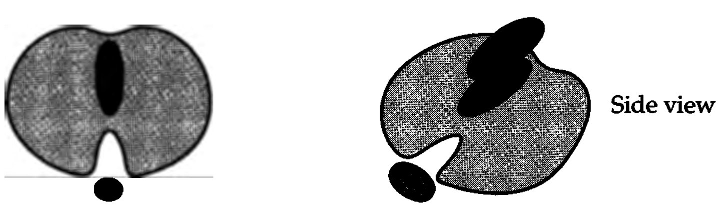


Figure 4.54 The circles bound disks at different heights.

Now we would like to connect the disks to one another at the crossings of the knot by twisted bands (Figure 4.55). The result is a surface with one boundary component *such that the boundary component is the knot.* Pretty amazing! Let's try that again. (See Figure 4.56.)

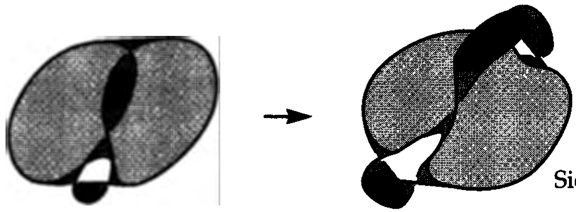


Figure 4.55 Connect the disks by twisted bands.

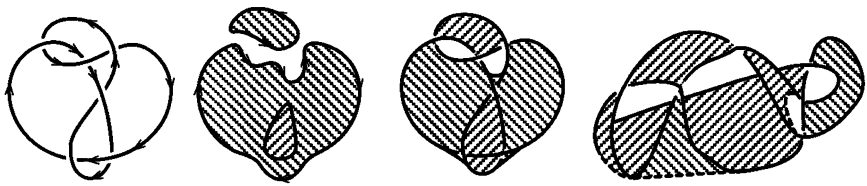


Figure 4.56 This surface has boundary the 6_3 knot.

Exercise 4.18 Show that the surface that we get doesn't depend on the direction we choose on the knot.

In fact, the surfaces that we are generating are always orientable. To see this, we need to show that each surface has two distinct sides, which can be painted two different colors. Let's give each Seifert circle the orientation that it inherits from the knot, either clockwise or counterclockwise. For each disk that has a clockwise orientation on its bounding Seifert circle, we paint its upward pointing face white and its downward pointing face black. For each disk that has a counterclockwise orientation on its bounding Seifert circle, we paint its upward pointing face black and its downward pointing face white.

At each crossing in the knot, we connect two of the disks bounded by the Seifert circles by a band containing a half-twist. If one of the two disks is adjacent to the other, the two disks have opposite orientations on their boundaries. Hence, the twist in the band allows us to extend the black and white paint across the two faces of the band so that they match up consistently on the disks. If one of the two disks is on top of the other, the two disks have the same orientation on their boundaries. Again, the twist in the band allows us to extend the paint consistently across the band. Thus, the entire surface can be painted black and white so that no black paint touches any white paint, and therefore the surface is orientable (Figure 4.57).

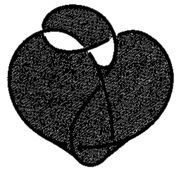


Figure 4.57 The surface has two sides that can be painted different colors.

Exercise 4.19 Note that if we add an edge and two vertices to the surface across each crossing, cutting each band in half, we cut the surface up into valid faces. Use this fact to show that if c is the number of crossings and s is the number of Seifert circles, then $\chi = s - c$ and the genus of the surface is $g = (s - c - 1)/2$.

Exercise 4.20 Use Seifert's algorithm to find surfaces bounding the knots in Figure 4.58. Use Euler characteristic to identify the surfaces.

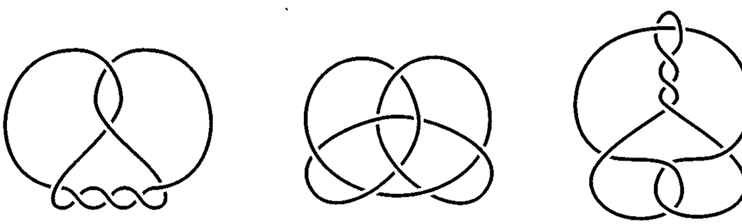


Figure 4.58 What are the surfaces?

Exercise 4.21 Show that Seifert's algorithm applied to a nontrivial projection always generates at least two Seifert circles.

Notice that Seifert's algorithm can be used to generate lots of different surfaces for the same knot (Figure 4.59). We could alter the projection of the knot and then obtain a surface that at least looks different.

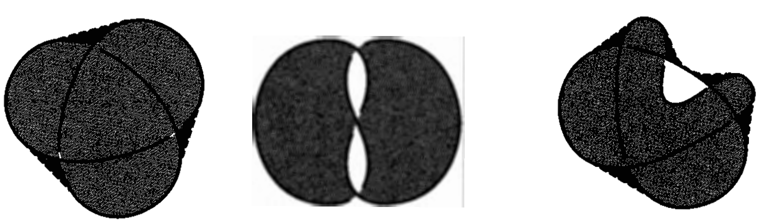


Figure 4.59 Other Seifert surfaces for the same knot.

Given a knot K, a **Seifert surface for** K is an orientable surface with one boundary component such that the boundary component of the surface is the knot K. We have just described one way to obtain a Seifert surface for a knot. However, there may be other Seifert surfaces for the same knot.

We define the **genus** of a knot to be the least genus of any Seifert surface for that knot. For example, the unknot bounds a disk. When we cap off the disk, we get a sphere, which has genus 0; therefore the unknot has genus 0 (Figure 4.60). Note that the unknot is the only knot with genus 0.

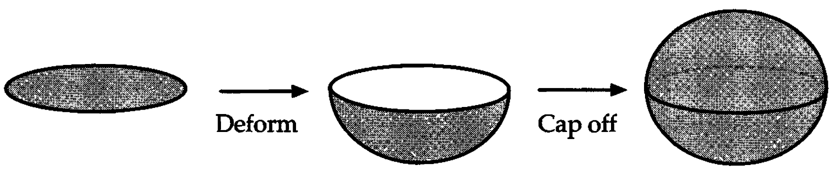


Figure 4.60 The unknot has genus 0.

What about the figure-eight knot? By Seifert's algorithm, we obtain a Seifert surface with genus 1. Since the figure-eight knot is not trivial, it cannot bound a surface of genus 0, so 1 is the least genus of a Seifert surface for the figure-eight knot. Thus, the genus of the figure-eight knot is 1.

Exercise 4.22 The twist knots are the knots shown in Figure 4.61. They include the trefoil and figure-eight knots. Show that all of the twist knots have genus 1.

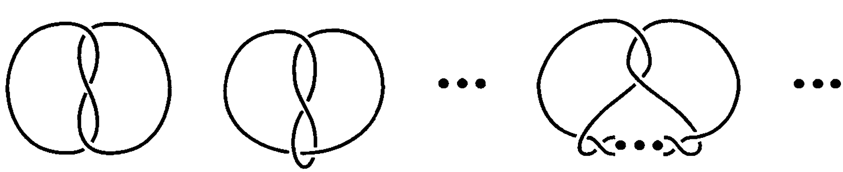


Figure 4.61 Show that all the twist knots have genus 1.

The definition for an incompressible surface from Section 4.1 applies without change to surfaces with boundary.

Exercise 4.23 Show that a minimal genus Seifert surface for a knot K must be incompressible. (Consider Euler characteristic.)

Is it true that Seifert's algorithm will always yield the Seifert surface of minimal genus? That's a little too much to hope for. We could take a very nasty projection of a knot and we couldn't expect to get the minimal genus Seifert surface by applying Seifert's algorithm to this knot. However, in the case of alternating knots, we can use Seifert's algorithm to find the minimal genus.

Theorem Applying Seifert's algorithm to an alternating projection of an alternating knot or link does yield a Seifert surface of minimal genus.

There are several proofs of this, the easiest of which is due to David Gabai (see Gabai, 1986), a professor at Caltech. According to the theorem, it is easy to calculate the genus of an alternating knot or link.

Exercise 4.24 Calculate the genus of the knots 6_3 and 7_6.

Let's look at what effect composition has on genus. Let $g(K)$ be the genus of a knot K.

Theorem $g(J\#K) = g(J) + g(K)$.

So if we know the genus for each of two knots, we can simply add them together to get the genus of the composition of the knots. Let's go through the proof of this, as it utilizes techniques that occur often in knot theory.

Proof. It's easy to see that $g(J\#K) \leq g(J) + g(K)$. We can just take a Seifert surface Q of genus $g(J)$ for J and a Seifert surface R of genus $g(K)$ for K, remove a little piece of each along their boundaries, and sew them together to obtain a Seifert surface of genus $g(J) + g(K)$ for $J\#K$ (Figure 4.62). However, it is conceivable that $J\#K$ has a Seifert surface with smaller genus than this. We will show that in fact it does *not*.

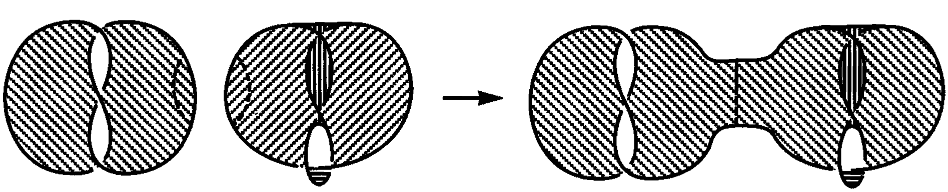


Figure 4.62 $g(J\#K) \leq g(J) + g(K)$.

Let S be a Seifert surface of minimal genus for $J\#K$. Since $J\#K$ is a composite knot, there is a twice-punctured sphere F that separates the J part of the knot from the K part of the knot. This twice-punctured sphere will intersect the Seifert surface S (Figure 4.63). We deform the surfaces through space (perform an isotopy in math lingo) in order to rearrange the way the two surfaces intersect. When S just touches F at a point, we can move S slightly to eliminate the intersection (Figure 4.64). We can also move S slightly so that the intersection consists entirely of loops and/or arcs (Figure 4.65).

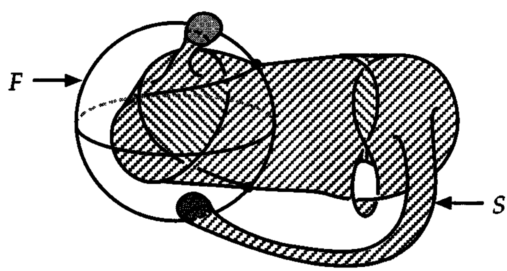


Figure 4.63 S and F.

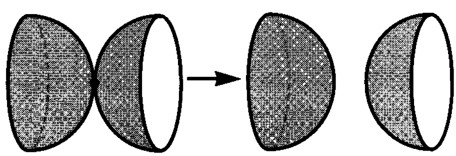


Figure 4.64 Remove single point intersections.

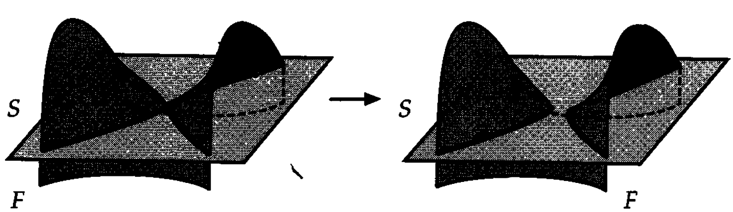


Figure 4.65 All intersection curves are either circles or arcs.

Although we can imagine much nastier situations, where our intersection set was even worse (say the two surfaces intersect in a disk or in an infinite number of discrete points), all of these situations can be remedied by a slight deformation of S, resulting in an intersection set containing only arcs and loops. This slight movement of S to simplify the intersection is called putting the surfaces in **general position**. There is an entire theory of mathematics that says that it can be done, but we won't get into that. Intuitively, it sounds reasonable, and we will go with that feeling.

There is an arc of intersection between F and S that begins and ends at the punctures of F. Since we can assume that the boundary of S intersects the punctures of F exactly twice, there can be only one arc of intersection between F and S (Figure 4.66). All other intersection curves between S and F are loops. We eliminate each of the loops of intersection one by one until none remain. Notice that there are three places where we can think of these intersection curves as lying (Figure 4.67). We can think of them as curves in the knot complement, floating around in three-space. We can also think of them as curves lying on the Seifert surface S. There, we have a set of intersection loops lying on S and one intersection arc that begins and ends on the one boundary component of S. We can also think of the intersection curves as lying on the twice-punctured sphere F. The one intersection arc begins at one of the punctures on F and ends at the other puncture.

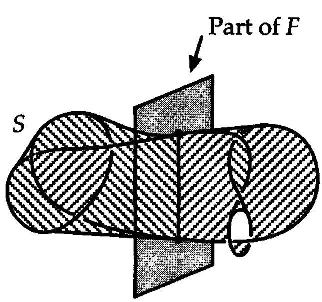


Figure 4.66 There is only one arc of intersection.

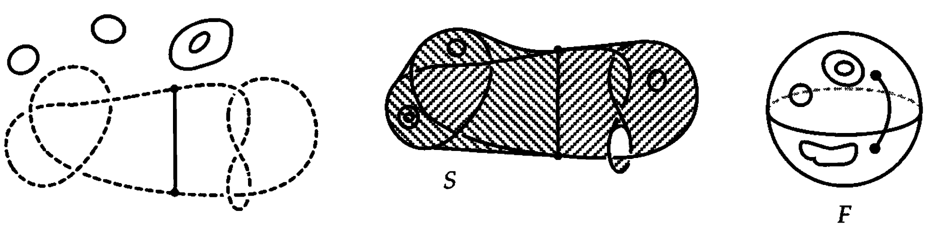


Figure 4.67 Three ways to think about intersection curves.

Given a particular intersection loop on F, it must either separate the two punctures from one another on F or it must have both of the punctures on the same side of it on F. However, since the single intersection arc connects the one puncture to the other on the surface of F and since that arc cannot intersect the loop, it must be that both of the punctures are on the same side of the loop on F. The other side of the loop must then be an unpunctured disk. Therefore, every intersection loop on F bounds an unpunctured disk on F (Figure 4.68).

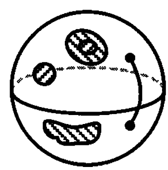


Figure 4.68 Every intersection loop bounds an unpunctured disk on F.

There must be an intersection loop that is *innermost* on F, that is to say, it bounds a disk on F containing no other intersection curves. We call this curve C. Cut S open along C, obtaining two copies of C in the cut open S. Glue disks to each of the new curves, where each disk is parallel to the disk bounded by C in F (Figure 4.69). Now, F and the new S do not intersect along C at all. This new S may or may not be connected. If it is not connected, throw away the piece of S that does not touch the knot. The resulting S is still a Seifert surface for $J\#K$, but it intersects F in one less intersection circle.

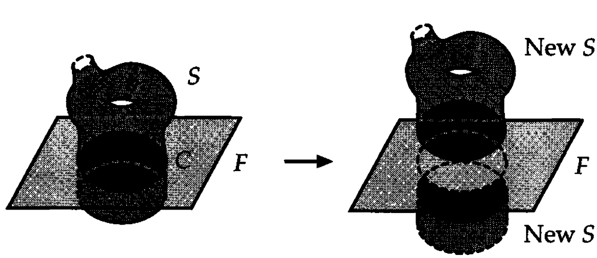


Figure 4.69 Forming a new S.

Exercise 4.25 Show that the "surgery" we just did to S could not have increased its genus (by looking at how the surgery changes the Euler characteristic of S).

Exercise 4.26 Show that the fact that S is a minimal genus Seifert surface implies that after cutting and then pasting in the two disks, the surface has two pieces, and the piece that we throw away is a sphere.

We repeat this surgery operation until S and F do not have any intersection circles left. Since S is still a Seifert surface for $J\#K$ and since the surgeries did not increase genus, S must still be a minimal genus Seifert surface for $J\#K$. There is now only one intersection arc between F and S. Thus, F divides S into a Seifert surface for J and a Seifert surface for K (Figure 4.70).

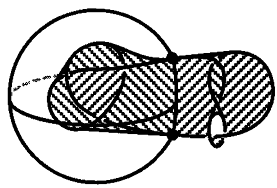


Figure 4.70 F divides S into a Seifert surface for J and a Seifert surface for K.

The sum of the genera of these two Seifert surfaces must then be the genus of S. Therefore $g(J) + g(K) \leq g(J\#K)$, since the genera of J and K are each less than or equal to the genera of these two Seifert surfaces. As we have already seen the reverse inequality, we have that

$$g(J) + g(K) = g(J\#K)$$

as we set out to prove.

Isotoping surfaces to clean up the intersections, and then performing surgeries to lower the number of intersection curves between the surfaces until none or one curve remains is a procedure that is relatively common in knot theory and in the more general field of topology. We can use this theorem to prove a fact we stated way back in Section 1.2, namely, that *the trivial knot cannot be the composition of two nontrivial knots* (Figure 4.71). Why not? Any nontrivial knot has genus at least 1 (genus 0 means the knot bounds a disk and is therefore trivial). So the composition of two nontrivial knots has genus at least 2 and therefore cannot be the trivial knot.

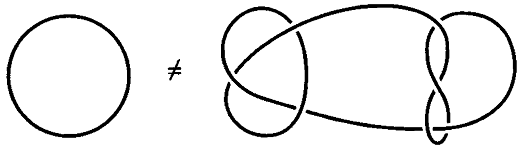


Figure 4.71 The trivial knot is not the composition of two nontrivial knots.

That was satisfying. It really makes you appreciate surfaces.

Exercise 4.27 Show that genus number 1 knots are prime.

Exercise 4.28 Prove that if we take the composition of n copies of the same nontrivial knot, calling the result J_n, then as n approaches ∞, the crossing number of J_n approaches ∞. (*Hint*: Use the fact that there are only a finite number of knots with a given crossing number [but first use the Dowker notation to prove this fact].)

Exercise 4.29 Let's see if we can improve on the result from the last exercise. Prove that if J_n is the composition of n copies of the same nontrivial knot, then the crossing number of J_n is at least n. (*Hint*: Use the Euler characteristic of a Seifert surface and Exercises 4.19 and 4.21.)

Let's talk a little more about this fact that Seifert's algorithm, when applied to an alternating projection, yields a minimal genus Seifert surface. We might wonder if this is true for any other types of knots. In fact, Louis Kauffman from the University of Illinois at Chicago extended the class of alternating links to the class of "alternative links" [see (Kauffman, 1983)]. He showed that the genus of any link in this larger class is also given by Seifert's algorithm. The class of alternative links includes all alternating links and all torus links, the topic of Section 5.1.

At the very least, we might hope that the minimal genus Seifert surface can be obtained by applying Seifert's algorithm to some projection of the knot. But surprisingly enough, an English mathematician named Hugh Morton proved (Morton, 1986, in references for Chapter 6) that there are in fact knots for which the minimal genus Seifert surface cannot be obtained by applying Seifert's algorithm to *any* projection of the knot.

Unsolved Question

Determine exactly which knots have a projection such that Seifert's algorithm applied to that projection yields a minimal genus Seifert surface. Perhaps Morton's examples are a very small subset of the set of all knots.

We mentioned in Section 2.3 that mutant knots are difficult to distinguish. In particular, the Kinoshita–Terasaka mutants of Figure 2.32 stumped knot theorists for awhile. Francis Bonahon (University of Southern California) and Lawrence Siebenmann (Institutes des Hautes Etudes Scientifiques) did find a way to tell them apart in 1981. Subsequently, David Gabai from Caltech managed to show that these two mutants do not have the same genus and hence must be distinct. Genus was enough to capture the essence of their difference.

5

Types of Knots

5.1 Torus Knots

We have already looked at several particular types of knots. For instance, we worked with alternating knots in Section 1.1 and rational knots in Section 2.3. In this section, we look at torus knots, that is, knots that lie on an unknotted torus, without crossing over or under themselves as they lie on the torus. Figure 5.1 is a picture of the trefoil knot on a torus.

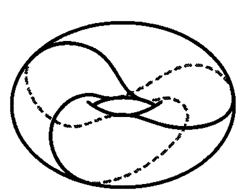


Figure 5.1 A trefoil knot on a torus.

We call a curve that runs once the short way around the torus a **meridian curve**. A curve that runs once around the torus the long way is called a **longitude curve** (Figure 5.2). The trefoil knot in Figure 5.1 wraps three times meridionally around the torus and twice longitudinally. We can see that these wrapping numbers are correct by adding the meridian and longitude curves to the torus on which the trefoil sits, and then counting the number of times the trefoil crosses each. The trefoil crosses the longitude three times. In order to do so, it must wrap around the torus in the meridional direction three times. The trefoil crosses the meridian twice, so it must wrap around the torus in the longitudinal direction two times. We call the trefoil knot a (3, 2)-torus knot (Figure 5.3). Figure 5.4 is a (4, 3)-torus knot. Every torus knot is a (p, q)-torus knot for some pair of integers. In fact, the two integers will always be relatively prime (that is, their greatest common divisor is 1).

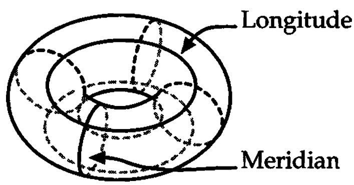


Figure 5.2 A meridian and longitude on a torus.

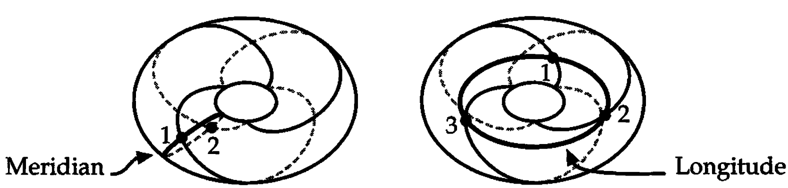


Figure 5.3 The trefoil is a (3, 2)-torus knot.

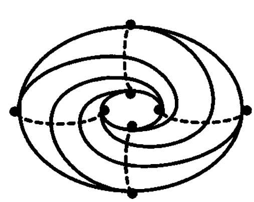


Figure 5.4 A (4, 3)-torus knot.

This next knot (Figure 5.5) lies on a torus, but it doesn't look like the (p, q)-torus knots we have drawn. However, we can deform it until it looks more like the torus knot that it is.

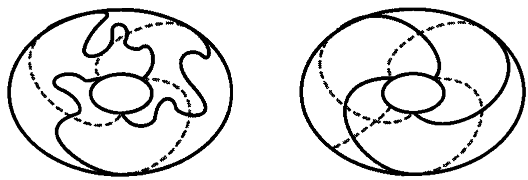


Figure 5.5 Two pictures of the same torus knot.

Exercise 5.1 What torus knot is shown in Figure 5.6?

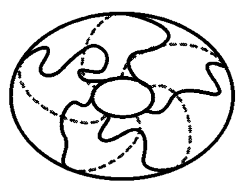


Figure 5.6 Mystery torus knot.

How do we go about drawing a (p, q)-torus knot? Well, suppose we want to draw a (5, 3)-torus knot. It wraps five times meridionally around the torus so it should cross the longitude five times. We mark five points on the outside equator of the torus and five points on the inside equator (Figure 5.7). We also want the knot to wrap three times longitudinally around the torus. We attach each point that we marked on the outside equator of the torus to the corresponding point on the inside equator, utilizing a strand that runs directly across the bottom of the torus (Figure 5.8). Now, we attach each point on the outside equator to the point on the inside equator that is a 3/5 turn clockwise from the outside point (meaning we jump ahead three points), utilizing a strand that runs over the top of the torus (Figure 5.9). The result is a knot that travels three times longitudinally around the torus and five times meridionally.

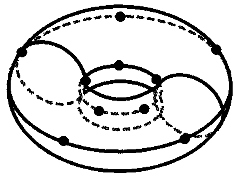


Figure 5.7 Marking points on the equators.

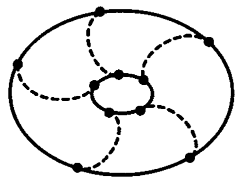


Figure 5.8 Attach points by strands across the bottom of the torus.

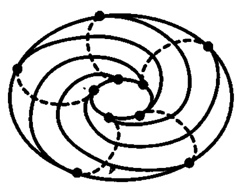


Figure 5.9 Constructing a (5, 3)-torus knot.

Similarly, if we want to draw a (p, q)-torus knot, we just place p points around the inside and outside equators of the torus, attach the inside and outside points directly across the bottom of the torus, and then attach each outside point to the inside point that is clockwise q points ahead, using a strand that goes over the top of the torus.

Exercise 5.2 Draw a (4, 3)-torus knot.

Exercise 5.3 What would happen if we tried to draw a torus knot where p and q were not relatively prime? Say a (3, 6)-torus knot?

Exercise 5.4 Show that a (p, q)-torus knot always has a projection with $p(q - 1)$ crossings.

In fact, *every (p, q)-torus knot is also a (q, p)-torus knot.* Say for instance that we have the trefoil knot, which we have seen is a (3, 2)-torus knot.

Remove a disk from the torus that the knot is sitting on, where that disk does not touch the knot. As we saw in Figure 4.34, we can deform a torus with one boundary component into two bands that are attached to one another. As the deformation occurs, we carry along the knot. The shorter band corresponds to a meridian of the torus, while the longer band corresponds to a longitude of the torus.

Take the longer band and turn it inside out. Then take the shorter band and turn it inside out also. We can now deform our two attached bands back out into a torus with one boundary component, but now with the roles of the two bands reversed. The band that originally corresponded to a longitude on the old torus now corresponds to a meridian on the new torus, and the band that originally corresponded to a meridian on the old torus now corresponds to a longitude on the new torus. Since the meridian and the longitude have been exchanged, the knot is now a (2, 3)-torus knot on the new torus (Figure 5.10).

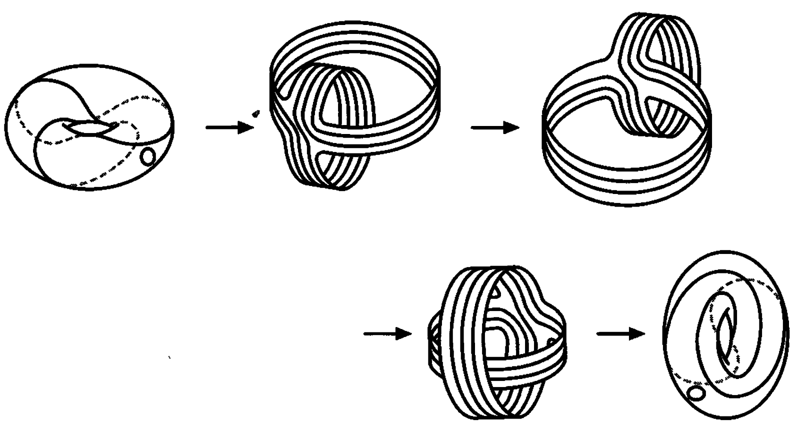


Figure 5.10 A (3, 2)-torus knot is a (2, 3)-torus knot.

This process works just as well to show that any (p, q)-torus knot is also a (q, p)-torus knot. In conjunction with Exercise 5.4, this implies that a (p, q)-torus knot has a projection with $p(q - 1)$ crossings *and* a projection with $q(p - 1)$ crossings. Therefore, the crossing number for a (p, q)-torus knot is at most the smaller of $p(q - 1)$ and $q(p - 1)$. It has recently been proved by Kunio Murasugi of the University of Toronto that in fact the smaller of $p(q - 1)$ and $q(p - 1)$ *is exactly* the crossing number of a (p, q)-torus knot. [See (Murasugi, 1991).]

A **solid torus** is a doughnut where we include both the interior of the doughnut as well as the surface. The **core curve** of a solid torus is the trivial knot that runs once around the center of the doughnut. A **meridional disk** of the solid torus is a disk in the solid torus that has a meridian curve as its boundary (Figure 5.11).

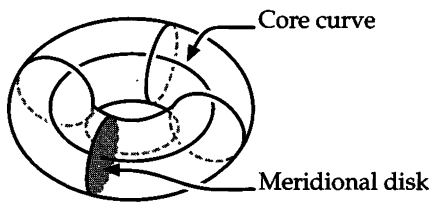


Figure 5.11 The core curve and a meridional disk in a solid torus.

Exercise 5.5 Let K be a (p, q)-torus knot sitting on a torus that is the boundary of a solid torus. Let J be the core curve at the center of the solid torus. Determine the linking number of J and K. What if K sits on the torus as a (q, p)-torus knot?

Figure 5.12 shows the 8_{19} knot.

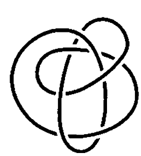


Figure 5.12 The 8_{19} knot.

Exercise 5.6 Show that the 8_{19} knot is the (3, 4)-torus knot (using either a series of pictures or by making it out of string).

This last example demonstrates that it is not at all obvious when a given knot is a torus knot.

Unsolved Question 1

Find an algorithm that will determine whether or not a given projection is the projection of a torus knot. This is hard since it assumes that you can tell whether or not a projection is a projection of the unknot, a difficult problem that was solved by Wolfgang Haken, but that has yet to be put in the form of an algorithm that can be implemented on a computer, as we mentioned in Section 1.1.

Unsolved Question 2

Show that $c(J\#K) = c(J) + c(K)$ holds when J and K are torus knots.

Unsolved Question 3

How about if J is a torus knot and K is an alternating knot? See Section 6.2.

Unsolved Question 4

Is the unknotting number of a (p, q)-torus knot equal to $(p - 1)(q - 1)/2$? [*Hold the presses*: Peter Kronheimer (at Oxford) and Tom Mrowka (at Caltech) recently announced a positive solution to this question!]

Exercise 5.7 Show that in the specific case that $(p, q) = (3, 4)$, the (p, q)-torus knot can be unknotted with $(p - 1)(q - 1)/2$ crossing changes.

What about determining the genus of a minimal Seifert surface spanning a given torus knot? In fact, torus knots are like alternating knots in that Seifert's algorithm applied to a projection as earlier will yield a minimal genus Seifert surface. Both torus knots and alternating knots fall in the category of alternative knots that we mentioned in Section 4.3.

Exercise 5.8 (a) Applying the result from the previous paragraph, use a standard projection as we described before to determine the genus of a (p, q)-torus knot.

(b) Does it matter if the projection comes from the knot represented as a (q, p)-torus knot rather than a (p, q)-torus knot?

We can generalize the notion of a torus knot. By definition, a torus knot is a nontrivial knot that can be placed on the surface of a standardly embedded torus without crossing over or under itself on the surface. By *standardly embedded*, we mean that the torus is unknotted in space. But certainly, there will be knots that cannot be placed on a standardly embedded torus but that can be placed on a standardly embedded genus two surface.

For lack of a better name, let's call these **two-embeddable knots**, since they can be embedded (placed without any crossings) on a standardly embedded genus two surface. For instance, the figure-eight knot is a two-embeddable knot (Figure 5.13). More generally, we will say that a knot K is an **n-embeddable knot** if K can be placed on a genus n standardly embedded surface without crossings, but K cannot be placed on any standardly embedded surface of lower genus without crossings.

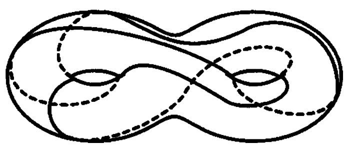


Figure 5.13 The figure-eight knot is a two-embeddable knot.

Exercise 5.9 Determine the minimal genus standardly embedded surface that the 5_2 knot can be embedded on, given that it is not a torus knot.

Exercise 5.10 Show that a knot with bridge number b is an n-embeddable knot where $n \leq b$.

Exercise 5.11 Show that any knot is an n-embeddable knot for some n. (*Hint*: Take a projection for the knot and have the strands at a crossing run over and under a handle of the surface. The surface that you use need only be isotopic to a standardly embedded surface.)

This last exercise shows that we can have a hierarchy of knots, depending on the minimal genus of a standardly embedded surface that they lie on. This is one measure of the complexity of a knot. This particular measure of complexity does not get mentioned much in knot theory. Instead, knot theorists use an invariant called *tunnel number*, which is closely related to this invariant. We will not have time to discuss tunnel number in this book.

Exercise 5.12 Suppose K is an n-embeddable knot and suppose that K can be embedded on a genus n surface such that the surface is cut into two pieces by the knot. Show that the genus of K (that is to say, the minimal genus of a Seifert surface for K) is at most $n - 1$.

5.2 Satellite Knots

A second set of knots that has become very important in recent years is the set of satellite knots. Let K_1 be a knot inside an unknotted solid torus (Figure 5.14). We knot that solid torus in the shape of a second knot K_2 (Figure 5.15). This will take the knot K_1 that lies inside the original solid torus to a new knot inside the knotted solid torus. We call this new knot, K_3, a **satellite knot**. The knot K_2 is called the **companion knot** of the satellite knot. We always assume that the companion knot is a nontrivial knot, since otherwise the resulting satellite knot would just be K_1 back again. We also always assume that the knot K_1 hits every meridional disk of the solid torus, and it cannot be isotoped to miss any of them. We think of the satellite knot as a knot that stays within a solid torus that has the companion knot as its core curve, just as a satellite stays within orbit around a planet.

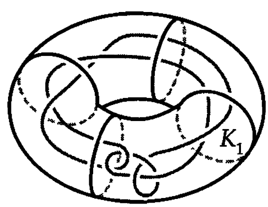


Figure 5.14 A knot K_1 inside a solid torus.

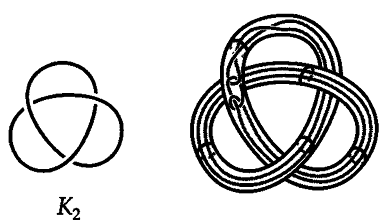


Figure 5.15 Knot the solid torus like K_2.

Notice that there is a knotted torus in space that misses the satellite knot, lying in the complement of the knot. In fact, this knotted torus is always incompressible, but proving this would take a substantial amount of work. If, on the other hand, we take the original knot K_1 to be an unknot, but sitting inside the solid torus twisted up as in Figure 5.16, then the resulting satellite knot is called a **Whitehead double** of the companion knot. The name refers to the fact that the knot K_1 here resembles the Whitehead link.

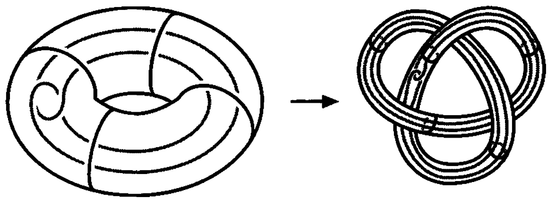


Figure 5.16 A Whitehead double of the trefoil.

A Whitehead double is not unique. We can cut the solid torus open along a meridional disk, twist one end some number of times, and then glue the meridional disks back together again, to obtain a homeomorphic copy of the solid torus (Figure 5.17). But now two strands of K_1 are twisted around each other. Then, when we knot this solid torus as a trefoil knot, we obtain a second Whitehead double of the trefoil. Both of these Whitehead doubles have the property that if we cut open three-space along the knotted torus, we get two pieces, one of which is the solid torus with K_1 in it, and one of which is three-space with the interior of a solid torus knotted as a trefoil knot removed from it.

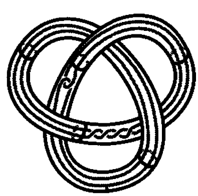


Figure 5.17 A second Whitehead double of the trefoil.

If the original knot K_1 is again unknotted, but sitting inside the solid torus as in Figure 5.18, then the resulting satellite knot is called a **two-strand cable** of the companion knot. It's as if we had a cable that ran twice around the companion knot. Again, the two-strand cable will not be unique, as we can add twists to it.

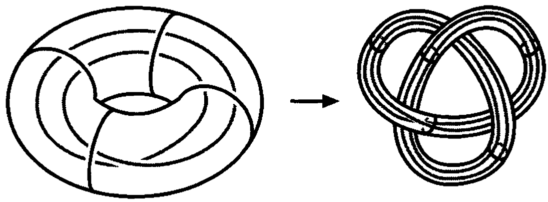


Figure 5.18 The two-strand cable of a knot.

Exercise 5.13 Draw a satellite knot corresponding to K_1 and K_2 from Figure 5.19. Now draw a second one. (You needn't prove that the second one is distinct.)

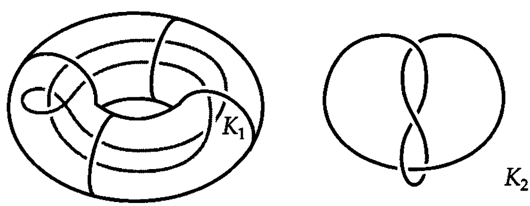


Figure 5.19 Draw a satellite knot corresponding to K_1 and K_2.

The operation of forming a satellite knot can be thought of as a generalization of the idea of composition. If K_1 only has one strand that reaches longitudinally around the solid torus as in Figure 5.20, then the satellite knot formed by knotting the solid torus like K_2 is in fact the composite knot $K_1 \# K_2$. (Notice the swallow-follow torus that we mentioned in the previous chapter.)

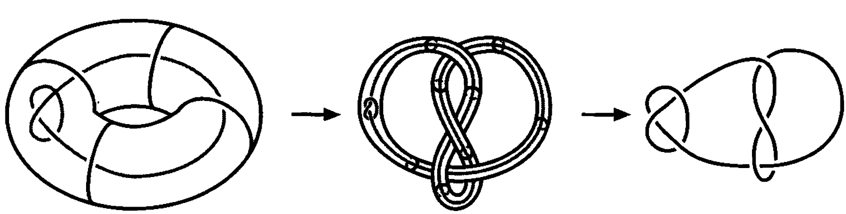


Figure 5.20 Satellite knots are sometimes just composites.

We could form a satellite knot so that the companion itself is also a satellite knot. Just as every composite knot factors into a unique set of prime factor knots, we can ask if it's true that every satellite knot is obtained from a unique sequence of taking satellites. In fact, the answer is *yes*, but it was only proved in 1987 [see (Soma, 1987)].

Exercise 5.14 Determine how this satellite knot was made (Figure 5.21). In particular, identify its companion and draw it inside the solid torus before the solid torus is knotted.

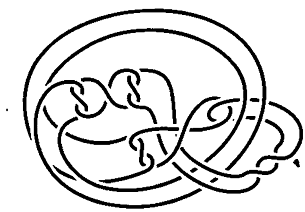


Figure 5.21 How was this satellite knot constructed?

Unsolved Question

It is amazing that the answer to this question has not yet been found. Show that the crossing number of a satellite knot is greater than the crossing number of the companion that it was constructed from. This certainly seems like it ought to be true but no one has been able to prove it.

If the knot K_1 that we start with is a torus knot, then we call the resulting satellite knot with companion K_2 a **cable knot** on K_2 (Figure 5.22). We can think of it as taking a cable that wraps around the knot K_2 a total of p times meridionally and q times longitudinally. In one field of mathematics called *algebraic geometry*, the most prevalent types of knots are cable knots. Sometimes the cable knots are cables on cables on cables on torus knots.

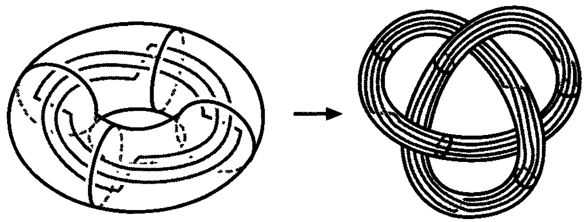


Figure 5.22 A cable knot.

5.3 Hyperbolic Knots

It was not until 1974 that anyone realized there was such a thing as a hyperbolic knot. Now, it appears that among prime knots, the overwhelming majority are hyperbolic knots. (*Note*: What do we mean when we talk about the "vast majority" of an infinite set of prime knots? We mean that for all of the prime knots of n or fewer crossings, a certain large percentage of them are hyperbolic. As n approaches ∞, that percentage is expected to approach 100%. However, this has not been proved. It is an interesting open conjecture.) In fact, William Thurston proved in 1978 that the only knots that are not hyperbolic knots are torus knots and satellite knots. (Here we are including all composite knots as satellite knots, contrary to a few authors.) So in the three sections in this chapter so far, we have defined three categories of knots, such that *every* knot falls into exactly one of the three categories (Figure 5.23).

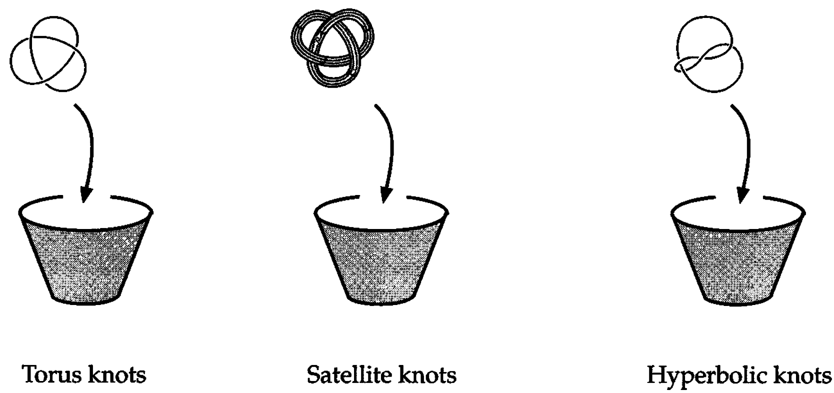


Figure 5.23 Every knot ends up in one of these three baskets.

Before 1974, no one realized that a knot complement could be hyperbolic. At that time, Robert Riley, who was an American working on his Ph.D. at Southampton in England, showed that the figure-eight knot was hyperbolic. After much effort, he also showed that two other knots were hyperbolic. He then proceeded to write an immense computer program that was designed to attempt to show that there were additional hyperbolic knots. William Thurston, who was then a professor at Princeton, had been thinking about related ideas. In the summer of 1976, he went to visit England and he met Robert Riley in the coffee room at the University of Warwick. After discussions with Riley about his work, Thurston's ideas crystallized. He realized that in fact, almost all knots are hyperbolic. The field of hyperbolic three-manifolds came into existence and has since become an essential part of topology. Thurston received the Fields medal (the mathematics version of a Nobel prize) in 1982.

But what is a hyperbolic knot? First, let's look at the official definition.

A **hyperbolic knot** is a knot that has a complement that can be given a metric of constant curvature -1.

Now, that's a mouthful. What does it mean? A metric is simply a way to measure distance. Thus, we can measure distance in the knot complement, that is, in three-space minus the knot. Given two points in three-space minus the knot, we can determine the distance between them. Usually, we measure the distance between two points $P_0(x_0, y_0, z_0)$ and $P_1(x_1, y_1, z_1)$ in three-space using the formula

$$d(P_0, P_1) = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}$$

This method for measuring distance is called the **Euclidean metric.**

But now, in the complement of this knot, we will be measuring distance in a different way, using a distance measure that has curvature -1. What do we mean by that? Here is a two-dimensional analog. A sphere has positive curvature (Figure 5.24a). If we pick a point on the surface of the sphere and take cross sections in several directions through that point, all of the cross sections are circles that curve in the same direction. A plane, however, has zero curvature (Figure 5.24b). If we pick a point and take cross sections in several directions through the point, we always get a line, which has no curvature. But a saddle has negative curvature (Figure 5.24c). If we take the central point, and take cross sections in two different directions through that point, we obtain two parabolas, one of which opens up and one of which opens down. That is the essence of negative curvature.

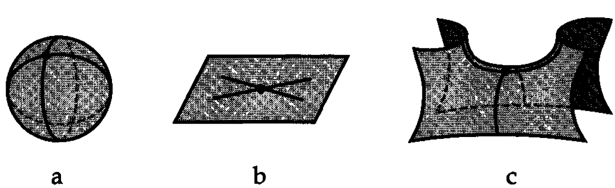


Figure 5.24 (a) Positive curvature. (b) Zero curvature. (c) Negative curvature.

We are interested in three-dimensional space (since the complement of a knot is three-dimensional), so we can't draw the pictures like we could of the sphere, plane, and saddle. But the Euclidean metric for three-space that we gave before is an example of a metric with curvature zero. It is a so-called **flat metric**, having no curvature, just like the plane is flat. The metric that we want to put on the complement of the knot is not flat, but rather has curvature -1. The geometry that results is called *hyperbolic geometry*, and the metric is called a **hyperbolic metric**.

We describe the simplest example of a three-dimensional space that has a hyperbolic metric. It is called **hyperbolic three-space**, and is denoted by H^3. The particular model of H^3 that we will look at is called the *Poincaré model*, after the French mathematician Henri Poincaré (1854–1912). The points in this model are the points in three-space inside the unit ball. So

$$H^3 = \{(x, y, z) : x^2 + y^2 + z^2 < 1\}$$

Now we need to describe how to measure the distance between two points in H^3. Let P_1 and P_2 be two such points. First, we describe a particular path through H^3 from P_1 to P_2 (Figure 5.25). Let C be part of a circle in H^3 that has both of its endpoints on the unit sphere, is perpendicular to the unit sphere at its endpoints, and passes through the two points P_1 to P_2. Assuming P_1 and P_2 do not both lie on a line segment that is a diameter of the unit sphere, there is always a unique such arc of a circle. If P_1 and P_2 do lie on the same diameter, we will replace the arc of a circle with that line segment that is a diameter passing through P_1 and P_2.

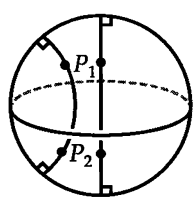


Figure 5.25 Paths between points in H^3.

It will turn out that the shortest path in hyperbolic three-space from P_1 to P_2 is the path within the arc of a circle or vertical line from P_1 to P_2. Let's call this path w. Any arc of a circle or diameter in H^3 that is perpendicular to the unit sphere is called a **geodesic** in H^3. A geodesic is a curve that has the property that for any two points P_1 and P_2 within it, the shortest path from P_1 to P_2 also lies in the curve. Geodesics in H^3 play the role that straight lines play in Euclidean space. Notice that straight lines are geodesics in Euclidean space, as the shortest path between any two points in a line also lies in the line.

In order to measure the distance between P_1 and P_2, we integrate the function $2/(1 - r^2)$ along the shortest path from P_1 to P_2, where r is the distance to the origin. Therefore, the official definition of the distance from P_1 to P_2 is

$$d(P_1, P_2) = \int_w \frac{2\,ds}{1 - r^2}.$$

If you are unfamiliar with a path integral of this type, it's just one more reason for wanting to learn calculus. If you are familiar with a path integral of this type, try the following exercise.

Exercise 5.15 Find the hyperbolic distance between the two points (0, 0, 0) and (0, 0, $\frac{1}{2}$) in the Poincare model. (Note that as we are integrating along a horizontal line, the path differential ds just becomes dx and r^2 becomes x^2.) What about the distance between (0, 0, 0) and (0, 0, a) as a approaches the value 1?

Hyperbolic space has lots of interesting properties. For example, note that if we form a triangle in hyperbolic space such that each of its three edges comes from segments of geodesics, the sum of the angles of the triangle is less than the sum of the angles of the corresponding Euclidean triangle with the same vertices. Since the sum of the angles of the Euclidean triangle is exactly 180°, this means that the sum of the angles of the hyperbolic triangle is strictly less than 180° (Figure 5.26). This amazing fact will always be the case. *The angles of any triangle in hyperbolic three-space add up to* less *than 180°*. Strange, but true.

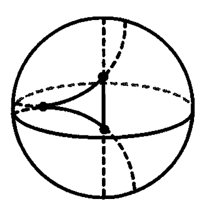


Figure 5.26 The angles of a hyperbolic triangle add up to less than 180°.

There are numerous other interesting properties of hyperbolic three-space that we will not get into here, but what we will do is describe how we can use pieces of hyperbolic three-space in order to obtain so-called hyperbolic manifolds. The pieces that we use are tetrahedra (Figure

5.27). The edges of the tetrahedra are geodesics in hyperbolic space and the faces are geodesic planes in hyperbolic space. It's no big surprise that these geodesic planes are pieces of spheres that are perpendicular to the unit sphere bounding H^3 or disks contained in planes that pass through the origin in H^3.

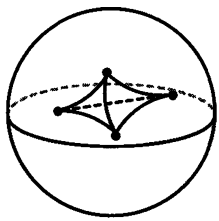


Figure 5.27 Geodesic planes and tetrahedra in H^3.

Remember how we constructed surfaces in Section 4.1 by gluing together pairs of edges in a set of triangles until every edge had been glued to some other edge. Analogously, it is possible to glue together pairs of faces in a set of tetrahedra until every face has been glued to some other face. When done correctly, the result can sometimes be a knot complement. If the tetrahedra that we glue together are actually hyperbolic tetrahedra, in that they sit inside hyperbolic space, and if we glue them together along their faces so that their faces match without distortion, in order that the hyperbolic method of measuring distances within the individual tetrahedra match, the result is a hyperbolic knot complement. We can use the hyperbolic method for measuring distance within the individual tetrahedra in order to obtain a hyperbolic method for measuring distance in the entire knot complement. We then say that the knot is a **hyperbolic knot**.

Every hyperbolic knot has a **hyperbolic volume**. This is a positive real number that can be computed out to as many decimal places as are needed. It is simply the sum of the volumes of the individual hyperbolic tetrahedra that make up the knot complement, and it gives the volume of the complement of the knot, as measured by our hyperbolic metric. Although it appears that the volume of three-space minus the knot would be infinite, it is in fact finite when we measure it using this hyperbolic method of measuring volume. The hyperbolic volume is an invariant for the hyperbolic knots, as it depends only on the knot itself and not on any particular projection of the knot. Figure 5.28 shows the volumes of some knots.

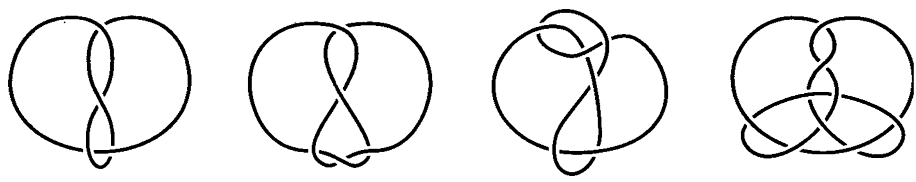


Figure 5.28 Volumes of hyperbolic knots.

In the tables of knots and links at the end of the book, we have listed the volumes of all of the hyperbolic knots and links next to the pictures of the knots and links. Notice how few knots and links are not hyperbolic. (We have listed the hyperbolic volume of a nonhyperbolic knot as 0.) In fact, in the tables of prime knots, all but six of the nontrivial knots of 10 or fewer crossings are hyperbolic.

Two hyperbolic knots with distinct volumes must be distinct knots. At least in our table through 10 crossing knots, volume turns out to be an excellent way to tell knots apart, distinguishing all but the 6 nonhyperbolic knots (all of which happen to be torus knots). In fact, there *are* distinct knots that have the same volume. For example, Figure 5.29 depicts the 5_2 knot and a 12-crossing knot, both of which have the same volume.

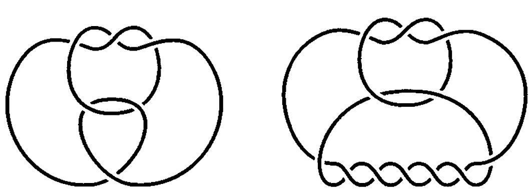


Figure 5.29 Two knots with the same volume.

More generally, if we flip a tangle in a hyperbolic knot to produce a mutant knot, the mutant will also be a hyperbolic knot *and* it will have the same volume. So we can't tell the knots in Figure 5.30 apart by volume. But in some sense, these examples of knots with the same volume are exceptional. Almost all knots can be distinguished by their volume.

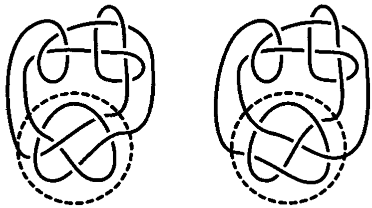


Figure 5.30 Mutants have the same volume.

Unsolved Question 1

What is the smallest volume of any knot? (That there is a smallest volume is a difficult fact to prove.) It is conjectured that the figure-eight knot has the smallest volume among all hyperbolic knots, namely 2.0298. . . . Note that neither the trivial knot nor the trefoil knot are hyperbolic knots, so the figure-eight knot is the simplest knot that we know of that is hyperbolic.

This conjecture was made by William Thurston in 1978, and has remained open since. I have worked on this problem on and off for the last ten years, but I haven't been able to solve it.

Unsolved Question 2

Is any one of the volumes a rational number a/b, where a and b are integers? Is any one of the volumes an irrational number (not of the form a/b where a and b are integers)? Amazingly enough, even though we can calculate the volume of a knot out to as many decimal places as we want, we cannot tell whether any one of the volumes is either rational or irrational.

How do we actually compute the volume of a knot? We first cut the complement of the knot into a finite set of tetrahedra. We then place the tetrahedra in hyperbolic space. In order that these hyperbolic tetrahedra glue together correctly to give the hyperbolic metric, a set of equations must be satisfied. If there are n tetrahedra, we obtain n polynomial equations in n variables. The variables are in fact complex variables. We then use numerical methods to solve this system of equations numerically. The solution to the system determines the hyperbolic metric on each tetrahedron. We can then compute the hyperbolic volume of each of the tetrahedra and add the volumes to get the volume of the knot complement (Figure 5.31).

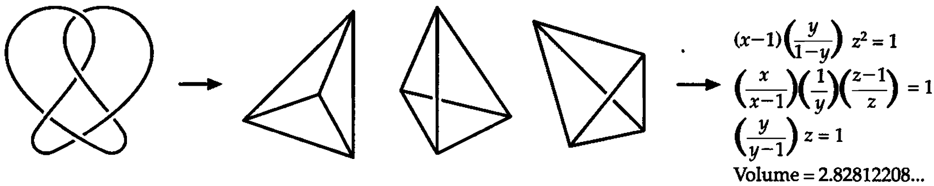


Figure 5.31 Finding the hyperbolic volume of a knot.

In fact, it is difficult and arduous to do this procedure by hand in all but the simplest cases of knots. Luckily, we don't have to do it. Jeff Weeks (a 1985 Ph.D. student under William Thurston) has written a computer program that does it for us. We simply input a knot by drawing it on the screen with a mouse. The mouse can then be used to click on any crossings that we want changed in order to obtain the knot that we want. The computer takes over from there, cutting the knot complement into tetrahedra, generating the set of equations that must be satisfied, finding the numerical solution of the set of equations, and then calculating the corresponding volume.

Given two knots that are hyperbolic and that have a reasonable number of crossings (say, no more than 100), Jeff Weeks's program is able to determine whether or not they are the same knot. This is the original question we discussed in Section 1.1, which can now be solved by computer, at least when we restrict ourselves to hyperbolic knots with few enough crossings.

Unsolved Question

Write a computer program that will determine whether or not any two knots with a reasonable number of crossings are the same. What's the idea here? Start with a knot. First, the computer program needs to recognize whether or not the knot is prime. It's enough for the computer to check whether the knot is satellite or not, since we have seen that composite knots are a special case of satellite knots.

If it is a satellite knot, cut the knot complement open along the knotted torus (which, remember, is incompressible) that always exists in a satellite knot complement. To the outside of the torus will be one knot complement (the complement of the companion knot) and to the inside will be a link complement. Continue to cut each of these two pieces open along incompressible tori until you have a set of knot and link complements, each of which is either a torus link complement or a hyperbolic link complement.

Given two knots, perform this decomposition of each of their complements into torus and hyperbolic link complements. Jeff Weeks's program can currently determine whether or not the hyperbolic link complements in the two decompositions are the same. A program needs to be written to determine whether or not the torus link complements in the two decompositions are the same. Then the program needs to check how the various pieces are glued back together to decide if the two knots are the same. The biggest open question is how to write a program that finds these incompressible tori.

This section has been a bit sketchier than some of the other sections, since the level of mathematical background necessary to dive deeper into the topic of hyperbolic knots is somewhat higher than for other topics that we have covered. However, the concept of hyperbolic knots is so interesting and has proved to be so valuable for knot theory, that it was worth discussing, even if we did not go to great depths.

5.4 Braids

This section is not entirely appropriate to this chapter, since braids are not a particular type of knot. However, every knot can be described by a braid. Since we can easily restrict ourselves to certain types of braids, which then correspond to certain types of knots, braids will generate types of knots. Besides, braids are so beautiful that we can't put them off any longer.

A **braid** is a set of n strings, all of which are attached to a horizontal bar at the top and at the bottom (Figure 5.32). Each string always heads downward as we move along any one of the strings from the top bar to the bottom bar. Another way to say the same thing is that each string intersects any horizontal plane between the two bars exactly once.

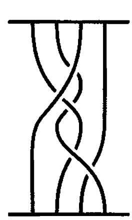


Figure 5.32 A braid.

We consider two braids to be equivalent (Figure 5.33) if we can rearrange the strings in the two braids to look the same without passing any strings through one another or themselves while keeping the bars fixed and keeping the strings attached to the bars. We are not allowed to pull the strings over the top of the upper bar or beneath the bottom bar. It is probably helpful to think of there being huge horizontal plywood sheets at the level of the top and bottom bars, in order to discourage us from trying to pull the strings over and around the bars.

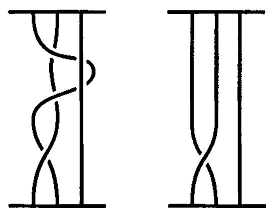


Figure 5.33 These two braids are equivalent.

What do braids have to do with knots and links? We can always pull the bottom bar around and glue it to the top bar, so that the resulting strings form a knot or link, called the **closure** of the braid (Figure 5.34). Therefore every braid corresponds to a particular knot or link. We can think of there being an axis coming right out of the page, around which the closure of the braid is wrapped. We then have a **closed braid representation** of the knot if there is a choice of orientation on the knot so that, as we traverse the knot in that direction, we always travel clockwise around the axis without any backtracking. In Figure 5.35, we see two projections of the trefoil with axes, the first of which is not a closed braid around its axis, and the second of which is a closed braid around its axis.

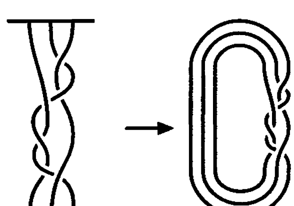


Figure 5.34 The closure of a braid.

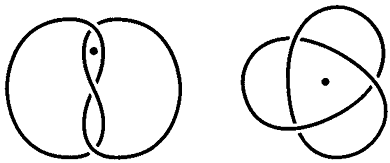


Figure 5.35 (a) *Not* a closed braid. (b) A closed braid.

Exercise 5.16 Draw closed braid representations of each of these knots (Figure 5.36). (*Hint*: You may want to make them out of string or an extension cord and then see if you can rearrange them to travel clockwise around a pencil. Another option is to make a choice of an axis and then start to have the knot travel around the axis clockwise. Whenever it starts to travel counterclockwise, pass that portion of the knot through the axis to fix the problem. It is recommended that you do not try to use the technique that follows this problem in order to do it.)

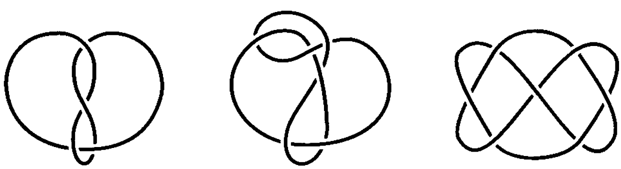


Figure 5.36 Draw these knots as closed braids.

What knots and links can be represented as closed braids? Amazingly enough, *they all can. Every knot or link is a closed braid.* This was first proved by J. W. Alexander in 1923 (before he started fooling around with polynomials; see the next chapter).

We will use the idea of bridges from Section 3.2 to prove this. Let L be our knot or link in a particular projection, and let's orient each of the components of L. For any strand of the knot between an overcrossing and an undercrossing, choose a point on the strand. As we traverse the knot or link in the direction of the orientation, label these chosen points P_1 through P_n (Figure 5.37), where the first point P_1 was a point that occurred after an undercrossing. We can think of these labeled points as being the intersection of the projection plane with the knot or link, and the strands above and below the projection plane as the bridges. The strand of the knot connecting P_1 to P_2 lies above the projection plane; the strand connecting P_2 to P_3 lies beneath the projection plane. In fact, the strand connecting P_{2i-1} to P_{2i} always lies above the projection plane, and the strand connecting P_{2i} to P_{2i+1} always lies below the projection plane.

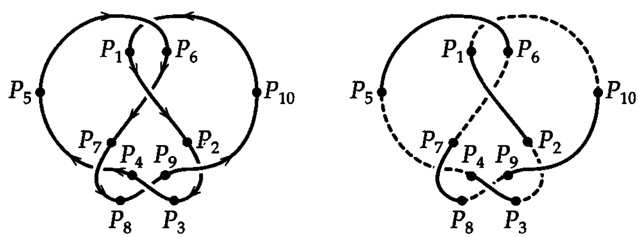


Figure 5.37 Label points $P_1, \ldots, P_n$.

Let's isotope (rearrange without cutting and pasting) the projection so that the n strands beneath the projection plane are lined up as in Figure 5.38. We arrange the strands so that the even-numbered points are all next to one another. We don't have any problem performing this rearrangement under the projection plane, since we are just sliding nonintersecting arcs around a bit. On the top side of the plane, the bridges are getting messier, but they never cross one another, so we still have perfectly good bridges. Let A be a straight line in the projection plane that is a perpendicular bisector of all of the lower bridges. Each of the upper bridges will then cross A an odd number of times, since an upper bridge starts at a point P_{2i-1} that is north of segment A and ends at the point P_{2i} that is south of segment A. We would like the upper bridges to cross A only once, so we will do a little bridge work to make it so.

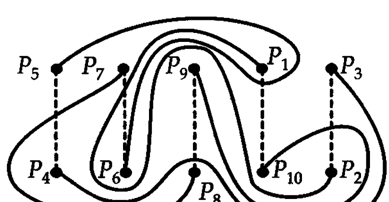


Figure 5.38 Rearranging the projection to line up the lower bridges.

If one of the upper bridges does cross A more than once, simply take the second point on the upper bridge where it crosses A after leaving P_{2i-1} and push the upper bridge a little below the projection plane at that point (Figure 5.39). The one upper bridge now splits into two new upper bridges and a new lower bridge. The first new upper bridge crosses A once, while the second new upper bridge crosses A two fewer times than the original upper bridge did. We can repeat this process with the new upper bridges and eventually with the other remaining upper bridges until every upper bridge crosses the line A exactly once.

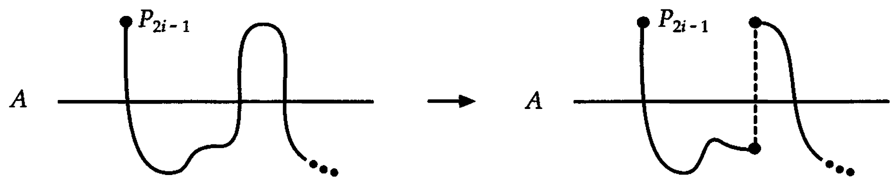


Figure 5.39 Making upper bridges cross A once.

We now draw our link as a closed braid with axis A (Figure 5.40). We begin each of the upper bridges at its starting point P_{2i-1} in the projection plane. As we draw the bridge, we have it increase in height above the projection plane until it is directly above the line A, at which time we are at the point on the bridge that used to be where it crossed A. After that point, we have the bridge decrease in height until it reaches the point P_{2i} back in the projection plane. Similarly, starting from the even-numbered points, we have the lower bridges decrease in height until they are directly beneath A, at which time we are at the point on them where they used to cross under A. We then have them increase in height until they reach the odd-numbered points back in the projection plane.

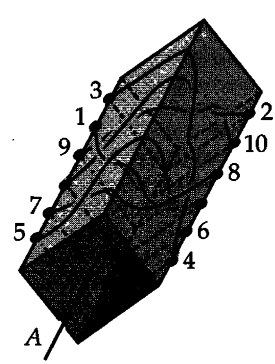


Figure 5.40 A link is a closed braid.

There we have it, a braid representation for any knot or link. But now, if every link can be represented as a closed braid, we are probably interested in representing the link as a simple braid with as few strings as possible.

Let's define the **braid index** of a link to be the least number of strings in a braid corresponding to a closed-braid representation of the link. For example, the braid index of the unknot is 1, and the braid index of the trefoil knot is 2.

Exercise 5.17 Describe all of the knots and links of braid index 2.

The braid index is an invariant for knots and links, but in general it is difficult to compute. Putting a knot or link in braid form and then counting the strings gives an upper bound on the braid index, but how do we know there isn't a braid form of the knot or link with fewer strings? In the next chapter, we see one way to get a lower bound on the braid index using knot polynomials.

Recently, Shuji Yamada, of Ehime University in Japan, related the braid index to the number of Seifert circles (as in Section 4.3). In particular, he proved the amazing fact that the braid index of a knot is equal to the least number of Seifert circles in any projection of the knot. [See (Yamada, 1987).] In an even more recent paper, Yoshiyuki Ohyama of Waseda University in Japan proved that if L is a nonsplittable link, $c(L)$ is its crossing number, and $b(L)$ is its braid index, then $c(L) \geq 2(b(L) - 1)$. Notice that for the figure-eight knot, which has braid index 3, this is actually an equality.

How should we describe a given braid? A projection of a braid can be described by listing which of the strings cross over and under each other as we move down the braid. We can arrange it so that no two crossings in the braid occur at exactly the same height.

For instance, let's look at three-string braids. If the first string crosses over the second, we call that crossing a σ_1 crossing. If the first string crosses under the second, we call it an σ_1^{-1} crossing. If the second string crosses over the third, we call that an σ_2 crossing, and if it crosses under the third string, it's an σ_2^{-1} crossing (Figure 5.41). Hence the braid shown in Figure 5.42 is described completely by listing the crossings in order from top to bottom as $\sigma_2 \sigma_1 \sigma_1 \sigma_2^{-1} \sigma_1 \sigma_1$. We call this a **word** for the braid. More generally, if we have a braid with n strings, we denote the i^{th} string crossing over the $i + 1^{\text{st}}$ string by σ_i and the i^{th} string crossing under the $i + 1^{\text{st}}$ string by σ_i^{-1}.

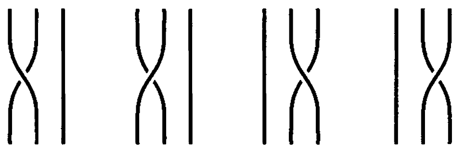


Figure 5.41 Crossings.

Figure 5.42 This braid is described by the word $\sigma_2\,\sigma_1\,\sigma_1\,\sigma_2^{-1}\,\sigma_1\,\sigma_1$.

Exercise 5.18 Draw the five-string braid given by the word $\sigma_1\,\sigma_2\,\sigma_3\,\sigma_2^{-1}\,\sigma_4\,\sigma_1^{-1}$.

Exercise 5.19 Identify the knot that has four-string braid word $(\sigma_1^{-1}\,\sigma_2\,\sigma_3^{-1})^2\sigma_3^{\,3}$. (It is one of the six-crossing knots in the appendix table.)

Exercise 5.20 Find a word such that the closure of the corresponding braid gives the knot 6_3.

Exercise 5.21 Show that the closure of the n-string braid $(\sigma_1\,\sigma_2 \cdots \sigma_{n-1})^m$ is a knot if and only if n and m are relatively prime.

Exercise 5.22 Check that the closure of the n-string braid $(\sigma_1\,\sigma_2 \cdots \sigma_{n-1})^m$ is simply the (m, n)-torus knot (assuming m and n are relatively prime).

This is a handy way to denote knots and links. Say we want to describe a particular knot to a friend over the phone. We just say, "It's the knot that has the closed braid representation corresponding to the word $\sigma_1\,\sigma_2\,\sigma_3\,\sigma_2^{-1}\,\sigma_4\,\sigma_1^{-1}$."

Besides simplifying phone conversations about knots and links, there are some other advantages to this notation. For instance, suppose a braid has $\sigma_i^{-1}\ \sigma_i$ as part of the word that describes it. Then geometrically, this pair of crossings looks like Figure 5.43.

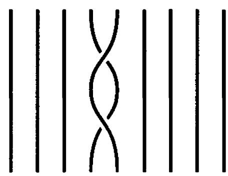


Figure 5.43 The crossings corresponding to $\sigma_i^{-1}\ \sigma_i$.

A simple Type II Reidemeister move eliminates both crossings, but leaves us with an equivalent braid. The effect on the word is to eliminate $\sigma_i^{-1}\ \sigma_i$. The same phenomenon occurs for $\sigma_i\ \sigma_i^{-1}$ also. For example the word $\sigma_1\ \sigma_2\ \sigma_3\ \sigma_2^{-1}\ \sigma_2\ \sigma_3^{-1}\ \sigma_2^{-1}\ \sigma_1^{-1}$ collapses down to nothing, meaning that the braid that it represents is equivalent to the trivial braid of three vertical strings that do not cross.

Notice that a closed braid has an orientation given by always choosing to orient the braid so that the direction on the strings runs from the top to the bottom.

Exercise 5.23 If a given word represents a knot with a particular orientation, what word gives the same knot but with the opposite orientation?

Let's return to open braids for a second. Notice that if we have two n-string braids, we can stack them on top of each other to create a new braid that is the product of the original two braids (Figure 5.44).

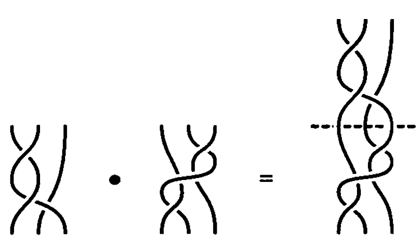


Figure 5.44 Multiplying two braids.

Exercise 5.24 (a) Describe an n-string braid I_n that acts as an identity when it multiplies any other n-string braid. That is to say, multiplying a braid B by this braid yields the braid B back again.

(b) Describe an inverse braid for a given braid B. That is to say, find a braid B^{-1} that has the property that $BB^{-1} = I_n$ and $B^{-1}B = I_n$.

(c) Show that the multiplication of n-string braids is associative, namely $B_1(B_2B_3) = (B_1B_2)B_3$ for any three n-string braids B_1,B_2, and B_3.

What we have constructed in this exercise is known in mathematical circles as a **group**. Namely, it is a set of elements, in this case the n-string braids, and a way to multiply elements such that:

1. There exists an identity element that, when multiplying any element, doesn't change it;
2. Every element has an inverse;
3. The multiplication is associative, that is, $a(bc) = (ab)c$ for any three elements a, b, c of the group.

Let B_n denote the group of n-string braids. It is a very interesting example of a group.

Exercise 5.25 If you haven't seen the theory of groups before, convince yourself that the integers with the "multiplication" given by addition form a group, the real numbers without zero and with the usual multiplication form a group, and the integers with the usual multiplication do not form a group.

Note that a particular element of B_n is a braid together with all the other braids that are equivalent to it. We say that an element is an equivalence class of braids, although we will sometimes refer to it as a single braid. An element of B_n has many different projections and many different words that represent it. We would like to know when two different words written out in the letters $\sigma_1^{\pm 1}, \ldots, \sigma_n^{\pm 1}$ represent the same braid. We have already seen one rule that gives equivalence between words, namely, we can add or delete $\sigma_i \sigma_i^{-1}$ or $\sigma_i^{-1} \sigma_i$ from a word. The next exercise gives us a second rule that we can apply.

Exercise 5.26 Using a picture, show that the two braids $\sigma_i \sigma_{i+1} \sigma_i$ and $\sigma_{i+1} \sigma_i \sigma_{i+1}$ are equivalent.

The first rule is a certain kind of Type II Reidemeister move that we can apply to the braid projection. This second rule is a kind of Type III Reidemeister move that we can apply to the braid projection. There is a third rule that we can apply, but it does not come from the Reidemeister moves. If a braid contains $\sigma_i \sigma_j$, where $|i - j| > 1$, then we can switch the order of σ_i and σ_j, replacing $\sigma_i\sigma_j$ in our word with $\sigma_j\sigma_i$. See Figure 5.45.

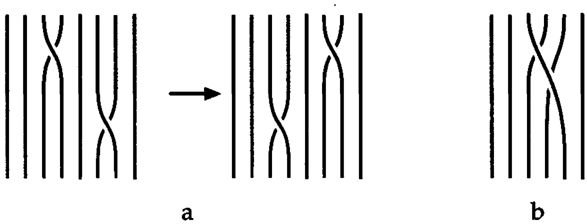


Figure 5.45 (a) Switching σ_i and σ_j when $|i - j| > 1$. (b) No such switch when $|i - j| = 1$.

Two words w_1 and w_2 represent the same braid if and only if we can get from the one word to the other by a sequence of these three operations. For instance, $w_1 = \sigma_1\sigma_2\sigma_4^{-1}\sigma_1\sigma_2\sigma_4$ and $w_2 = \sigma_2\sigma_1\sigma_2^2$ represent the same five-string braid since we can get from the first word to the second word by the following set of applications of the three rules.

$$w_1 = \sigma_1\sigma_2\sigma_4^{-1}\sigma_1\sigma_2\sigma_4 \xrightarrow{\text{Rule 3}} \sigma_1\sigma_2\sigma_4^{-1}\sigma_1\sigma_4\sigma_2 \xrightarrow{\text{Rule 3}} \sigma_1\sigma_2\sigma_4^{-1}\sigma_4\sigma_1\sigma_2$$

$$\xrightarrow{\text{Rule 1}} \sigma_1\sigma_2\sigma_1\sigma_2 \xrightarrow{\text{Rule 2}} \sigma_2\sigma_1\sigma_2\sigma_2 = w_2$$

Exercise 5.27 Show that the two words $w_1 = \sigma_2\sigma_1\sigma_2\sigma_1\sigma_2^{-1}\sigma_4$ and $w_2 = \sigma_2^2\sigma_3\sigma_1\sigma_3^{-1}\sigma_4$ represent the same five-string braid, up to equivalence.

Exercise 5.28 Find a sequence of these three equivalence moves that gives the Type III Reidemeister move depicted in Figure 5.46.

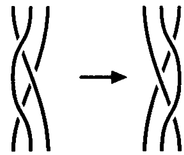


Figure 5.46 A Reidemeister move on a three-string braid.

In fact, we are interested in knowing even more than which words represent the same braid. To make braids useful for knot and link theory, we would like to be able to determine when the closures of two braids represent the same oriented link. Let's say that two braids are **Markov equivalent** if their closures yield the same oriented link. Just as we saw that the three Reidemeister moves gave all equivalences between different projections of the same link, we would like to have a set of moves on braids that give all equivalences on the corresponding closed braids.

Remarkably enough, in a 1935 paper, A. A. Markov outlined a proof of the theorem now known as **Markov's theorem**. It says that two braids are Markov equivalent if and only if they are related through a sequence of the three operations that we have already seen, which are operations that obviously give us back the same open braid, and two additional operations. The first new operation is called **conjugation**. On the word for a braid, conjugation is the operation of multiplying the word at the beginning by σ_j and at the end by σ_j^{-1}. Or we could multiply at the beginning by σ_j^{-1} and at the end by σ_j. Geometrically, this has the effect shown in Figure 5.47.

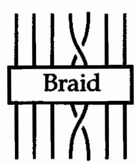


Figure 5.47 Conjugation by σ_j.

It's pretty easy to see that conjugation does not change the oriented link corresponding to the closed braid. (In fact, it corresponds to a Type II Reidemeister move on the link projection.) Notice that the need for at least one more operation is clear, since none of the operations so far change the number of strings in the corresponding braid. However, it's easy to come up with two closed-braid representations of the same link that do not have the same number of strings. The next operation, called **stabilization**, remedies this problem. Here, we add or delete a loop in the closed braid. In terms of the word describing a braid, this operation takes a word w corresponding to an n-string braid and replaces it with the word $w\sigma_n$ or $w\sigma_n^{-1}$, each of which corresponds to an n + 1-string braid. We also allow the inverse operation, where a word of the form $w\sigma_n$ or $w\sigma_n^{-1}$ is replaced with just the word w, assuming w does not contain the letters σ_n or σ_n^{-1}

within it (Figure 5.48). The resulting word w then corresponds to a braid with one less string.

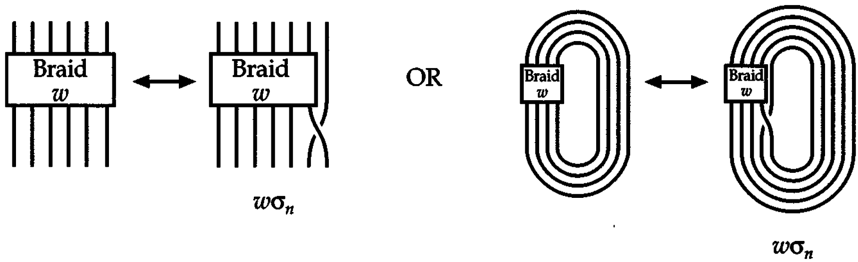


Figure 5.48 Adding or deleting loops.

As is apparent from the figure, the oriented link corresponding to the closed braid remains unchanged by either of these operations. (Note that the operations correspond to a Type I Reidemeister move on the link projection.) What is surprising is that the two operations of conjugation and stabilization, together with the three operations mentioned previously, suffice to get us from any one closed-braid representation of an oriented link to any other closed-braid representation of the same oriented link. We will not go through the proof because it is quite difficult. [A proof appears in (Birman, 1976).] However, let's try our hand at a particular example.

In Figure 5.49, we see two closed-braid representatives of the figure-eight knot. Hence, there must be a sequence of these Markov moves (together with the three equivalence moves for a given open braid) that take us from the first braid to the second braid. A possible sequence appears in Figure 5.50.

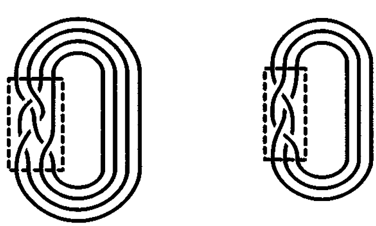


Figure 5.49 Two closed braids giving the figure-eight knot.

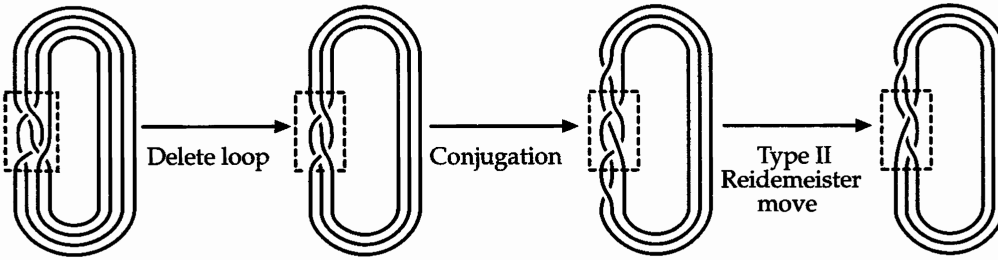


Figure 5.50 A sequence of Markov moves to get from one braid to the other.

Exercise 5.29 Find a sequence of Markov moves that demonstrate the equivalence of the two closed braids in Figure 5.51.

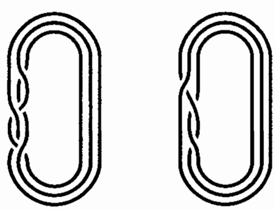


Figure 5.51 Find the Markov moves from one braid to the other.

More on braids later.

5.5 Almost Alternating Knots

The idea of almost alternating knots is a new one, which at the time of this writing (1993) is only two years old. Only time and research will determine if it is a good idea. We call a projection of a link an **almost alternating projection** if one crossing change in the projection would make it an alternating projection. We call a link an **almost alternating link** if it has an almost alternating projection, and if it does not have an alternating projection. For instance, the 8_{19} knot in Figure 5.52 is an almost alternating knot since it is known to be nonalternating (a not at all obvious fact), but it has the almost alternating projection pictured.

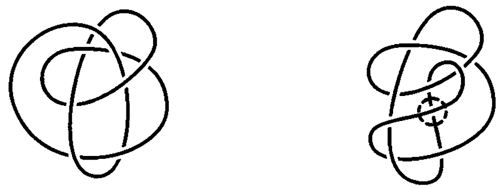


Figure 5.52 The 8_{19} knot is almost alternating.

Exercise 5.30 Show that the knots 8_{20} and 8_{21}, shown in Figure 5.53, are almost alternating. (It is known that they are not alternating knots, so you may assume that.)

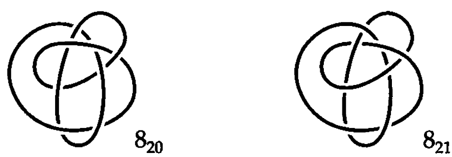


Figure 5.53 Show that these knots are almost alternating.

Exercise 5.31 Show that all of the nonalternating nine-crossing knots in Table 1.1 at the end of the book are almost alternating.

That last cited exercise takes a little more time, but it shows that every prime nonalternating knot of nine or fewer crossings is almost alternating. In fact, of the 393 nonalternating knots and links of eleven or fewer crossings, all but at most five are almost alternating. The remaining five don't look almost alternating, but we can't yet prove that they aren't.

Unsolved Question

Either show that the three knots shown in Figure 5.54 are almost alternating (by finding a projection that is almost alternating) or prove that they are not almost alternating.

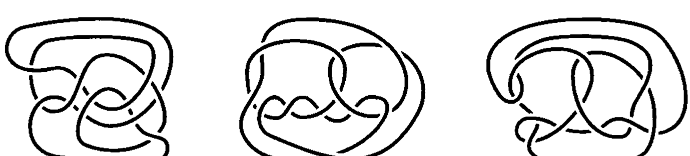


Figure 5.54 Three knots that could be almost alternating.

How do we show that all but three of the nonalternating knots of 11 or fewer crossings are almost alternating? John Conway used his notation (which we discussed in Section 2.4) in order to list all knots of 11 or fewer crossings. In particular, he listed the nonalternating ones. Since then, it has been proved that these knots are definitely not alternating. In Exercise 2.15, you showed that an algebraic knot that contains no negative signs in its Conway notation is in fact an alternating knot.

Exercise 5.32 Show that an algebraic link that has exactly one negative sign in its Conway notation has an almost alternating projection.

The result from working Exercise 5.32 immediately tells us that all but 17 of the nonalternating knots in Conway's list of 11-crossing prime knots are almost alternating. Similar tricks allow us to finish off the remaining knots in the list with only three exceptions.

Okay, so we are generally agreed that there are lots of almost alternating knots. If we can prove any theorems about almost alternating knots, we know that our results will apply to a lot of knots. We won't be wasting our time. We also know that there are a lot of results known for alternating knots (in fact, we will see some of them in the next chapter). The idea here is to try to generalize those results so they apply to almost alternating knots.

An amazing array of knots and links have almost alternating projections. In fact, the unknot has an almost alternating projection (Figure 5.55). Even more amazing is the fact that every alternating knot or link has an almost alternating projection. We can simply perform a Type II Reidemeister move to an alternating projection to obtain an almost alternating projection (Figure 5.56).

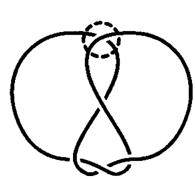


Figure 5.55 An almost alternating projection of the unknot.

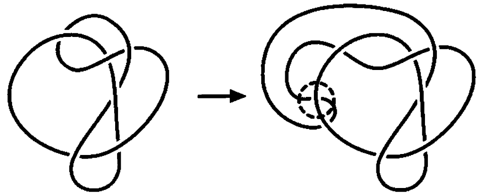


Figure 5.56 Alternating knots have almost alternating projections.

In the summer of 1990, six students and I proved that *a prime almost alternating knot is either a torus knot or a hyperbolic knot.* Therefore, since we discussed in Section 5.2 the fact that all knots fall into one of the three categories of torus, satellite, or hyperbolic knots, it immediately follows that no satellite knot can ever be an almost alternating knot. However, this theorem does not extend to links.

This theorem was a direct extension of the previously known fact for alternating links, namely that a prime alternating link is either a torus link or a hyperbolic link [see (Menasco, 1984)]. In fact, the only torus links that are alternating are the closures of the two-string braids, so just about every prime alternating link is hyperbolic (Figure 5.57).

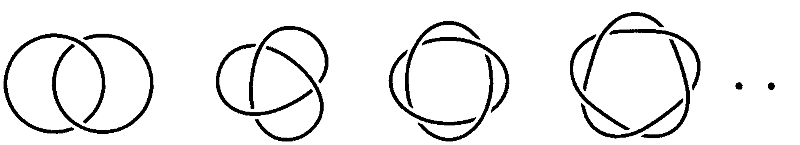


Figure 5.57 The only alternating torus links are the closures of the two-string braids.

Unsolved Question

Which torus knots are almost alternating knots? The (2, q)-torus knots are closures of the two-string braids and are therefore all alternating knots. Both the (3, 4) and (3, 5)-torus knots are almost alternating (the (3, 4)-torus knot is the 8_{19} knot, which we have seen to be almost alternating). We conjecture that these are the only almost alternating torus knots.

Given an almost alternating projection, we can always complicate it by a "tongue move," as in Figure 5.58. We just push a part of the link up over the nonalternating crossing. The result is a new almost alternating projection with two more crossings.

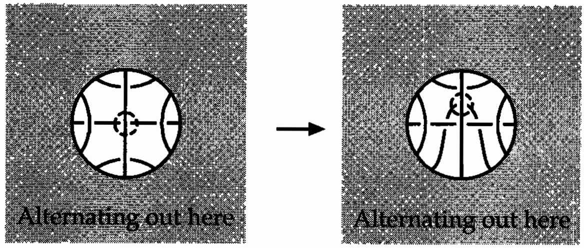


Figure 5.58 A tongue move.

ꝏ *Unsolved Question*

Determine exactly which almost alternating projections are projections of the unknot. Figure 5.59 shows one such projection of the unknot.

Figure 5.59 There may be other almost alternating projections of the unknot.

ꝏ *Unsolved Question*

When does the projection of an almost alternating link represent a splittable link? A connected alternating projection of a link is never splittable. This was first proved by Robert Aumann in 1956. There *are* projections of almost alternating links that are splittable. For instance, projections such as in Figure 5.60 are splittable.

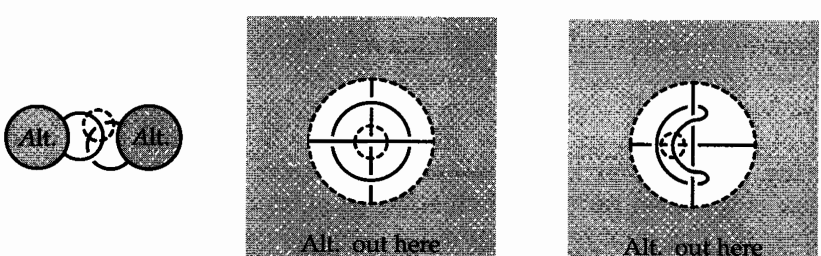


Figure 5.60 These almost alternating links are splittable.

What we would like is a complete list of all phenomena that can occur in the almost alternating projections of splittable links. Then we could look at such a projection and immediately tell if the link was splittable or not. We might conjecture that a "reduced" almost alternating projection of a link is splittable if and only if it is obviously so (that is to say, it consists of a projection that can be nontrivially separated by a circle in the projection plane that misses the link). The question is to decide what "reduced" means here.

We can also take this idea of almost alternating links and extend it. Define an **m-almost alternating knot** to be a knot that has a projection where m crossing changes would make the projection alternating, and the knot has no projection that could be made alternating in fewer crossing changes. We then consider alternating knots to be 0-almost alternating and almost alternating knots to be one-almost alternating. An example of a two-almost alternating knot is the following Whitehead double of the trefoil. It is not alternating or almost alternating because it is a satellite knot. In Figure 5.61, we display a projection that is two-almost alternating.

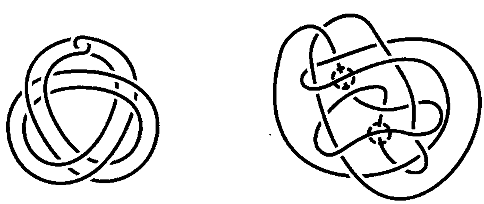


Figure 5.61 Two projections of a Whitehead double of the trefoil.

Exercise 5.33 Show that every knot is m-almost alternating for some m.

Exercise 5.34 Show that if a knot has an m-almost alternating projection, then it has an $(m + 1)$-almost alternating projection.

Exercise 5.35 Show that if a knot K has a projection with n crossings, then it is m-almost alternating for some $m \leq n/2$.

We have now divided all knots into separate categories, depending on their value of m. This number m measures how far a knot is from being alternating. It is similar to the unknotting number in that the unknotting number is the least number of crossing changes necessary in any projection to make the knot into the unknot. The unknotting number measures how far a knot is from being the unknot. In some sense, these two measurements, the almost alternating number and the unknotting number, are the two extremes. The alternating knot is "the most complicated" knot we can create by changing crossings in a projection, and the trivial knot is the simplest knot we can create by changing crossings in a projection.

Unsolved Question

Find a relation between the unknotting number and the almost alternating number. Note that Bleiler and Nakanishi's example from Chapter 3 (Figures 3.7 and 3.8) shows that a single knot need not realize its unknotting number (two in this case) and its almost alternating number (zero in this case) in the same projection.

Finally, let's discuss one other extension of almost alternating links that I have been thinking about recently. Instead of projecting a knot onto a plane, we project our knot onto a torus. Let T be an unknotted torus in space. Let K be a knot that can be projected onto the torus so that the projection is alternating, when viewed from outside (or inside) the torus. We also require that every closed curve on the torus that doesn't bound a disk on the torus intersects the projection. Then we say that K is a **toroidally alternating knot**.

For example, Figure 5.62 shows a knot on the torus T that is toroidally alternating. Obviously, it is in an alternating projection on the torus. It is not quite so obvious that every closed curve on the torus that doesn't bound a disk on the torus intersects the projection. But notice that if we treat the crossings as vertices, so that the projection of the knot becomes a graph on the torus, and then if we remove the graph from the torus, all of the remaining regions are disks. Any closed curve on the torus that doesn't intersect the projection must lie in one of these disks. But then it will bound a disk on the torus. Hence, this is a toroidally alternating knot.

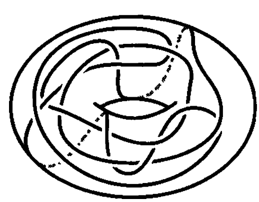


Figure 5.62 A toroidally alternating knot.

Exercise 5.36 Show that the trivial knot is toroidally alternating. (The hard part is finding a projection such that the closed curves on the torus that don't bound disks on the torus intersect the projection. See the preceding paragraph.)

Exercise 5.37 Draw your own nontrivial toroidally alternating knot.

Exercise 5.38 Show that any almost alternating knot is toroidally alternating. (*Hint*: Put the almost alternating projection on the torus, and push the funny crossing through to the other side of the torus.)

*Exercise 5.39** Show that any alternating knot is toroidally alternating.

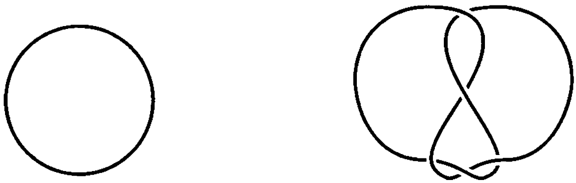
In fact, it turns out that toroidally alternating knots behave similarly to alternating and almost alternating knots. In particular, a prime toroidally alternating knot is either a torus knot or a hyperbolic knot. [See (Adams, 1992).] In a preprint from 1993, a Japanese mathematician, Chuichiro Hayashi, has independently come up with and examined the concept of toroidally alternating knots and links. Since this idea of toroidally alternating knots is very new, it will be a while before we know how useful it is. In the meantime, think about ways to generalize it.

6 Polynomials

6.1 The Bracket Polynomial and the Jones Polynomial

In this chapter, we talk about one of the most successful and interesting ways to tell knots apart. To each knot, we associate a polynomial. We compute the polynomial from a projection of the knot, but any two different projections of the same knot yield the same polynomial. So the polynomial is an invariant of the knot.

If we have two pictures of two knots and the computed polynomials are different, that tells us immediately that the two knots have to be distinct. For instance, we show that for one of the polynomials that we compute [which we denote by $V(t)$] the unknot has polynomial 1 while the 5_2 knot has polynomial $V(t) = -t^{-1} + 3t^{-2} - 2t^{-3} + 3t^{-4} - t^{-5} - t^{-6}$ (Figure 6.1). Therefore, the unknot and the 5_2 knot are distinct, a fact that we have not been able to prove until now. Note also that by a polynomial, we mean a **Laurent polynomial**, which can have both positive and negative powers of t.



$V(t) = 1$ $\qquad$ $V(t) = -t^{-1} + 3t^{-2} - 2t^{-3} + 3t^{-4} - t^{-5} - t^{-6}$

Figure 6.1 Polynomials of the unknot and the 5_2 knot.

Before we introduce any of the polynomials, let's take a look at the history of polynomial invariants for knots and links. The first polynomial associated to knots and links was due to J. Alexander in about 1928. This polynomial invariant was very good at distinguishing between knots and links. Mathematicians utilized the Alexander polynomial to distinguish knots and links for the next 58 years. In 1969, John Conway found a way to calculate the Alexander polynomial of a link using a so-called **skein relation.** This is an equation that relates the polynomial of a link to the polynomial of links obtained by changing the crossings in a projection of the original link. We will see that skein relations form the basis of much that is to follow.

In 1984, Vaughan Jones, a mathematician from New Zealand, discovered a new polynomial for knots and links. The polynomial came out of work he was doing on operator algebras, an area of mathematics previously unrelated to knot theory. He happened to notice that a relation that came up in operator algebras looked very much like a relation that occurred in knot theory. This led him to the discovery of a new polynomial for knots.

Jones's discovery generated immense excitement among knot theorists. Many knot theorists started working on polynomials. Four months after Jones announced his new polynomial, the discovery of the HOMFLY polynomial was announced. The name HOMFLY comes from the first letters of the names of the discoverers Hoste, Ocneanu, Millett, Freyd, Lickorish, and Yetter. Amazingly enough, this same polynomial was discovered by these people while they worked in four different independent groups. (A pair of Polish mathematicians named Przytycki and Traczyk also developed the same polynomial independently, but their work didn't arrive in the mail until several months later.) Since then, numerous other polynomials have been discovered, a few of which we will discuss.

We start with a discussion of the Jones polynomial as understood by Louis Kauffman, a mathematician at the University of Illinois in Chicago. In order to define the Jones polynomial, we first develop a polynomial associated to knots that is called the **bracket polynomial**. In our development of the bracket polynomial, we take the approach that a mathematician would take if he or she were trying to discover a polynomial that was an invariant for knots and links. Although it does not appear that way at the start, ultimately we obtain a Laurent polynomial of a single variable A.

Let's suppose that we are trying to create this polynomial invariant for knots and links, and that we have a few requirements for the polynomial. First of all, we would like the polynomial of the trivial knot to be 1. If we use the notation $<K>$ to denote the bracket polynomial of a knot K, then our first rule becomes:

$$\text{Rule 1:} \qquad <\bigcirc> = 1$$

Second, we want a method for obtaining the bracket polynomial of a link in terms of the bracket polynomials of simpler links. We use the following skein relation. Given a crossing in our link projection, we split it open vertically and horizontally, in order to obtain two new link projections, each of which has one fewer crossing. We make the bracket polynomial of our link projection a linear combination of the bracket polynomials of our two new link projections, where we have not yet decided on the coefficients, so we just call them A and B.

$$\text{Rule 2:} \qquad <\times> = A <)(> + B <\asymp>$$
$$<\times> = A <\asymp> + B <)(>$$

The second equation here is just the first equation, but looked at from a perpendicular view. If you bend your neck so that your head is horizontal and look at the first tangle in the second equation, it will appear the same as the first tangle in the first equation. Applying the first equation in Rule 2 to this tangle gives us exactly the second equation, so we don't actually consider these two equations as distinct rules. Finally, we would like a rule for adding in a trivial component to a link (the result of which will always be a split link). So we will say:

$$\text{Rule 3:} \qquad <L \cup \bigcirc> = C <L>$$

Each time we add in an extra trivial component that is not tangled up with the original link L, we just multiply the entire polynomial by C. As with A and B, we consider C a variable in the polynomial, for the time being.

The most important criterion for our polynomial is that it be an invariant for links. That is to say, the calculation of the polynomial cannot depend on the particular projection that we start with. *It must be unchanged by the Reidemeister moves.* Well, let's see what happens to our polynomial when we apply the Reidemeister moves. We'll begin with a Type II Reidemeister move. We want $<\,)\!\!(\,> = <)(>$ (see Figure 6.2).

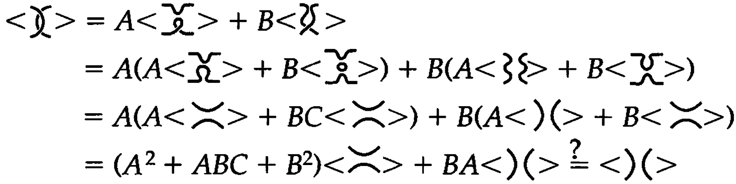
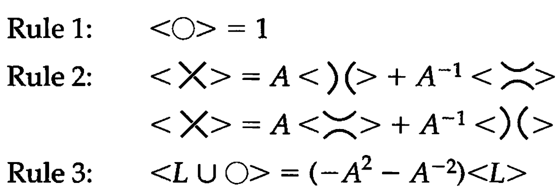


Figure 6.2 Effect on bracket polynomial of Type II move.

In order that the polynomial be unchanged by this move, we are forced to make $B = A^{-1}$, so that the coefficient in front of the vertical tangle is one. But that's okay. We weren't committed to having a B in the final polynomial anyway. Once we have replaced B by A^{-1}, it is apparent that we also need $A^2 + C + A^{-2} = 0$, so that the coefficient in front of the horizontal tangle is zero. This means we should make $C = -A^2 - A^{-2}$. Then, the bracket polynomial will be unchanged by a Type II Reidemeister move. Hence, from now on, our three rules for computing the bracket polynomial become:



Note that our polynomial now has a single remaining variable A.

Now, let's see what effect the third Reidemeister move has on the polynomial (Figure 6.3). Thus, Type III Reidemeister moves have no effect on the polynomial. Once we have fixed it so that the Type II moves leave the polynomial unchanged, the Type III move comes for free.

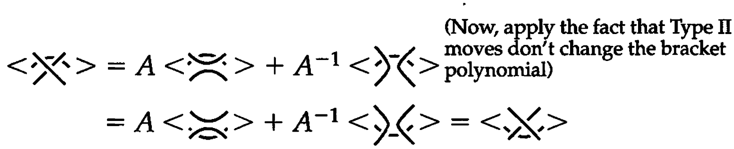


Figure 6.3 Effect on bracket polynomial of Type III move.

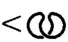
Before we discuss the Type I Reidemeister move, let's do a couple of quick calculations with our polynomial. First we just use Rules 1 and 3 to calculate the polynomial for the usual projection of the trivial link of two components. By Rule 3, where we let L be the unknot, we have

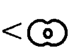
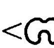
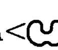
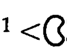
$$<\bigcirc \cup \bigcirc> = -(A^{-2} + A^{2})<\bigcirc> = -(A^{2} + A^{-2})1$$

the last equality coming from Rule 1.

Exercise 6.1 What would the bracket polynomial of the usual projection of the trivial link of n components be?

Let's try computing the bracket polynomial of a projection of the simplest nontrivial link on two components, the Hopf link. This time, we will use all three rules.

<  > $= A$ <  > $+ A^{-1}$ <  >

$= A\,(A$<  > $+ A^{-1}$ <  >$) + A^{-1}\,(A$<  > $+ A^{-1}$ <  >$)$

$= A\,(A(-(A^{2} + A^{-2})) + A^{-1}\,(1)) + A^{-1}\,(A(1) + A^{-1}\,(-(A^{2} + A^{-2})))$

$= -A^{4} - A^{-4}$

Exercise 6.2 Find the bracket polynomial for the projection of the trefoil knot, shown in Figure 6.4.

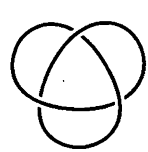


Figure 6.4 Find the bracket polynomial.

We obviously saved the best for last. Let's see what happens to the polynomial when we apply a Type I Reidemeister move (Figure 6.5). This looks *bad*. The polynomial has been changed by a Type I move. Our polynomial *does* depend on what projection of the knot we have. But don't despair, maybe we can fix the problem.

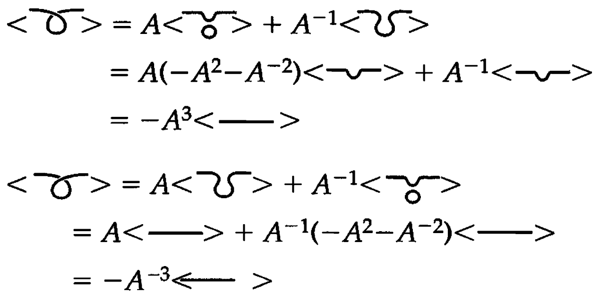


Figure 6.5 Effect on bracket polynomial of Type I move.

Let's give an orientation to our knot or link projection L. At every crossing of the projection, we have either a $+1$ or -1, as we saw in Section 1.4 (Figure 6.6). We call the sum of all these $+1$s and -1s the **writhe** of the oriented link projection L and denote it $w(L)$. (This is also sometimes called the *twist* of the projection.) Thus, for instance, we can calculate the writhe of the oriented link projection shown in Figure 6.7.

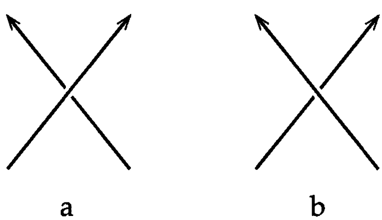


Figure 6.6 (a) +1 crossing. (b) −1 crossing.

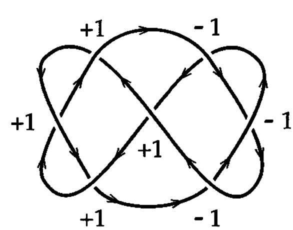


Figure 6.7 $w(L) = +4 - 3 = 1$.

Exercise 6.3 Calculate the writhe of the oriented link projection in Figure 6.8.

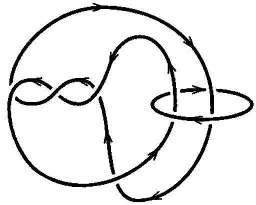


Figure 6.8 Determine the writhe of this link projection.

Exercise 6.4 Show that the writhe of a link projection is invariant under Reidemeister moves II and III.

Notice that Reidemeister move I always changes the writhe by ± 1. We are now going to define a new polynomial called the X polynomial. It is a polynomial of oriented links and it is defined to be

$$X(L) = (-A^3)^{-w(L)} <L>$$

Since both $w(L)$ and $<L>$ are unaffected by moves II and III, $X(L)$ is unaffected by moves II and III.

What happens to $X(L)$ when we do a Reidemeister move of Type I to L? Suppose first we had a strand as in Figure 6.9 and we took out the twist. Then $w(L') = w(L) + 1$, so

$$\begin{aligned} X(L') &= (-A^3)^{-w(L')} <L'> \\ &= (-A^3)^{(-w(L)+1)} <L'> \\ &= (-A^3)^{-(w(L)+1)} ((-A)^3 <L>) \\ &= (-A^3)^{-w(L)} <L> = X(L) \end{aligned}$$

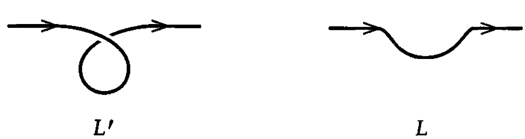


Figure 6.9 Effect on $X(L)$ of Type I move.

Thus, $X(L)$ is unaffected by this Type I Reidemeister move. Similarly, it is unaffected by the other version of a Type I move. *Therefore, $X(L)$ is an invariant for knots and links.* It does not depend on the particular projection, but rather depends only on the knot itself. As an example, consider the link ○ ○. As we previously computed, <○ ○>= $-A^2 - A^{-2}$. Since the writhe of this link is 0, we have that X(○ ○)= $-A^2 - A^{-2}$. This result is independent of the projection of the link. We could take a really nasty projection of this link, like the one in Figure 6.10. If we calculated the X polynomial for this projection (not something that I am recommending you do), we would find that the answer was exactly the same, namely $-A^2 - A^{-2}$.

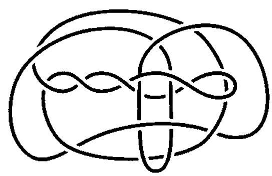


Figure 6.10 A nasty projection of ○ ○.

Exercise 6.5 Compute $X(L)$ for each of the following oriented links. What happens to $X(L)$ if we change the orientation on one or both components of the links?

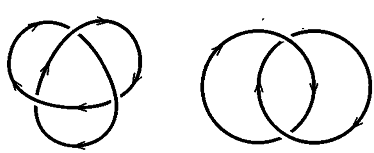


Figure 6.11 Compute $X(L)$.

*Exercise 6.6** (This problem assumes a familiarity with complex numbers.) We utilize the writhe of the link projection in order to deal with the fact that the bracket polynomial is not invariant under a Type I Reidemeister move. If we were not after a polynomial invariant of knots and links, but rather would be satisfied with a numerical invariant, we could use the fact that A is still a variable when we confront the problematic Type I moves. Show that there is a choice of complex number for A that will generate a complex-valued numerical invariant for knots and links, which is unchanged by the three Reidemeister moves. Use it to show that the figure-eight knot is not a trivial knot. (This requires the calculation of the bracket polynomial for the figure-eight knot.)

The original Jones polynomial is obtained from $X(L)$ by replacing each A in the polynomial with $t^{-1/4}$. The resulting polynomial with variable t

(and powers that are not necessarily integers) is *exactly* the polynomial that Jones first came up with in 1984. We denote this polynomial by $V(L)$, and sometimes $V(t)$, when the link involved is clear. How good is the Jones polynomial (or equivalently the X polynomial) at distinguishing knots? Every prime knot of 9 or fewer crossings has a distinct Jones polynomial. So we can distinguish between them all.

Exercise 6.7 Use the solution to Exercise 6.5 to write the Jones polynomial of the trefoil knot.

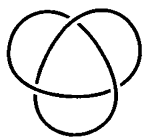


Figure 6.12 Determine the Jones polynomial.

Unsolved Question 1

Does the Jones polynomial distinguish every other knot from the unknot? That is to say, is there a nontrivial knot with Jones polynomial equal to 1? No one has yet found such a knot, but no one can prove such a knot doesn't exist. This is an important question.

Unsolved Question 2

Vaughan Jones has proved that the Jones polynomial of an (m, n)-torus knot is $t^{(m-1)(n-1)/2}(1 - t^{m+1} - t^{n+1} + t^{m+n})/(1 - t^2)$. The only known proof, however, relies on algebras and is relatively difficult. Find a simple proof of this fact. (Maybe relate the Jones polynomial of an (m, n)-torus knot to the Jones polynomial of a simpler torus knot.)

The Jones polynomial can be shown to satisfy a skein relation of its own. Let L_+, L_-, and L_0 be three oriented link projections that are identical except where they appear as in Figure 6.13.

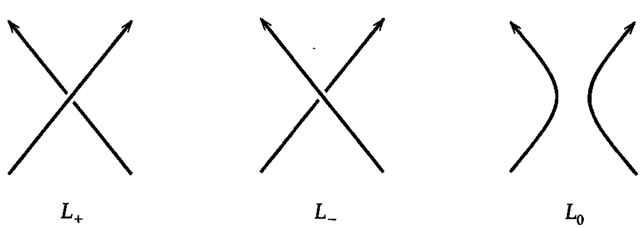


Figure 6.13 Three link projections that are almost identical.

Exercise 6.8 Use the skein relation of the bracket polynomial in order to show that the Jones polynomials of the three links in Figure 6.13 are related through the equation:

$$t^{-1}V(L_+) - tV(L_-) + (t^{-1/2} - t^{1/2})V(L_0) = 0$$

This was the original skein relation that Vaughan Jones recognized to hold for the Jones polynomial. We utilize it in Section 6.3.

6.2 Polynomials of Alternating Knots

We would like to have a second way to think about the bracket polynomial. We focus on Rule 3 for its computation. Four regions of the projection plane come together at a crossing. We label two of them with an A and two of them with a B by the following simple rule. Rotate the overstrand counterclockwise, passing over two of the regions. Label these two regions with an A and the remaining two regions with a B, as in Figure 6.14.

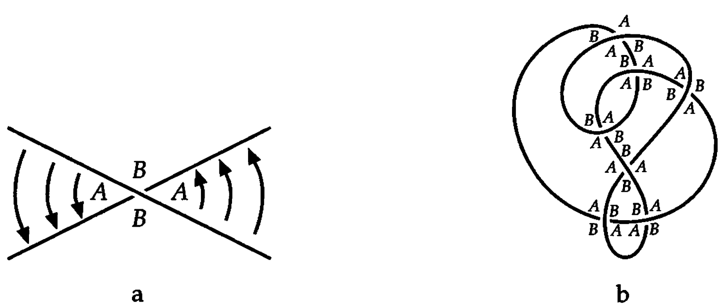


Figure 6.14 (a) Labeling a crossing. (b) Labeling a projection.

In Rule 2 for calculating the bracket polynomial, we split open a crossing (Figure 6.15a) in two different ways. When that crossing is labeled, the first splitting opens a channel between the two regions labeled A at the crossing. We call this an **A-split** (Figure 6.15b). The second splitting opens a channel between the two regions labeled B at the crossing. We call this a **B-split** (Figure 6.15c).

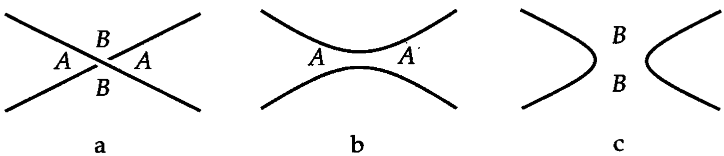
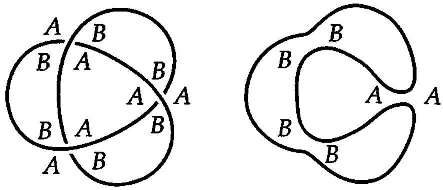


Figure 6.15 (a) Crossing. (b) A-split. (c) B-split.

Suppose L is a link in a projection of n crossings. Rule 3 allows us to determine the bracket polynomial of L using the bracket polynomials for two links L_1 and L_2, each of which has one fewer crossing than L. The links L_1 and L_2 are obtained by splitting a particular crossing in the projection of L as an A-split and then as a B-split. Since we can use Rule 3 to determine the bracket polynomials of each of L_1 and L_2 in terms of the bracket polynomials for a pair of links, each of which has one fewer crossing, the bracket polynomial of L depends now on four links, each of which has two fewer crossings than L. Continuing in a similar manner, we eventually have the bracket polynomial for L in terms of the bracket polynomials for 2^n links, all of which have *no* crossings. Each of these 2^n links simply comes from making a choice of an A-split or B-split at each of the crossings in the projection of L. Since there are n crossings, and we have two choices of how to split each crossing, there will be exactly 2^n links.

We call a choice of how to split all of the n crossings in the projection of L a **state**. The bracket polynomial of L then depends on the bracket polynomials for all of the possible states of the projection of L. Given a particular state of L, what is the bracket polynomial for the corresponding link L' that this state turns L into? The link L' has no crossings. Hence, L' must be a set of nonoverlapping unknotted loops in the plane. We will let $|S|$ be the number of loops in L'. Then by Exercise 6.1, we know that the bracket polynomial of L' is simply $(-A^2 - A^{-2})^{|S|-1}$.

But what factor is this polynomial multiplied by when we add it into the bracket polynomial of the original link? Each time we split at a crossing, the polynomials of the two resultant links were multiplied by either an A or an A^{-1}, depending on whether the split was an A-split or a B-split. So the polynomial of L' is multiplied by $A^{a(S)}A^{-b(S)}$, where $a(S)$ is the number of A-splits in S and $b(S)$ is the number of B-splits in S. Hence, the total contribution to the bracket polynomial by the state S is $A^{a(S)}A^{-b(S)}(-A^2 - A^{-2})^{|S|-1}$. For example, the particular state of the trefoil knot shown in Figure 6.16 contributes $A^1 A^{-2}$ to the bracket polynomial of this projection of the trefoil knot.



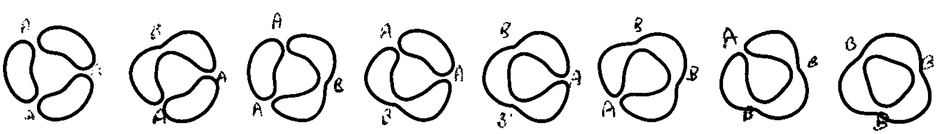
$|S| = 1$

Figure 6.16 One state of this projection of the trefoil.

The bracket polynomial of the projection of the link L will now be simply the sum over all of the possible states of these contributions. We write this as

$$<L> = \sum_S A^{a(S)} A^{-b(S)} (-A^2 - A^{-2})^{|S|-1} \tag{6.1}$$

This point of view has some advantages. In particular, if we want to compute the bracket polynomial of a given projection of L, we can simply list all of the links obtained by splitting all of the crossings of L in every possible combination and then compute the contribution to the polynomial of each term. As an example, let's recompute the bracket polynomial of the trefoil projection in Figure 6.16. Since there are three crossings in the projection, there will be $2^3 = 8$ states. For each of the eight states, we have to compute $|S|$, which we do by simply counting how many circles there are in the corresponding link (Figure 6.17).



$|S| = 3$ $|S| = 2$ $|S| = 2$ $|S| = 2$ $|S| = 1$ $|S| = 1$ $|S| = 1$ $|S| = 2$

Figure 6.17 Computing the bracket polynomial for this projection of the trefoil knot.

Hence,

$$\begin{aligned}<K> &= A^3A^0(-A^2-A^{-2})^{3-1} + A^2A^{-1}(-A^2-A^{-2})^{2-1} + A^2A^{-1}(-A^2-A^{-2})^{2-1} \\ &\quad + A^2A^{-1}(-A^2-A^{-2})^{2-1} + A^1A^{-2}(-A^2-A^{-2})^{1-1} + A^1A^{-2}(-A^2-A^{-2})^{1-1} \\ &\quad + A^1A^{-2}(-A^2-A^{-2})^{1-1} + A^0A^{-3}(-A^2-A^{-2})^{2-1} \\ &= A^3(-A^2-A^{-2})^2 + 3A(-A^2-A^{-2}) + 3A^{-1} + A^{-3}(-A^2-A^{-2}) \\ &= A^7-A^3-A^{-5}\end{aligned}$$

Exercise 6.9 Compute the bracket polynomial for this projection of the figure-eight knot using Equation (6.1) (see Figure 6.18).

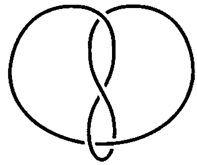


Figure 6.18 Find the bracket polynomial.

We have defined an alternating knot to be any knot that has a projection such that if you traverse the knot in a particular direction, you alternately pass over and then under crossings, one after the other. We call the projection an *alternating projection*. We will call an alternating projection **reduced** if there are no unnecessary crossings in the projection, as in Figure 6.19. Note that if we ever had an unreduced alternating projection, we could simplify it to a reduced alternating projection, thereby lowering the number of crossings in the projection. But if an alternating projection is reduced, there is no obvious way to lower the number of crossings. This fact formed the basis for two conjectures dating back to the last century.

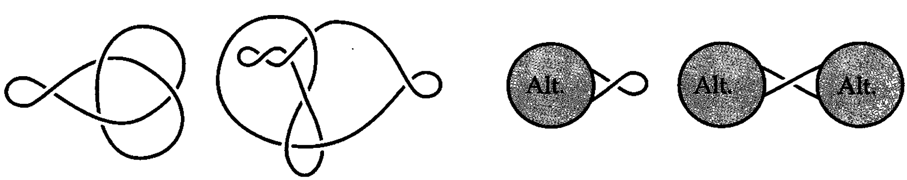


Figure 6.19 Unreduced alternating projections.

Conjecture 1

Two reduced alternating projections of the same knot have the same number of crossings.

Conjecture 2

A reduced alternating projection of a knot has the least number of crossings for any projection of that knot.

Both of these conjectures withstood the concerted efforts of numerous mathematicians, but neither one could withstand attack by the new polynomials. Both were shown to be true by Louis Kauffman, Morwen Thistlethwaite, and Kunio Murasugi in 1986. Together, these conjectures imply that we can determine the crossing number for any alternating knot. We simply take an alternating projection and simplify it until it is reduced. Since all reduced alternating projections for this knot have the same number of crossings, and since the least number of crossings occurs in a reduced alternating projection, the least number of crossings *is* the number of crossings in this projection.

We will prove the first conjecture; the second one is a bit more difficult, but not outrageously so. We refer you to (Kauffman, 1988) for a proof of the second. Let's begin with a definition. The **span** of a polynomial is the difference between the highest power that occurs in the polynomial and the lowest power that occurs in the polynomial. For instance, the span of the polynomial

$$A^3 - 2A^2 + 1 - A^1 - 7A^{-2}$$

is $3 - (-2) = 5$.

Let's look at the span of the bracket polynomial. Even though the bracket polynomial is not an invariant for knots, it is true that the span of the bracket polynomial is an invariant. That is, for a given knot K, if we calculate the bracket polynomial from *any* projection whatever of the knot, and then take the span, we will always get the same answer. Let's see why this is the case. Suppose that we have two different projections P_1 and P_2 of the knot K. Then there is a series of Reidemeister moves that take us from P_1 to P_2. We have already seen that the Reidemeister moves of Types II and III do not change the bracket polynomial at all, so they must both leave the span of the bracket polynomial unchanged.

We *do* know that Type I Reidemeister moves can change the bracket polynomial. But what do they do to the span? We saw in Figure 6.5 that a Type I move multiplies the entire polynomial by A^3 or A^{-3}. If we multiply by A^3, this increases the highest power in the polynomial by 3 and increases the lowest power in the polynomial by 3. Hence the difference of those two, which gives the span, is unchanged. Similarly, multiplying the entire polynomial by A^{-3} also leaves the span unchanged. Thus, all three Reidemeister moves leave the span of the bracket polynomial unchanged. Thus, the span of the bracket polynomial must be the same for all projections of the knot K, and the span of the bracket polynomial is an *invariant of the knot*.

We now prove the lemma that is the key to proving Conjecture 1.

Lemma If K has a reduced alternating projection of n crossings, then span $(<K>) = 4n$.

Proof. We know that the span of the bracket polynomial of K doesn't depend on the projection of the knot that we use, so we might as well use the reduced alternating projection given to us in the statement of the lemma. Since the span of the bracket polynomial is simply the difference between the highest power that occurs in the bracket polynomial and the lowest power that occurs, we look at each of these two quantities in turn.

First, let's figure out what the highest power will be in the bracket polynomial. Each state contributes a term of the form $A^{a(S)}A^{-b(S)}(-A^2 - A^{-2})^{|S|-1}$. If we expand this out, the highest power of A occurring in this term will be $A^{a(S)}A^{-b(S)}(A^2)^{|S|-1}$. Among all the states we therefore want to find the one that has the highest value of $a(S) - b(S) + 2(|S| - 1)$. That highest value will be the highest power of A that occurs in the bracket polynomial.

In order to make $a(S) - b(S) + 2(|S| - 1)$ as large as possible, we want to pick a state where $|S|$ and $a(S)$ are large but $b(S)$ is small. For $|S|$ to be large, we need there to be many disjoint circles in the link corresponding to S. For $a(S)$ to be large and $b(S)$ to be small, we want as many of the splits as possible to be A-splits, and consequently, as few of the splits as possible to be B-splits. Let's try taking all A-splits and no B-splits. Since we have n crossings, this means $a(S) = n$ and $b(S) = 0$. What happens to $|S|$? Since the knot is alternating, when we place A's and B's around a crossing, we see that the vertices in any region of the projection are either all labeled with A's or all labeled with B's. Let's shade the A regions gray while leaving the B regions white (Figure 6.20).

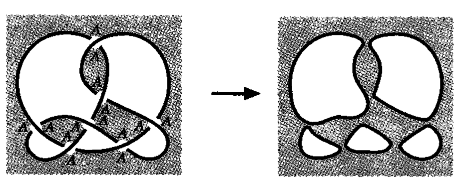


Figure 6.20 Shade the A regions gray while leaving the B regions white.

What happens when we open all of the A-channels? The gray waters flood the projection, leaving only a set of white islands in the middle of the gray lake. How many circles are there in the resulting link? Each circle is either the boundary of an island or the boundary of the lake (if the outermost region is white). Thus, if W is the number of white regions in the original projection, including possibly the outer region, then $|S| = W$. Therefore, the highest power of A corresponding to this particular state is $a(S) - b(S) + 2(|S| - 1) = n + 2(W - 1)$. We claim that every other state has a highest power that is lower than this. Why?

Any other state has some B-splits. We show that if we have a state S_1 and we go to a state S_2 by changing one A-split to a B-split, the highest power cannot go up. The highest power in the term corresponding to S_1 is $a(S_1) - b(S_1) + 2(|S_1| - 1)$. Then the highest power of the term corresponding to S_2 is of the form $(a(S_1) - 1) - (b(S_1) + 1) + 2(|S_2| - 1)$ since we have decreased the number of A-splits by one and increased the number of B splits by one. So the question remaining is how different $|S_2|$ can be from $|S_1|$. But S_2 differs from S_1 in only one split. Either that change in split increases the number of circles by one or it decreases the number of circles by one. Hence $|S_2| = |S_1| \pm 1$. Thus the highest power of the term corresponding to S_2 is $a(S_1) - b(S_1) - 2 + 2((|S_1| \pm 1) - 1)$. This is certainly no greater than the highest term corresponding to S_1, as we wanted to show.

Thus, any time we change an A-split to a B-split, we do not increase the highest power. Since every state can be obtained from the all-A-split state by a sequence of such changes, no other state has a higher power than the all-A-split state. In fact, it's not hard to see that the other states have highest powers that are strictly less than the highest power in the all-A-split state.

Exercise 6.10 Show that the highest power of A that occurs in an all-A-split state is strictly greater than the highest power of A that occurs in a state with exactly one B-split. (*Hint:* Use the fact that the projection is reduced and alternating.)

Therefore, the highest power that occurs in the bracket polynomial for K is in fact $n + 2(W - 1)$. By a similar argument, we can also show that the lowest power that occurs is $-n - 2(D - 1)$, where D is the number of darkened regions. This lowest power occurs in the term of the polynomial coming from the all-B-split state. We therefore have that span($<K>$) = highest power − lowest power = $n + 2(W - 1) - (-n - 2(D - 1)) = 2n + 2(W + D - 2)$. But $W + D$ is the total number of regions in the projection, and the total number of regions is $n + 2$. (See the next exercise.) Hence span($<K>$) = $2n + 2n = 4n$, as we set out to prove. □

Exercise 6.11 Show that the number of regions R in a connected knot or link projection is always two more than the number of crossings (including the region outside the knot). (*Hint:* Either use the fact that the

Euler characteristic of a disk is always 1 or simply draw a knot, keeping count of the number of regions created whenever you create a new crossing.)

Exercise 6.12 As a simple corollary to the lemma, show that if K has a reduced alternating projection of n crossings, then the span of its Jones polynomial is exactly n. (A pretty amazing fact, which when taken together with Conjecture 2, says that the crossing number of an alternating knot is exactly the span of its Jones polynomial.)

We are now almost done with the proof of Conjecture 1.

Theorem Two reduced alternating projections of the same knot have the same number of crossings.

Proof. If the first projection has n crossings, then by the lemma, the span of the bracket polynomial of that projection is $4n$. But since the span of the bracket polynomial is an invariant of the knot, it doesn't change when we change projections. So the span of the bracket polynomial corresponding to the second projection is also $4n$. But the lemma then implies that the number of crossings in the second projection is also n. Hence both projections have the same number of crossings. □

In 1983, William Menasco proved that if $K_1\#K_2$ is an alternating knot, then it appears composite in any alternating projection. That is to say, there is a circle in the projection plane that intersects the knot twice, such that the factor knots on either side of the circle are themselves alternating. In particular, $K_1\#K_2$ must look something like Figure 6.21. Together with Conjecture 2, Menasco's result has the following corollary.

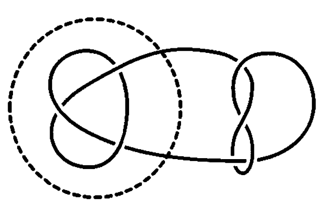


Figure 6.21 A composite alternating knot.

Corollary $c(K_1\#K_2) = c(K_1) + c(K_2)$ for $K_1\#K_2$ an alternating knot.

Proof. Choose a reduced alternating projection for $K_1 \# K_2$. By Menasco's result, K_1 appears as part of this projection. Hence, we have a reduced alternating projection of K_1. Conjecture 2 then says that the least number of crossings for K_1 is the number appearing in this picture. Since K_2 is alternating, Conjecture 2 also says that its least number of crossings is the number appearing in this picture. Since $K_1 \# K_2$ is alternating, its least number of crossings also occurs in this picture. Hence $c(K_1 \# K_2) = c(K_1) + c(K_2)$. □

As we mentioned in Section 4.1, it is still an open conjecture that $c(K_1 \# K_2) = c(K_1) + c(K_2)$ holds for all knots. Alternating knots are the first major category of knots for which this conjecture has been shown to be true. There is a larger class of knots, called *adequate knots*, that contains all of the alternating knots, and for which the arguments that were applied to alternating knots can be extended. In particular, we can define an adequate projection of a knot and then show that a reduced adequate projection of a knot has the minimal crossing number of that knot.

Unsolved Questions

What other knots besides alternating and adequate knots have minimal crossing number in a particular type of projection? As we mentioned in Section 5.1, Kunio Murasugi proved that the least number of crossings for a torus knot occurs in one of the two standard projections for a torus knot (depending on which of p and q is larger). Are there other categories of knots for which it holds? Is it true $c(K_1 \# K_2) = c(K_1) + c(K_2)$ if K_1 and K_2 are both torus knots? What if one is an alternating knot and the other is a torus knot?

A conjecture for alternating knots that has been around since the days of Peter Guthrie Tait (1831–1901) in the 1890s is the **Tait Flyping Conjecture**. This is a conjecture about projections of knots where we project to the sphere (as we did in Section 2.2), rather than to a plane. The conjecture then says that if we have two reduced alternating projections of the same knot, they are equivalent on the sphere if and only if they are related through a sequence of moves called **flypes**. A flype is a 180° rotation of a tangle, as in Figure 6.22. Although we have drawn the flype as if it is occurring on a plane, think of it as occurring on the surface of a large sphere, so that to us the sphere looks like a plane.

Figure 6.22 Flypes.

Exercise 6.13 Find the sequence of flypes that get us from the first projection in Figure 6.23 to the second projection.

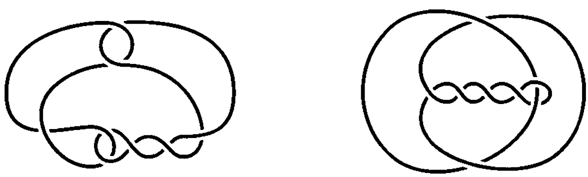


Figure 6.23 Find the sequence of flypes.

If true, the Tait Flyping Conjecture allows us to draw all possible reduced alternating projections of a given alternating knot. They are all obtained by doing all the possible flypes on any one projection. This process generates at most a finite number of projections. The conjecture also implies that unlike the process of using Reidemeister moves to get from one projection to another, flypes allow us to get from any one reduced alternating projection of a knot to any other reduced alternating projection of the same knot *without ever increasing the number of crossings.*

In 1990, William Menasco (State University of New York at Buffalo) and Morwen Thistlethwaite (University of Tennessee) together proved the Tait Flyping Conjecture. The proof uses a blend of geometric techniques and the new polynomials.

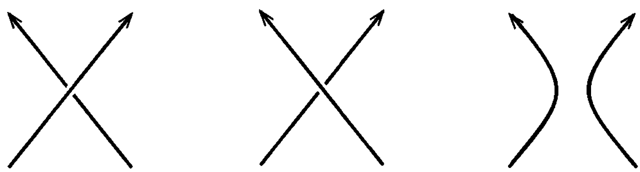
6.3 The Alexander and HOMFLY Polynomials

The very first polynomial for knots was the Alexander polynomial, invented back in 1928. It is a polynomial for oriented links, and we describe it so that its variable is t. At the time of its invention, it was defined in terms of relatively abstract mathematical concepts beyond the scope of this book. It wasn't until 1969 that John Conway showed that the Alexander polynomial Δ can be computed using just two rules. The first rule is the usual one, namely that the trivial knot has trivial polynomial equal to 1.

$$\text{Rule 1:} \qquad \Delta(\bigcirc) = 1.$$

But a difference here is that this holds true for *any* projection of the trivial knot, not just the usual one.

The second rule is similar to the skein relation that we saw was satisfied by the Jones polynomial in Exercise 6.8. Again, we take three projections of links L_+, L_-, and L_0 such that they are identical except in the region depicted in Figure 6.24. Then the polynomials of these three links are related through our second rule:

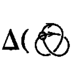
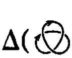
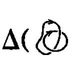


L_+ $\qquad$ L_- $\qquad$ L_0

Figure 6.24 Three links that are identical except at this crossing.

$$\text{Rule 2:} \qquad \Delta(L_+) - \Delta(L_-) + (t^{1/2} - t^{-1/2})\,\Delta(L_0) = 0$$

Although we do not prove it here, these two rules are enough to ensure that the Alexander polynomial is an invariant for knots and links. In particular, this means that if we are given a projection of a knot, we can compute the Alexander polynomial of the knot in any projection, and we will get the same answer. We do not need to keep the projections frozen throughout the calculation, as we had to do with the bracket polynomial. For example, let's compute the Alexander polynomial of the trefoil knot. Treating the trefoil knot as L_+, with the circled crossing as the one that appears in Figure 6.24, we obtain

$\Delta($  $) - \Delta($  $) + (t^{1/2} - t^{-1/2})\,\Delta($  $) = 0$

where

$\Delta($  $) = \Delta(\bigcirc) = 1$

and

$$\Delta($$$$) - \Delta($$$$) + (t^{1/2} - t^{-1/2})\,\Delta($$$$) = 0$$

By Exercise 6.15, which follows,

$\Delta($$) = 0$, so $\Delta($$) = -t^{1/2} + t^{-1/2}$ and

$$\Delta($$$$) = (t^{1/2} - t^{-1/2})^2 + 1 = t - 1 + t^{-1}$$

Exercise 6.14 Compute the Alexander polynomial of the figure-eight knot.

Exercise 6.15 Show that the Alexander polynomial of a splittable link is always 0. (*Hint:* Picture the splittable link as L_0.)

Unlike the Jones polynomial, there are known examples of nontrivial knots with Alexander polynomial equal to 1. This is one of the disadvantages of the Alexander polynomial: It cannot distinguish all knots from the trivial knot. For instance, the $(-3, 5, 7)$-pretzel knot has Alexander polynomial 1 (Figure 6.25).

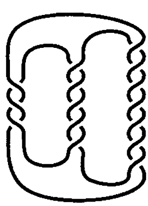


Figure 6.25 The $(-3, 5, 7)$-pretzel knot has Alexander polynomial 1.

One question that comes up is whether or not Rule 2 always allows us to calculate the Alexander polynomial. When we were calculating the bracket polynomial, it was always clear that the application of the skein relation to a crossing resulted in two link projections that were simpler than the original link projection, since they each had fewer crossings. Hence, we knew that the process would eventually lead to a set of trivial links, for which we could calculate the polynomials. Here, it is less clear that we can always express the Alexander polynomial of a given link in terms of the Alexander polynomials of two simpler links.

You will remember, however, that when we discussed unknotting number, we proved that any projection can be turned into a projection of a trivial link by changing some subset of the crossings. Therefore, suppose we have a knot or link for which we would like to compute the Alexander polynomial. Given a particular projection, we could choose a crossing, such that it is one of the crossings that we would like to change in order to turn the projection into a trivial projection. Letting the original projection correspond to either L_+ or L_-, we can use the skein relation in order to obtain the polynomial of our original link in terms of the polynomial of a link with a projection with one fewer crossing and the polynomial of a link with a projection that is one crossing closer to the trivial projection. Iterating this procedure allows us to obtain the polynomial of the original link in terms of the polynomials of a set of trivial links.

This process of repeatedly choosing a crossing, and then applying the skein relation to obtain two simpler links, yields a tree of links called the **resolving tree**. At the top is our original link; at the bottom, we find all of the trivial links that result from repeatedly applying the skein relation. For instance, here is a resolving tree for the trefoil knot (Figure 6.26).

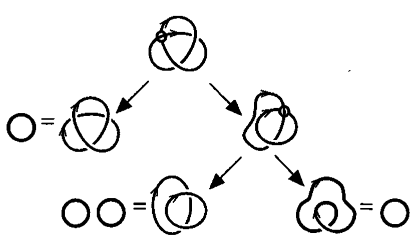


Figure 6.26 Resolving tree for the trefoil knot.

Exercise 6.16 Find a resolving tree for the knot 6_3 (shown in Figure 6.27).

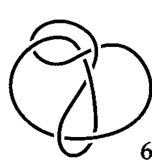

6_3

Figure 6.27 A 6_3 knot.

These resolving trees are useful for the calculation of several of the new polynomials. Define the depth of a resolving tree to be the number of levels of links in the tree, not including the initial level at the top. So the resolving tree shown for the trefoil in Figure 6.26 has depth two.

Define the **depth of a link L** to be the minimal depth for any resolving tree for that link. Here is an invariant for links that measures the complexity of the calculation of the Alexander polynomial. Let's see what is known. The only links of depth zero are the trivial links. It has been proved by Bleiler and Scharlemann that a knot of depth one is always a trivial knot, and that the links of depth one are all Hopf links, possibly with a few extra disentangled trivial components added in. The links of depth two have also been classified, this time by Abigail Thompson from the University of California at Davis, working jointly with Martin Scharlemann.

Unsolved Problems

1. Classify the links of depth n for any $n > 2$.
2. Show that there are only a finite number of links with a given number of components and a given depth.

Remember that the Alexander polynomial was the only polynomial for over 50 years. But once Vaughan Jones discovered the Jones polynomial in 1984, quite a few mathematicians started to look for a polynomial with *two* variables instead of one, a polynomial that would generalize both the Jones polynomial and the Alexander polynomial. The first such polynomial to be discovered was the HOMFLY polynomial, which is a two-variable Laurent polynomial, the variables being m and l. As an example, the oriented link in Figure 6.28 has HOMFLY polynomial $P = (-l^3 - l^5)m^{-1} + (2l^3 - l^5 - l^7)m + (-l^3 + l^5)m^3$.

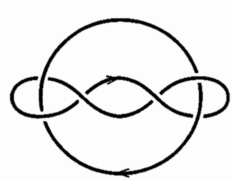


Figure 6.28 A link with HOMFLY polynomial as above.

What are the rules for calculating the HOMFLY polynomial?

Rule 1: $P(\bigcirc) = 1.$

As before, we want the unknot to have polynomial 1. As in the case of the Alexander polynomial, this holds true for *any* projection of the unknot.

Rule 2: If L_+, L_-, and L_0 are again three oriented links that are identical except in the region that appears in Figure 6.24, then

$$lP(L_+) + l^{-1}P(L_-) + mP(L_0) = 0$$

Notice how similar this relationship is to both Rule 2 for calculating the Alexander polynomial and to the skein relation that we showed was satisfied by the Jones polynomial in Exercise 6.8.

Let's use these rules to calculate the HOMFLY polynomials for some links. In Figure 6.29, we see three links that are identical except at the one crossing, and thus form a triple of links L_+, L_-, and L_0. Hence, we have that $lP(L_+) + l^{-1}P(L_-) + mP(L_0) = 0$. But both L_+ and L_- are simply slightly twisted pictures of the unknot. Hence $P(L_+) = P(L_-) = 1$, and therefore $mP(L_0) = -(l + l^{-1})$. Thus, we have shown that $P(L_0) = -m^{-1}(l + l^{-1})$.

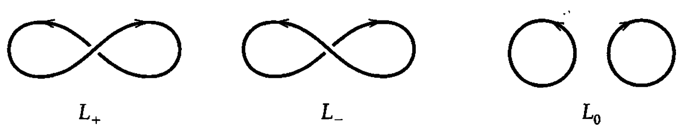


Figure 6.29 Three related links.

Exercise 6.17 Compute the polynomial of the link L_- in Figure 6.30 utilizing the rules for the HOMFLY polynomial together with the HOMFLY polynomial that we have already computed.

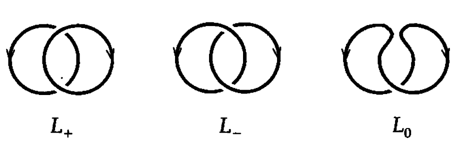


Figure 6.30 Compute the HOMFLY polynomial of L_-.

Exercise 6.18 Determine the polynomial of the trefoil in Figure 6.31.

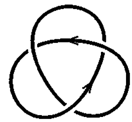


Figure 6.31 Find the HOMFLY polynomial of the trefoil.

Exercise 6.19 Show that the HOMFLY polynomial of a knot is identical to the HOMFLY polynomial of the same knot, but with the opposite orientation.

This last exercise demonstrates that we need not distinguish between orientations when we are discussing the HOMFLY polynomial of a knot. If we are dealing with a link, however, changing some but not all of the orientations on the components can have an effect on the polynomial.

In general, we *can* always compute the HOMFLY polynomial of a link. As with the Alexander polynomial, all that we need is a resolving tree in order to do the calculation. However, the calculation can be very slow. In particular, most of the current computer programs are effective for computing the polynomials of simple knots and links, but we wouldn't want to try them on a 100-crossing projection. Even the fastest computers would take too long to do the computation. (The link would eventually reduce to 2^{100} links, none of which have any crossings. But 2^{100} is larger than the estimated number of particles in the universe.)

There is a computer program written by Hugh Morton and Hamish Short that can be applied to closed braids of nine or fewer strings and that can successfully compute the polynomials of knots with several hundred crossings. It utilizes a different algorithm.

The HOMFLY polynomial has several very interesting properties. For instance, suppose $L_1 \cup L_2$ is the so-called **split union** of the two links L_1 and L_2. This is simply the link obtained by moving L_1 over near L_2, but not overlapping them or linking them in any way (Figure 6.32). (L_1 and L_2 can be separated by a sphere, so the resulting link is splittable.) Then $P(L_1 \cup L_2) = -(l + l^{-1})m^{-1}P(L_1)P(L_2)$. In particular, if we apply this union to the trivial knot on two components, we obtain the same HOMFLY polynomial that we computed using the two rules for the polynomial. We won't bother to prove this fact. Although the proof is not difficult, it's a little on the messy side. You might try to prove this yourself, using the rules for calculating the HOMFLY polynomial.

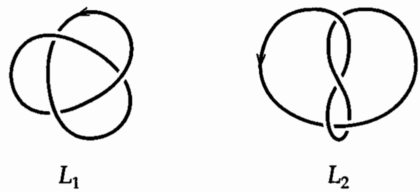


Figure 6.32 The disjoint union of L_1 and L_2.

Exercise 6.20 (a) Show that $P(L \cup \bigcirc) = -(l + l^{-1})m^{-1}P(L)$.
(b)* Show that $P(L_1 \cup L_2) = -(l + l^{-1})m^{-1}P(L_1)P(L_2)$.

However, this equation does lead to a second interesting property of the polynomial:

$$P(L_1 \# L_2) = P(L_1)P(L_2)$$

That's right, the polynomial of the composition of two links is simply the product of the polynomials of the factor links. This seems too good to be true. For instance, the polynomial of the composite of two trefoils is just $(-2l^2 - l^4 + l^2m^2)^2$ (Figure 6.33), which is the square of the polynomial for the trefoil.

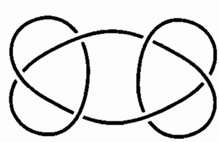
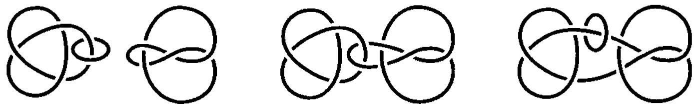


Figure 6.33 This composite has polynomial $(-2l^2 - l^4 + l^2m^2)^2$.

Another amazing fact here is that we didn't say how to take the composition of a link. We didn't specify which component of the first link should be connected up to which component of the second link. In fact, it doesn't matter. All those possibly distinct composite links will have the same polynomial (Figure 6.34). This is our first example of links that are certainly distinct, but that cannot be distinguished by the HOMFLY polynomial.



a b c d

Figure 6.34 (a) L_1. (b) L_2. (c) First choice for $L_1\#L_2$. (d) Second choice for $L_1\#L_2$.

Let's see how to prove the formula for $P(L_1\#L_2)$ from the formula for $P(L_1 \cup L_2)$. The composite link $L_1\#L_2$ has a projection that appears as in Figure 6.35. Without cutting the strands to L_1, let's flip that part of the projection corresponding to L_2 in two different ways, to get the two links L_+ and L_-. Note that both of these projections are still projections of $L_1\#L_2$. In addition, L_0 is simply the disjoint union $L_1 \cup L_2$ (Figure 6.36).

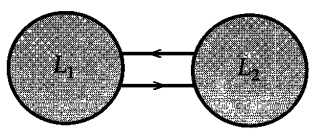
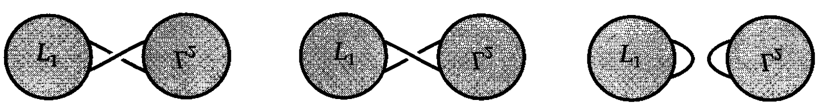


Figure 6.35 A projection for $L_1\#L_2$.



L_+ L_- L_0

Figure 6.36 Three related links.

The second rule for calculation of the P polynomial then says

$$lP(L_1\#L_2) + l^{-1}P(L_1\#L_2) + mP(L_1 \cup L_2) = 0$$

But we know that $P(L_1 \cup L_2) = -(l + l^{-1})m^{-1}P(L_1)P(L_2)$; hence we have

$$lP(L_1\#L_2) + l^{-1}P(L_1\#L_2) + m(-(l + l^{-1})m^{-1}P(L_1)P(L_2)) = 0$$
$$(l + l^{-1})P(L_1\#L_2) + (-(l + l^{-1})P(L_1)P(L_2)) = 0$$
$$P(L_1\#L_2) = P(L_1)P(L_2)$$

This is an amazing rule. Remember back in Section 1.2 when we said that composition of prime knots was analogous to multiplication of prime integers? This rule says that the polynomials of the knots behave exactly as the integers do. The polynomial of a composite knot factors into the polynomials of all of its factor knots.

How good is the HOMFLY polynomial at telling apart knots and links? Better than either the Jones polynomial or the Alexander polynomial, since we will see that both of those are simply special cases of this polynomial. But we have already seen examples of links that it will not distinguish, particularly the links coming from different ways to take the composition of two links. However, perhaps it does better with knots. Maybe every distinct knot has a distinct HOMFLY polynomial and maybe we could tell all knots apart simply by looking at their polynomials.

We should be so lucky. The HOMFLY polynomial is not what is called a **complete invariant** for knots. It cannot distinguish all knots. In particular, a pair of mutant knots (which we discussed in Section 2.3) will always have the same HOMFLY polynomial (Figure 6.37). Mutants are big trouble in general. As we saw in Section 5.3, they cannot be distinguished by hyperbolic volume either. We did see that the Kinoshita–Terasaka mutants were distinguishable because their minimal genus Seifert surfaces had different genera, but this is not a general technique for distinguishing mutants. Many pairs of mutants have the same genus.

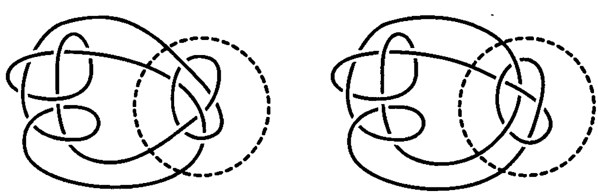


Figure 6.37 Two mutant knots have the same polynomial.

Unsolved Question 1

Find a good way to distinguish between mutants. You want to find an invariant for knots that is affected by mutation. Maybe the following question gives us a possibility.

Unsolved Question 2

Given a pair of mutant knots, is there always a choice of integers p and q such that the (p, q)-cable knots on each of the two mutant knots have distinct HOMFLY polynomials?

How is the HOMFLY polynomial related to the Jones polynomial? We simply replace l by it^{-1} and m by $i(t^{-1/2} - t^{1/2})$ in the HOMFLY polynomial. Here, $i = \sqrt{-1}$. Note that the skein relation for the HOMFLY polynomial then becomes the skein relation for the Jones polynomial that we discussed in Exercise 6.8. For instance, we have seen in Exercise 6.18 that the HOMFLY polynomial of the trefoil knot is $P(K) = -2l^2 - l^4 + l^2m^2$. Substituting for m and l, we have that

$$\begin{aligned} V(K) &= -2(it^{-1})^2 - (it^{-1})^4 + (it^{-1})^2(t^{-1/2} - t^{1/2})^2 \\ &= 2t^{-2} - t^{-4} - t^{-2}(t^{-1} - 2 + t) \\ &= t^{-4} - t^{-3} + t^{-1} \end{aligned}$$

This is exactly the Jones polynomial that we computed earlier for the trefoil.

Exercise 6.21 Show that the substitution $l = i$ and $m = i(t^{1/2} - t^{-1/2})$ turns the HOMFLY polynomial into the Alexander polynomial (by showing that the resulting polynomial obeys the rules for the Alexander polynomial). Thus, the HOMFLY polynomial is a more powerful invariant than either the Jones polynomial or the Alexander polynomial. It carries their information within it.

Finally, we mention that the HOMFLY polynomial can be very helpful in determining the braid index of a knot. Robert Williams and John Franks, and independently Hugh Morton, proved that if we let E be the largest exponent of l in the HOMFLY polynomial of the oriented link L and e the smallest exponent of l, then the braid index n of the link L satisfies the inequality

$$n \geq \frac{(E - e)}{2} + 1$$

Amazingly enough, this inequality is sharp for all but five prime knots of 10 or fewer crossings (the exceptions being 9_{42}, 9_{49}, 10_{132}, 10_{150}, and 10_{156}).

Unsolved Question

Determine what is special about these five knots. Why are they the only prime knots of 10 or fewer crossings with braid index strictly greater than $(E - e)/2 + 1$?

There are several other polynomials that we will not discuss, each of which has its own advantages and disadvantages. See the references for readable articles that explore these other polynomials.

6.4 Amphicheirality

A while back, we defined the notion of an amphicheiral knot, namely a knot such that it is ambient isotopic to its mirror image. That is to say, a knot is amphicheiral if it can be deformed through space to the knot obtained by changing every crossing in the projection of the knot to the opposite crossing. We also insist that an orientation on the knot is taken to the corresponding orientation on the mirror image of the knot under the ambient isotopy. Let K^* be the mirror image of K.

Exercise 6.22 Show that the bracket polynomial of K^* is just the bracket polynomial of K where the variable A is replaced everywhere by the variable A^{-1}. Show that the same is true for the X polynomial.

If K is an amphicheiral knot, then K is in fact the same knot as K^*; they are simply in distinct projections. Hence, it must be the case that

$$X_K(A) = X_{K^*}(A)$$

[The notation $X_K(A)$ means the X polynomial of K with variable A.] However, Exercise 6.22 shows that

$$X_K(A) = X_{K^*}(A^{-1})$$

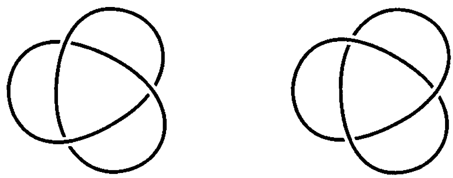
Thus, if K is an amphicheiral knot, it must be that

$$X_K(A) = X_{K^*}(A^{-1}) = X_K(A^{-1})$$

Hence the polynomial of an amphicheiral knot must be *palindromic*, that is to say, the coefficients must be the same backwards or forwards, where we list *all* of the coefficients, including all the zeros.

What about the figure-eight knot? We showed that the figure-eight knot was amphicheiral in Section 1.3, utilizing the Reidemeister moves. Therefore its polynomial should be palindromic. In fact, its polynomial is $A^8 - A^4 + 1 - A^{-4} + A^{-8}$, which is palindromic. Replacing every A by an A^{-1} gives us the same polynomial back again.

On the other hand, the trefoil knot has polynomial $A^4 + A^{12} - A^{16}$. This polynomial is not palindromic. If we replace every A by an A^{-1}, we get $A^{-4} + A^{-12} - A^{-16}$, which is not the same polynomial. Hence, this shows that the trefoil knot is not amphicheiral. *The trefoil knot is distinct from its mirror image.* This means that even though all along we have been discussing the trefoil as if it were a single knot, it is actually two knots, one called the *right-hand trefoil* and the other called the *left-hand trefoil* (Figure 6.38). In fact, the first proof that the left-hand and right-hand trefoils are distinct was offered by Max Dehn (1878–1952) in 1914. However, he didn't have knot polynomials to work with then, so his proof was completely different.



a b

Figure 6.38 The left-hand trefoil (a) and the right-hand trefoil (b) are distinct.

Let's look at the implications of this discussion for alternating knots. We have already seen that if K is an alternating knot in a reduced alternating projection of n crossings, then

$$\text{max deg} <K> = n + 2(W - 1)$$

and

$$\text{min deg} <K> = -n - 2(D - 1)$$

Since $X(K) = (-A)^{-3w(K)} <K>$, we have that

$$\text{max deg } X(K) = n + 2(W - 1) - 3w(K)$$

and

$$\text{min deg } X(K) = -n - 2(D - 1) - 3w(K)$$

But we have already seen that for an amphicheiral knot

$$X_K(A) = X_K(A^{-1})$$

In particular, this means that

$$\text{max deg } X(K) = -\text{min deg } X(K)$$

So we obtain the formula

$$n + 2(W - 1) - 3w(K) = -(-n - 2(D - 1) - 3w(K))$$

This yields $3w(K) = W - D$.

Hence, in order for an alternating knot to be amphicheiral, it must be that the difference in the number of white regions and darkened regions is exactly three times the writhe, and this equality must be satisfied in *any* reduced alternating projection.

Exercise 6.23 Show that if the absolute value of the writhe in a reduced alternating projection of an alternating knot is greater than or equal to one third of the numbers of crossings in that projection, then the knot cannot be amphicheiral. (In fact, Morwen Thistlethwaite has proved the stronger result, that if K is amphicheiral, $w(K) = 0$.)

Exercise 6.24 Exactly one of the knots of six or seven crossings in the appendix table is amphicheiral. Determine which one it must be.

In 1890, Peter Guthrie Tait made the following conjecture:

Unsolved Conjecture

If the crossing number of a knot K is odd, that knot is not amphicheiral.

The trefoil knot is an example of a knot for which the conjecture holds. The least number of crossings for the trefoil is three, an odd number, and it is not amphicheiral.

Exercise 6.25 Prove the preceding open conjecture when it is restricted to alternating knots.

Since the information of the Jones and X polynomials is embedded within the HOMFLY polynomial, it should also provide us with information about amphicheirality.

Exercise 6.26 Show that the HOMFLY polynomial of K^* is obtained by replacing each l in the HOMFLY polynomial of K with an l^{-1}. Use this fact to show that the left-hand and right-hand trefoil knots are distinct.

Although surprisingly effective at determining the amphicheirality of knots, the HOMFLY polynomial is not infallible. For instance, the knot 9_{42} has HOMFLY polynomial

$$P(9_{42}) = (-2l^{-2} - 3 - 2l^2) + (l^{-2} + 4 + l^2)\, m^2 - m^4$$

Note that the polynomial is unchanged when every l is replaced by an l^{-1}. Hence, $P(9_{42}) = P(9_{42}^*)$. However, there exists a "signature" invariant coming out of algebraic topology that proves that 9_{42} is not amphicheiral.

Unsolved Question

Find a complete invariant for amphicheirality. That is to say, find an invariant that will definitively determine whether or not a knot is amphicheiral.

7 Biology, Chemistry, and Physics

7.1 DNA

As we mentioned in Section 1.1, much of the initial interest in knot theory was motivated by the possibilities of applications to chemistry. However, it wasn't until the 1980s that applications to chemistry were actually realized. In particular, we start by discussing applications of knot theory to DNA, beginning with some background.

In the 1950s, it was realized that the genetic code appeared in the double helix structure of DNA. Deoxyribonucleic acid (DNA) is a molecule that is formed by pairs of long molecular strands that are bonded together by ladder rungs and that spiral around each other, forming the so-called double helix. The molecular strands are made up of alternating sugars and phosphates. Each sugar is bonded to one of four bases, A = Adenine, T = Thyamine, C = Cytosine, and G = Guanine (Figure 7.1). The rungs of the ladder are formed by hydrogen bonding between pairs of bases, where A always bonds to T and C always bonds to G. Note that the sequence of bases as we move down one strand is then mimicked by the other strand, except that the As and Ts have been exchanged and the Cs and Gs have been exchanged. The sequence of As, Ts, Cs, and Gs as we run down one of the strands is the genetic code, giving a blueprint for life. These molecules contain on the order of millions of individual atoms, all of which are packed into the tiny nucleus of a cell. In fact, if the nucleus of a cell were the size of a basketball, the DNA within it would be equivalent to 200 kilometers of fishing line. And it's not as if we carefully wound the fishing line up before we stuffed it into the basketball. It's a tangled mess.

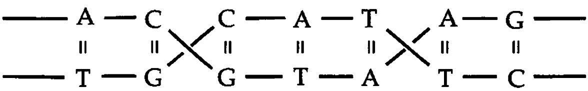


Figure 7.1 The DNA double helix.

But the DNA has to be utilized in order to perform various biological functions, such as replication, transcription, and recombination. These are the processes of reproducing a given DNA molecule, copying segments of DNA, and modifying DNA molecules, respectively. All three of these are necessary for life. The knotting and tangling in the DNA molecules make the performance of these processes difficult. In order for these biological mechanisms to function, there must be some way of manipulating the tangled masses of DNA molecules.

Nature gets around this problem by providing enzymes called **topoisomerases**. These enzymes manipulate the DNA topologically. In Figure 7.2 we see three of the possible actions the enzymes can take. However, a particular enzyme may have a much more sophisticated action. Conceivably, it could take two strands of DNA and replace them with a nontrivial tangle. Once a particular enzyme has been isolated, biochemists would like to determine how it acts on the DNA. Since much of the DNA in the cell is not circular DNA, the enzyme could cause a knot to be formed in a strand of DNA, but because the two ends of the strand are free, the knot might slip off the end of the DNA strand. The scientists would not be able to see what effect the enzyme has had. To solve this problem, scientists utilize circular DNA molecules. Letting the enzyme act on this DNA, they then examine the result. If the enzyme is causing knotting, that knotting will be captured on the circular DNA.

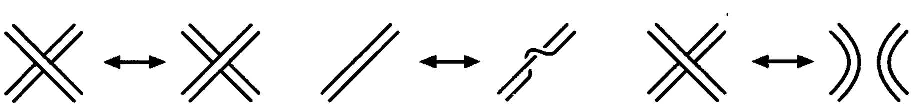


Figure 7.2 Three actions that enzymes can take on DNA.

In fact, circular DNA occurs in nature. In the early 1960s, it was discovered that DNA in a certain bacteriophage appeared as a single stranded ring on the order of 15,000 atoms. (By single stranded, we mean that this type of DNA is not double stranded, but rather consists of a single strand of alternating phosphates and sugars.) Since then, biochemists have discovered that both single-stranded cyclic DNA and duplex (the usual two-stranded double helix) cyclic DNA are prevalent, appearing not only in many bacteria and viruses but also in the mitochondria of human cells. More recently, biochemists have discovered how to artificially create cyclic DNA. It is to these synthetic molecules that they can then apply the enzymes, in order to determine their effect.

We focus on the duplex cyclic DNA. Each of the phosphates along the ladder edge is bonded to two different sugar molecules. Each of the sugars is bonded to a base molecule, being one of the C, T, G, or A molecules, and also to two phosphates, which occur at two different sites on the sugar molecule, called the 3′ and the 5′ sites (Figure 7.3). A single phosphate will be bonded to the 3′ site of one sugar and to the 5′ site of another sugar. Hence we can think of a phosphate as the connector, sticking a 3′ site on one sugar to a 5′ site on a second sugar. Thus, a linear strand of DNA will have two ends, one of which is a sugar with an open 3′ site and one of which is a sugar with an open 5′ site. This gives an orientation to a single strand of DNA, determined by the convention that we start at the 5′ end of the strand and head toward the 3′ end.

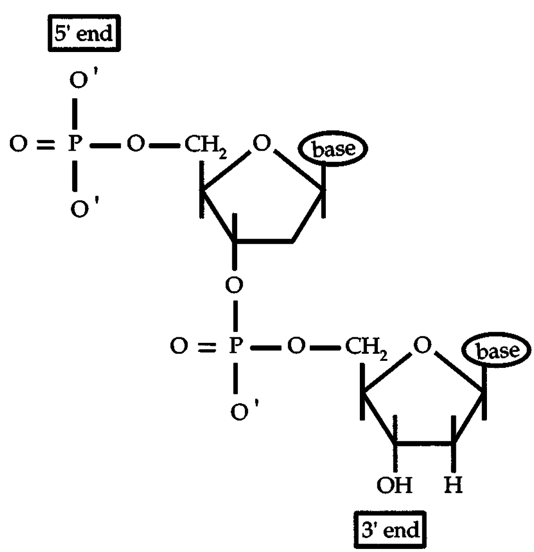


Figure 7.3 Two sugar molecules and their bonds.

Interestingly enough, linear duplex DNA has no such orientation. If an end of one strand has an open 3′ site, the corresponding end of the parallel strand will have an open 5′ site. In particular, each end of the linear duplex DNA will have both an open 3′ and an open 5′ site. The two strands are oppositely oriented, giving us no way to orient the duplex linear molecule (Figure 7.4).

Figure 7.4 Linear single-strand DNA is oriented, linear duplex DNA is not.

For our considerations, there is an even more important consequence to these sites. Namely, if the ends of the linear duplex DNA are brought together to form a cyclic duplex DNA molecule, the 3′ site must be glued to a 5′ site and vice versa. This forces each strand of the DNA to glue its head to its own tail rather than to the tail of the other strand. Hence we get two linked strands rather than a single strand running around twice. Put another way, there must be an even number of half-twists in the cyclic duplex DNA, when it is laid out flat in the plane.

The geometry of cyclic duplex DNA is very interesting. It can be modeled as a ribbon in three-space, with the two ends of the ribbon glued together (Figure 7.5). The two boundaries of the ribbon correspond to the two edges of the DNA ladder. Since there must be two distinct edges of the ribbon, we know that the ribbon never takes the form of a Möbius band. The curve that runs along the center of the ribbon is called the **axis** of the ribbon. Although it doesn't model a part of the molecule, it does tell us how contorted the molecule is in space. We can choose an orientation on the axis and then give the two boundaries of the ribbon orientations that match it.

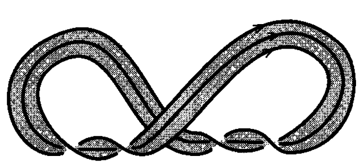


Figure 7.5 A ribbon modeling cyclic duplex DNA.

Let's look at these ribbons in space in more detail. Suppose that we have such a ribbon in space, but instead of treating it as if it were made of rubber, we assume that it is fixed rigidly in space. We compute some invariants that depend on this particular placement in space, and that would vary if we did move the ribbon. First, we define the **twist** of the ribbon, denoted **Tw(*R*)**. It measures how much the ribbon twists around its axis. When the axis lies flat in the plane, without crossing itself, the twist of the ribbon is simply one-half of the sum of the +1s and −1s occurring at the crossings between the axis and a particular one of the two link components bounding the ribbon. It doesn't matter which link component we use. We get the same answer with either one. The +1s and −1s are determined by the convention that we utilized in Section 1.4 and that appears in Figure 7.6.

When the axis is not flat in the plane, we must define the twist of the ribbon more abstractly. It is the so-called integral of the incremental twist of the ribbon about the axis, integrated as we traverse the axis once. (If you don't know what an integral is, you now have at least one good reason to take calculus.) It simply measures how much the ribbon twists about the axis from the frame of reference of the axis. It need not be an integer.

Next, we define the **writhe** of the ribbon, denoted **Wr(*R*)**. It measures how much the axis of the ribbon is contorted in space. For any particular projection of the axis, define the **signed crossover number** to be the sum of all the ±1s occurring at crossings where the axis crosses itself. Notice that we do not divide by 2, as we did for linking number. Again, we use the convention from Section 1.4 to decide which crossings are +1 crossings and which are −1 crossings (Figure 7.6).

Now, for the writhe of the ribbon, we take the average value of the signed crossover number, over every possible projection of the axis. Keep in mind though that the axis remains fixed in space, so when we talk about all the possible projections of the axis, we mean the planar pictures that would result as we looked at the fixed axis from all possible vantage points on a sphere that surrounds it (Figure 7.7). Such an average value is

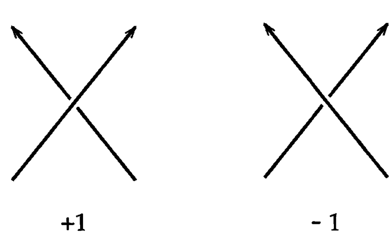


Figure 7.6 Convention for determining the writhe of a ribbon.

determined by utilizing integrals. We take the integral of the signed crossover numbers, integrating over all vantage points on the unit sphere, and then divide by the integral of one, integrating over the unit sphere.

$$\text{Average value} = \frac{\int \text{signed crossover number } dA}{\int dA} = \frac{\int \text{signed crossover number } dA}{4\pi}$$

since $\int dA$ is just the surface area of the unit sphere.

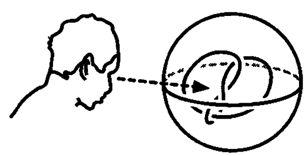


Figure 7.7 Looking at axis from all possible vantage points.

Note that if the ribbon axis lies in a plane, the signed crossover number is 0 for all projections except for those where our eye is in the plane. These last projections do not have a well-defined crossing number because we are looking at the axis edge on. But we ignore these projections since they form 0% of the total set of projections. (What percent of the surface of a sphere is the equator?) Hence the writhe of the ribbon axis would be 0 when it lies on a plane.

It becomes trickier to compute the writhe when the axis is not in a plane, as some projections will have crossings and others will not. See Figure 7.8, for instance. In order to compute the writhe of a particular ribbon in space, we would need to have detailed equations that described exactly where the axis was. The writhe is also not necessarily an integer.

Figure 7.8 The writhe of this knot is difficult to compute. (The thicker parts of the knot are closer to your eye than the thin parts.)

Finally, we can treat the two boundaries of the ribbon as components of a link and then compute the linking number of the two components, denoting the result by **Lk(*R*)**. Remember the linking number is just one-half of the sum of the ± 1s occurring at the crossings between the two components. This last invariant does not depend on the particular placement of the link in space. It would remain the same if we treated the ribbon as if it were made of rubber, and isotoped it to a different position.

James White of UCLA, Brock Fuller of Caltech, and G. Călugăreanu, a Czech mathematician, all working independently, discovered the following remarkable relation between these three invariants:

$$\mathrm{Lk}(R) = \mathrm{Tw}(R) + \mathrm{Wr}(R) \tag{8.1}$$

In simple cases, we can use this equation to find one of the invariants, knowing the other two. For example, we see the values of these three invariants in the two cases shown in Figure 7.9. In Figure 7.9a, the axis of the ribbon lies flat in the plane, giving $\mathrm{Wr}(R) = 0$. The linking number is easily computed to be $+1$, and the twist is then forced to be $+1$ by Equation 8.1. In Figure 7.9b, the ribbon doesn't twist around its axis at all, giving $\mathrm{Tw}(R) = 0$. Since we can compute $\mathrm{Lk}(R) = +1$, it must that $\mathrm{Wr}(R) = +1$.

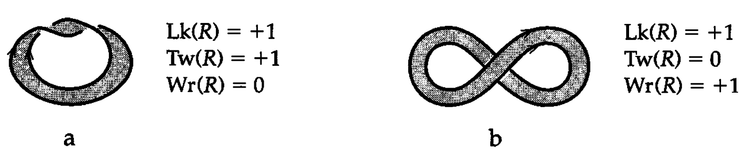


Figure 7.9 Twist, writhe, and linking number.

Equation 8.1 implies that if we have a ribbon that we isotope to a new position in space, any change in twist has to be exactly balanced by the change in writhe, since the linking number is unchanged by the isotopy. In Figure 7.9, we see this effect. These two ribbons are in fact isotopic, as you can easily check with your belt. Buckle your belt together with one full twist in it (well, take it off, first). Now, see if you can place it flat in the plane like Figure 7.9b. Unless your belt has a lot of elastic in it, you won't succeed, but you will see a projection of it that looks right. Your other option is to go buy a more elastic belt.

In its relaxed state, DNA twists around its axis at a rate of 10.5 base pairs per helical twist. This relaxed rate of twisting is caused by the way the sugars, phosphates, and base pairs bond. Thus, a cyclic duplex DNA with exactly 105 base pairs could lie flat in the plane as a two-braid link with Tw(R) = 10, Lk(R) = 10, and Wr(R) = 0 (Figure 7.10).

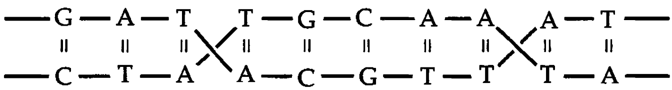


Figure 7.10 A relaxed cyclic duplex DNA.

However, sometimes a cyclic duplex DNA is more tightly twisted than 10.5 base pairs per twist, having fewer base pairs per twist, and hence twisting more over the same length of molecule. For example, suppose we have a cyclic duplex DNA that has an axis in the plane, so Wr(R) = 0. But now suppose that both the twist and the linking numbers have doubled, so that Tw(R) = Lk(R) = 20, while the number of base pairs has remained at 105. Then we have 5.25 base pairs per twist. The DNA is uncomfortably overwound (Figure 7.11). Since the total number of base pairs in the molecule is fixed, the only option for the DNA is to reduce the number Tw(R) to 10. Then there will be 10.5 base pairs per twist. However, Equation 8.1 says that decreasing Tw(R) must cause an increase in Wr(R). Hence, Wr(R) must now go to 10. This means that the axis of the ribbon will now become contorted in space. This effect is known as **supercoiling.**

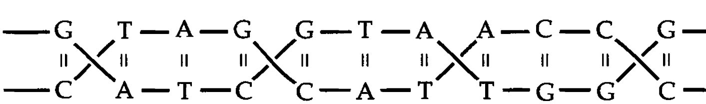


Figure 7.11 Overwound DNA.

We are all familiar with supercoiling. It is the same effect we see with the coiled cord that attaches the phone receiver to the phone box. That cord likes to twist at a rate of about five coils per inch. In its relaxed state (when it's not attached at one end), it will lie flat, coiled at that rate. But if we hold both ends and start to add twists to the cord, it remains straight only as long as we stretch it out. As soon as we let it go slack while still holding the ends, we immediately see it go to a supercoiled position (Figure 7.12). The cord gets twisted up with itself, its axis "writhing" through space.

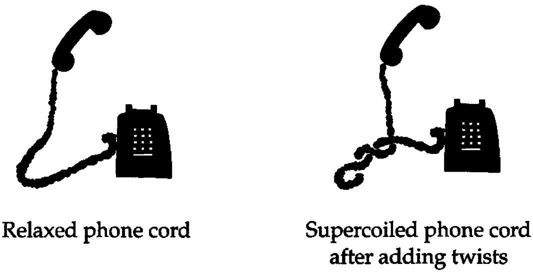


Figure 7.12 Adding twists to the phone line to create supercoiling.

This phenomenon has interesting implications for biochemistry. Suppose we have a cyclic duplex DNA molecule that is relaxed, lying with its axis in a plane and twisting happily at 10.5 base pairs per twist. Suppose now that an enzyme comes along and nicks one of the two strands open, twists it once around the other strand, and reglues the two ends together. Assuming the axis still lies in a plane, we have increased each of Lk(R) and Tw(R) by one (Figure 7.13). One would expect such a change to be virtually unnoticeable, since it happens at one point on the molecule, a localized effect.

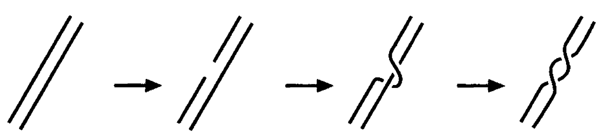


Figure 7.13 An enzyme adds a full twist to a DNA molecule.

However, now the molecule is too tightly wound to be comfortable. So instead of staying in the plane, it will decrease Tw(R) by one and therefore increase Wr(R) by one. This means the axis of the ribbon will tangle with itself. It will no longer lie flat. Biochemists can discern such a change in the molecule. They can place the molecules in a gel and then pass electricity through the gel to attract the molecules toward an electrode. The molecules with greater supercoiling are more compact, and hence move more quickly through the gel, allowing their separation (Figure 7.14). Once the molecules have been separated, they can be examined under an electron microscope. Using recently developed techniques, the DNA can be coated to thicken it, making it possible to see the actual crossings and tangling. In Figure 7.15, we see a picture from an electron microscope of actual DNA.

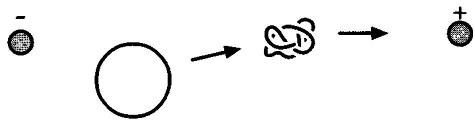


Figure 7.14 Supercoiled molecules move faster through the gel.

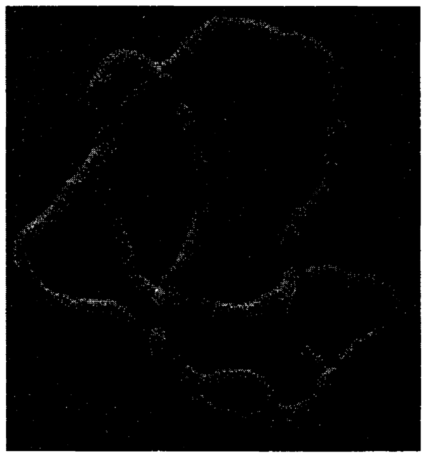


Figure 7.15 Knotting in DNA under an electron microscope. (From Wasserman et al., 1985.)

We now return to the original question, which was how to determine the action of an enzyme on DNA. We discuss a particular type of action by an enzyme called **site-specific recombination**, which is a process whereby an enzyme attaches to two specific sites on two strands of DNA, called **recombination sites**, each of which corresponds to a particular sequence of base pairs that the enzyme recognizes. After lining the sites up, the enzyme cuts the two strands open and recombines the four ends in some manner. In Figure 7.16, we show one of the simplest actions.

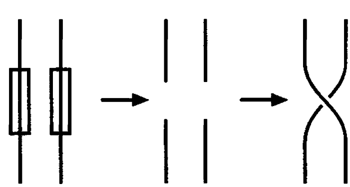
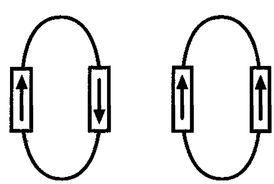


Figure 7.16 A possible site-specific recombination.

When the two strands are on two different molecules, it is hard to determine exactly what operation has taken place. So instead biochemists create a single circular DNA molecule that contains a copy of each of the two recombination sites necessary for the reaction. Then, when the enzyme acts on this molecule, the result can be analyzed to determine the effect of the enzyme. Since a recombination site is a nonpalindromic sequence of base pairs, we can choose an orientation for the site. When a pair of sites is utilized in an enzyme action, we pick the orientations of the two sites so that they will match when the enzyme pulls the two sites together. When both sites appear on the same circular DNA molecule, these orientations can either point in the same direction as we traverse the molecule, in which case we say that we have **direct repeats**, or their orientations can point in opposite directions as we traverse the molecule, this case being known as **inverted repeats** (Figure 7.17). Before the reaction takes place, we call the DNA molecule the **substrate**. During the reaction, either the enzyme or random thermal motion lines the two sites up so that their orientations match. Depending on the action of the enzyme, and on whether we have direct repeats or inverted repeats, the resulting molecule, called the **product**, can be a knot, an unknot, or a two-component link.



a b

Figure 7.17 (a) Direct repeats and (b) inverted repeats.

Exercise 7.1 Suppose the enzyme acts by adding one crossing as in Figure 7.16. In the case of direct repeats, determine whether the product will be a knot or a link. Similarly, determine which it is in the case of inverted repeats.

We use the concept of tangles from Chapter 2 to analyze the effect of a given enzyme. We think of the circular molecule before the reaction (the substrate) as being made up of two tangles, the **substrate tangle**, denoted S, which is unchanged by the enzyme, and the **site tangle**, T, where the enzyme acts. So far, the site tangle has always been trivial, consisting of two vertical strands. However, this need not always be the case. The enzyme replaces the site tangle with a new tangle called the **recombination tangle** *R* (Figure 7.18). We assume that we know what knot the substrate is in, and we can determine what knot the product becomes. The three variables that we do not know are the three tangles *S*, *T*, and *R*.

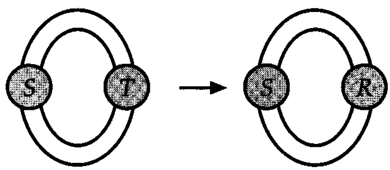
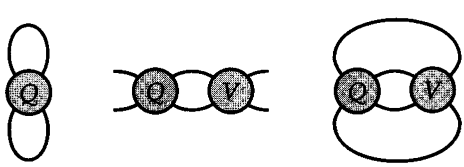


Figure 7.18 The enzyme replaces the site tangle with the recombination tangle.

Let's establish some notation. Let **N(Q)** denote the knot or link obtained by connecting the top two strands of a tangle *Q* to each other and the bottom two strands of *Q* to each other. Let ***Q* + *V*** denote the tangle obtained by adding the two tangles *Q* and *V* together, this addition being the addition for tangles that we defined in Chapter 2 (Figure 7.19). In this notation, the facts that the substrate comes from the tangles *S* and *T* and the product comes from the tangles *S* and *R* can be written in two equations in the three unknowns *S*, *T*, and *R*:

$$N(S + T) = \text{substrate}$$
$$N(S + R) = \text{product}$$

Since we have more variables than we have equations, we can never hope to determine all three of *S*, *T*, and *R* from knowing the knotting of the substrate and the product. If we happen to know one of the three, however, we should be able to determine the other two.



a b c

Figure 7.19 (a) $N(Q)$. (b) $Q + V$. (c) $N(Q + V)$.

We make one assumption that does not come out of the mathematics, but rather, is supported by biological observation. Namely, the tangles T and R do not depend on the tangle S. They depend only on the enzyme that is acting, and not on the knottedness of the molecule it acts on.

One example of a topoisomerase is the enzyme **Tn3 resolvase**. This enzyme acts on a particular duplex cyclic DNA molecule with direct repeats. Once it has matched up the two sites, it replaces the T tangle with a single R tangle and releases the molecule. Once in a while, however, it will repeat the tangle replacement a second time before releasing the molecule. Even more rarely, it can repeat the tangle replacement a number of times, yielding even more complicated molecules. In a series of experiments, biochemists established what products resulted when the enzyme acted, and determined the following equations, where we use the notation for rational knots from Section 2.2:

$$N(S + T) = N(1) \quad \text{(the unknot)}$$
$$N(S + R) = N(2) \quad \text{(the Hopf link)}$$
$$N(S + R + R) = N(211) \quad \text{(the figure-eight knot)}$$
$$N(S + R + R + R) = N(11111) \quad \text{(the Whitehead link)}$$

From this set of equations, De Witt Sumners, of Florida State University, and Claus Ernst, of Western Kentucky University, proved that $S = (-3, 0)$ and $R = (1)$ (Figure 7.20). Moreover, they proved that it should then be the case that $N(S + R + R + R + R) = N(12111)$ (the 6_2 knot). This last knot has been observed as a product. For more details on the proof, see (Sumners, 1993).

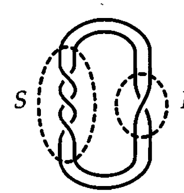


Figure 7.20 $S = (-3, 0)$ and $R = (1)$.

A second example of a topoisomerase is the **Int** enzyme. This is an enzyme utilized by the bacteriophage λ virus. The virus inserts its own genetic material into a DNA molecule using the site-specific recombination of the Int enzyme. The Int enzyme chooses a specific site on a given DNA molecule and a specific site on the circular viral DNA molecule. When the virus cuts open the two molecules and reglues the ends, the result is a new

DNA molecule containing the viral DNA (Figure 7.21). Although Figure 7.21 depicts the Int enzyme acting by simply cutting open the two molecules at the two sites, and then gluing the ends together, its action may be much more complicated. It may insert a complicated tangle into the resulting molecule.

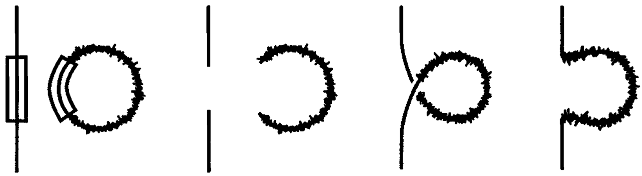
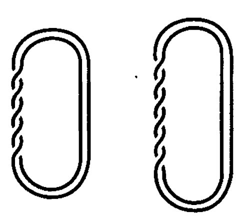


Figure 7.21 Int enzyme inserts bacteriophage λ viral DNA.

To determine how the Int enzyme acts, biochemists synthesized single DNA molecules containing both sites, some of which were relaxed and some of which were supercoiled, but all of which were topologically unknotted. When the Int enzyme was allowed to act on them, the resulting molecules were all (2, q)-torus knots or links, also known as a two-braid. (When q is odd, we get a knot, and when q is even, we get a two-component link.) In fact, for inverted repeats (Figure 7.22a) the products were always (2, q)-torus knots, where $q = 1, 3, 5, \ldots, 23$. [Note the (2, 1)-torus knot is the unknot and the (2, 3)-torus knot is the trefoil knot.] For direct repeats (Figure 7.22b), the result was always a two-component link. As an example of the kind of theorem mathematicians can prove, we have the following [see (Sumners, 1993)].



a b

Figure 7.22 Result from the action of Int for (a) inverted repeats and (b) direct repeats.

Theorem Suppose that we have two different substrate tangles S_0 and S_1 and a site tangle T such that $N(S_0 + T) =$ unknot, and $N(S_1 + T) =$ unknot. Suppose that after the enzyme Int acts on these two molecules, $N(S_0 + R)$ = unknot and $N(S_1 + R)$ = trefoil. Then T and S_0 are rational tangles (see Figure 7.23). This is surprisingly difficult to prove, using several very recent results in knot theory. It's also a little disappointing, since we would have liked the mathematics to say exactly what the tangles must be. But in fact, there is not enough empirical information to nail them down. With more experimental data, mathematicians will hopefully be able to say exactly what topological effect the Int enzyme has. Currently, biochemists believe the Int enzyme acts as in Figure 7.24.

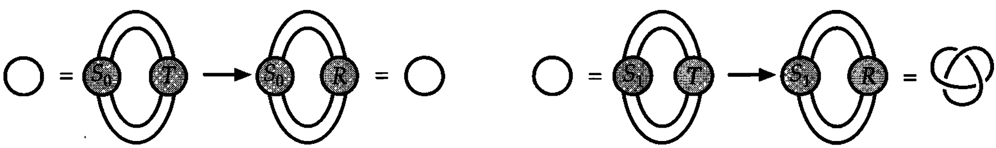


Figure 7.23 In this situation, T and S_0 are rational tangles.

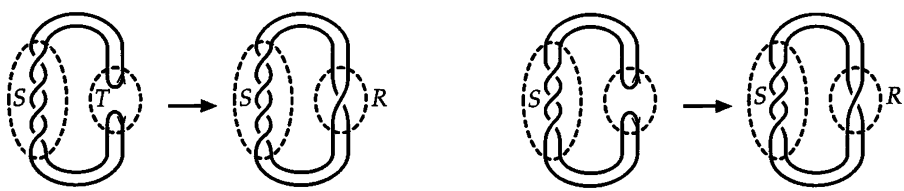


Figure 7.24 The conjectured action of the Int enzyme.

7.2 Synthesis of Knotted Molecules

It's one thing to find knots and links in DNA. DNA is a molecule built up out of millions of individual atoms, and is an extremely complicated molecule. But we might wonder if much simpler molecules can knot or link. Perhaps we could take a chain of atoms that bond together to form a circle. However, that same chain of atoms with the same bonds may in fact form a knotted chain, rather than the unknot. As a chemist, should we distinguish between these two molecules? After all, they are made up of the same set of atoms bonded together in exactly the same sequence. In fact, we do have to treat the two molecules as distinct, since it is possible that they will have different properties. One might behave like an oil, while the other behaves like a gelatin.

Actually, two molecules made up of the same set of atoms, bonded in the same way, can form distinct molecules, even if knots or links are not present. For instance, Figure 7.25 shows two molecules, each of which consists of the same four atoms bonded to the same central atom. However, one is the mirror image of the other. There is no way to rotate the first molecule through space to make it match the second.

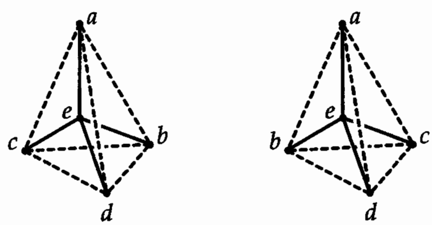


Figure 7.25 Two distinct molecules in space.

Exercise 7.2 If the four atoms H, H, C, and C are bonded to the central atom C, how many distinct atoms can be constructed? (The four outer atoms will appear as the vertices of a tetrahedron, the center point of which will be the central atom. Rotate the possible molecules through space to try to match them up with one another.)

In the example in Figure 7.25, the fact that we considered the two molecules to be different depended on our knowing what each of the individual atoms was. If we simply knew what the molecular graph looked like, we would not be able to distinguish the two molecules (Figure 7.26).

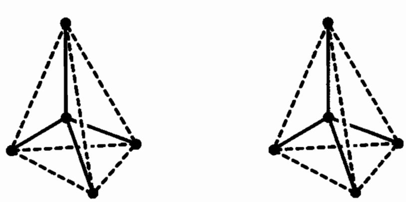


Figure 7.26 Both molecules have the same molecular graph.

We are interested in molecules that are made out of the same atoms and bonds, but that can be distinguished by their molecular graphs. As an example, Figure 7.27 shows two molecules that are both made from the same set of atoms, bonded in the same way. However, the first resembles a Möbius band ladder with four rungs and a right-hand twist, while the second resembles a Möbius band ladder with four rungs and a left-hand twist. Note that the second molecule is the mirror image of the first. This means that the two molecules have exactly the same molecular graph, only the graphs are embedded in space in two different ways. We cannot deform the first embedding of the graph to the second embedding of the graph through three-space. We say that the two molecules are homeomorphic, but they are not isotopic. We call a pair of molecules that are homeomorphic but not isotopic a pair of **topological stereoisomers**.

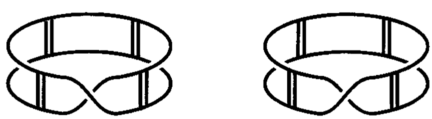


Figure 7.27 Two molecules made from the same set of atoms and bonds.

As a second example, if we have the same atoms bonded in the same sequence to form three molecules, only the first is the unknot, the second is the left-hand trefoil, and the third is the right-hand trefoil, all three of the molecules will be topological stereoisomers with each other (Figure 7.28). (In Section 6.4, we showed that the left-hand trefoil was distinct from the right-hand trefoil.)

Figure 7.28 Three topological stereoisomers.

Chemists are very interested in topological stereoisomers because they may provide substances never before seen. A good way of obtaining topological stereoisomers is to synthesize molecules that can be knotted or linked. Then the unknotted or unlinked version will be a topological stereoisomer with the knotted or linked version. In fact, as early as the first decade of the 1900s, Willstatter discussed the possibility of synthesizing a pair of linked molecular rings at a seminar in Zurich. However, it wasn't until the late 1950s that chemists began to make progress toward this goal, with five groups independently working on synthesizing linked molecular rings.

Chemists call a set of linked molecular rings a **catenane** (the Latin word *catena* means chain). The first successful synthesis of a catenane with verification thereof was accomplished by Wasserman in 1960. The idea was to use **macrocyclization**, which is the formation of cyclic molecules with at least 20 atoms. Good techniques for macrocyclization became available in the 1950s. Once a large enough cyclic molecule had been created, the goal was to thread a second linear molecule through it and then preserve the threading long enough to allow the two ends of the linear molecule to be glued together (see Figure 7.29). Since Wasserman's successful synthesis of a catenane in 1960, chemists have come up with several other techniques for synthesizing such molecules. Having successfully created nontrivial links, they then set out to create a nontrivial knot.

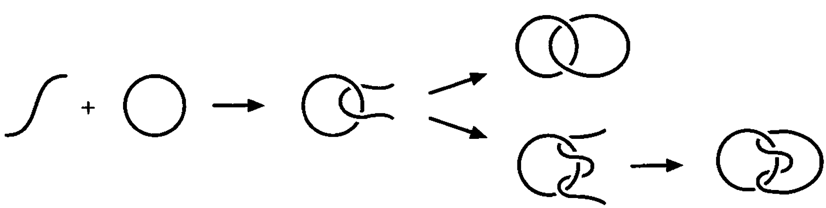


Figure 7.29 Synthesizing a link.

How does one go about trying to synthesize a knotted molecule? To make a knot out of string, we simply hold one end of the string while we tie a knot in the other end, perhaps by putting a loop in the string, and then passing the end of the string through the loop. We then glue the ends of the string together, and voila, a knot (Figure 7.30).

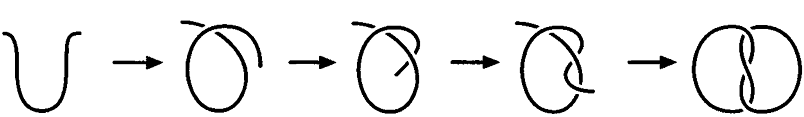


Figure 7.30 Making a knot.

Great, but how are we going to do this on the molecular level? Unlike DNA knotting, knotting in synthesized molecules is complicated by the fact that so few atoms are involved. When two atoms bond, there is not much movement possible at the joint, largely because there is a preferred bonding angle. Thus, unless a large enough number of atoms are involved, the molecular strand is too inflexible to tie in a knot. (This is related to the stick number of a knot, which we discussed in Section 1.6.) Moreover, even if the molecular strand is flexible enough, how do we get a chain of bonded atoms to form a loop and then get the end of the chain to pass through the loop?

Instead, it might be easier to have a template that holds strands of molecules in place until the knot can be formed. Then the template can be removed and a knot results. This idea had been utilized by Christina Dietrich-Buchecker and Jean-Pierre Sauvage to synthesize catenane. At the University of Strasbourg in France, they developed a technique for interlacing two molecular threads, using a central transition metal to form the template. Then, by connecting the ends of the two loops, and removing the central transition metal, they had a reliable method for producing catenane (Figure 7.31). Dietrich-Buchecker and Sauvage then realized that if they could double the template, they could in essence create three crossings, so that when the ends of the loops were connected, and the transition metals removed, a trefoil knot would result. In 1988, they announced the first successful synthesis of a knotted molecule. A schematic of the molecule appears in Figure 7.32.

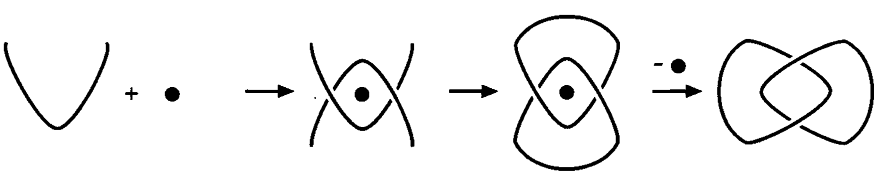


Figure 7.31 Using a template to synthesize catenane.

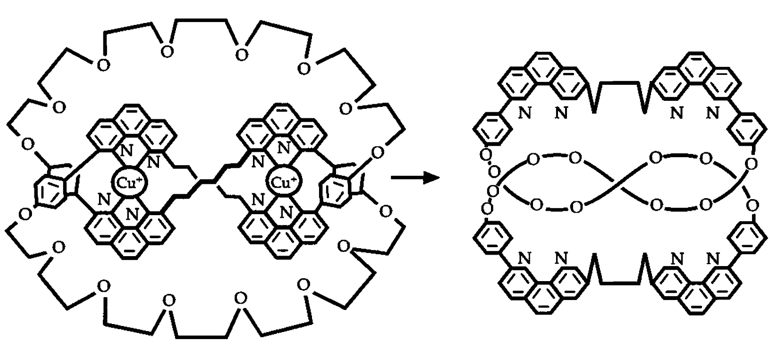


Figure 7.32 The first synthesis of a knotted molecule.

At the same time, David Walba and his colleagues at the University of Colorado were approaching the problem of the synthesis of a knotted molecule from a different direction. Independently, in the late 1950s, Wasserman and van Gulick had suggested that if one could synthesize a Möbius band ladder with extra twists and then break the rungs on the ladder, a knot or link would result. In Figure 7.33, we show the cases of 1, 2, 3, and 5 twists, yielding a trivial knot, the Hopf link, the trefoil knot, and the (5, 2)-torus knot, respectively. Unfortunately, the molecules utilized to form the Möbius band proved to be too rigid to allow the requisite number of twists needed in order to obtain a knot. However, Qun Yi Zheng, working under Walba, managed to add a clasp to the Möbius band. In the fall of 1990, Zheng announced the successful synthesis of a knotted molecule (Figure 7.34). At the time of this writing, he was still working on the determination of which knot it is.

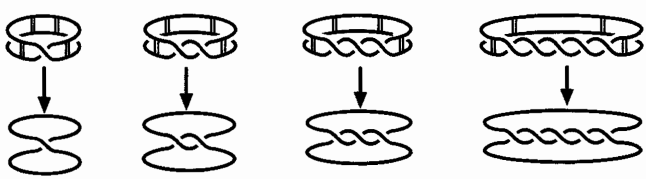


Figure 7.33 Obtaining knots from twisted Möbius ladders.

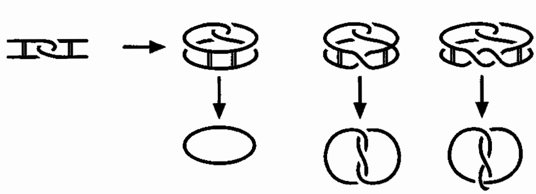


Figure 7.34 Zheng's knotted molecule.

There are numerous ways to attempt generalizations of Walba's approach. Van Gulick suggested three-strand ladders. After twisting and gluing the two ends of the ladders together, and then breaking the rungs, any of the products shown in Figure 7.35 may result. In fact, if the three strands can be made to braid, chemists might successfully synthesize any three-braid knot or link. In Figure 7.36, we show a possible scheme for synthesizing the figure-eight knot or the Borromean rings. With the addition of clasps, numerous knots and links are possible.

Figure 7.35 Twisted three-strand ladders.

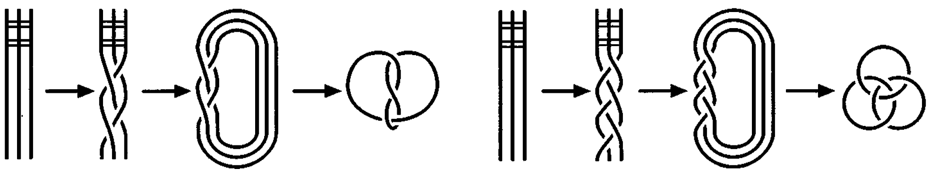
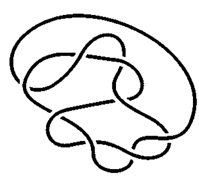


Figure 7.36 Possible methods for synthesizing the figure-eight knot or the Borromean rings.

Exercise 7.3 Utilizing three-strand rings and clasps, show how chemists might make a knot like the one below:



7.3 Chirality of Molecules

Mathematical questions abound in the theory of topological stereochemistry. As an example, we mentioned earlier that the left-hand and right-hand Möbius bands with four rungs are topological stereoisomers. This means that the left-hand Möbius band ladder with four rungs cannot be deformed through space to its mirror image, the right-hand Möbius ladder with four rungs. A molecular graph in space that cannot be deformed through space to its mirror image is called **topologically chiral**, while a molecular graph in space that can be deformed to its mirror image is called **topologically achiral**. (How to remember which is achiral and which is chiral? The best I have come up with is that achiral means "able to be deformed to its mirror image." Both "achiral" and "able" begin with an a.)

You will remember that in Section 1.3, we defined a knot or link to be amphicheiral if it could be deformed to its mirror image. Hence, for knots, amphicheiral and topologically achiral mean the same thing. Note that we are ignoring such real properties of a molecule as the bond angles and the bond length. Even if a given molecule can be topologically deformed to its mirror image, it may not be possible to deform the actual molecule through space to its mirror image, since the rigidity of the bonds won't allow it. A given molecule may be topologically achiral but not "chemically achiral." However, a topologically chiral molecule *must* be chemically chiral.

In 1986, Jonathan Simon, a mathematician at the University of Iowa, proved that a Möbius ladder with four or more rungs is always topologically chiral. This is a mathematical result that has chemical consequences. It says that *any* molecule with a molecular graph in the form of a Möbius ladder with four or more rungs has a topological stereoisomer, namely its mirror image. This is true even if the edges of the ladder are exactly the same as the rungs. Chemists are currently attempting to synthesize Möbius ladders with indistinguishable rungs and edges (Figure 7.37).

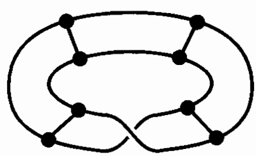


Figure 7.37 A Möbius ladder with indistinguishable rungs and edges.

What about a Möbius band ladder with three rungs? This pair of molecules was synthesized by Walba, Richards, and Haltiwanger in 1982 (Figure 7.38). Are they topological stereoisomers? Not quite. Here is a nice figure due to John Simon that demonstrates that if we do not distinguish between the rungs and the edges, the first embedding of the graph can be deformed to the second embedding of the graph (Figure 7.39). Hence this graph is topologically achiral. However, if we do distinguish between rungs and edges, say by coloring all rungs red and edges blue, there is no deformation taking the first graph to the second, so that red rungs go to red rungs and blue edges go to blue edges. This fact is also due to Simon, and appears in the same paper.

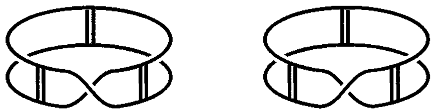


Figure 7.38 Möbius band ladders with three rungs.

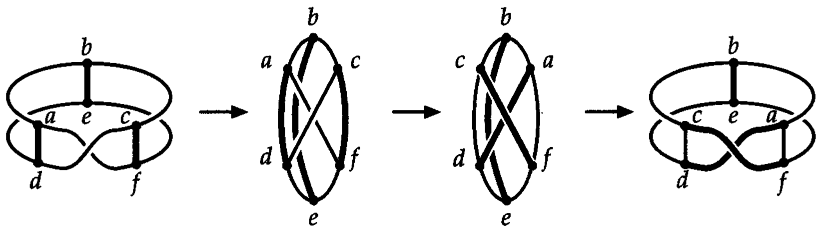


Figure 7.39 A three-rung Möbius ladder. (From John Simon, 1986.)

Since chemists would like to find interesting pairs of topological stereoisomers, they would like to know which knots are chiral and which are achiral. The first knot that was shown to be chiral was the trefoil knot. (See Section 6.4 for a proof.) As we mentioned in Section 6.4, the following conjecture remains open.

Unsolved Conjecture of Tait (1890)

Any knot whose minimum crossing number is odd must be topologically chiral.

If knotted molecules can be synthesized, what about molecules in the form of various graphs? Every molecule corresponds to some graph, but usually the graph is relatively simple. In Figure 7.40, we see several molecules with graphs that are easily identified. They are not the kind of graphs that mathematicians can get excited about. In fact, all of the graphs shown are planar graphs, meaning they can be drawn in the plane with no crossings. They are essentially flat.

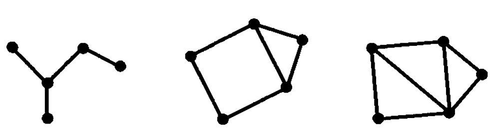


Figure 7.40 Molecules in the form of graphs.

What about finding molecular graphs that are not flat? Certainly, one place to look is at metals. Metals will form huge three-dimensional lattices, each atom of which will be bonded to all of its nearest neighbors. This is in fact why metals have the properties that they do.

But excluding metals, can we synthesize nonplanar molecular graphs? In 1981, Howard Simmons III and Leo Paquette independently managed to synthesize molecules in the form of the complete graph on five vertices. Denoted K_5, this is the graph that has five vertices, where each vertex is connected to every other vertex by an edge. This graph is nonplanar (Figure 7.41). That is to say, there is no way to embed this graph in the plane.

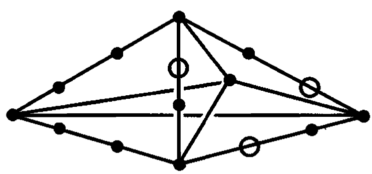


Figure 7.41 The graph K_5 is realized by a molecule.

Exercise 7.4 Prove that K_5 is nonplanar. [*Hint:* Suppose that there was an embedding of K_5 in the plane. Use the fact that every circle in the plane divides the plane into two regions (something that is quite difficult to prove, but that we will accept as true) and the fact that you can't get from one region to the other without crossing the circle.]

Exercise 7.5 Prove that $K_{3,3}$ is also nonplanar. ($K_{3,3}$ is the graph with six vertices, where we have separated the vertices into two subsets, each of size three. Every vertex in one subset is connected by edges to all of the vertices in the other subset but to none of the vertices in its own subset.)

If all of the edges were the same in a K_5 molecule, the molecule would be topologically achiral. However, the Simmons–Paquette K_5 molecule has three types of edges: C–C single bonds, $-Ch_2Ch_2-$ chains, and $-Ch_2O-$ chains.

Unsolved Question

Prove that if the three types of edges are distinguished this molecule is topologically chiral.

7.4 Statistical Mechanics and Knots

Until very recently, statistical mechanics and knot theory had nothing to do with one another. However, in the process of discovering the new polynomial invariants for knots, Vaughan Jones also established a connection between these two fields. It is currently an area of extremely active research.

Let's start with a little statistical mechanics. In statistical mechanics, we are dealing with large systems of particles. Instead of keeping track of the characteristics of each particle separately, we keep track of the aggregate behavior. For instance, we might measure the average energy of the system (known as the temperature). We are only interested in those quantities that do not depend on the number of particles, given that enough particles are present. For example, cutting an ice cube in half will not change the temperature of the two resulting pieces. The temperature isn't dependent on the number of particles, assuming there are enough of them.

However, even when we only consider the average behavior of the system, strange effects can occur. One example is a phase transition, where a system of particles transforms from a gas to a liquid, a liquid to a solid, or vice versa. Such a transition does not occur for just one molecule at a time, but instead occurs for the whole system over a very short period of time. Suddenly, when the appropriate temperature is reached, the liquid freezes.

A second example is magnetization, where a bar of metal can be held in a magnetic field and the magnetic axes of all of the molecules line up, resulting in the magnetization of the bar. Even when the surrounding magnetic field is turned off, the bar remains magnetized. Reversing the surrounding magnetic field results in flipping the axes of all of the molecules. The reversal of all the axes occurs almost simultaneously.

Mathematically modeling such systems has been one of the most difficult problems in physics. We discuss a particular model known as the **Ising model**, developed by E. Ising in 1925. It works well when modeling a system where only particles near one another interact. Two particles that are not neighbors have no effect on one another. Let's look at the model as it applies to the magnetization of a metal. We consider each molecule of the metal to be a vertex of a graph. The edges of the graph denote the interactions between adjacent molecules. Only two molecules connected by an edge can interact.

A particular type of graph that we will look at is called a **lattice**, where the vertices and edges form a regular repeating pattern in space. In fact, metals consist of molecules that are at the vertices of a lattice in three-dimensional space (Figure 7.42), and therefore lattices are relevant to the real world. The square lattice in the plane is a particularly simple example of a lattice (Figure 7.43). Each vertex has four nearest neighbors with which it interacts. All of the other particles are too far away to affect it.

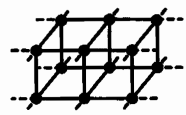


Figure 7.42 A metal.

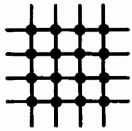


Figure 7.43 The planar square lattice.

Of course, a planar square lattice doesn't seem like a particularly good model for a substance that is in a liquid, gas, or solid form, since the molecules of the substance will occur in three-dimensional space rather than in the plane. The three-dimensional versions of this model, however, have so far proved too complex to solve. The two-dimensional models have been successfully solved, and demonstrate the hoped for behaviors, such as phase transitions.

In the Ising model, each particle can be in one of two different states, which we denote with a $+1$ or -1 at the appropriate vertex. In the example of magnetization, the $+1$ state corresponds to when the magnetic axis (often called the **spin vector**) points up and the -1 state corresponds to when the spin vector points down. Figure 7.44 depicts a particular state for a 3×3 square lattice.

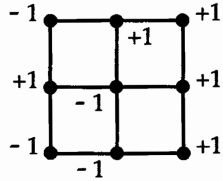


Figure 7.44 A particular state of a 3×3 square lattice.

If we have a finite set of particles, we can number them $1, 2, \ldots, n$ and then write the state of particle i as s_i. A choice of state for each particle in the system gives us a state S for the whole system, which we write as $S =$

$(s_1, s_2, \ldots, s_n)$, listing the states for each of the particles in order. Each edge of the graph has an energy associated to it, corresponding to the energy of interaction between the particles at its ends. This energy of interaction depends on the states of the two particles at the ends of the edge, and so we will write it as $E(s_i, s_j)$. In the Ising model, there are two possibilities for this energy. When $s_i = s_j$, we have one energy of interaction, and when $s_i \neq s_j$, we have another.

$$E_= = E(+1, +1) = E(-1, -1)$$

and

$$E_{\neq} = E(+1, -1) = E(-1, +1)$$

We do not specify particular values for $E_=$ and $E_{\neq}$ because the values depend on the particular system that we are modeling. For each edge, it will be handy to define a term $w(s_i, s_j)$, which is given by

$$w(si, sj) = \exp\left(\frac{-E(s_i, s_j)}{kT}\right)$$

(Note: exp (z) denotes e^z.) The value k is the so-called Boltzmann constant, which in case you were wondering is 1.38×10^{-23} joules/degrees Kelvin: The variable T is simply the temperature of the system, given in degrees Kelvin. Notice that in this model, w also takes on one of two values, depending on whether s_i and s_j agree or disagree. We denote these two possibilities by $w_=$ and $w_{\neq}$.

The energy of the entire system of particles in a particular state can be calculated as the sum of the energies of all the edges. We write this as

$$E(S) = \Sigma E(s_i, s_j)$$

We are interested in a function $Q(S)$ that depends on the energy of the system when it is in state S. This function is given by

$$Q(S) = \exp\left(\frac{-E(S)}{kT}\right)$$

Then, we see that

$$Q(S) = \exp\left(\frac{-E(S)}{kT}\right) = \exp\left(\frac{-\Sigma E(s_i, s_j)}{kT}\right) = \Pi\exp\left(\frac{-E(s_i, s_j)}{kT}\right) = \Pi w(s_i, s_j)$$

(Note: Πz_i denotes the product $z_1 z_2 \ldots z_n$.)

We then define a function called the **partition function** denoted by P, which is equal to the sum over all possible states of the system of these terms. We write this as:

$$P = \Sigma_s Q(S) = \Sigma_s \exp\left(\frac{-E(S)}{kT}\right) = \Sigma_s \Pi\, w(s_i, s_j)$$

We will primarily be working with P in this last form.

The partition function is a useful quantity to determine. From it, we can calculate the value of any observable property of the system. In particular, we can calculate the probability that the system of particles is in a particular state S_0 as

$$\frac{\exp(-E(S_0)/kT)}{P}$$

Exercise 7.6 Determine the partition function of the Ising model corresponding to a 2×2 planar lattice, also known as a square. Note that there are 2^4 possible states of the system, generating 16 terms in the partition function. However, many of them are the same. Write the resulting function in terms of $w_=$ and $w_{\neq}$.

Although up to now we have been looking at graphs coming from lattices, we will not restrict ourselves to these graphs any longer. The one restriction that we will retain is that the graphs that we look at should be planar graphs, that is to say, graphs that lie in the plane with no edges crossing over one another. In general, if the number of particles is at all large, computing the partition function becomes extremely difficult. For a graph with n vertices, we must sum over the 2^n possible states of the system (since each of the n particles has two possible states, and therefore the number of states of the set of particles is the product of these n twos). For instance, 56 particles would mean 2^{56} possible states of the system, which is approximately 7.2×10^{16}. If our computer can calculate the terms of the partition function for one million states per second, then it will only take our computer 2283 years to come up with the whole partition function. Hence, clever means are needed to avoid having to do such a calculation. That brings us to the so-called **Yang–Baxter equation**, also known as the **star–triangle relation**.

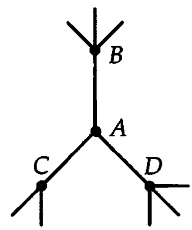


Figure 7.45 A star in the graph.

Suppose we have a so-called star in our graph, that is to say a vertex with three edges coming out (Figure 7.45). Let's label the vertex A. Let B, C, and D be the three vertices that are attached to A. Then the term of the partition function $Q(S)$ corresponding to a particular state S can be written as the product of the terms for the edges incident to A times the product of the terms for all the other edges. Hence, this term becomes

$$Q(S) = \Pi w(s_i, s_j) = w(s_A, s_B)\, w(s_A, s_C)\, w(s_A, s_D)\, \Pi w(s_i, s_j)$$

In the second product $\Pi w(s_i, s_j)$ above, and in all the following places that it occurs, we are taking the product over all the terms corresponding to edges other than the three edges between A and B, A and C, and A and D. Let S be a particular state of the system such that $s_A = 1$. Let S' be a state that is identical to S except that $s'_A = -1$. Then, adding together the two terms in the partition function for these two states, we have

$$\begin{aligned} Q(S) + Q(S') &= w(+1, s_B)\, w(+1, s_C)\, w(+1, s_D)\, \Pi w(s_i, s_j) + \\ &\qquad w(-1, s_B)\, w(-1, s_C)\, w(1, s_D)\, \Pi w(s_i, s_j) \\ &= (w(+1, s_B)\, w(+1, s_C)\, w(+1, s_D) + w(-1, s_B)\, w(-1, s_C)\, w(-1, s_D))\, \Pi w(s_i, s_j) \end{aligned}$$

Since we can do this for any pair of states that differ only in their values at A, and since every state has a corresponding state that does differ from it only in its value at A, we can write the entire partition function as

$$\begin{aligned} P = \Sigma_S(&w(+1, s_B)\, w(+1, s_C)\, w(+1, s_D) + \\ &w(-1, s_B)\, w(-1, s_C)\, w(-1, s_D))\, \Pi w(s_i, s_j) \end{aligned} \quad (7.1)$$

where now we sum over states S of the system of particles excluding A.

We would like to exchange the term

$$w(+1, s_B)\, w(+1, s_C)\, w(+1, s_D) + w(-1, s_B)\, w(-1, s_C)\, w(-1, s_D)$$

for a term that depends on the energies of edges between B, C, and D. In this way, we will have replaced the star that was centered at A by the triangle through B, C, and D (Figure 7.46). In general, we would not be able to do this, except for the fact that the energies of interaction along the three new edges need not be the same energies of interaction along the rest of the edges in the graph.

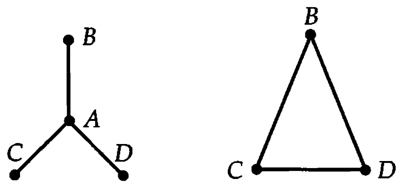


Figure 7.46 The star is replaced by the triangle.

Let $w'(s_i, s_j)$ denote the new energies of interaction along the three edges between B, C, and D. If we can find values for $w'(s_B, s_C)$, $w'(s_C, s_D)$, and $w'(s_D, s_B)$ such that the following equation holds, then we will be able to replace the star with a triangle.

$$w(1, s_B)\, w(1, s_C)\, w(1, s_D) + w(-1, s_B)\, w(-1, s_C)\, w(-1, s_D) = \\ w'(s_B, s_C)\, w'(s_C, s_D), w'(s_D, s_B) \quad (7.2)$$

For example, when s_B, s_C, and s_D are all equal to $+1$, we obtain the equation

$$w_=^3 + w_{\neq}^3 = (w'_=)^3$$

Exercise 7.7 Show that as we plug in various choices of ± 1 for s_B, s_C, and s_D in Equation 7.2, two distinct equations are generated. Solve these two equations to find R, $w'_=$, and $w'_{\neq}$ in terms of $w_=$ and $w_{\neq}$.

Once we know R, $w'_=$, and $w'_{\neq}$, we can replace the partition function (7.1) by the equation

$$P = \Sigma w'(s_B, s_C)\, w'(s_C, s_D)\, w'(s_D, s_B)\, \Pi w(s_i, s_j)$$

This is exactly the partition function of the original graph with the star replaced by the triangle, only taking the interaction energies along the three edges of the triangle to correspond to w' instead of w.

So this is the star–triangle relation, also known as the Yang–Baxter equation. Instead of calculating the partition function of a graph containing a star, we replace the star with the corresponding triangle, and calculate the partition function for a graph with one less vertex. Remember, one less vertex will halve the total number of terms in the partition function, making our job of determining the partition function a lot easier. Great, but what does any of this have to do with knot theory?

In Section 2.3, we demonstrated how to turn a knot projection into a signed planar graph. In particular, the goal of Exercise 2.24 was to see what happened to the Reidemeister moves under this transformation. In Figure 7.47, we see that a Type III Reidemeister move on a knot projection becomes a star–triangle exchange on the corresponding signed planar graphs. Note that we were originally looking at partition functions of planar graphs without signs on the edges. However, we can define a partition function for a signed planar graph. Instead of having a single energy of interaction $E(s_i, s_j)$ and the corresponding term $w(s_i, s_j)$, we have two types of energies of interactions and two types of terms $w_+(s_i, s_j)$ and $w_-(s_i, s_j)$.

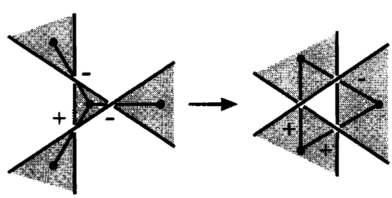


Figure 7.47 Type III Reidemeister move becomes star–triangle exchange.

Each of w_+ and w_- takes on two possible values depending on whether $s_i = s_j$ or $s_i \neq s_j$. Given an appropriate choice for these values, the partition function of a signed planar graph will also satisfy the star–triangle relation, thereby making it an invariant of the Type III Reidemeister move. In fact, given the extra freedom of assigning four different values rather than just two, we no longer need to assume that the values for the w's along the three new edges are distinct from the values of the w's corresponding to the original graph.

If we can make a choice of interaction energies on the edges so that the partition function also satisfies relations corresponding to the Type I and Type II Reidemeister moves, then the partition function becomes an invariant for knots and links. This forms the basis for the connection between knot theory and statistical mechanics.

In order that the partition function remain invariant under the Type I and Type II Reidemeister moves, certain conditions must be satisfied by w_+ and w_-. For instance, a Type II move on a knot projection can translate into either of the following changes on the signed planar graph.

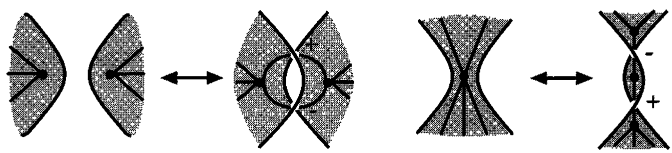


Figure 7.48 Translating Type II moves to the signed planar graphs.

Exercise 7.8 Show that in order for the partition function to satisfy the first translation of a Type II Reidemeister move as above, w_+ and w_- must satisfy the equation

$$w_+(a, b)\, w_-(a, b) = 1 \tag{7.3}$$

for each of the possible values of a and b. (In particular, note that once we know w_+, w_- is completely determined.)

In order that the partition function be left invariant under the second of these translations, we also need it to be true that

$$w_-(a, 1)\, w_+(1, b) + w_-(a, -1)\, w_+(-1, b) = 2\,\delta(a, b) \tag{7.4}$$

for $a = \pm 1$ and $b = \pm 1$, where

$$\delta(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$

is the so-called **Kronecker delta function** (also sometimes called the *Dirac delta function*).

Exercise 7.9 Show that a partition function corresponding to values for w_+ and w_- that satisfy Equation 7.4 will be left invariant by the second translation of a Type II Reidemeister move.

Exercise 7.10 Show that if

$$w_+(a, b) = \begin{cases} 1 & \text{if } a = b \\ i & \text{if } a \neq b \end{cases}$$

and if

$$w_-(a, b) = \begin{cases} 1 & \text{if } a = b \\ -i & \text{if } a \neq b \end{cases}$$

then both Equations 7.3 and 7.4 are satisfied.

Similarly, the translations of the Type III moves generate equations for w_+ and w_- that can be satisfied by the appropriate choice of their values. For instance, one of the Type III Reidemeister moves translates into

$$w_+(1, s_B)\, w_+(1, s_C)\, w_-(1, s_D) + w_+(-1, s_B)\, w_+(-1, s_C)\, w_-(-1, s_D) = \sqrt{2}w_+(s_B, s_C)\, w_-(s_C, s_D)\, w_-(s_D, s_B) \quad (7.5)$$

Exercise 7.11 Show that the values of w_+ and w_- that were given in Exercise 7.10 also satisfy Equation 7.5.

As with the case of the bracket polynomial, the Type I Reidemeister move does cause the partition function to vary. However, just as we did for the bracket polynomial, we can place a factor in front of the partition function that accommodates this variation and causes the resultant "partition function" to be invariant under all three Reidemeister moves. Thus, we obtain a partition function that is an invariant for the corresponding knot or link.

In the case of the Ising model, the resulting partition function yields a knot invariant known as the Arf invariant. We use the Arf invariant in Section 8.2, although we develop it from a completely different point of view. In Section 8.3, we generalize the Ising model to the **Potts model**, where we allow q different states at each vertex of the graph. With appropriate choices for the values of w_+ and w_-, the partition function generates a knot invariant, which turns out to be the Jones polynomial $V(t)$, where q and t are related by the equation $q = 2 + t + t^{-1}$.

8 Knots, Links, and Graphs

8.1 Links in Graphs

Graph theory is an area of mathematics that traditionally had very little to do with knots or links. But we will look at a recent area of research that ties the two together.

As we saw in Chapter 2, a graph consists of a set of vertices and a set of edges that connect the vertices. Figure 8.1 contains some pictures of graphs. A graph is simply defined by the number of vertices it has and by which vertices are connected by edges. So, the two graphs in Figure 8.2 are considered to be equivalent graphs even though they sit in space in different ways. We say that the two graphs are **isomorphic**, or that they have the same **isomorphism type**. Sometimes we will talk about an abstract graph, meaning the isomorphism type of the graph, rather than a particular way of realizing the graph in space.

The graph K_6, called the **complete graph on six vertices**, is the graph where every one of the six vertices is connected to every other one by exactly one edge. Figure 8.2 shows two different ways to place K_6 in space.

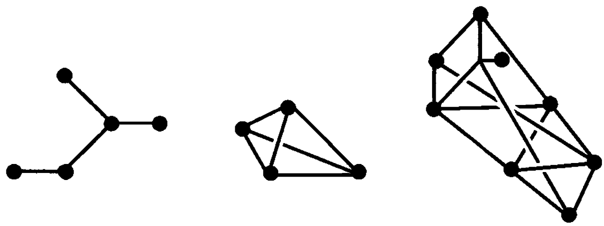


Figure 8.1 Graphs.

Although these two graphs are isomorphic, they are not isotopic, since there is no way to deform one of them through space to look like the other, without allowing edges to pass through themselves or each other. Just as we did for surfaces in Chapter 4, we call a particular way to place K_6 in space an **embedding** of K_6. Figure 8.3 shows a much nastier embedding of K_6.

Figure 8.2 Two ways to place K_6 in space.

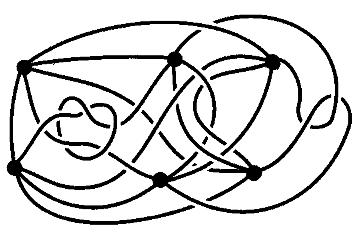


Figure 8.3 A nasty embedding of K_6.

Let's call a **triangle** in an embedding of K_6 any set of three consecutive edges that form a triangle in the graph. Notice that if we choose any three vertices, we can form a triangle from the edges connecting them. We can also form a second triangle from the remaining three vertices (Figure 8.4). Thinking of this pair of triangles as two components of a link, we are interested in whether they are linked or not in the embedding (Figure 8.5).

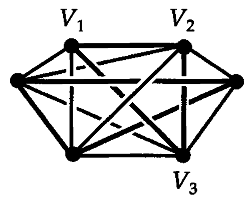
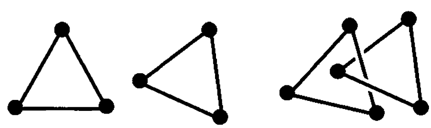


Figure 8.4 A pair of disjoint triangles defined by three vertices.



a b

Figure 8.5 (a) Unlinked triangles. (b) Linked triangles.

In 1983, John H. Conway (the same Conway who catalogued knots) and Cameron Gordon (who cosolved one of the oldest problems in knot theory; see Section 9.3) published a paper entitled, "Knots and Links in Spatial Graphs" (Conway and Gordon, 1983). In that paper, they proved the following theorem.

Theorem Every embedding of K_6 contains at least one pair of linked triangles.

This is an amazing fact. No matter how we place K_6 in space, there will always be a link contained within it. Even if we change the embedding by letting one edge pass through another specifically in order to destroy a link in the original embedding, we can't help but either create a new link in the process or at least leave another link in the embedding. In particular, all three of the pictures in Figures 8.2 and 8.3 contain nontrivial links. Figure 8.6 shows a nontrivial link in the first picture.

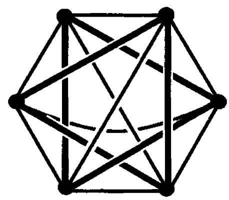


Figure 8.6 A nontrivial link in the first embedding.

Exercise 8.1 Find a nontrivial link in the second embedding. How about the third?

In fact, there may be more than one pair of linked triangles in an embedding, but there is always at least one by the theorem. Let's go through the proof of the theorem. It's surprisingly easy. Each choice of three vertices gives us a pair of disjoint triangles, one that passes through the three vertices that we chose and one that passes through the other three vertices. How many pairs of disjoint triangles are there in K_6?

Well, first let's see how many ways there are to pick a set of three vertices from a set of six. We have six choices for the first vertex, then five remaining choices for the second vertex, and finally four remaining choices for the last vertex. Thus, there are a total of 120 choices here. But we don't care about the order in which the vertices were picked. That is, if 1, 3, and then 2 were picked, that gives the same triangle as if 2, 1, and then 3 were picked. Therefore, since there are six different ways we could order 1, 2, and 3, we have to divide our total number of choices by 6. This gives 20 different ways to pick three vertices where the order of the vertices doesn't matter. But in fact, if {1, 2, 3} and {4, 5, 6} were two different choices for our set of three vertices, they would both yield the same pair of triangles, namely the triangle through 1, 2, and 3 paired with the triangle through 4, 5, and 6. Thus, we have to divide our total of 20 possibilities by 2, leaving us with a total of 10 pairs of disjoint triangles.

We need a way to distinguish embeddings, a so-called **invariant for the embeddings**. Suppose we have a particular embedding of K_6. Each pair of disjoint triangles in the embedding has a linking number once we orient the two triangles (see Section 1.4). But notice that changing an orientation on one of the triangles only changes the sign of the linking number, not the absolute value of the linking number. Since we don't want to bother with orientations, we just look at the absolute values of the linking number for each pair of disjoint triangles in the embedding.

Given an embedding of K_6, let's define $\boldsymbol{U}$ to be the positive integer obtained by taking the sum of the absolute values of the linking numbers for all 10 pairs of disjoint triangles in the embedding. In the first picture in

Figure 8.1, only one of the 10 pairs of disjoint triangles is linked, and the absolute value of the linking number for that pair of triangles is equal to 1. So, for that embedding, $U = 1$.

Exercise 8.2 Calculate U for the second embedding of K_6 in Figure 8.2.

If we deform our embedding around in space without letting any edges pass through each other, or as mathematicians put it, if we isotope the embedding through space, then the linking numbers of all of the triangles remain the same. Hence U remains unchanged. But we want to understand how this number U changes as we go from one embedding of K_6 to a different embedding. What's the best way to think about changing embeddings? If we allow edges to pass through one another, and treat all edges as if they were made of rubber, and can be deformed accordingly, it shouldn't be any problem to get from one embedding to any other. For instance, we can get from the one embedding at the left in Figure 8.7 to the one at the right by the sequence of crossing changes and isotopies depicted.

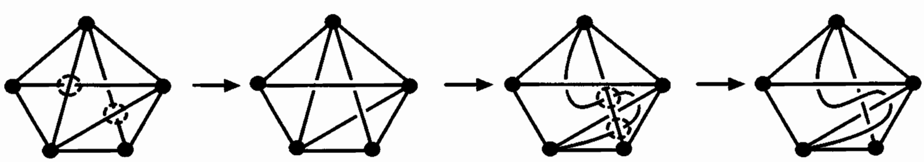


Figure 8.7 Changing one projection into another.

Keep in mind that we are dramatically changing the structure of our object when we allow edges to pass through one another. For instance, if we took the graph K_3, which is simply a triangle, and embedded it in space in all the different possible ways, we would simply be looking at all the possible knots in space. But once we allow edges to pass through one another, every one of those knots could be changed into the unknot. Therefore, all of the techniques of knot theory are useless once we decide to let edges pass through one another.

Let's define a new number V to be equal to zero if U is even and one if U is odd. We call V the **mod 2 reduction of U** since V is simply the remainder when we divide U by 2. We show that even though changing embeddings destroys the basic properties of how our graph sits in space, it does not change V. The argument is relatively simple. We already know that isotoping the embedding doesn't affect U, and hence can't affect V. So all we have to check is that crossing changes do not affect V.

Suppose we do change a crossing. If that crossing is between an edge and itself, the crossing cannot be between two disjoint triangles since two disjoint triangles cannot share the same edge. Hence U is unaffected by the crossing change. In fact, if the crossing is between two edges coming out of the same vertex, then the two edges cannot be on disjoint triangles and the crossing change again leaves the number U unchanged. Therefore, the only time U and V might be affected is if we change a crossing between two nonadjacent edges E_1 and E_2. Any pair of disjoint triangles such that one of the triangles contains E_1 and the other contains E_2 will have their linking number changed by ± 1 when we change this crossing. But together, the two edges E_1 and E_2 end at four of the vertices in the graph. There are only two vertices v_1 and v_2 that they don't intersect (Figure 8.8). If we pair v_1 with E_1 and v_2 with E_2, we obtain a pair of disjoint triangles. If we pair v_2 with E_1 and v_1 with E_2, we obtain a second pair of disjoint triangles. These two pairs of disjoint triangles are the only pairs that pass through the two edges E_1 and E_2. Since changing the crossing between E_1 and E_2 changes each of the linking numbers for these two pairs of disjoint triangles by either $+1$ or -1, the crossing change alters U by either -2, 0 or $+2$. Most importantly, U is changed by an even number. Adding or subtracting an even number from U will leave V unchanged.

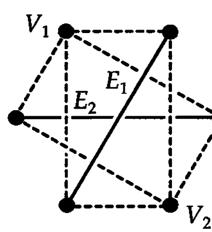


Figure 8.8 Two nonadjacent edges in K_6.

Thus, as we go from one embedding of K_6 to another by changing crossings, V remains unaffected. But we have already seen $U = 1$, and hence $V = 1$ in the first embedding of K_6 in Figure 8.2. Therefore, $V = 1$ in every projection of K_6. In particular, if $V = 1$, U can never be zero. So, in every projection of K_6 there is at least one pair of disjoint triangles with a nontrivial linking number. That is to say, every projection of K_6 contains a nontrivial link.

Note that any graph containing K_6 as a subgraph will also contain a link in any embedding of it into three-space. We say that a graph is **intrinsically linked** if it has the property that any embedding of it in three-space contains a nontrivial link.

*Exercise 8.3** The graph $K_{3,3,1}$ is the graph given by taking three sets of vertices, the first set having three vertices, the second set having three vertices, and the third set having one vertex. All of the vertices in any one of the sets are connected by edges to all of the vertices in the other two sets but to none of the other vertices in their own set. Figure 8.9 shows a particular embedding of this graph. Prove that $K_{3,3,1}$ is intrinsically linked.

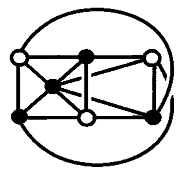


Figure 8.9 An embedding of $K_{3,3,1}$.

Exercise 8.4 Define an **expansion** of a graph G to be a new graph obtained from G by "splitting a vertex of G." By this, we mean replacing a particular vertex v of G by two vertices u and w connected by a new edge, and replacing each of the old edges that ended at v by a new edge that begins where the old edge began and ends at either u or w. A picture of an expansion appears in Figure 8.10. Notice that there are lots of choices for expansions, even if we have already chosen the vertex to expand. Prove that if G is intrinsically linked, so is any expansion of G.

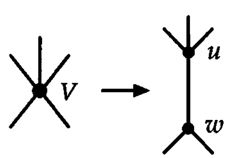


Figure 8.10 An expansion of the graph G.

Very recently, three mathematicans, Neil Robertson (at Ohio State University), P. D. Seymour (at Bellcore), and Robin Thomas (at Georgia Tech), proved that a graph is intrinsically linked if and only if it contains one of seven special graphs called the *Petersen graphs* (Figure 8.11), or an expansion of one of them (Robertson et al., 1993). The Petersen graphs are exactly the graphs obtainable from K_6 by repeated triangle-Y exchanges, where three edges that form a triangle in a graph are replaced by three edges and a new vertex that form a Y. We also allow the reverse operation, replacing a Y with a triangle, which incidentally is the star-triangle relation we saw in Section 7.4.

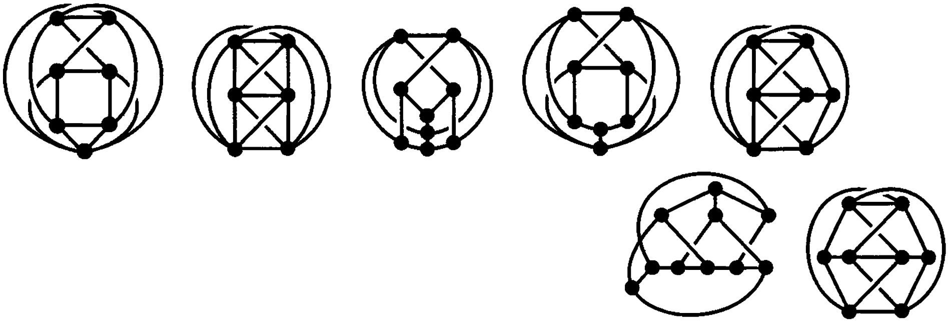


Figure 8.11 The Petersen graphs.

Exercise 8.5 Show that $K_{3,3,1}$ is a Petersen graph by demonstrating that it can be obtained from K_6 by triangle-Y exchanges. Determine which of the graphs in Figure 8.11 represents $K_{3,3,1}$.

8.2 Knots in Graphs

In the previous section, we showed that the complete graph on six vertices always contains a pair of linked triangles, no matter how we embed the graph in space. We could also ask if there are graphs such that they always contain a knot, no matter how we embed them in space. But first, we need to decide how we want the knot to sit in the graph.

A **Hamiltonian cycle** in a graph is a sequence of edges in the graph such that any two consecutive edges share a vertex, the last edge and the first edge share a vertex, and every vertex is hit by a pair of consecutive edges exactly once. Together the edges in the Hamiltonian cycle make up a loop in the graph that hits every vertex exactly once (Figure 8.12). Such a loop may be either knotted or unknotted. In the same paper in which they proved K_6 is intrinsically linked, Gordon and Conway (1983) also proved that if the graph K_7 is embedded in space in any manner whatsoever, it will always contain a Hamiltonian cycle that is knotted (Figure 8.13).

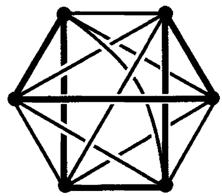


Figure 8.12 A Hamiltonian cycle in K_6.

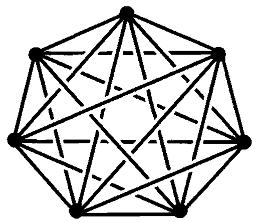


Figure 8.13 An embedding of K_7.

Exercise 8.6 Find a knotted Hamiltonian cycle in this embedding of K_7.

Exercise 8.7 Find an embedding of K_7 containing no trefoil knots. (*Hint:* Use the fact that the trefoil knot is prime and make sure all of the knots in your embedding are composite.)

We are not able to go through the entire proof that an embedding of K_7 always contains a knotted Hamiltonian cycle, as it is a little too time-consuming. However, the original paper is readable and it is a good place to obtain more information. Instead, we talk a bit about the idea of the proof.

First, we need to look at a new invariant for knots and links called the **Arf invariant**. Like the variable V we defined in the last section, the Arf invariant will always have a value of 0 or 1. There are several ways to define the Arf invariant. We take a point of view due to Louis Kauffman (1983). Let's define a new type of move on an oriented knot or link, namely, let's define a **pass-move** to be a change in a projection as in Figure 8.14. A pair of oppositely oriented strands can be passed through another pair of oppositely oriented strands. Such a move certainly can change the knot we are dealing with. We call two knots **pass equivalent** if there exists a sequence of pass-moves that takes us from the one knot to the other, where we can rearrange the projection of the knot any way that we want after each pass-move.

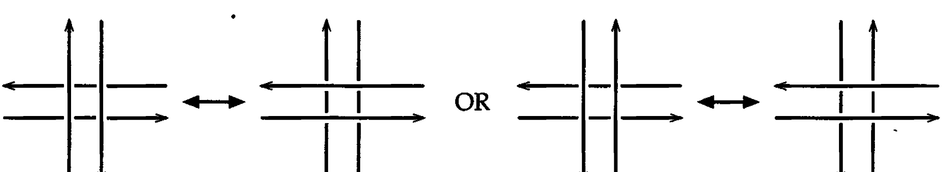


Figure 8.14 A pass-move.

Exercise 8.8 Show that a knot with part of its projection as in Figure 8.15a is pass equivalent to a knot with that part of its projection as in Figure 8.15b. (A belt may help you see this, where the edges of the belt correspond to the strands of the knot.)

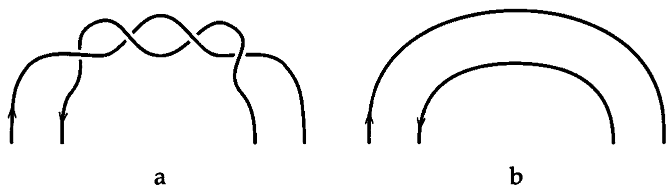


Figure 8.15 These two knots are pass equivalent.

Here is the amazing part. *Every knot is either pass equivalent to the unknot or to the trefoil knot.* Let's spend some time showing that this is the case. We utilize the Seifert surfaces that we discussed in Chapter 4 to show this. Given a knot, choose a projection of the knot and then apply Seifert's algorithm to obtain a Seifert surface for the knot. Remember that the resulting surface is orientable with a single boundary component, such that the boundary is knotted into the knot in question.

We show that a Seifert surface can be deformed through space so that it appears as a single disk with bands attached. For instance, in Figure 8.16 we see the Seifert surface for a projection of the figure-eight knot and the deformed Seifert surface that is isotopic to the original. More generally, since Seifert's algorithm always produces a surface obtained by connecting a set of disks by twisted bands, we can always choose a sequence of the connecting bands to untwist and widen, so that the set of disks becomes a single disk with a set of bands attached (Figure 8.17). Note that none of the resulting bands has an odd number of twists in it, since the Seifert surface is always orientable.

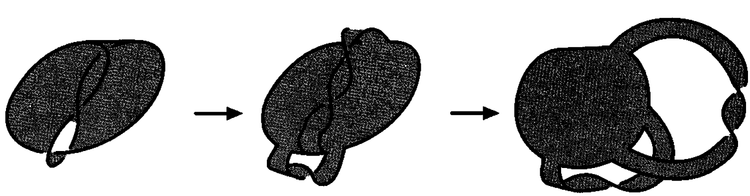


Figure 8.16 The figure-eight Seifert surface is a disk with bands.

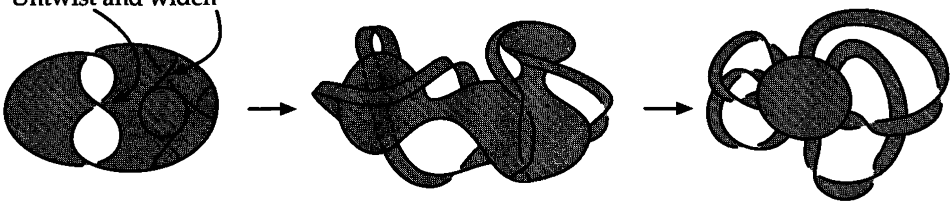


Figure 8.17 Every Seifert surface is a disk with bands.

Exercise 8.9 Show that if we put an orientation on the boundary of the Seifert surface, then the two edges of each band are always oriented oppositely.

Therefore, if we pass one band through another, we are simply doing a pass-move on the knot. Not only do we get a new knot that is pass equivalent to the old one, we also get a Seifert surface for the new knot (Figure 8.18). Notice that by utilizing pass-moves in the way illustrated in Figure 8.19, we can unhook bands that are linked with one another.

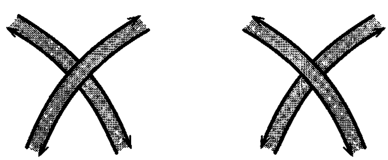


Figure 8.18 Passing bands through one another is a pass-move on the knot.

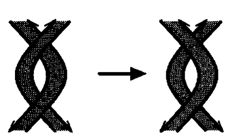


Figure 8.19 Pass-moves allow us to disentangle the bands.

We also saw in Exercise 8.8 that we can remove four half-twists in a band. Since the original surface was orientable and since this means that each band has an even number of half-twists in it, we can lower the number of half-twists in each band until it has zero or two half-twists in it. The bands with two half-twists in them can then be deformed to replace each twist with a curl (Figure 8.20). (This is exactly the same as when we were looking at DNA in Chapter 7, where we replaced some twist in a DNA ribbon with writhe.) Now, we have no twists in the bands.

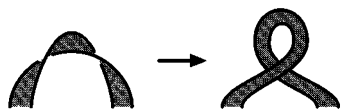


Figure 8.20 Two half-twists can be replaced by a curl.

Finally, note that if one of the ends of each of two distinct bands B_1 and B_2 lie between the two ends of a third band B_3 on the edge of the disk, we can slide the end of B_1 along one edge of B_2 to move it outside the two ends of B_3 (Figure 8.21). If necessary, we will repeat the disentangling step and the untwisting step after this sliding. In this way, we can make sure that there is at most one end of a band between the two ends of any single band on the edge of the disk.

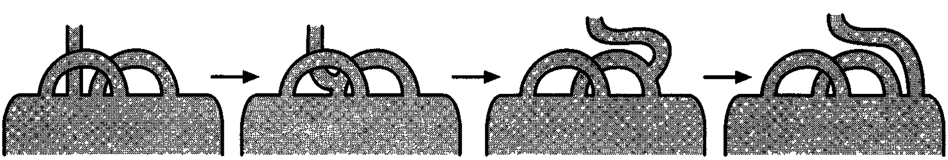


Figure 8.21 Moving one of two ends out from between the ends of a third band.

Exercise 8.10 Explain why every band must have another end of a band between its two ends on the boundary of the disk.

In particular, this means that the bands match up in pairs and there are an even number of them. Thus, the Seifert surface now appears as in Figure 8.22. Now we can cut the Seifert surface into pieces, each of which has two bands attached to it. The original knot gets cut into a set of factor knots. Each of the resulting factor knots is pass equivalent to one of the three knots shown in Figure 8.23.

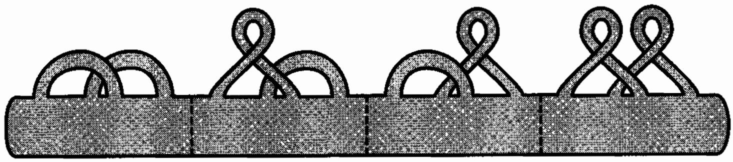


Figure 8.22 The Seifert surface after disentangling, untwisting, and sliding.

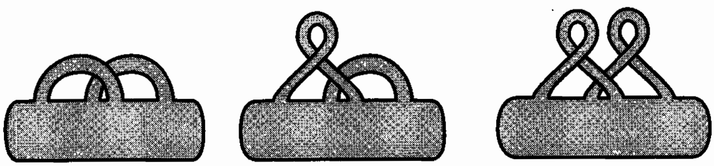
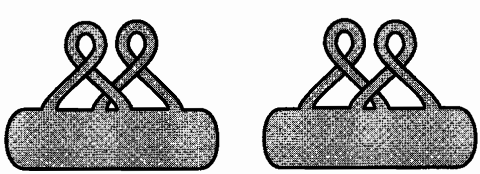


Figure 8.23 Cut the Seifert surface into three types of pieces.

Exercise 8.11 Show that the knots bounding the first and second types of Seifert surfaces in Figure 8.23 are trivial knots, while the knot bounding the third type of Seifert surface is a trefoil knot.

We have therefore just shown that every knot is pass equivalent to a composition of trivial knots and trefoil knots. However, since the composition of any knot K with the trivial knot just gives the knot K back again, we have shown that *every knot is pass equivalent to either the trivial knot or a composition of trefoil knots*. You are probably wondering which trefoil knot we mean here, since we showed that there was both a left-hand trefoil and a right-hand trefoil in Section 6.4. Actually, it doesn't matter, because the left-hand trefoil and the right-hand trefoil are pass equivalent to one another. Starting with the right-hand trefoil, appearing in the particular projection shown in Figure 8.24a, we can obtain its mirror image by passing all of the overlapping bands through each other.



a b

Figure 8.24 The trefoil is pass equivalent to its mirror image.

Great, we are getting closer to our goal. Now we need to look at the composition of a set of trefoil knots. Let's begin with a composition of two trefoil knots. Since one of them is pass equivalent to its mirror image, we can replace $K\#K$ with $K\#K^*$, where K^* denotes the mirror image of K, obtained by changing every crossing in a projection of K to its reverse crossing.

*Exercise 8.12** Show that if K is a trefoil knot, then $K\#K^*$ is pass equivalent to the trivial knot (Figure 8.25). (It takes a bit of playing around to find the right pass-move, but there is a single pass-move that will do the trick.)

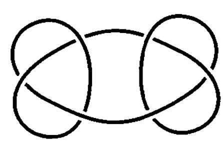


Figure 8.25 Show that this knot is pass equivalent to the trivial knot.

Now let's recap. We have shown that any knot is pass equivalent to either the trivial knot or to a composition of a number of trefoils. If the number of trefoils is even, we can pair them up and then, using the fact that a pair of trefoils is pass equivalent to the trivial knot, show that the original knot is pass equivalent to the trivial knot. On the other hand, if our original knot is pass equivalent to the composition of an odd number of trefoil knots, we can eliminate pairs of trefoils by pass-moves, and thereby show that the original knot is pass equivalent to a single trefoil. Therefore, we have proved what we set out to prove. *Every knot is either pass equivalent to the trivial knot or to the trefoil knot!*

The one thing that we didn't prove is that the trefoil knot and the unknot are not pass equivalent to each other. This is a bit more difficult, so we will take it on faith. See Kaufmann, 1983 for a proof.

We now will define the **Arf invariant** $a(K)$ of a knot K to be 0 if the knot is pass equivalent to the unknot and to be 1 if the knot is pass equivalent to the trefoil knot. That's a pretty straightforward definition. At the very least, we know the Arf invariant for two knots, namely the unknot and the trefoil knot.

Exercise 8.13 Determine the Arf invariant of the figure-eight knot in Figure 8.26.

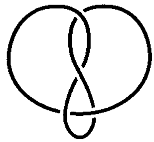


Figure 8.26 A figure-eight knot.

The Arf invariant has one very nice property, namely, if K_+, K_-, and L are projections that are identical outside the region shown, and if K_+ and K_- are knots, while L is a two-component link where each of the strands shown in the picture of L in Figure 8.27 corresponds to a distinct component, then the Arf invariants of the two knots are related through the equation:

$$a(K_+) = a(K_-) + lk(L_1, L_2)$$

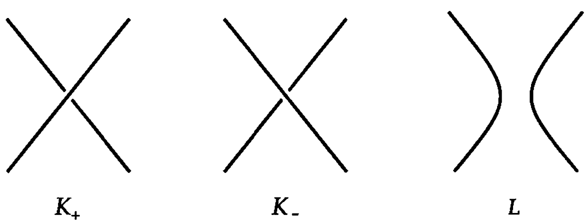


Figure 8.27 The Arf invariants.

We do not prove this, as it would be too time-consuming. It turns out, however, that the Arf invariant can be determined from the Alexander polynomial that we discussed in Section 6.3, and the skein relation satisfied by that polynomial can be used to verify this equation.

Unsolved Question

Find a direct proof that does not utilize the Alexander polynomial to show that $a(K_+) = a(K_-) + lk(L_1, L_2)$.

The idea of the proof that every embedding of K_7 contains a knotted Hamiltonian cycle is similar in spirit to the proof that every embedding of K_6 contains a pair of linked triangles. Given a particular embedding of K_7, we first define ω to be the sum of the Arf invariants, summing over every Hamiltonian cycle in the graph.

Exercise 8.13 Show that there are 360 Hamiltonian cycles in K_7.

We actually don't care about ω itself, but rather, we care about whether it is even or odd. Therefore, just as we did when we were working with linked triangles, we define Ω to be 0 if ω is even and to be 1 if ω is odd. (As we did with V, we call Ω the mod 2 reduction of ω since it is simply the remainder when we divide ω by 2.) Then, just as we did when we were looking at linked triangles, we can see what effect a crossing change has on Ω. Conway and Gordon prove that a crossing change leaves Ω unaffected. This is a slightly more difficult argument that we will not go into. (See Conway and Gordon, 1983 for the details.)

But what does this mean? Since Ω is unaffected by crossing changes, Ω must be the same for every embedding of K_7. In particular, if Ω is equal to 1 for any specific embedding, Ω is equal to 1 for every embedding. In fact, it is tedious but not too difficult to show that for the embedding of K_7 in Figure 8.28, all of the Hamiltonian cycles except one are unknotted, and the last Hamiltonian cycle is a trefoil knot. Hence, $\Omega = 1$ for this embedding, and therefore for all embeddings.

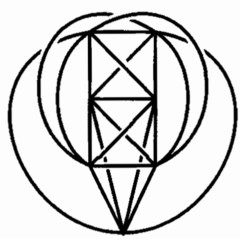


Figure 8.28 An embedding of K_7.

Exercise 8.14 Find the one knotted Hamiltonian cycle in this embedding of K_7.

Finally, if $\Omega = 1$ for every embedding, then for a given embedding, it cannot be the case that all the Hamiltonian cycles in that embedding are unknotted. Therefore, every embedding of K_7 contains a knotted Hamiltonian cycle.

In 1988, Miki Shimabara proved that any embedding of the graph $K_{5,5}$ also contains a knotted Hamiltonian cycle. The graph $K_{5,5}$ is called a bipartite graph. It is obtained by taking two sets of five vertices and attaching each vertex in the first set to every one of the vertices in the second set by edges.

We say that a graph is **intrinsically knotted** if every embedding of the graph in three-space contains a knotted cycle (not necessarily a Hamiltonian cycle). Note that if a graph contains a subgraph that is intrinsically knotted, it also must be intrinsically knotted.

Unsolved Question 1

Find other graphs besides K_7 and $K_{5,5}$ that are intrinsically knotted. So far, these two graphs and the graphs that contain them are the only graphs known to be intrinsically knotted.

Unsolved Question 2

Determine a finite set of graphs such that every graph that is intrinsically knotted either contains one of them or an expansion of one of them.

Unsolved Question 3

Is it true that if G is intrinsically knotted, and any one vertex and the edges coming into it are removed, the remaining graph is intrinsically linked? This holds true for K_7 and $K_{5,5}$.

By their characterization of intrinsically linked graphs, Robertson, Seymour, and Thomas (1993) did prove that an intrinsically knotted graph is always intrinsically linked.

8.3 Polynomials of Graphs

Polynomials were pretty handy when dealing with knots and links. Let's see how we can compute some polynomials of graphs. We start with the so-called **dichromatic polynomial**, $Z_G(q, v)$, a polynomial in two variables q and v. It's a polynomial for abstract graphs. That is to say, the polynomial does not depend on how the graph is embedded in three-space, but rather on the isomorphism type of the graph. In this sense, it differs dramatically from the polynomials of knots and links that we discussed in Chapter 6. Note that we are allowing more than one edge to share the same pair of endpoints and we are allowing edges that begin and end at the same vertex.

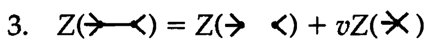
The dichromatic polynomial is defined by the following three formulas:

1. $Z(\bullet) = q$

This says that a graph consisting of just a single vertex has polynomial equal to just q.

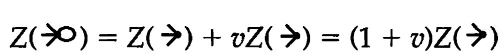
$$2.\quad Z(\bullet G) = qZ(G)$$

This says that adding a new vertex to a graph, such that it is not attached by any edges, causes the polynomial of the graph to be multiplied by q.



This says that if we pick a particular edge of a graph G, then the polynomial for G is obtained by adding the polynomial of the graph with that edge deleted to v times the polynomial of the graph with that edge collapsed down to a single vertex.

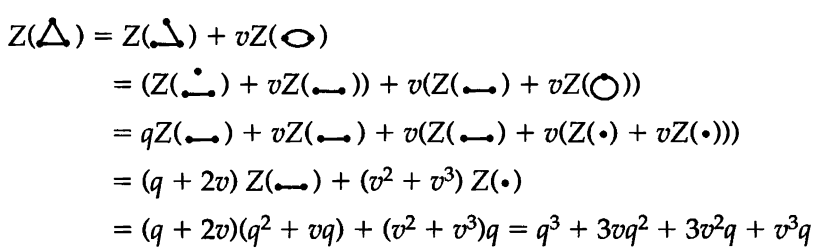
Note that if we apply this rule to an edge that begins and ends at the same vertex, we obtain:



Let's try our luck at computing with these three rules. In our first example, let's compute the dichromatic polynomial of the graph $\bullet\!\!-\!\!\bullet$.

$$Z(\bullet\!\!-\!\!\bullet) = Z(.\ .) + vZ(.) = q^2 + vq$$

Let's try a harder one. How about a triangle?



Exercise 8.16 Find the dichromatic polynomial of a square graph  and of the complete graph K_4 .

Of course, just writing down the rules does not guarantee that the polynomial is well defined. How do we know that if we calculated the polynomial of a graph in two different ways, by choosing different edges to remove at various stages, that we would get the same answer? In fact, although we won't take the time to prove it here, we will always get the same answer. Note also that unlike the knot polynomials, which depend entirely on the particular knot, these polynomials of graphs do not depend on how the graph sits in space. All that matters is which of the vertices are hooked to which other vertices by edges.

Okay, so we have a new polynomial. But what good is it? Well, one example of what we could use this polynomial for is the following. Suppose that we have a particular graph and we want to color each of the vertices of the graph one of q possible colors, so that no two vertices connected by an edge have the same color. We call such a choice of colorings of the graph a **vertex coloring.** In fact, vertex colorings come up in the real world. For instance, suppose there is a set of VHF television stations in a particular region of the country, and some of their signals overlap with one another. There are only 13 channels, but many more stations than that. Then one forms a graph where each station is a vertex, and an edge between two stations means that those stations are located close enough together that their signals will interfere with one another unless they are given distinct channels. The goal is then to successfully color the graph (with channel numbers 1 through 13, instead of colors) so that each station gets a channel, but no two stations that share an edge get the same channel.

Surprisingly enough, if we set $v = -1$ in the dichromatic polynomial of the graph, we get exactly the number of distinct vertex colorings of the graph. For example, since the dichromatic polynomial of the triangle graph is $q^3 + 3vq^2 + 3v^2q + v^3q$, it should be the case that the number of vertex colorings of the triangle graph is

$$q^3 + 3(-1)q^2 + 3(-1)^2q + (-1)^3q = q^3 - 3q^2 + 2q$$

Let's see if that is right. Given a triangle and q possible colors to color the vertices with, we can color the first vertex with any of the q colors. However, since the second vertex is connected to the first by an edge, we can only color it with one of the remaining $q - 1$ colors. The third vertex is connected to both of the first two, and therefore it can only be colored with one of the $q - 2$ colors that hasn't yet been utilized. Therefore, the total number of ways that the vertices of the graph can be colored so that no two connected vertices have the same color is $q(q - 1)(q - 2) = q^3 - 3q^2 + 2q$. This is exactly the same result that came from plugging $v = -1$ into the dichromatic polynomial.

In the case of our television stations, we plug $v = -1$ and $q = 13$ into the dichromatic polynomial for our graph corresponding to the station interferences, and if the result is greater than 0, we know there is at least one vertex coloring of the graph, and hence there is at least one choice of channel assignments that will prevent interference. So let's prove this fact that the dichromatic polynomial of a graph yields the number of vertex colorings when evaluated at $v = -1$. First we should prove it for graphs that have no edges.

Exercise 8.17 Show that any graph G that consists only of vertices and no edges has its number of vertex colorings given by $Z_G(q, -1)$. (In fact, for a graph of this type, this will be true no matter what value is given to v.)

Now we want to prove it for graphs that have edges. Suppose that we have proved it for any graph with m edges. If we can then show that it holds for a graph with $m + 1$ edges, induction will imply that it holds for any graph. Let G be a graph with $m + 1$ edges. Let E be an edge of G that connects two distinct vertices A and B. When $v = -1$, Rule 3 for the dichromatic polynomial says that Z(>—<) = Z(> <) −Z(><), where we let G' and G'' be the two new graphs appearing in the equation. Both G' and G'' have m edges. Therefore Z(> <) gives the number of vertex colorings of G', while Z(><) gives the number of vertex colorings of G''. But the number of vertex colorings of G will be the number of vertex colorings of G' minus the number of those colorings where both A and B have the same coloring. But the number of vertex colorings of G' where A and B have the same colors is exactly the same as the number of vertex colorings of G''. Hence $Z(G') - Z(G'')$ gives the number of vertex colorings of G. But the equation says that this is Z(G). Hence Z(G) is the number of vertex colorings of G.

This is an amazing fact. For instance, one of the most difficult theorems proved in the last century is the so-called Four-Color Theorem, which says that any map of countries in the plane can be colored with four colors, so that the resulting map will never have two countries of the same color sharing an edge. This theorem was proved by Wolfgang Haken (yes, the same guy who worked on knots) and Kenneth Appel at the University of Illinois, using a computer to eliminate thousands of cases.

If we take the dual graph to the map of countries as in Figure 8.29, then we are just asking whether the vertices of this planar graph can be colored with four or fewer colors so that no two vertices that share an edge get the same color. But the planar graph G has a coloring with four colors if and only if $Z(G) \neq 0$ when $q = 4$ and $v = -1$. Hence the Four-Color Thereom is equivalent to proving that $Z(G) \neq 0$ when $q = 4$ and $v = -1$ for every planar graph G.

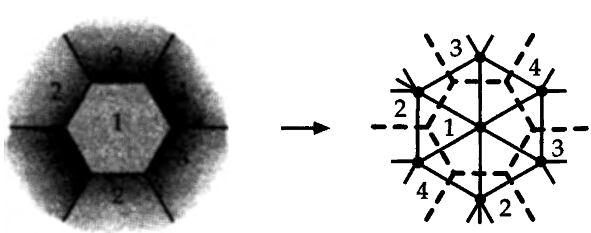
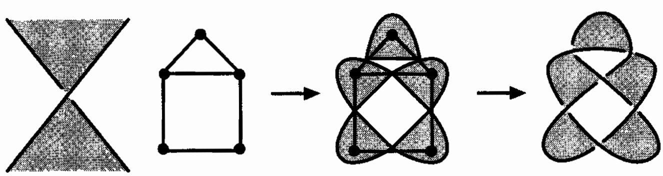


Figure 8.29 A coloring of a map becomes a coloring of a planar graph.

ꝏ *Unsolved Problem*

Find a simple proof of the Four-Color Theorem. This is still one of the biggest open questions in all of mathematics.

We would like to show that the dichromatic polynomial is related to the polynomials that come from knots and links. Remember that in Section 2.3, we saw that a planar graph can always be reinterpreted as a link. However, each edge of the graph generates a crossing in the link, and we have two choices of how to put the crossing in. If we shade the regions of the plane created by the link in a checkerboard fashion, so that the region outside the link is not shaded, we can choose all of the crossings so that if the shaded regions are placed north and south, the overcrossing strand goes from southwest to northeast, as in Figure 8.30. The planar graph then turns into an alternating link. If G is our planar graph, we will denote the resulting alternating link by $L(G)$. We include an example in Figure 8.30.



Crossing

Figure 8.30 A planar graph becomes an alternating link.

We show that the dichromatic polynomial of a planar graph G can be obtained as a "bracket" polynomial of the corresponding alternating link $L(G)$. However, the bracket polynomial that we use will be a variation on the one we saw in Section 6.1, where we defined the bracket polynomial for knot and link projections by the equations:

Rule 1: $\langle \bigcirc \rangle = 1$

Rule 2: $\langle L \cup \bigcirc \rangle = -(A^{-2} + A^{2})\langle L \rangle$

Rule 3: $\langle \times \rangle = A\langle)(\rangle + A^{-1}\langle \asymp \rangle$

We now define the so-called **square bracket polynomial** for a knot or link projection. It has two variables q and v, and it is defined by the same three equations, only with different coefficients:

$$\begin{array}{ll} \text{Rule 1:} & [\bigcirc] = q^{1/2} \\ \text{Rule 2:} & [L \cup \bigcirc] = q^{1/2}\,[L] \\ \text{Rule 3:} & [\times] = q^{-1/2}\, v\, [)(] + [\asymp] \end{array}$$

The resulting square bracket polynomial is not necessarily an invariant for knots and links. Given a projection of a knot or link, however, we can calculate the square bracket, being careful not to isotope away crossings in the projections of links in the process, as this could change the result. For instance, even though the knot shown in Figure 8.31 is the trivial knot, the square bracket polynomial of this projection is not $q^{1/2}$. Rather, using Rule 3 followed by Rules 1 and 2, we calculate it as follows:

$$\begin{aligned} [8] &= q^{-1/2} v [8] + [\,{}^{\circ}_{\circ}\,] \\ &= q^{-1/2} v q^{1/2} + q^{1/2} [\circ] \\ &= v + q^{1/2} q^{1/2} \\ &= v + q \end{aligned}$$

Figure 8.31 A nontrivial projection of the trivial knot.

Exercise 8.18 Calculate the square bracket polynomial for the two projections in Figure 8.32.

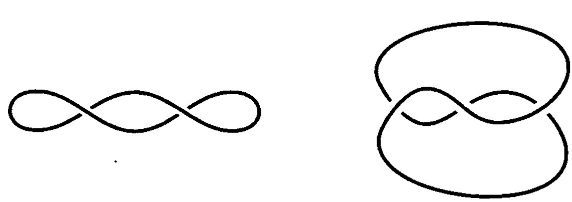


Figure 8.32 Find the square bracket polynomials for these projections.

Amazingly enough, we can now see the dichromatic polynomial of a planar graph G realized by the square bracket of the associated alternating link $L(G)$.

Theorem $Z_G(q, v) = q^{N/2}[L(G)]$, where N is the number of vertices of G.

Proof: We want to show that the left side of this equation equals the right side. First we prove it for graphs without any edges. Suppose, first of all, that we have a graph that is just a single vertex (Figure 8.33). Then the associated link $L(G)$ is just a trivial projection of the trivial knot. Hence, the square bracket polynomial for $L(G)$ is $q^{1/2}$ by the first rule for computing the square bracket polynomial. Multiplying this by $q^{N/2}$, where $N = 1$, gives us q. But this is exactly the dichromatic polynomial of a graph consisting of a single vertex. Therefore, we have proved the theorem in the very simple case that the graph G is a single vertex.

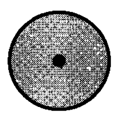


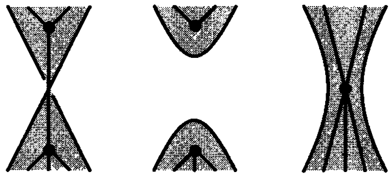
Figure 8.33 Proving the theorem for a very simple graph.

Exercise 8.19 Show that the theorem is true for any graph consisting only of vertices and no edges. (Note that if G is a graph consisting only of vertices, $L(G)$ is the link obtained by taking a set of trivial link components, such that one surrounds each of the vertices.)

Now, let's see if we can use the third rule for computing the square bracket polynomial in order to prove the theorem for any graph G. We use induction on the number of edges in the graph. We have already proved the theorem for graphs with no edges. Let's suppose we have proved it for all graphs with fewer edges than G. We then prove that it holds for G also.

Let N be the number of vertices in G. Define G' and G'' to be the two graphs depicted in Figure 8.34. Since both G' and G'' have fewer edges than G, we know that

$$Z(G') = q^{N/2}[L(G')] \quad \text{and} \quad Z(G'') = q^{(N-1)/2}[L(G'')]$$



G G' G''

Figure 8.34 The dual graphs.

Our goal is to show that $Z(G) = q^{N/2}[L(G)]$. However, by the third rule for computing the dichromatic polynomial

$$\begin{aligned} Z(G) &= Z(G') + vZ(G'') \\ &= q^{N/2}[L(G')] + vq^{(N-1)/2}[L(G'')] \\ &= q^{N/2}([L(G')] + vq^{-1/2}[L(G'')]) \\ &= q^{N/2}[L(G)] \end{aligned}$$

by the third rule for computing the square bracket polynomial. This proves our theorem. □

Let's look at an example. The triangle graph G in Figure 8.35 becomes a trefoil knot. According to the theorem, the dichromatic polynomial of the triangle should be given by

$$\begin{aligned} q^{N/2}[\ ] &= q^{3/2}[\ ] = q^{3/2}(q^{-1/2}v[\ ] + [\ ]) \\ &= q^{3/2}(q^{-1/2}v(q^{-1/2}v[\ ] + [\ ]) + (q^{-1/2}v[\ ] + [\ ])) \\ &= q^{3/2}(q^{-1/2}v(q^{-1/2}v(q^{-1/2}v[\ ] + [\ ]) + (q^{-1/2}v[\ ] + [\ ])) \\ &\quad + (q^{-1/2}v(q^{-1/2}v[\ ] + [\ ]) + (q^{-1/2}v[\ ] + [\ ]))) \\ &= q^{3/2}(q^{-1/2}v(q^{-1/2}v(q^{-1/2}vq + q^{1/2}) + (q^{-1/2}vq^{1/2} + q)) \\ &\quad + (q^{-1/2}v(q^{-1/2}vq^{1/2} + q) + (q^{-1/2}vq + q^{3/2}))) \\ &= v^3q + 3v^2q + 3vq^2 + q^3 \end{aligned}$$

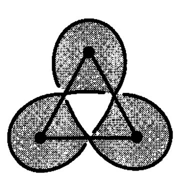


Figure 8.35 $L(G)$ is a trefoil knot.

We get exactly the dichromatic polynomial that we computed earlier for the triangle graph.

So here we have the idea of a polynomial formed as a "bracket." Depending on the coefficients we choose for the "bracket" relations, we can compute polynomials for knots or for graphs. Moreover, this has implications for statistical mechanics.

In Section 7.4, we discussed the Ising model. There, each vertex of a planar graph was allowed to have one of two states. Here, we generalize that to a new model called the **Potts model** (Figure 8.36). We still work with a planar graph, where each vertex can be thought of as a particle, only now, instead of only two states, each vertex can have one of q states, where q is some positive integer. It is sometimes helpful to think of the q states that a particle can be in as q possible colors that a particle can have.

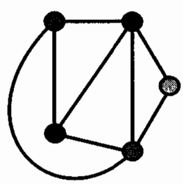


Figure 8.36 A particular state of the Potts model.

Amazingly enough, we show that if we compute the dichromatic polynomial of the planar graph, and then plug in the right substitution for the variable v, *the dichromatic polynomial of the graph will become exactly the partition function of the Potts model.* As in Section 7.4, the partition function of our graph is defined to be

$$P = \sum e^{-E(S)/kT}$$

where we sum over all possible states of the system. As before, k is the Boltzmann constant and T is the temperature of the system. The energy of the system in a particular state S is again given by $E(S)$. This energy is the sum of the interaction energies of the edges:

$$E(S) = \sum E(S_i, S_j)$$

where we sum over all pairs of vertices that are connected by edges in the graph. The state S is given by the individual states of the vertices, so $S = s_1, s_2, \ldots, s_n)$. For this particular model, we choose the interaction energy along an edge to be $E(s_i, s_j) = \delta(s_i, s_j)$, where δ is again the so-called **Kronecker delta function**, defined by

$$\delta(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$

Thus, an edge in the graph contributes to the energy of the particular state only when the vertices at its two endpoints are in the same state. (That is to say, they have the same color associated to them.)

Theorem Let $v = e^{-(1/kT)} - 1$. Then the dichromatic polynomial $Z_G(q, v)$ for the planar graph G becomes the partition function of the Potts model on the graph G, with the variable q from the polynomial giving the number of possible states at each vertex.

Proof: Starting with the partition function, we rewrite it in order to show it is the dichromatic polynomial.

$$\begin{aligned} P &= \sum e^{-E(S)/kT} \\ &= \sum e^{-\Sigma\delta(S_i, S_j)/kT} \\ &= \sum \prod (e^{-1/kT})^{\delta(S_i, S_j)} \\ &= \sum \prod (1 + v\delta(S_i, S_j)) \end{aligned}$$

This last line follows from the previous one by writing out the two possibilities for the values of the function δ ($\delta = 0$ or $\delta = 1$), and seeing that the equality between these two expressions holds in either case. We claim that $\sum \prod (1 + v\delta(s_i, s_j))$ is exactly the dichromatic polynomial in q and v. To see this, we show that this polynomial satisfies the three rules for computing the dichromatic polynomial. Since we already have said that the dichromatic polynomial is well defined, any polynomial satisfying the same set of defining rules must itself be the dichromatic polynomial.

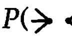
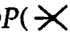
First note that if we have a graph G consisting of just a single vertex and no edges, the partition function P is given by

$$\sum \prod (1 + v\delta(s_i, s_j)) = \sum 1$$

summing over all states s. Since the vertex can take on q states, we have $P(.) = \sum 1 = q$. Hence the partition function does satisfy the first rule for the dichromatic polynomial. Let's check the second rule by finding the partition function of a graph with one extra vertex that is not connected to the rest of the graph by an edge.

Exercise 8.20 Show that $P(.G) = qP(G)$.

Finally, we want to check the last rule for the dichromatic polynomial. That is, we need to show that the partition function satisfies

$P($$) = P($$) + vP($$)$

Suppose we have labeled all of the vertices in G by the integers $\{1,2, \ldots, n\}$. Let a and b be two of the integers, which label two vertices in our graph G that are connected by an edge e. As in Figure 8.34, suppose G' is the graph G, except that e has been deleted, and suppose G'' is the graph G, except that e has been contracted so that the two vertices labeled by a and b have been identified. Our goal is to show that $P(G) = P(G') + vP(G'')$. But

$$P(G) = \Sigma\Pi(1 + v\delta(s_i, s_j))$$

For a given state s of the entire graph, the term in this sum that corresponds to s is

$$\Pi(1 + v\delta(s_i, s_j)) =$$

$$(1 + v\delta(s_a, s_b)) \underset{\substack{(i,j) \neq \\ (a,b)}}{\Pi} (1 + v\delta(s_i, s_j)) = \underbrace{\underset{\substack{(i,j) \neq \\ (a,b)}}{\Pi}(1 + v\delta(s_i, s_j))}_{\text{Term 1}} + \underbrace{v\delta(s_a, s_b)) \underset{\substack{(i,j) \neq \\ (a,b)}}{\Pi} (1 + v\delta(s_i, s_j))}_{\text{Term 2}}$$

(just applying the distributive law)

The first term in the preceding equation is exactly the term in the partition function of G'corresponding to the state s. For the second term, note that when the two vertices a and b have the same color in the state s, $\delta(s_a, s_b) = 1$. The second term is then just v times the term in the partition function of G'' corresponding to the state s. When the two vertices a and b have different colors in the state s, $\delta(s_a, s_b) = 0$ and the second term disappears. But it's just as well that it disappears, since there is no corresponding state for G''; the two vertices have collapsed to one, and cannot have different colors.

Thus, summing over all possible states, we have that $P(G) = P(G') + vP(G'')$, as we wanted to show. Therefore, since P satisfies the rules for computing the dichromatic polynomial, *it is the dichromatic polynomial*. Thus, we have seen that the dichromatic polynomial of a graph, which can be computed utilizing a skein relation on the corresponding alternating link, is in fact the partition function for the statistical mechanical model known as the Potts model. □

Mathematicians and physicists are still working on the connections between statistical mechanics and knot theory. There remains lots of interesting work to be done.

9

Topology

9.1 Knot Complements and Three-Manifolds

All the knots that we have looked at have lived in three-dimensional space. Let's call this three-dimensional space R^3. Then R^2 is a plane and R^1 is a line. We discuss R^4 in the next chapter.

In Chapter 4, we defined the **complement of a knot**, namely, all of space minus the knot (Figure 9.1). It's as if we had drilled a wormhole

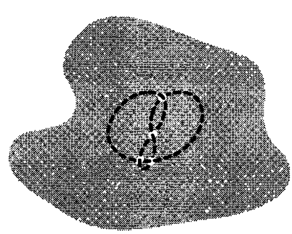


Figure 9.1 The complement of a knot.

through space where the knot had been. However, keep in mind that this wormhole is very thin, its thickness being the thickness of the missing knot, which is only one point thick. We denote the complement of the knot by $R^3 - K$. The complement of a knot is an example of a **three-manifold.** Remember that when we talked about surfaces in Chapter 4, we said that a surface was a two-manifold. The defining property of a two-manifold was that around each point in the two-manifold, there was a disk (not necessarily flat) of points, also in the two-manifold. So, we can say that a three-manifold satisfies the property that around each point in the three-manifold, there is a ball of points that is also in the three-manifold. The simplest example of a three-manifold is R^3, three-dimensional space (Figure 9.2). If we pick any point in three-dimensional space, there is a ball of points around it that is also in three-dimensional space.

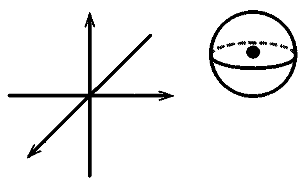


Figure 9.2 R^3 is a three-manifold.

As a second example, you are sitting in a three-manifold right now. The spatial universe is a three-manifold (Figure 9.3). Pick any point in front of you. Then there is a ball of points around it. The same is true for any other point in the universe that we care to pick. Hence, the spatial universe is a three-manifold. (However, there are two caveats here. First of all, we can't check whether points that are very far away have this property, but it is a reasonable assumption that they do. Secondly, one possible explanation for the phenomenon of black holes is that they are exceptional points in the universe that do not have three-dimensional balls surrounding them. They are so-called **singularities** in the three-manifold structure of the universe.)

Just as there are many different possible two-manifolds, there are many different possible three-manifolds. Any one of them could be a model for the universe we live in. Most people picture the universe as R^3, namely as three-dimensional space that just continues off forever in every direction. But of course, at one time people believed that the surface of the earth was flat, like the plane R^2. In fact, the surface of the earth turned out to be a sphere. (It might have been interesting if it had turned out to be a torus.) So we can't assume that the universe is as uninteresting a three-manifold as R^3. What are other possibilities?

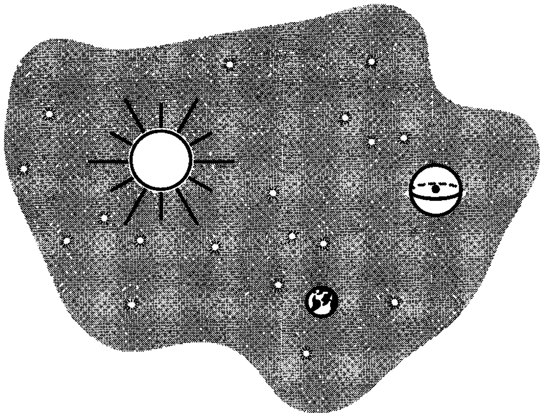


Figure 9.3 The spatial universe is a three-manifold.

A complement of a knot is also an example of a three-manifold (Figure 9.4). In order to see this, we have to show that for any point in the complement of the knot, there is a ball of points surrounding it that is also in the complement of the knot. If we pick a point in three-space that is far away from the knot, then it is easy to see that there is a ball of points in the complement of the knot surrounding that point. Any ball in three-space that contains the point and avoids the knot will work. If the point we pick in the complement is very close to the missing knot, we will just pick a ball around the point that is very small and still avoids the knot. The ball will then be in the complement of the knot. Thus, the knot complement $R^3 - K$ is a three-manifold. In particular, this means that the universe could be a knot complement.

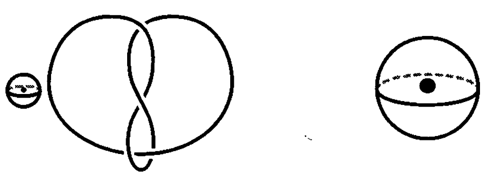


Figure 9.4 A knot complement is a three-manifold.

What would it mean for the universe to be a knot complement? Would it mean that somewhere out in space there was a giant knot?

No! If the universe really is a knot complement, then that means the knot is missing from space. So we can never see it. It's not there. It would be as if the knot were infinitely far away. As we headed for where the knot would be, distances that look small in the picture would actually be extremely large and we could never reach the knot (see Figure 9.5).

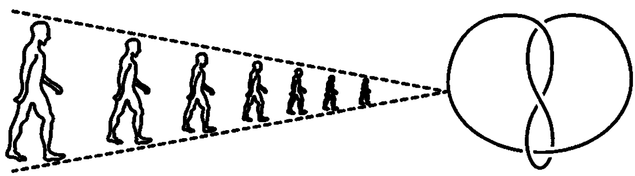


Figure 9.5 The knot is infinitely far away.

For the last 90 years, mathematicians have been trying to determine all of the possibilities for three-manifolds. We have seen that R^3 is a three-manifold and that knot complements are three-manifolds. Similarly, link complements are also three-manifolds. In fact, there are lots of other three-manifolds. Unfortunately, it is usually impossible to picture them in three-space. Just as most two-manifolds do not exist in a two-dimensional plane, but rather exist in three- or four-dimensional space, most three-manifolds exist in four- or higher dimensional space. We describe some interesting examples in the next section.

9.2 The Three-Sphere and Lens Spaces

We want to define another three-dimensional space called the **three-sphere**. It is the analog of the two-dimensional sphere, only one dimension up. Since the two-sphere lives in three-space, it makes sense that the three-sphere will have to live in four-space. That's going to make its description a little bit more difficult. It will sound strange at first. We describe it in two different ways, without mentioning four-space.

First, notice that the two-sphere can be described as two curved disks, usually called hemispheres, glued together along their boundaries (Figure 9.6). Since the analog of a disk one dimension up is a solid ball, we describe the three-sphere as two solid balls, with their boundary spheres glued together (Figure 9.7). Of course, we couldn't actually glue the boundary of the first ball to the boundary of the second ball in three-space, as we couldn't deform the one boundary onto the other. But that's not too surprising, as we said that the three-sphere doesn't live in three-space.

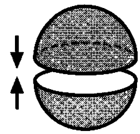


Figure 9.6 The two-sphere.

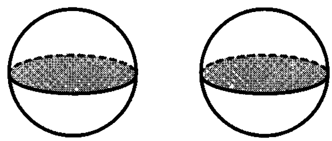


Figure 9.7 The three-sphere.

We think of the gluing abstractly. If we were actually standing inside the first ball, we could walk up to its boundary and then pass right through the boundary into the second ball. Since the two boundaries are glued together, we can pass back and forth between the two balls at will. Notice also that this description of the three-sphere satisfies the definition of a three-manifold (Figure 9.8). Certainly any point that is in the interior of either one of the two balls is contained in a ball. Also, a point x on the boundary of one of the two balls is surrounded by a ball B in the manifold, half the ball B coming from the first of the two balls and the other half coming from the second. These two half-balls are glued together to form the whole ball B surrounding the point x.

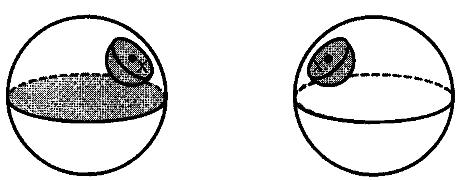


Figure 9.8 The three-sphere is a three-manifold.

Let's give a second description of the three-sphere. We take the points in R^3 together with one extra point, which we think of as off at infinity, since it does not sit in R^3. We denote this one extra point by ∞, the symbol for infinity. But just think of it as an extra point. We then say that the three-sphere is $S^3 = R^3 \cup \{\infty\}$. In this second description of S^3, we can see the two balls making up the first description (Figure 9.9). The points a distance less than or equal to one from the origin in R^3 become the first ball. The points a distance greater than or equal to one together with the one extra point $\{\infty\}$ become the second ball.

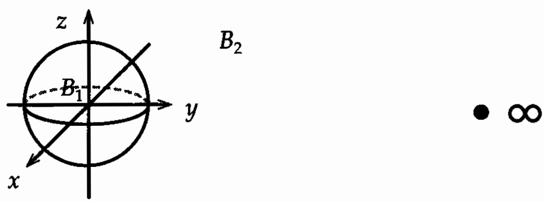


Figure 9.9 The three-sphere.

If a point is inside the ball of radius one from the origin, its distance to the origin is the usual distance. If a point is a distance d that is one or greater from the origin, we measure its distance from $\{\infty\}$ to be $1/d$. So the farther out a point is in the usual measure of distance, the closer that point is getting to the point $\{\infty\}$. The $\{\infty\}$ becomes the center of a second ball made up of all of the points outside the ball of radius one in R^3.

Exercise 9.1 One potential model for our three-dimensional spatial universe is S^3. Assuming that we have fast space travel, how might we discover that our universe is S^3? What properties of the universe that we could check might tell us it is S^3? (There is no one answer to this question; it's a vague essay-type question.)

This last description of S^3 makes it clear that S^3 and R^3 are very similar, differing in only one point $\{\infty\}$. However, the advantage that S^3 has over R^3 is that S^3 is compact.

When we discussed surfaces in Chapter 4, we said that a surface was compact if it could be triangulated with finitely many triangles. But what does it mean to talk about a triangulation of a three-manifold? We simply replace the triangles with their analogs, one dimension up, tetrahedra. So a triangulation (sometimes called a *tetrahedralization*) of a three-manifold means a decomposition of the three-manifold into tetrahedra, so that pairs of tetrahedra either don't intersect or they intersect in a face or an edge or a vertex.

Thus, for example, here are triangulations of R^3 and S^3 (Figure 9.10). Note that when we glue the two balls together on their boundaries to form

S^3, we have to glue so that vertices go to vertices, edges go to edges, and faces go to faces. As with a triangulation, we assume that the tetrahedra are rubber and can therefore appear misshapen. Notice also that the triangulation of R^3 has infinitely many tetrahedra, while the triangulation of S^3 is made up of a finite number of tetrahedra. Given a choice between these two options, it's not surprising that mathematicians usually choose to work with the three-manifolds that can be triangulated with finitely many tetrahedra, since it's always easier to work with finite sets.

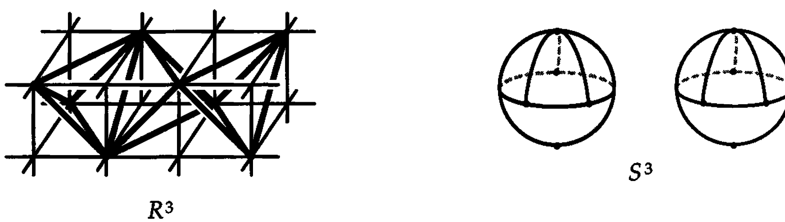


Figure 9.10 Triangulations of R^3 and S^3.

We say that a three-manifold is **compact** if it can be triangulated with finitely many tetrahedra. Most of the time, it's convenient to think of knots as living in S^3 rather than in R^3, throwing in the extra point $\{\infty\}$. Then the knot lies in a compact space. However, once we remove the knot K from S^3, the three-manifold $S^3 - K$ is no longer compact (Figure 9.11). There is no way to triangulate the result without using infinitely many tetrahedra. Note that we could triangulate it with infinitely many tetrahedra, by utilizing smaller and smaller tetrahedra to fill up the space as we approached the missing knot.

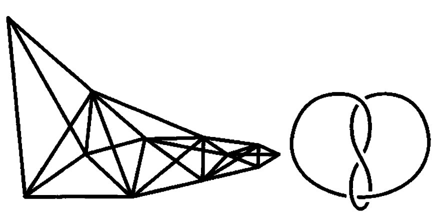


Figure 9.11 $S^3 - K$ is not compact.

How can we visualize the complement of a knot in S^3? We could either take $R^3 - K$ and throw in the one extra point $\{\infty\}$. Or we could instead have the knot pass through the point $\{\infty\}$, so that when we remove the knot from S^3, we remove the point $\{\infty\}$. But how do we make the knot pass through this point at infinity? We place the knot in space so that it goes off to infinity in two directions, resembling a knotted line (Figure 9.12). Remember, since the two strands of the knotted line go further and further away from the origin in three-space, they must be getting closer and closer to the point $\{\infty\}$. So we include the point $\{\infty\}$ as part of the knot. Now, when we remove the knot from S^3, the point at infinity is removed and $S^3 - K$ is exactly the set of points that we see, namely R^3 minus this knotted line.

Figure 9.12 The complement of a knot when the knot goes off to $\{\infty\}$.

Exercise 9.2 Draw the knot 5_2 so that it passes through $\{\infty\}$.

Let's look at some other three-manifolds. We generalized our first description of the three-sphere in Section 9.2. There, we described the three-sphere as two balls glued together along their boundaries. In fact, S^3 is the only manifold that can be obtained by gluing together two balls along their spherical boundaries. This fact is difficult to prove, so we will accept it on faith.

Now, instead of two balls, we glue two solid tori together along their boundaries (Figure 9.13). This time, different choices of how to glue the two boundaries together generates different three-manifolds. Notice that all of the manifolds that we generate in this way will be compact, since we can triangulate each of the solid tori with finitely many tetrahedra.

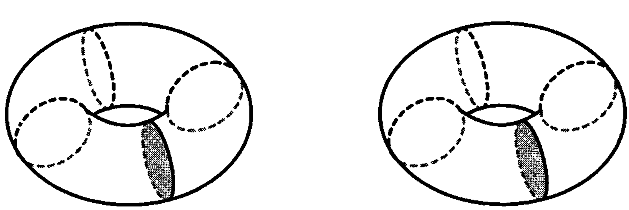


Figure 9.13 Gluing the boundaries of two solid tori together.

One way of gluing the two boundaries together is to glue a meridian curve on the first torus boundary to a longitude curve on the second torus boundary. In order that our gluing of the two boundary tori identify points one-to-one, each meridian curve on the boundary of the first solid torus will be forced to be glued to a longitude curve on the boundary of the second solid torus. Surprisingly, this gluing generates S^3 again (Figure 9.14). We can see this if we cut the first torus open along two meridional disks. Glue one of the resulting two pieces to the second solid torus so that a meridian curve glues to the longitude curve of the boundary of the second torus. The resulting object is a ball. The remaining piece of the first solid torus is also a ball. Hence, the manifold obtained by gluing in the remaining piece is equivalent to a manifold obtained by gluing two balls together along their boundary. But the only such manifold is S^3.

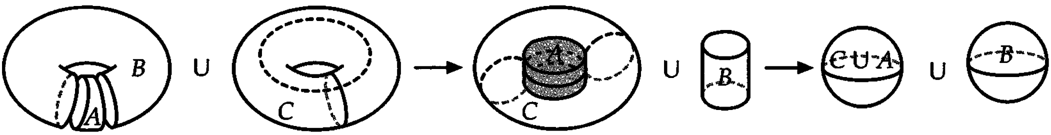


Figure 9.14 Gluing two solid tori together in this way yields S^3.

Let's try some other way to glue two solid tori together along their boundaries. For instance, we can send a meridian of the boundary of the first solid torus to the meridian of the boundary of the second solid torus. At the same time, we also send the longitude on the first torus to the longitude on the second torus. This means that each meridian curve on the first torus is sent to the corresponding meridian curve on the second torus. Hence, the boundary of each meridional disk in the first torus is glued to the boundary of a meridional disk in the second torus (Figure 9.15). But we know that if we glue two disks together on their boundary, the result is a two-sphere. Hence, the pairs of meridional disks form two-spheres in the resultant manifold.

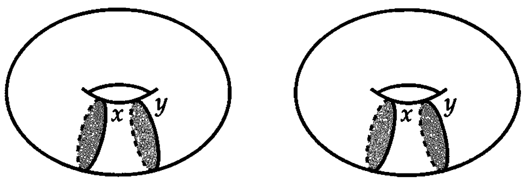


Figure 9.15 Gluing two meridional disks together on their boundaries.

Let L be a longitude on the first solid torus. At each point x along L we have a two-sphere that is perpendicular to L in our new manifold, the two-sphere coming from the union of the meridional disk cutting through the first solid torus at x and the meridional disk in the second solid torus that is glued to its boundary. Hence we have a circle's worth of two-spheres, the circle being L. We denote the resulting manifold by $S^2 \times S^1$, and say that it is the product of a two-sphere S^2 with a circle S^1. This manifold is distinct from the three-sphere S^3.

We can think of this description of $S^2 \times S^1$ as simply being obtained by taking a solid torus and "reflecting it in its boundary" to get a second solid torus attached to the first. We treat the boundary as a mirror with two carbon copies of the solid torus on each side. Note that the meridional disks on the two solid tori are then glued together along their boundaries. The manifold without boundary that is obtained by reflecting a manifold M in its boundary is called the **double of M**.

Here is a second way to picture $S^2 \times S^1$. Starting with a solid ball, hollow out a smaller ball from its interior. Now glue the inner boundary sphere to the outer boundary sphere by identifying each point on the inside sphere to the point radially outward from it on the outside sphere (Figure 9.16). Of course, the gluing is again done abstractly in our heads rather than in a picture, as the resulting manifold doesn't exist in three-space. The concentric spheres in this picture form the spheres from the previous description. We have a circle's worth of spheres, since, as we travel out along a radius, at each point we have a concentric sphere. But at the last point, the sphere there is glued to the sphere corresponding to the first point, making the radial interval into a circle.

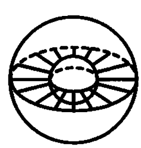


Figure 9.16 Glue inner sphere to outer sphere to obtain $S^2 \times S^1$.

Exercise 9.3 If this new description of $S^2 \times S^1$ is giving us the same manifold as in the previous description, then we should be able to see the two solid tori that make up the first description in the picture that gives the second description in Figure 9.16. Find a torus in Figure 9.16 that cuts the manifold into those two solid tori.

This new manifold is pretty amazing in its own right. For instance, the theory of knots in $S^2 \times S^1$ is quite different from the theory of knots in S^3. As an example, take a look at the knot in Figure 9.17. Note that it is a knotted loop since the north pole of the inner sphere is glued to the north pole of the outer sphere, making the knotted arc into a loop. Surprisingly, even though it appears knotted it isn't really knotted at all. We can undo the snarl within it (Figure 9.18).

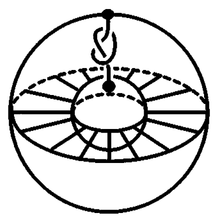


Figure 9.17 A knot in $S^2 \times S^1$.

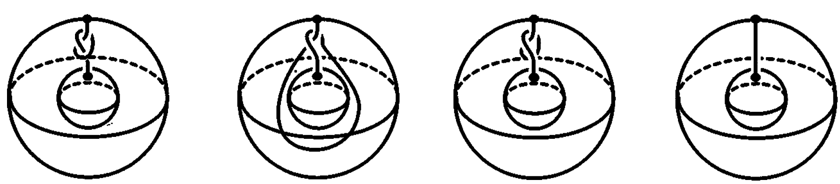


Figure 9.18 Unknotting a knot in $S^2 \times S^1$.

Exercise 9.4 Show that the snarl within the knot in $S^2 \times S^1$ depicted in Figure 9.19 can also be undone.

Exercise 9.5 Show that if the snarl appearing in the knot in Figure 9.19 were replaced with *any* other snarl, it could still be undone.

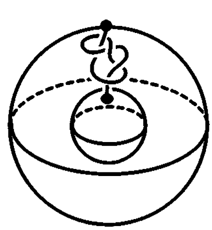


Figure 9.19 Untangle this knot in $S^2 \times S^1$.

However, even though we have undone the snarls in these particular knots, we would not say that they are trivial knots, as they do not bound disks within $S^2 \times S^1$. Moreover, there are knots that pass through the spheres more than once that cannot be untangled in this manner. Hence, there is a nontrivial theory of knots in $S^2 \times S^1$.

Unsolved Question

Figure out how to define polynomial invariants of knots, when those knots live in $S^2 \times S^1$.

So now we have seen two different manifolds, S^3 and $S^2 \times S^1$, both of which can be obtained by gluing together two solid tori along their boundaries. In the first case, we glued a meridian of the first torus to a longitude of the second torus. In the second case, we glued a meridian of the first torus to a meridian of the second torus. In general, in order to obtain new three-manifolds, we can glue the meridian of the first solid torus to any nontrivial curve that doesn't intersect itself on the second solid torus. Such a curve is a (p, q)-curve, where p is the number of meridians and q is the number of longitudes.

A **lens space** $L(p, q)$ is the three-manifold obtained by gluing the boundaries of two solid tori together, so that the meridian of the first solid torus goes to a (p, q)-curve on the second solid torus. We might think that it matters where the longitude of the first solid torus goes, but in fact it does not. Once we decide where the meridian is going, the resultant manifold is determined. The lens spaces are all examples of compact three-manifolds. (See (Rolfsen, 1976) (in the references for Chapter 1) for lots more on lens spaces.)

Exercise 9.6 Given a lens space $L(p, q)$, what compact three-manifold with boundary do we get if we remove the interior of a neighborhood of the core curve of one of the two solid tori that make up the lens space? (*Hint*: We are removing a solid torus from $L(p, q)$.)

Unsolved Problem

Determine a theory for knots and links in lens spaces by projecting the knot or link onto the torus that splits the lens space into two solid tori. Extend the knot invariants such as the polynomials to this more general setting.

Instead of gluing two solid tori together along their boundaries, we could generalize and glue two solid "handlebodies" along their boundaries (Figure 9.20). The two handlebodies must have the same genus in order for us to be able to glue their boundaries together. Just as we had for two solid tori, there are infinitely many different ways to glue the boundaries of two solid handlebodies together.

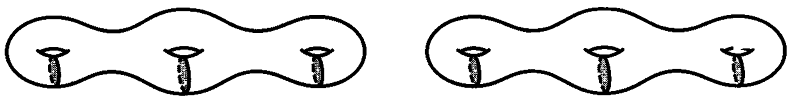


Figure 9.20 Glue two solid handlebodies together along their boundaries.

Amazingly enough, we have just described *every* compact orientable three-manifold. Each one can be obtained by gluing together the boundaries of a pair of handlebodies. For some of them, we need to glue two handlebodies of very high genus, but if we allow arbitrarily high genus, we will obtain all three-manifolds. Put another way, every compact orientable three-manifold contains a surface of some genus n such that when the manifold is cut open along the surface, two handlebodies of the same genus are the result. Such a splitting of the manifold is called a **Heegaard splitting**.

Here is the idea behind the proof. Every compact three-manifold has a finite tetrahedralization, namely a decomposition into a finite number of tetrahedra. Given a three-manifold M and such a tetrahedralization of M, define the **one-skeleton** of the tetrahedralization to be the union of all of the vertices and edges of the triangulation. It is a graph embedded in the three-manifold. If we thicken it up, we obtain a handlebody contained within the three-manifold, which we will call H_1 (Figure 9.21).

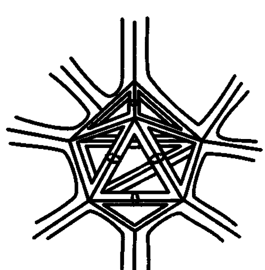


Figure 9.21 Thickening up the one-skeleton yields a handlebody.

Let's look at a second graph that we can embed in the three-manifold. At the center of each tetrahedron, place a vertex. If two tetrahedra are glued together along a face, put in an edge connecting the two vertices at the centers of the two tetrahedra, the edge passing through the shared face. The resulting graph is called the **dual graph** to the tetrahedralization (Figure 9.22). If we now thicken up this graph, we again obtain a handlebody, and if we thicken it up enough, this new handlebody will fill up all of the original three-manifold other than the points already contained in H_1. Together, the handlebody, call it H_2, and the original handlebody H_1 fill all of M, and they share a common boundary. In other words, they form a Heegaard splitting of M.

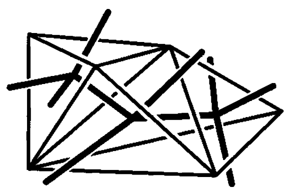


Figure 9.22 The dual graph to a tetrahedralization.

Great! It sounds like we are just about done with three-manifold theory. We can now "list" every compact three-manifold. We first list all of the manifolds with genus zero Heegaard splittings. That is to say, we write down the three-sphere. Then we list all of the manifolds with genus one Heegaard splittings. This is exactly the lens spaces. Then we list all of the manifolds with genus two splittings. Continuing in this manner, we can list all manifolds with Heegaard splittings up to some genus n.

Unfortunately, we cannot give a complete list of all of the possibilities without repeats, since many of these descriptions yield the same three-manifold, and it is difficult to tell which ones do. In fact, no one has successfully listed all of the three-manifolds with genus two Heegaard splittings. Moreover, if someone hands us a three-manifold described in some manner other than by a splitting, we have no way to determine which of the three-manifolds in the list it is.

Unsolved Problem

Determine a method for listing all of the three-manifolds with genus two Heegaard splittings, such that each manifold is listed exactly once and such that one can determine which of the manifolds in the list a given manifold with a genus two Heegaard splitting is.

9.3 The Poincaré Conjecture, Dehn Surgery, and the Gordon-Luecke Theorem

Around the turn of the century, a French mathematician named Henri Poincaré (1854–1912) asked a mathematical question that we can paraphrase as follows: "If an object looks like a ball and acts like a ball, is it a ball?"

What do we mean by this? Well, let's look at some properties of a ball. A ball has one boundary component, which is a sphere. We can get from any one point in the ball to any other without leaving the ball (we say that the ball is connected). Its interior (excluding the points on the boundary) is a three-manifold (around each point in the interior, there is a ball of points in the interior). It is compact, since we can triangulate it with finitely many tetrahedra (in fact, one tetrahedron will suffice). Finally, notice that if we take any loop inside the ball, that loop can be shrunk to a point inside the ball (Figure 9.23). This property that all loops can be shrunk down to points won't hold in most manifolds. For example, take the three-manifold given by a solid torus. The core curve cannot be shrunk to a point without leaving the solid torus. (See Figure 9.24.) We call a space **simply connected** if it has the property that all loops within the space shrink to points within the space.

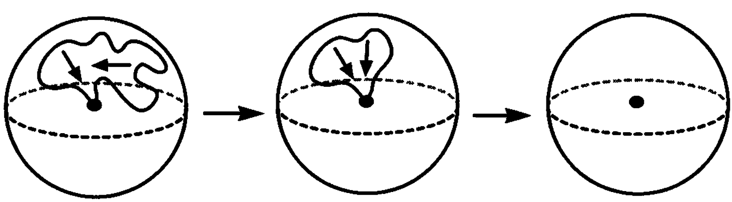


Figure 9.23 Loops shrink to points inside the ball.

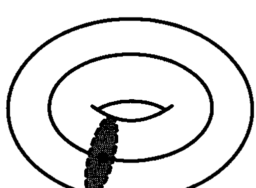


Figure 9.24 Not all loops shrink to points within a solid torus.

Poincaré's conjecture can then be stated as follows:

If there is a three-manifold M with one boundary component, that boundary component being a sphere, and if that three-manifold is compact, connected, and simply connected, then that three-manifold is a ball.

This conjecture is certainly the biggest open question in topology, and one of the biggest open questions in all of mathematics. Most of the best topologists have tried at one time or another to prove or disprove it. Wolfgang Haken worked on this problem for 10 years before he switched to the other most well-known problem of the time, the Four-Color conjecture, which, together with Kenneth Appel, he solved in 1974. Note that all we have to do to disprove the Poincaré conjecture is to come up with an example of a three-manifold that has all of those properties in the hypothesis of the conjecture, but that is not a ball.

Poincaré's conjecture is more commonly stated in the following form:

Let M be a compact, connected, simply connected, three-manifold without boundary. Then M must be the three-sphere S^3.

Exercise 9.7 Show that these two formulations of Poincaré's conjecture are equivalent. (*Hint*: Remember, around every point in a three-manifold, there is a ball in the three-manifold. You can also use the fact that if the interior of a ball is removed from S^3, a ball remains.)

So how would we go about proving Poincaré's conjecture? Well, remember that every compact connected three-manifold without boundary can be constructed by taking the union of two handlebodies glued together along their boundaries. So all we need to show is that if a three-manifold constructed in this way has the property that loops shrink to points, that three-manifold must be S^3. In fact, this approach has succeeded in the case of genus one or genus two handlebodies. But no one has yet shown that a counterexample to the Poincaré conjecture couldn't come from gluing together two handlebodies of genus three or more.

The concept of gluing two handlebodies together along their boundaries generalized the idea of gluing two solid tori together along their boundaries. We would now like to generalize the gluing of two solid tori in a different manner. Here, we glue one solid torus to one knot exterior along their torus boundaries (Figure 9.25). [A **knot exterior** is just the complement of an open solid torus knotted like the knot. The open solid torus that we remove is called a tubular neighborhood of the knot.]

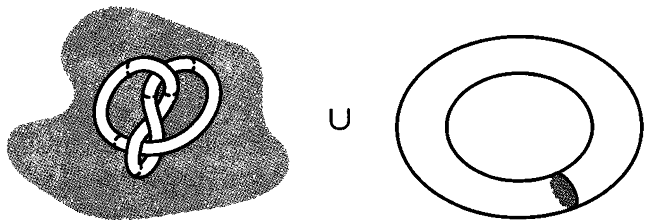


Figure 9.25. Glue a solid torus to a knot exterior along their torus boundaries.

If we glue in the solid torus so that a meridional curve on the solid torus goes to a meridional curve on the knot exterior, we will have just filled in the missing solid torus, and the resulting manifold will again be S^3. We simply drilled out a solid torus from S^3 to get the knot exterior, and then we replaced it to get back S^3. However, as we saw in our construction for lens spaces, we can glue the meridian of the solid torus to *any* (p, q)-curve that doesn't intersect itself on the torus boundary of the knot exterior. For different choices of (p, q)-curves, we can glue in the solid torus to get various different three-manifolds.

Given a knot K in S^3, the operation of drilling out a tubular neighborhood of the knot and then gluing in a solid torus so that its meridian curve goes to a (p, q)-curve on the torus boundary of the knot exterior is called **Dehn surgery** (Figure 9.26). The result is a compact three-manifold without boundary. Notice that gluing two solid tori together is a special case of Dehn surgery. If we do Dehn surgery on a trivial knot, we remove a tubular neighborhood of a trivial knot from S^3. The knot exterior that we are left with is simply a solid torus. Hence, when we glue a second solid torus to the knot exterior, we are actually gluing two solid tori together along their boundaries. So the lens spaces come from Dehn surgery on the trivial knot.

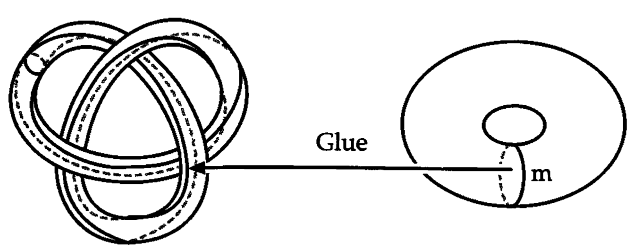


Figure 9.26. Glue meridian of solid torus to (p, q)-curve.

We could also do Dehn surgery on a link in S^3 (Figure 9.27). For each component of the link, we drill out a tubular neighborhood of the link and then glue in a solid torus. Obviously, since we have so many choices of links and (p, q)-curves to glue to, we expect to get a lot of different compact three-manifolds this way. In fact, we get *all* of them. *Every* compact connected three-manifold comes from Dehn surgery on a link in S^3. (This was proved by two different mathematicians, Raymond Lickorish and Andrew Wallace, in the early 1960s, working independently and using entirely different methods.) This is the basic connection between knots and links and three-manifolds. If we could really understand knots and links and the Dehn surgeries on them, we would really understand *all* compact three-manifolds!

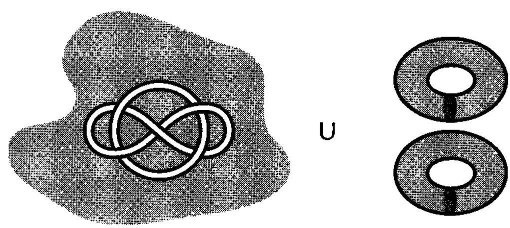


Figure 9.27 Dehn surgery on a link in S^3.

But, just as in the case of gluing together pairs of handlebodies to get all compact three-manifolds, knowing that we get a complete list of compact three-manifolds by Dehn surgery is not sufficient to understand all such manifolds. We have no way of knowing which of the manifolds in the list are actually the same. For instance, just restricting ourselves to Dehn surgery on knots, we could ask whether there could be more than one surgery on a particular knot K in S^3 that yielded S^3. Of course, as we already noted, there is always at least one surgery on a knot that yields S^3, namely the surgery where we send a meridian of the solid torus to the meridian of the knot exterior. But could it be that a second surgery sending the meridian of the solid torus to some (p, q)-curve on the boundary of the knot exterior also yields S^3?

Suppose that there was such a second surgery. Then, after gluing the solid torus to the knot exterior of K, the result is S^3, in both cases. Therefore, S^3 is made up of this particular knot exterior and a solid torus, but in two different ways. Put another way, there exist two different solid tori in S^3 such that when we remove either one, we are left with the exterior of K. Since a solid torus in S^3 is always a tubular neighborhood of the knot that is its core curve, this means that there are two different knots in S^3 with the same exteriors. So what we are asking is, "Can there be two different knots K_1 and K_2 in S^3 with the same exterior?"

Here is a more visceral way to visualize the question. Suppose that the three-sphere S^3 is filled with green Jell-O. Suppose now we drill out a wormhole through the green Jell-O where the knot K_1 is. Could the green Jell-O that we are left with be in the same shape as the green Jell-O we would be left with if we drilled out the knot K_2 instead? The answer is an emphatic *no*, much to our relief. Two mathematicians, Cameron Gordon and John Luecke, both at the University of Texas, proved that two distinct knots cannot have the same exterior. Put another way, a knot is completely determined by its exterior. If we know what the exterior is, we essentially know what the knot is.

This question had first been posed by H. Tietze (1880–1964) in 1908. The solution didn't come until 1988. At the same time that Gordon and Luecke were working on the problem, David Gabai of Caltech was also closing in on the solution. Gordon and Luecke just barely won the race (Gordon and Luecke, 1989). Their proof utilizes results that are due to Gabai. They first proved their result for knots with low bridge numbers, solving it when the bridge number was 2, 3, 4, 5, and 6. From this, they saw how to do it for all bridge numbers.

It's hard to understand the significance of this result since it is hard to even conceive of the possibility of two different knots with the same exterior. Amazingly enough, there are examples of two different links in S^3 with the same exterior. In Figure 9.28, we see two links, which are in fact distinct. There is no deformation of the one link through space to the other. However, the two link exteriors are not distinct, they are homeomorphic. To see this, we utilize the idea of homeomorphism that we introduced back in Section 4.1. Two objects X and Y are homeomorphic if we can cut X open, rearrange the pieces, and then glue it back together so that any points that started together end together, and the result is Y. If we cut the exterior of the first link open along the disk with two holes, twist one copy of the disk 360°, and then reglue the two copies together, the result is the second link complement (Figure 9.29).

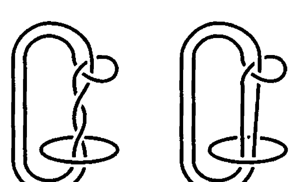


Figure 9.28 Two distinct links with homeomorphic exteriors.

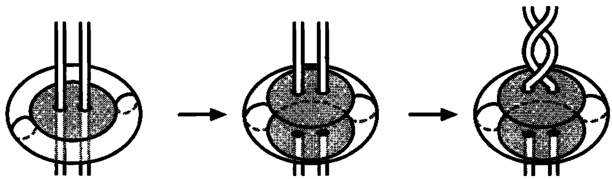


Figure 9.29 Seeing the homeomorphism between the two link exteriors.

Exercise 9.8 Show that the two links in Figure 9.30 have homeomorphic exteriors.

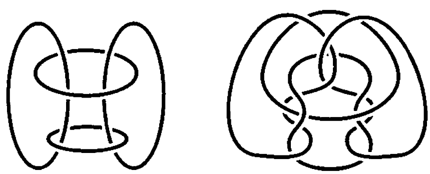


Figure 9.30 These links have homeomorphic exteriors.

Gordon and Luecke's result shows that exactly one Dehn surgery on a knot yields S^3. But could there be a second surgery on the knot that yields a manifold that is very similar to S^3? That is to say, could a second surgery yield a counterexample to the Poincaré conjecture, a compact, connected, simply connected, three-manifold that is not S^3? A knot that has the property that no surgery could possibly yield a counterexample to the Poincaré conjecture is said to satisfy **Property *P***.

Unsolved Question

Show that every knot satisfies Property P. If completed, this would be a partial result on the road to the Poincaré conjecture. (Unfortunately, the road is a long one, since, in order to prove the Poincaré conjecture, we would have to prove that no counterexamples to the Poincaré conjecture come from Dehn surgery on any link of any number of components.)

Property P is known to hold for certain classes of knots, including the rational knots. (Remember, rational knots are exactly the two-bridge knots.)

Let's return to this problem of deciding whether two different Dehn surgeries on two different links are actually yielding the same three-manifold. In fact, in 1978, Robion Kirby of University of California at Berkeley proved that there is a set of simple operations that one can do to a Dehn surgery description of a three-manifold, such that if two different Dehn surgery descriptions yield the same three-manifold, they must be related through a sequence of these operations. The manipulation of the Dehn surgery descriptions by these operations is known as the **Kirby calculus.**

The operations utilized in the Kirby calculus are reminiscent of the Reidemeister moves. Instead of getting us from one projection of a knot to another through a sequence of projections, each of which comes from the previous one by a single Reidemeister move, the operations get us from one Dehn surgery description of a three-manifold to another through a sequence of Dehn surgery descriptions, each of which comes from the previous one by a single operation. Unfortunately, just as in the case of the Reidemeister moves, where the number of moves necessary to get from one projection to another projection of the same knot has no apriori upper limit, the number of operations necessary to get from one Dehn surgery description of a three-manifold to another one also has no upper limit. Hence, the existence of these operations does not help us to determine whether or not two surgery descriptions yield the same manifold.

However, the Kirby calculus does form the basis for a means to generalize the polynomial invariants for knots and links to new invariants for distinguishing three-manifolds. First proposed by Edward Witten, a theoretical physicist at the Institute for Advanced Study in Princeton, the new invariants for three-manifolds come out of the theoretical area of physics known as quantum field theory. These new invariants can be realized as certain averages of link polynomials obtained from a given Dehn surgery representation of the manifold.

The approach to these invariants through averages of link polynomials is due to two Russian mathematicians named Nikolai Reshetikhin and V. G. Turaev, the first of whom is now at Berkeley and the second of whom is at the University of Strasbourg in France. The idea is to create the averages of the link polynomials in such a way that they are unchanged by the operations in the Kirby calculus. Then, since two different Dehn surgery representations of the same three-manifold will be related by a sequence of the Kirby operations, they will both yield the same value for the invariant. Here is an example where the field of knot theory, normally thought to be a subfield of the much larger field of topology, has had an impact felt well beyond its traditional boundaries. Much work is currently being done on these invariants.

10 Higher Dimensional Knotting

10.1 Picturing Four Dimensions

One of the basic mottos of mathematics is, "Generalize." Knot theory in three dimensions has been interesting, so why not go to four dimensions and see what happens?

In three dimensions, the theory of knots was the theory of knotted circles. A circle is essentially one-dimensional. If we were a tiny bug on a circle, it would appear one-dimensional to us. We could go either forward or backward on the circle and that's it. We say that a circle in three-space has **codimension two** since its dimensionality is two less than the dimension of the space that it lies in. When we go to four-space, we want to look at knotted objects that are the generalization of circles in three-space. We will still want the object to have codimension two. Since we will be in four-dimensional space, our object will need to be two-dimensional. So we want to look at knotting of surfaces (also called two-manifolds) in four-space. But what surfaces? What surface is the appropriate analog of a circle?

One example of a circle is the set of all points that are a distance of exactly one from the origin (0, 0) in the *xy* plane. To generalize this idea, let's take the set of all points that are a distance of exactly one from the origin (0, 0, 0) in three-space. The result is the sphere (Figure 10.1).

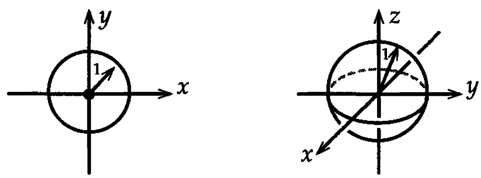


Figure 10.1 A sphere is the analog of a circle.

Exercise 10.1 If a sphere is the analog of a circle one dimension up, what's the analog of a circle one dimension down?

So now, instead of looking at how circles can knot in three-space, we want to look at how spheres can knot in four-space. First of all, let's see how we can picture four dimensions. It might help to think about ways to picture three dimensions first. The usual way to picture three dimensions is to use three spatial directions, often called the x, y, and z directions. This is the usual Cartesian coordinates (Figure 10.2).

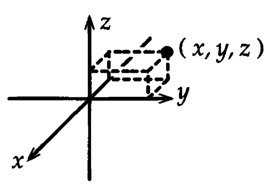


Figure 10.2 Picturing three-space using Cartesian coordinates.

But now, how about this alternative way to picture three-space? Instead, we'll use two spatial dimensions, x and z, and the third dimension will be time (Figure 10.3a). We think of three-space as if it had been sliced up into planes. As time passes, we move left to right through the sequence of planes, seeing what appears on that slice (Figure 10.3b). So how will a sphere look to us in this view of three-space? We just see a plane, and as time passes, we see the different two-dimensional slices of three-space. It's like watching a movie. First, the plane is empty. Then, at time $t = 1$, a point appears. As time passes, it grows into a circle. The circle continues to expand in radius until time $t = 3$. After that, it starts to shrink in size until time $t = 5$ when it becomes a point, and then disappears (Figure 10.4). We have just described a sphere in three-space using only two spatial dimensions.

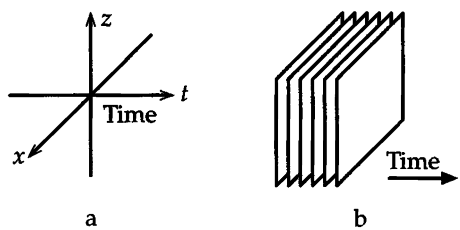


Figure 10.3 Time is a horizontal dimension.

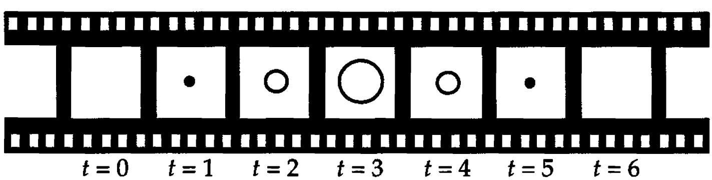


Figure 10.4 A sphere in three-space, when one of the dimensions is time.

Exercise 10.2 Describe the movie we see if we have a cube in three-space, where one of the dimensions is time. Do it both when the cube is parallel to the time direction and when it is not (see Figure 10.5).

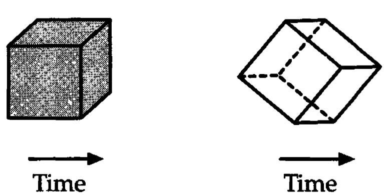


Figure 10.5 Describe the movies for these cubes.

What if we have a different embedding of the sphere? Our movies can become more interesting. For instance, Figure 10.6 depicts a sphere that is a bit deformed, and here is its movie. Great flick, huh? And what a plot.

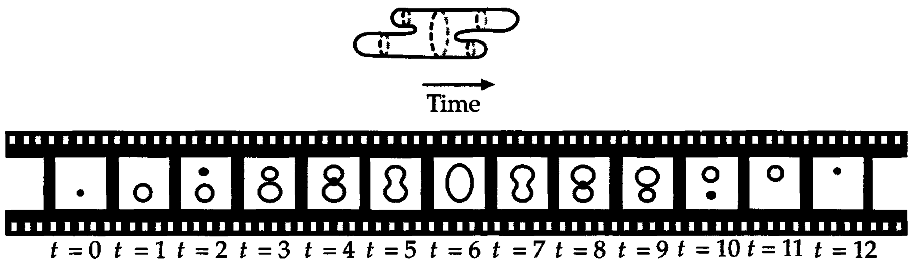


Figure 10.6 The movie for a deformed sphere.

Exercise 10.3 Draw the sequence of frames corresponding to the movie for the deformed sphere shown in Figure 10.7.

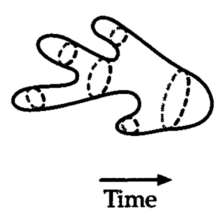


Figure 10.7 The sphere in Exercise 10.3.

Exercise 10.4 Draw the sequence of frames corresponding to the movie for the genus two surface shown in Figure 10.8.

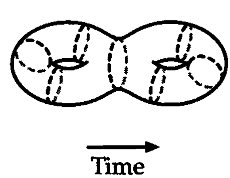


Figure 10.8 The genus two surface in Exercise 10.4.

We can now employ the same idea to describe four dimensions, using three spatial dimensions, x, y, and z, and one temporal dimension t. As time passes, we will see changes in three-space, but we will interpret this as if we are just seeing three-dimensional slices of four-space, one displayed after the other.

In four-dimensional space, instead of knotting circles that are one-dimensional, we knot spheres that are two-dimensional. First, let's look at an unknotted sphere in four-space. As we slice four-space, the sphere is sliced up into circles. But now the circles are in three-space, rather than in the xy plane (Figure 10.9).

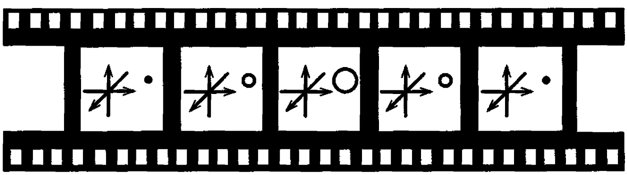


Figure 10.9. Movie for a sphere in four-space.

We can also describe the three-sphere S^3 in this model of four-space. Officially, the unit three-sphere is the set of all points a distance one from the origin in four-space. That is to say, it is the set of all points $\{(x, y, z, t : x^2 + y^2 + z^2 + t^2 = 1\}$ in four-space. The equation that defines the three-sphere can be rewritten as

$$x^2 + y^2 + z^2 = 1 - t^2$$

For any value of t less than -1 or greater than $+1$, this equation has no solutions. These values of t correspond to the time slices of four-space that miss the unit three-sphere entirely. Since we let the values of t vary from $t = -1$ to $t = +1$, where each value of t yields a three-dimensional slice of four-space, this equation simply describes two-spheres in three-space with radii varying between 0 and 1. For instance, when $t = \frac{1}{2}$, the equation of the three-sphere becomes $x^2 + y^2 + z^2 = \frac{3}{4}$, meaning that the three-sphere intersects this three-dimensional slice of four-space in a sphere of radius $\frac{\sqrt{3}}{2}$.

Exercise 10.5 Describe the movie of the unit three-sphere in four-space that results from the preceding paragraph.

It is not always convenient to make the fourth dimension time, since if we want to imagine ourselves in four-dimensional space, we would like time to be able to pass normally. So let's look at a second way to think about four-dimensional space that does not involve time. We will let the first three dimensions again be spatial, but this time the fourth dimension will be a **color dimension**. As we change color, we think of ourselves as moving in the color direction. For example, when we are at the color yellow, it's as if we had put yellow-colored glasses on, so the three-dimensional world around us looks yellow. Now imagine there is a knob on the glasses that allows color adjustment. As we change the color either toward orange on the one hand or green on the other, it's as if we are walking in the fourth possible direction in this four-dimensional world, namely the color direction. Say we move toward green. Now we are in the green three-dimensional world, which looks entirely different from the yellow world. There are green people here, who were not in the yellow world. Of course, just as we cannot suddenly disappear when we are in Massachusetts and reappear in California, we cannot jump colors either. If we want to get from red to yellow, we have to pass through orange and all of the shades of orange in between red and yellow to get there.

We will use the color model of four-space to show that there are *no* knotted loops in four-space. This is why no one studies knotted loops in four-space. Every knotted loop in four-space is equivalent to the unknot. Suppose we start with a knot in four-space that is entirely green. Then by passing the knot through itself some number of times, we can always make it into the unknot. In fact, the number of times that we need to let it pass through itself in order to obtain the unknot is exactly the unknotting number we introduced in Chapter 3.

Suppose we have a point x along the knot where we would like to pass the knot through itself. In a short strand of the knot containing this point, change the color of the knot to yellow. We are essentially pushing this strand of the knot in the fourth color direction. Note that we can't just make the knot yellow along this strand and leave it green everywhere else. Rather, the strand must gradually become yellow. The knot goes from green to green-yellow and then eventually to yellow as we approach the point x along the knot. After we pass through x, the knot goes back to green-yellow and eventually to green again. We depict this for the figure-eight knot in Figure 10.10 on a black and white scale. Think of the black as green and the white as yellow. Note that this slight change in the color of a strand of the knot is an isotopy of the knot in four-space. That is to say, it is a rubber deformation of the knot, with the resultant knot still equivalent to the original knot.

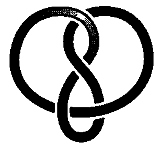


Figure 10.10 The figure-eight knot in colorized four-space.

But remember, the green and yellow three-dimensional worlds are distinct. If we have the glasses adjusted to green, we can't see yellow objects—they do not exist in our green world. Therefore we can move the yellow strand of the knot right through the green part of the knot. The two strands cannot see each other, they exist in different three-dimensional slices of four-space. We can repeat this operation until we have passed the appropriate strands through one another in order to unknot the knot. Then, we can push the yellow strands back into the green three-dimensional world. The result is a green unknot (Figure 10.11). Similarly, any knotted loop in four-space is equivalent to the unknot in four-space. The study of knotted loops in four-space is completely boring, since there is only one such loop, the trivial knot. Therefore, we will look at knotted two-spheres in four-space.

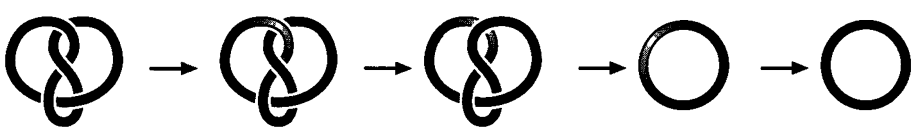


Figure 10.11. The figure-eight knot is equivalent to the unknot in four-space.

Exercise 10.6 Explain why a sphere in four-space does not separate four-space into two separate pieces (as it does in three-space). That is to say, show that a red bug that is outside a red sphere in four-space could get inside the red sphere. (Note that we are assuming that a bug that lives in four-dimensional space can change its color, but again it cannot skip colors in between, in going from one color to another.)

Exercise 10.7 Describe how to build a house in four-dimensional space that can keep out such bugs.

10.2 Knotted Spheres in Four Dimensions

The sphere in four-space that we looked at in Figure 10.6 was unknotted. It was also in a particularly nice position in space so that all of the slices other than the first and last were circles. It plays the role for spheres in four-space that the unknot played for loops in three-space.

If we have an object in four-space so that in the time model, a point initially appears, that point grows into a loop and then the loop shrinks to a point and disappears, that object is a sphere. However, there was no rule that said that the loop had to be unknotted. Let's look at another sphere in four-space with the movie shown in Figure 10.12. The slices of this sphere are simply knots. Although an interesting example in its own right, and the simplest way to knot a sphere in four-space, this example has one major flaw. Unfortunately, the sphere cannot be modeled with functions that are differentiable. At the endpoints, where the sequence of knots shrink down to point, the derivatives of the functions that describe the sphere cease to exist. Mathematically, this is a tremendous disadvantage, and so most mathematicians restrict themselves to manifolds that can be described by differentiable functions. In fact, the ideal situation is when we have a manifold described by functions, all of whose derivatives exist. Such manifolds are called **smooth** manifolds. We would like an example of a knotted sphere that is smooth. In particular, we do not want a sequence of nontrivial knots shrinking to a point. Let's take a look at the movie in Figure 10.13. At particular times, we see several curves come together at vertices, and then open up in the opposite direction. These are saddle points, just as we had in the movie for a sphere in three-space that was depicted in Figure 10.6.

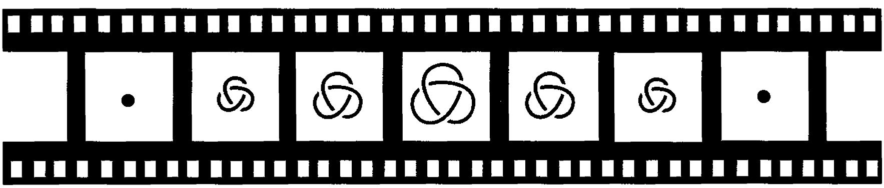


Figure 10.12. The movie of a knotted two-sphere in four-space.

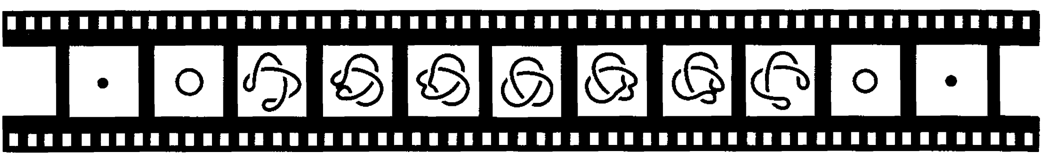


Figure 10.13. A smooth knotted two-sphere in four-space.

How do we know that this movie actually represents a sphere rather than some other surface? We can use Euler characteristic. (Check for yourself that the Euler characteristic of this object is 2. You need to think a bit about how to calculate the Euler characteristic for something like this.) Could it be that this sphere is unknotted? Maybe we could deform it through four-space in order to unknot it. In fact, the answer is no, there is no way to unknot it, but a proof is beyond the scope of this book.

Exercise 10.8 Draw a movie of a "knotted" genus two surface in four-space.

Exercise 10.9 Draw a movie of a sphere linked with a torus in four-space.

10.3 Knotted Three-Spheres in Five-Space

I know, it seems a little crazy. Two-dimensional spheres in four-space was getting out there, but this is ridiculous. Amazingly enough, however, if we use both the color model and the temporal model at the same time, we can picture five-space.

We will use three spatial dimensions x, y, and z, one time dimension t, and one color dimension c. So if we want to arrange a meeting with someone in this five-dimensional space, we would say, "Okay, I'll meet you in the building at the corner of 76th and Lexington (x and y coordinates), on the twenty-seventh floor (z coordinate) at 5:30 (t coordinate) in the color blue (color coordinate)." If we are to be able to move freely in this five-dimensional space, we have to think of ourselves as being able to travel in time and in color, as well as being able to travel in space. It's as if we could change our color to blue and thus be in the blue slice of five-space.

What does a three-sphere in five-space look like? We saw that the two-dimensional slices of a two-sphere in three-space start out as a point that grows into a circle and then shrinks to a point again, and in Exercise 10.5, we saw the slices of a three-sphere in four-space start as a point that grows into a sphere and then shrinks back down to a point again. When we go up to five-space, the slices of the three-sphere look exactly the same, only now they are sitting in four-dimensional slices of five-space. We let each of these four-dimensional slices have a different color. So there is a red slice denoted by R, an orange slice denoted by O, a yellow slice denoted by Y, and so forth. Each of these colored four-dimensional slices of five-space will have three spatial dimensions and one temporal dimension (Figure 10.14).

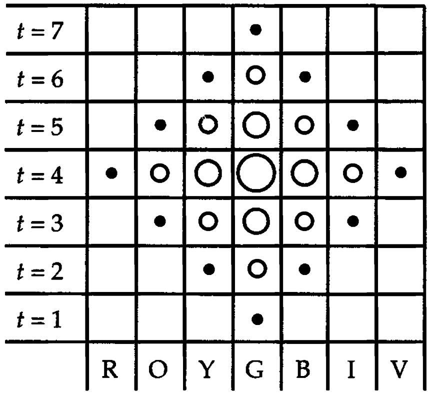


Figure 10.14. A three-sphere in five-dimensional space.

How about a knotted three-sphere in five-dimensional space? We will have each colored four-dimensional slice of five-space intersect our three-sphere in a sphere, point, or not at all. So for a particular color, say green, we will intersect the three-sphere in a set of time slices that together make up a two-sphere, and appear in the vertical column of pictures above G in Figure 10.15. Each of these colored slices that is represented by a column of pictures can intersect the three-sphere in either a point, as red and violet do, an unknotted two-sphere, as for instance orange and indigo do, or a knotted two-sphere, as green does. The yellow and blue columns are transitional stages between the colors corresponding to unknotted spheres and the colors corresponding to knotted spheres. This example of a knotted three-sphere in five-space is a smooth manifold (Figure 10.15).

Note that if we use time as one of our dimensions when representing five-space, we cannot then imagine ourselves moving around in the space. We would like to save time for our own usage. We could therefore pick some other attribute to use to keep track of a dimension. For instance, we could let brightness be a dimension. As brightness varied, we would be moving in the "brightness" direction of space. We could even use a hum that was constantly in the background and that got louder or softer as we moved in the directions corresponding to the sound. With attributes like these representing the extra dimensions, we can imagine time passing normally, and picture what it would be like to move around in five-dimensional space. For that matter, we could imagine moving around in space of any dimension.

Exercise 10.10 Figure out how to represent a four-sphere in six-space. "Draw" an unknotted four-sphere in six-space.

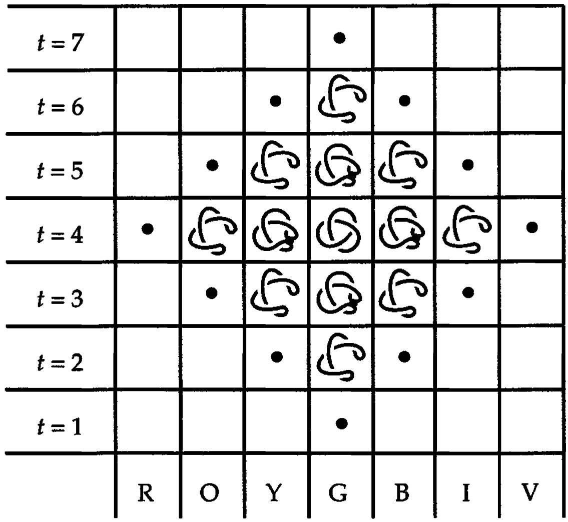


Figure 10.15. A knotted three-sphere in five-dimensional space.

Exercise 10.11 Draw a knotted four-sphere in six-space.

On that note, we will call it quits. The references at the end of the book contain numerous articles and books for further reading. Many of them can be read without any additional math background. However, some of them assume a course in topology and/or algebra, and a few of them assume one or two courses in algebraic topology. I have tried to be explicit about which these are. Have fun!

Knot Jokes and Pastimes

Jokes

Marty Scharlemann tells the story of a calculus student who came in for help, and after Marty had worked some problems, the student said, "So what kind of math do you like?"

Marty said, "Knot theory."

The student said, "Yeah, me either."

Three strings went into a bar and sat down at a table. The first string said to the others, "Is there a waitress here?"

The second one said, "No, you have to go up to the bar."

So the first one got up, went over to the bar, and said to the bartender, "I'll have three Scotches."

The bartender said, "We don't serve your kind in here."

"What kind?" said the string.

"Strings, we don't serve strings here."

So the string went back to the table and said to the other strings, "They won't serve us here."

The second string said, "Oh yeah, we'll see about that."

He got up, went over to the bar, pounded on the bartop, and said, "Hey bartender, I want three Scotches."

The bartender said, "I told your friend, and now I'm telling you, we don't serve strings here. Now beat it."

The string went back to the table and shrugged. The third string stood up. "Let me handle this," he said.

He tied himself into a nasty tangle and pulled the strands out at his end, creating a wild mop of a hairdo. Then he walked over to the bar, leaned over close, and said, "Bartender, I want three Scotches and I want them now."

The bartender turned around and looked at him. He looked him up and down. Then he said, "You're not fooling me, you're one of those strings, aren't you?"

The string looked him straight back in the eye and said, "Nope, I'm a frayed knot."

A woman walks into a bar with a cow and a dog. The bartender says, "Hey, we don't allow animals in here."

The woman says, "Oh, but these aren't your usual animals. These animals are knot theorists."

"Yeah, right," says the bartender. "I have known some knot theorists who I considered animals, but I have yet to meet an animal that I consider a knot theorist."

"Well, see for yourself," says the woman.

"Okay." He turns to the dog.

"Name a knot invariant," he says.

The dog says, "Arf, arf!!!" (See Section 8.2.)

The bartender is not impressed. He says to the cow, "Okay, you name a topological invariant."

The cow says, "μ, μ."

"Who are you trying to kid," says the bartender, "Get outta here."

As the three of them are dejectedly leaving the bar, the dog says to the cow, "Maybe I should have said the Jones polynomial."

(Thanks to Joel Hass for making up that one.)

Pastimes

Passing the Time of Day: Here is something tough to try. Take a long piece of string (3 ft)—nylon string works the best—about a quarter inch in diameter. Tie a weight at one end, something that is not too heavy. Now, holding one end of the string and letting the weighted end hang down, can you jerk the hand holding the string so that the free end of the string knots around itself?

Knot Games: You may have played this game at camp when you were a kid. A large group (5 to 10) of people stand in a circle shoulder to shoulder facing inward. Everyone puts their hands in the center and grabs random other hands. Once all of the hands have been paired up, the result is a human knot or link. The goal is then to untangle it without releasing any pairs of hands. But the interesting point that our counselors never mentioned (and maybe didn't realize) is that the people may have formed a nontrivial knot or link, in which case, they will never succeed in untangling it. This game also brings up some interesting questions. How often will a nontrivial knot be formed? How often will a link be the result?

A New Knot Invariant (or how to get to know people very well, very fast): Remember in Section 1.6 we discussed making a knot out of sticks, and we suggested trying your luck making a knot from the five sticks corresponding to your two forearms, your two upper arms, and an imaginary stick that runs from your left shoulder to your right. So we have these five sticks to work with. Given a knot K, define the human knot number of that knot, denoted $h(K)$, to be the least number of people necessary, when holding hands, to make that knot. Interestingly, even though a knot can be constructed from ten sticks, it does not mean that the human knot number is necessarily 2. Heads and bodies seem to get in the way.

Try showing $h(\text{trefoil knot}) = 2$, $h(\text{figure-eight knot}) = 2$. *Conjecture:* These are the only two knots with human knot number 2. (I haven't explored knots with human knot number 3—there aren't two other people that I know that well.)

An Unusual Way to Construct a Knot: Take a long thin piece of paper, put three half-twists in it, and then glue the ends together. (This is a particular embedding of the Möbius band.) Now, cut it open along the center line of the Möbius band. The result is a single band that is twice as long and that is tied into a trefoil knot. Note that this is exactly the method attempted by chemists to construct a trefoil molecule that we discussed in Section 7.2.

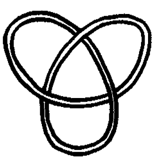
Appendix

Table of Knots, Links, and Knot and Link Invariants

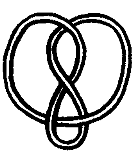
This table contains pictures, notation, and invariants for all the knots through nine crossings, the two-component links through eight crossings, and the three-component links through seven crossings. The pictures are from the book *Knots and Links*, by Dale Rolfsen (Berkeley, Calif.: Publish or Perish Press, 1976). They were drawn by Ali Roth.

On the first line following the picture of a given knot, we give the Alexander and Briggs notation for that knot (dating from 1926), followed by the Conway notation for the knot. (See Conway's paper, which is listed in the references for Chapter 2, for more details on this notation.) On the next line, we give the hyperbolic volume of the knot complement, without rounding, out to eight decimal places. A volume of 0.0 denotes the fact that the knot is not hyperbolic. These volumes come from the paper [A-H-W], which appears in the references to Chapter 5.

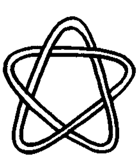
The last line contains a sequence of numbers that denote the Jones polynomial of the knot. The first number, which appears in the curly brackets, is the minimum degree of the polynomial. The next sequence of numbers gives the coefficients of the polynomial, beginning with the coefficient of the minimum degree term. For example, {−3} (1 −2 3 −3 3 −1 1) denotes the polynomial $t^{-3} - 2t^{-2} + 3t^{-1} - 3 + 3t^{1} - t^{2} + t^{3}$. The polynomials for knots were provided by Morwen Thistlethwaite. The polynomials for links come from a table produced by Helmut Doll and Jim Hoste in "A tabulation of oriented links," Mathematics of Computation, Vol. 57, No. 196(1991), 747–761. Note that many of the polynomials for links have fractional exponents. In the case of links, the given polynomial is for a particular choice of orientations on the components. Different orientations will often yield different polynomials. See the Doll-Hoste paper for a complete list of polynomials for all orientations.



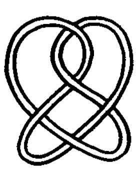
3_1 3
0.0
{−4} (−1 1 0 1)



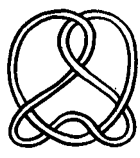
4_1 22
2.02988321
{−2} (1 −1 1 −1 1)



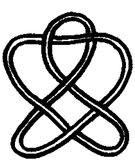
5_1 5
0.0
{−7} (−1 1 −1 1 0 1)



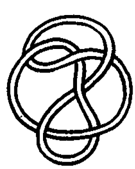
5_2 32
2.8281220
{−6} (−1 1 −1 2 −1 1)



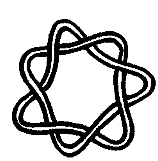
6_1 42
3.16396322
{−4} (1 −1 1 −2 2 −1 1)



6_2 312
4.40083251
{−5} (1 −2 2 −2 2 −1 1)

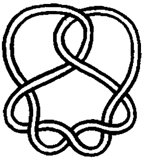


6_3 2112
5.69302109
{−3} (−1 2 −2 3 −2 2 −1)

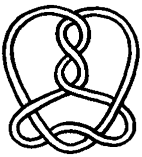


7_1 7
0.0
{−10} (−1 1 −1 1 −1 1 0 1)

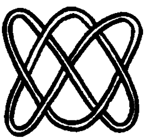
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



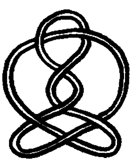
7_2 52
3.33174423
{−8} (−1 1 −1 2 −2 2 −1 1)



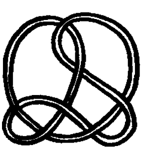
7_3 43
4.59212569
{−9} (−1 1 −2 3 −2 2 −1 1)



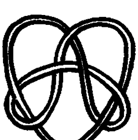
7_4 313
5.13794120
{−8} (−1 1 −2 3 −2 3 −2 1)



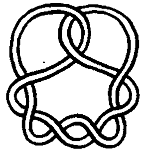
7_5 322
6.44353738
{−9} (−1 2 −3 3 −3 3 −1 1)



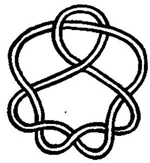
7_6 2212
7.08492595
{−6} (−1 2 −3 4 −3 3 −2 1)



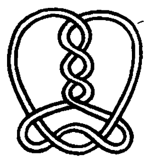
7_7 21112
7.64337517
{−3} (−1 3 −3 4 −4 3 −2 1)



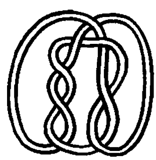
8_1 62
3.42720524
{−6} (1 −1 1 −2 2 −2 2 −1 1)



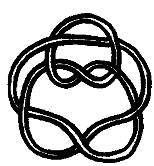
8_2 512
4.93524267
{−8} (1 −2 2 −3 3 −2 2 −1 1)



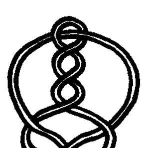
8_3 44
5.23868410
{−4} (1 −1 2 −3 3 −3 2 −1 1)



8_4 413
5.50048641
{−3} (1 −1 2 −3 3 −3 3 −2 1)

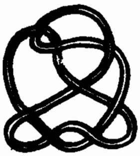


8_5 3,3,2
6.99718914
{−8} (1 −2 3 −4 3 −3 3 −1 1)

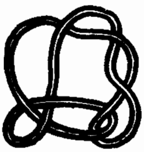


8_6 332
7.47523742
{−7} (1 −2 3 −4 4 −4 3 −1 1)

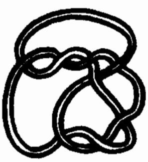
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



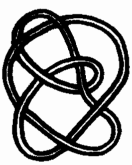
8_7 4112
7.02219658
{−2} (−1 2 −2 4 −4 4 −3 2 −1)



8_8 2312
7.801341224
{−3} (−1 2 −3 5 −4 4 −3 2 −1)

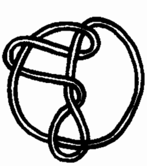


8_9 3113
7.58818022
{−4} (1 −2 3 −4 5 −4 3 −2 1)

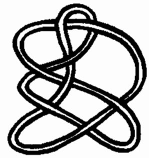


8_{10} 3,21,2
8.65114855
{−2} (−1 2 −3 5 −4 5 −4 2 −1)

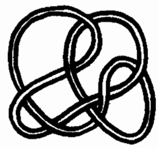
8_{11} 3212
8.28631681
{−7} (1−2 3 −5 5 −4 4 −2 1)



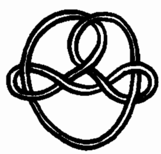
8_{12} 2222
8.93585692
{−4} (1−2 4 −5 5 −5 4 −2 1)



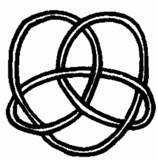
8_{13} 31112
8.53123220
{−3} (−1 3 −4 5 −5 5 −3 2 −1)



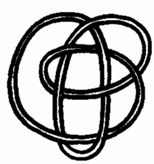
8_{14} 22112
9.21780031
{−7} (1 −3 4 −5 6 −5 4 −2 1)



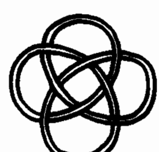
8_{15} 21,21,2
9.93064829
{−1 0} (1 −3 4 −6 6 −5 5 −2 1)



8_{16} .2.20
10.57902191
{−2} (−1 3 −4 6 −6 6 −5 3 −1)

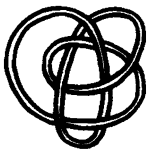


8_{17} .2.2
10.98590760
{−4} (1−3 5 −6 7 −6 5 −3 1)

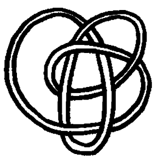


8_{18} 8*.
12.35090620
{−4} (1−4 6 −7 9 −7 6 −4 1)

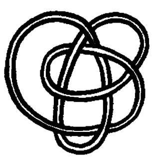
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



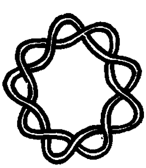
8_{19} 3,3,2−
0.0
{−8} (−1 0 0 1 0 1)



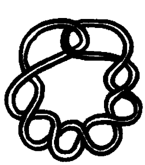
8_{20} 3,21,2−
4.12490325
{−1} (−1 2 −1 2 −1 1 −1)



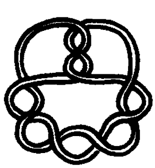
8_{21} 21,21,2−
6.78371351
{−7} (1−2 2 −3 3 −2 2)



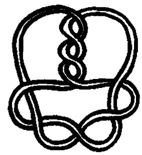
9_1 9
0.0
{−13} (−1 1 −1 1 −1 1 −1 1 0 1)



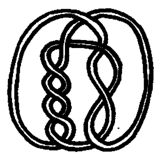
9_2 72
3.48666014
{−10} (−1 1−1 2 −2 2 −2 2 −1 1)



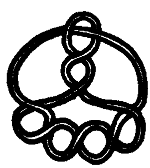
9_3 63
4.99485640
{−12} (−1 1 −2 3 −3 3 −2 2 −1 1)



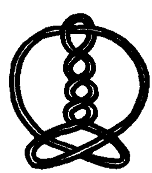
9_4 54
5.55651881
{−11} (−1 1 −2 3 −3 4 −3 2 −1 1)



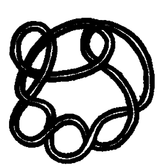
9_5 513
5.69844175
{−10} (−1 1 −2 3 −3 4 −3 3 −2 1)



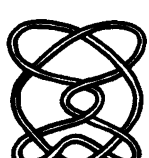
9_6 522
7.20360076
{−12} (−1 2 −3 4 −5 4 −3 3 −1 1)



9_7 342
8.01486145
{−11} (−1 2 −3 4 −5 5 −4 3 −1 1)

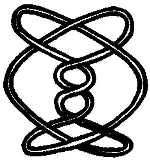


9_8 2412
8.19234796
{−6} (−1 2 −3 5 −5 5 −4 3 −2 1)

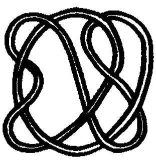


9_9 423
8.01681556
{−12} (−1 2 −4 5 −5 5 −4 3 −1 1)

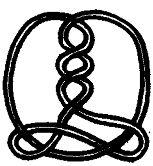
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



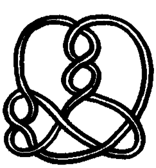
9_{10} 333
8.77345728
{−11} (−1 1 −3 5 −5 6 −5 4 −2 1)



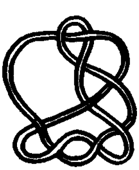
9_{11} 4122
8.28858904
{0} (1−2 3 −4 6 −5 5 −4 2 −1)



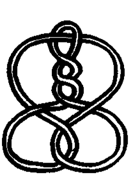
9_{12} 4212
8.83664234
{−8} (−1 2 −3 5 −6 6 −5 4 −2 1)



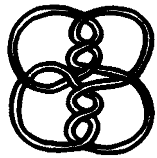
9_{13} 3213
9.13509403
{−11} (−1 2 −4 5 −6 7 −5 4 −2 1)



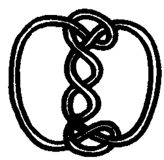
9_{14} 41112
8.95498926
{−3} (−1 3 −4 6 −6 6 −5 3 −2 1)



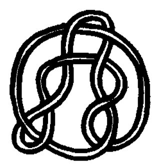
9_{15} 2322
9.88549866
{−1} (1−2 4 −6 7 −6 6 −4 2 −1)



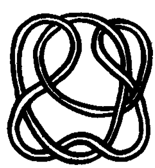
9_{16} 3,3,2+
9.88300696
{−12} (−1 3 −5 6 −7 6 −5 4 −1 1)



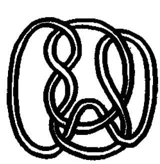
9_{17} 21312
9.47458045
{−6} (−1 3 −4 6 −7 6 −5 4 −2 1)



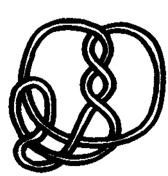
9_{18} 3222
10.05772963
{−11} (−1 2 −4 6 −7 7 −6 5 −2 1)



9_{19} 23112
10.03254744
{−5} (−1 3 −4 6 −7 7 −6 4 −2 1)

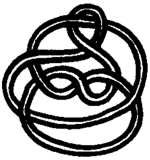


9_{20} 31212
9.64430407
{−9} (−1 3 −5 6 −7 7 −5 4 −2 1)

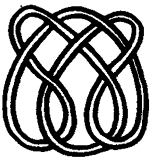


9_{21} 31122
10.18326553
{−1} (1−3 5 −6 8 −7 6 −4 2 −1)

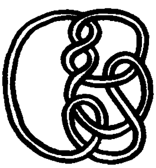
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



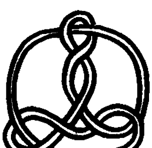
9_{22} 211,3,2
10.62072702
{−6} (−1 3 −5 7 −7 7 −6 4 −2 1)



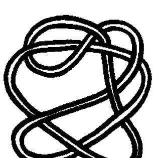
9_{23} 22122
10.61134829
{−11} (−1 3 −5 6 −8 8 −6 5 −2 1)



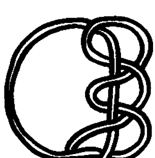
9_{24} 3,21,2+
10.83372910
{−5} (−1 2 −4 7 −7 8 −7 5 −3 1)



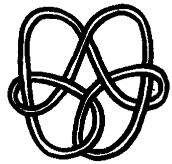
9_{25} 22,21,2
11.39030514
{−8} (−1 3 −5 7 −8 8 −7 5 −2 1)



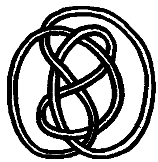
9_{26} 311112
10.59584051
{−2} (−1 3 −4 7 −8 8 −7 5 −3 1)



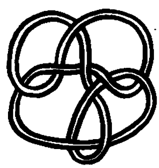
9_{27} 212112
10.99998095
{−5} (−1 3 −5 7 −8 9 −7 5 −3 1)



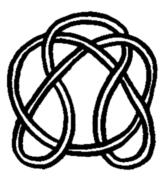
9_{28} 21,21,2+
11.56317701
{−7} (1 −3 5 −8 9 −8 8 −5 3 −1)



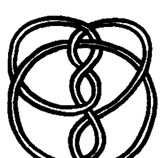
9_{29} .2.20.2
12.20585615
{−3} (1 −3 5 −7 9 −8 8 −6 3 −1)



9_{30} 211,21,1
11.95452696
{−4} (1 −3 6 −8 9 −9 8 −5 3 −1)

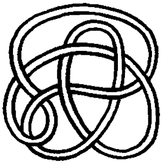


9_{31} 2111112
11.68631220
{−7} (1 −4 6 −8 10 −9 8 −5 3 −1)

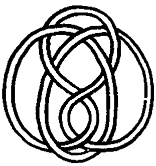


9_{32} .21.20
13.09989984
{−2} (−1 4 −6 9 −10 10 −9 6 −3 1)

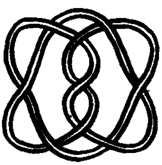
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



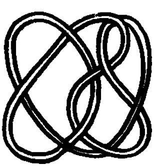
9_{33} .21.2
13.28045563
{−4} (1 −4 7 −9 11 −10 9 −6 3 −1)



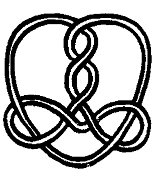
9_{34} 8*20
14.34458138
{−4} (1 −4 8 −10 12 −12 10 −7 4 −1)



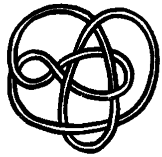
9_{35} 3,3,3
7.94057924
{−10} (−1 1 −3 4 −3 5 −4 3 −2 1)



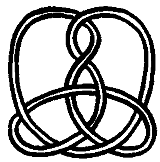
9_{36} 22,3,2
9.88457865
{0} (1 −2 4 −5 6 −6 6 −4 2 −1)



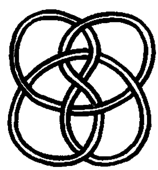
9_{37} 3,21,21
10.98944959
{−5} (−1 3 −4 7 −8 7 −7 5 −2 1)



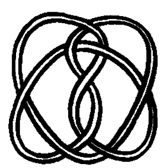
9_{38} .2.2.2
12.93285870
{−11} (−1 3 −6 8 −10 10 −8 7 −3 1)



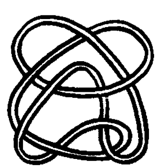
9_{39} 2:2:20
12.81031000
{−1} (1 −3 6 −8 10 −9 8 −6 3 −1)



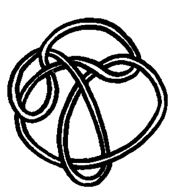
9_{40} 9*
15.01834285
{−7} (1 −4 8 −11 13 −13 11 −8 5 −1)



9_{41} 20:20:20
12.09893602
{−3} (−1 3 −5 8 −8 8 −7 5 −3 1)

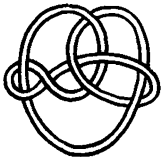


9_{42} 22,3,2−
4.05686022
{−3} (1 −1 1 −1 1 −1 1)

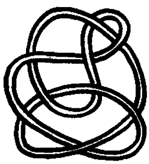


9_{43} 211,3,2−
5.90408585
{−7} (−1 2 −2 2 −2 2 −1 1)

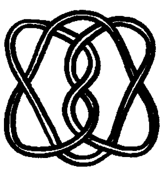
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



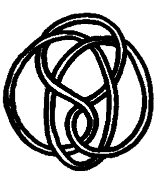
9_{44} 22,21,2−
7.40676757
{−5} (−1 2 −2 3 −3 3 −2 1)



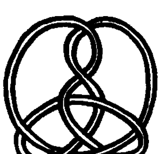
9_{45} 211,21,2−
8.60203116
{1} (2 −3 4 −4 4 −3 2 −1)



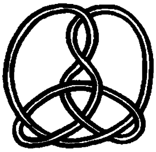
9_{46} 3,3,21−
4.75170196
{0} (2 −1 1 −2 1 −1 1)



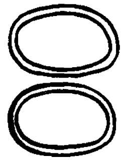
9_{47} 8*−20
10.04995786
{−5} (2 −4 4 −5 5 −3 3 −1)



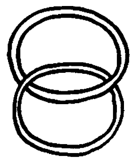
9_{48} 21,21,21−
9.53187983
{−1} (1 −3 4 −4 6 −4 3 −2)



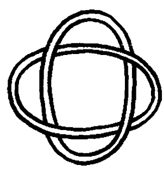
9_{49} −20:−20:−20
9.42707362
{−9} (−2 3 −4 5 −4 4 −2 1)



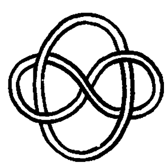
0^2_1 0
0.0
$\{\frac{-1}{2}\}$ (−1 −1)



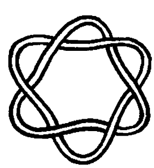
2^2_1 2
0.0
$\{\frac{1}{2}\}$ (−1 0 −1)



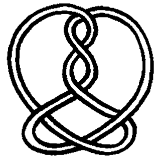
4^2_1 4
0.0
$\{\frac{1}{2}\}$ (−1 1 −1 0 −1)



5^2_1 212
3.66386237
$\{\frac{-7}{2}\}$ (1 −2 1 −2 1 −1)

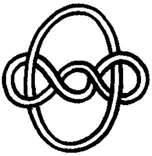


6^2_1 6
0.0
$\{\frac{5}{2}\}$ (−1 0 −1 1 −1 1 −1)

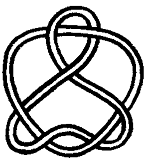


6^2_2 33
4.05976642
$\{\frac{3}{2}\}$ (−1 1 −2 2 −2 1 −1)

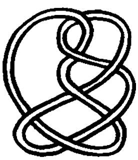
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



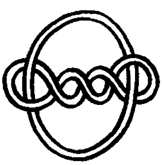
6^2_3 222
5.33348956
$\{\frac{3}{2}\}$ (−1 1 −3 2 −2 2 −1)



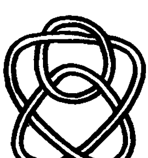
7^2_1 412
4.74949998
$\{\frac{-13}{2}\}$ (1 −2 2 −3 2 −2 1 −1)



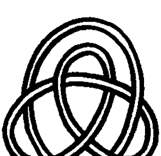
7^2_2 3112
6.59895153
$\{\frac{-3}{2}\}$ (−1 2 −3 3 −4 2 −2 1)



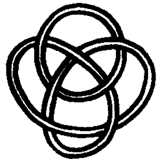
7^2_3 232
6.13813878
$\{\frac{-3}{2}\}$ (−1 1 −3 3 −3 2 −2 1)



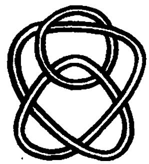
7^2_4 3,2,2
6.13813878
$\{\frac{-13}{2}\}$ (1 −2 3 −3 2 −3 1 −1)



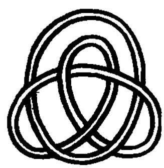
7^2_5 21,2,2
7.70691180
$\{\frac{3}{2}\}$ (−1 2 −4 3 −4 3 −2 1)



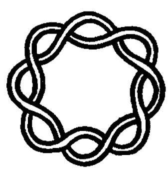
7^2_6 .2
8.99735194
$\{\frac{-9}{2}\}$ (−1 3 −4 4 −5 3 −3 1)



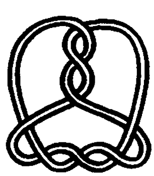
7^2_7 3,2,2−
0.0
$\{\frac{-3}{2}\}$ (1 −1 0 −1 0 −1)



7^2_8 21,2,2−
3.66386237
$\{\frac{-11}{2}\}$ (1 −1 1 −2 1 −2)

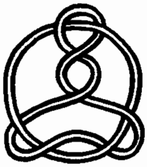


8^2_1 8
0.0
$\{\frac{7}{2}\}$ (−1 0 −1 1 −1 1 −1 1 −1)

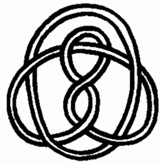


8^2_2 53
4.85117075
$\{\frac{5}{2}\}$ (−1 1 −2 2 −3 3 −2 1 −1)

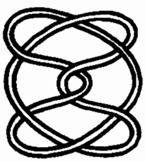
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



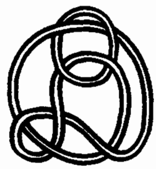
8^2_3 422
6.94755544
$\{\frac{5}{2}\}$ (−1 1 −3 3 −4 4 −3 2 −1)



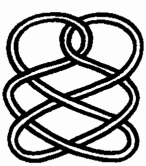
8^2_9 22,2,2
8.96736084
$\{\frac{-3}{2}\}$ (−1 2 −4 5 −5 4 −4 2 −1)



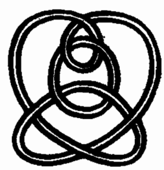
8^2_4 323
7.51768989
$\{\frac{3}{2}\}$ (−1 2 −4 4 −4 4 −3 1 −1)



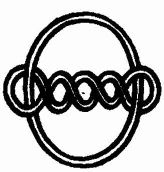
8^2_{10} 211,2,2
9.65949854
$\{\frac{-9}{2}\}$ (−1 3 −5 5 −6 5 −4 2 −1)



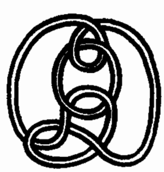
8^2_5 3122
7.89459448
$\{\frac{3}{2}\}$ (−1 2 −4 4 −5 4 −3 2 −1)



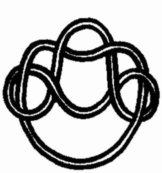
8^2_{11} 3,2,2+
8.79334560
$\{\frac{5}{2}\}$ (−1 1 −4 4 −5 5 −4 3 −1)



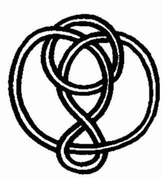
8^2_6 242
6.55174328
$\{\frac{3}{2}\}$ (−1 1 −3 3 −4 3 −2 2 −1)



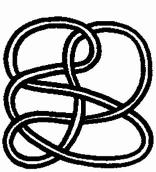
8^2_{12} 21,2,2+
9.65949854
$\{\frac{-11}{2}\}$ (1 −2 4 −6 5 −6 4 −3 1)



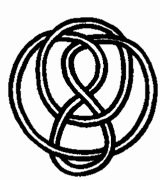
8^2_7 21212
8.83066495
$\{\frac{-7}{2}\}$ (−1 2 −4 4 −6 5 −4 3 −1)



8^2_{13} .21
11.37077417
$\{\frac{-5}{2}\}$ (1 −4 5 −7 7 −7 5 −3 1)

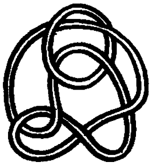


8^2_8 211112
9.67280773
$\{\frac{-5}{2}\}$ (1 −3 4 −6 6 −6 4 −3 1)

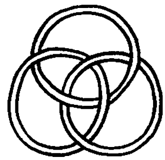


8^2_{14} .2:2
10.66697913
$\{\frac{3}{2}\}$ (−1 3 −6 5 −7 6 −4 3 −1)

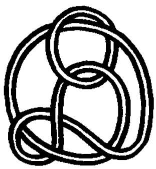
Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).



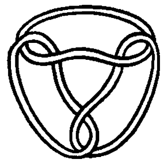
8^2_{15} 22,2,2−
3.66386237
$\{\frac{-5}{2}\}$ (−1 1 −2 1 −1 1 −1)



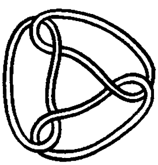
6^3_2 .1
7.32772475
$\{\frac{-6}{2}\}$ (−1 3 −2 4 −2 3 −1)



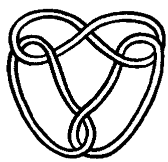
8^2_{16} 211,2,2−
5.33348956
$\{\frac{1}{2}\}$ (−2 2 −2 2 −2 1 −1)



6^3_3 2,2,2−
0.0
$\{\frac{-8}{2}\}$ (1 0 1 0 2)



6^3_1 2,2,2
5.33348956
$\{\frac{2}{2}\}$ (1 −2 3 −1 3 −1 1)



7^3_1 2,2,2+
7.70691180
$\{\frac{-6}{2}\}$ (−1 3 −3 4 −3 4 −1 1)

Illustrations from *Knots and Links* by Dale Rolfsen (Publish or Perish Press, 1976).

Suggested Readings and References

Chapter 1

Appel, K., and W. Haken. 1977. Every planar map is four colorable, I and II. *Illinois J. Math.* 21:429–567.

This pair of papers gives the solution of the renowned four-color theorem, that any planar map can be colored using only four colors so that two countries that share a border will never have the same color. The authors reduce the problem to thousands of special cases that are then checked by a computer. It does bring up the interesting question, how do you know that the computer program is completely free of bugs?

Appel, K., and W. Haken. 1977. The solution of the four-color-map problem. *Sci. Amer.* (September):108–121.

A readable account of the idea of the proof of the four-color theorem.

Ashley, C. 1944. *The Ashley Book of Knots*. New York: Doubleday.

An amazing book. Everything you ever wanted to know about knots from the non-mathematical point of view. Beautiful illustrations of innumerable knots. A treasure, if you can find it.

Burde, G., and H. Zieschang. 1986. *Knots*. Berlin: de Gruyter.

This is one of only a few books on the mathematical theory of knots. Lots of interesting material and a superb reference, but it does assume a mathematically sophisticated reader familiar with algebraic topology.

Crowell, R. H., and R. H. Fox. 1963. *Introduction to the Knot Theory*. New York/Berlin: Springer-Verlag.

This was one of the few books on knot theory for many years. It gives a wonderful introduction to the fundamental group of the complement of a knot.

Fox, R. H. 1962. A quick trip through knot theory. In *Topology of 3-manifolds and Related Topics*, 120–167. Englewood Cliffs, N.J.: Prentice-Hall.

An easy-to-read introduction to the state of knot theory at the time. Recommended reading for those who have a familiarity with algebraic topology.

Gordon, C. McA. 1978. Some aspects of classical knot theory. In *Lecture Notes in Mathematics*, 685:1–60. New York/Berlin: Springer-Verlag.

An interesting overview that predates the new polynomials.

Haken, W. 1961. Theorie der Normalflachen. *Acta Math.* 105:245–375.

Here is the paper where Haken gives an algorithm to determine whether a given knot is the unknot. Unfortunately, the alogorithm remains too complex to use in even the simplest cases.

Kauffman, L. 1987. On knots. In *Annals of Mathematical Studies*, No. 115. Princeton, N.J.: Princeton Univ. Press.

A compendium of interesting tidbits about knots. This book contains enough to keep you thinking for a long time.

Kirkman, T. P. 1885. The enumeration, description and construction of knots with fewer than 10 crossings. *Trans. R. Soc. Edinburgh* 32:281–309.

One of the very first papers on knots. Not an easy read, due to Kirkman's style.

Little, C. N. 1900. Non-alternate + − knots. *Trans. R. Soc. Edinburgh*, 39:771–778.

Another early paper on knots. Remember how hard Little worked.

Livingston, C. 1993. *Knot theory*. Carus Mathematical Monographs 24. Washington, D.C.: Math. Assoc. Amer.

This book is a readable introduction to knot theory, covering a predominantly distinct set of topics. Assumes a background through linear algebra.

Neuwirth, L. 1979. The theory of knots, *Sci. Amer.* 140 (June):84–96.

An easily read introduction to knot theory, preceding by five years the amazing discoveries of the new knot polynomials.

Przytycki, J. 1991. A history of knot theory from Vandermonde to Jones, *Proceedings of the Mexican National Congress of Mathematics*, November.

A short readable history of the field, including work done before the time of Lord Kelvin.

Reidemeister, K. 1932. Knotentheorie. In *Eregebnisse der Matematik und ihrer Grenzgebiete (Alte Folge 0, Band 1, Heft 1)*. Berlin:Springer. (Reprint Berlin: Springer-Verlag, 1974). (English transl., L. Boron, C. Christenson, and B. Smith, BCS Associates, Moscow, Idaho, 1983.)

Reidemeister introduces the Reidemeister moves and proves that two knots are equivalent if and only if we can get from a projection of the first to a projection of the second using the Reidemeister moves. It also includes some braids, the Alexander polynomial, and knot groups.

Rolfsen, D. 1976. *Knots and links*. Berkeley, Calif.: Publish or Perish Press.

The Bible of knot theory. Where many of the recent working knot theorists learned their knot theory. A fascinating book written in a readable style, although it does assume a familiarity with algebraic topology. It predates the new polynomials. A second edition has recently been published.

Tait, P. G. 1898. On Knots I, II, III. In *Scientific Papers*, Vol. 1: 273–347. London: Cambridge Univ. Press.

More early work on the classification of knots.

Thomson, W. H. 1869. On vortex motion. *Trans. R. Soc. Edinburgh* 25:217–260.

This is the paper where Lord Kelvin (aka W. H. Thomson) proposes that knotted vortices in the ether serve as a model for atoms. As Maxwell put it at the time, "It satisfies more of the conditions than any atom hitherto considered."

Chapter 2

Conway, J. H. 1970. On enumeration of knots and links, and some of their algebraic properties. *Computational Problems in Abstract Algebra, Proc. Conf. Oxford 1967*:329–358. Oxford: Pergamon.

Conway's tabulation of knots and links through eleven crossings. It includes descriptions of his notation for knots. A key paper in the history of knot theory.

Dowker, C. H., and M. B. Thistlethwaite. 1983. Classification of knot projections. *Topol. Appl.* **16**:19–31.

An explanation of the Dowker notation for knots.

Ernst, C., and D. W. Sumners. 1987. The growth of the number of prime knots. *Proc. Cambridge Phil. Soc.* 102:303–315.

Proves that the number of prime knots of crossing number n grows exponentially with n.

Thistlethwaite, M. B. 1985. Knot tabulations and related topics. In *Aspects of Topology*, edited by I. M. James and E. H. Kronheimer:1–76. London: Cambridge Univ. Press.

An interesting survey paper, with a lot of information about knot tabulation. Recommended reading.

Chapter 3

See the books in the readings for Chapter 1.

Kanenobu, T. and H. Murakami. 1986. Two-bridge knots with unknotting number one. *Proc. Amer. Math. Soc.* 98(3):499–502. November.

Kanenobu and Murakami determine exactly which two-bridge knots have unknotting number one. In particular, they give the first proof of the fact that 8_3 has unknotting number 2.

Nakanishi, Y. 1983. Unknotting numbers and knot diagrams with the minimum crossings. *Mathematics Seminar Notes*, 11:257–258. Kobe, Japan: Kobe University.

An early discussion of *k*-moves and related conjectures.

Scharlemann, M. 1985. Unknotting number one knots are prime, *Invent. Math.* 82:37–55.

Here is the first proof that a composite knot cannot be turned into a trivial knot by only one crossing change. This is a difficult to read, technical paper.

Chapter 4

Firby, P. A., and C. F. Gardiner. 1982. *Surface Topology.* Chichester, England: Ellis Horwood (distributed by Wiley).

This book gives a nice readable introduction to surfaces, going into a lot more depth than we do in Chapter 4.

Gabai, D. 1986. Genera of the alternating links. *Duke Math. Journal* 53(3):677–681.

The simplest proof of the fact that the genus of an alternating knot or link is realized by the Seifert surface obtained by applying Seifert's algorithm to a reduced alternating projection.

Massey, W. S. 1967. *Algebraic Toplogy: An Introduction.* Harbrace College Math Series. New York: Harcourt, Brace & World.

This book gives the classification of surfaces. Although the rest of the book assumes a certain amount of background, no previous background is necessary to read the chapter on surfaces. In addition, the book includes the fundamental group and covering space theory, if you are at the point where you want to learn this material.

Seifert, H. 1934. Uber das geschlect von knoten. *Math. Ann.* 110:571–592.

Here is where the idea of a Seifert surface for a knot is introduced.

Chapter 5

Adams, C., M. Hildebrand, and J. Weeks. 1991. Hyperbolic invariants of knots and links. *Trans. Amer. Math. Soc.* (1):1–56.

This paper gives the first list of hyperbolic volumes for knots through ten crossings and links through nine crossings. "Pictures" of the hyperbolic structures for some knots and links are included.

Adams, C., J. Brock, J. Bugbee, T. Comar, K. Faigin, A. Huston, A. Joseph, and D. Pesikoff. 1992. Almost alternating links. *Topol. Appl.* 46:151–165.

Here, we invented the concept of almost alternating links, and extended certain results known for alternating links to this new category.

Adams, C. 1992. Toroidally alternating knots and links. To appear in Topology.

Here is where the concept of a toroidally alternating link is introduced.

Birman, J. S. 1976. Braids, links and the mapping class groups. *Ann. Math. Studies No. 82.* Princeton, N.J.: Princeton Univ. Press.

The basic reference for braids, assuming an algebraic background and interest. Predates all of the new polynomials.

———. 1991. Recent developments in braid and link theory. *Math. Intell.* 13(1): 52–60.

A readable introduction to braids; however, with the assumption that the reader is familiar with groups, representations, and presentations. Discusses relations with the new polynomials.

Hayashi, C. 1994. Links with alternating diagrams on closed surfaces of positive genus. Preprint.

Independent work on toroidally alternating knots as in Adams, 1992 and their extensions to links that are alternating on higher genus surfaces.

Menasco, W. 1984. Closed incompressible surfaces in alternating knot and link complements. *Topology* 23(1):37–44.

Proves that a prime alternating link that is not a two-braid is hyperbolic. Assumes a technical proficiency in low-dimensional topology.

Meyerhoff, R. 1992. Geometric invariants for 3-manifolds. *Math. Intell.* 1:37–53.

This article is an excellent introduction to both the topology and geometry of surfaces and three-manifolds. It gives a careful description of hyperbolic space and what it

means for a surface or three-manifold to be hyperbolic. It then goes on to discuss invariants of hyperbolic manifolds such as the volume.

Murasugi, K. 1991. On the braid index of alternating links. *Trans. Amer. Math. Soc.* 326(1):237–260, July.

Among numerous other interesting results, this paper contains the proof that the crossing number of a (p, q)-torus knot is $p(q-1)$, where $p \geq q \geq 2$.

Ohyama, Y. 1993. On the minimal crossing number and the braid index of links. *Can. J. Math.* 45(1):117–131.

The author proves that $c(k) \geq 2(b(K)-1)$, where $c(k)$ is the minimal crossing number and $b(k)$ is the braid index.

Soma, T. 1987. On preimage knots in S^3, *Proc. Amer. Math. Soc.* 100(3):589–592.

This paper proves that if two satellite knots, both coming from the same knot in an unknotted solid torus, are equivalent, then the two corresponding companion knots are equivalent.

Thurston, W. *The Geometry and Topology of Hyperbolic 3-manifolds.* Princeton, N.J.: Princeton Univ. Press. In press.

In 1978, copies of the notes from a course that Thurston taught were disseminated to the mathematical community, and they have formed the basis for the field of topology known as hyperbolic three-manifold theory. Those notes have been due out in book form for some time now, and should appear in the near future.

Thurston, W., and J. Weeks. 1984. The mathematics of 3-dimensional manifolds. *Sci. Amer.* (July):108–120.

This article gives a nice introduction to the geometry of three-manifolds, in particular, hyperbolic three-manifolds. Recommended reading.

Weeks, J. 1985. *The Shape of Space.* New York: Dekker.

This book is an excellent elementary introduction to the topology and geometry of surfaces and three-manifolds. It explains the ideas behind hyperbolic surfaces. No background is assumed.

Yamada, S. 1987. The minimal number of Seifert circles equals the braid index. *Invent. Math.* 88:347–356.

This paper proves that the least number of Seifert circles in any projection of a knot is exactly the braid index of the knot.

Chapter 6

Alexander, J. W. 1928. Topological invariants of knots and links. *Trans. Amer. Math. Soc.* 30:275–306.

The invention of the very first polynomial invariant for knots.

Birman, J. S. 1985. On the Jones polynomial of closed 3-braids. *Invent. Math.* 81:287–294.

Looks at the Jones polynomials for closed braids of three strings.

Brandt, R. D., W. B. R. Lickorish, and K. C. Millett. 1986. A polynomial invariant for unoriented knots and links. *Invent. Math.* 84:563–573.

Introduces a new polynomial for knots and links (which we do not discuss in this book.)

Franks, J., and R. Williams. 1987. Braids and the Jones polynomial. *Trans. Amer. Math. Soc.* 12:303:97–108.

Relates braid index to the exponents of the HOMFLY polynomial (called here the generalized Jones polynomial).

Freyd, P., D. Yetter, J. Hoste, W. Lickorish, K. Millett , and A. Ocneau. 1985. A new polynomial invariant of knots and links. *Bull. Amer. Math. Soc.* 12:239–246.

This is the joint announcement by four different groups of their simultaneous discovery of a 2-variable polynomial that generalized the Jones polynomial.

Jones, V. F. R. 1985. A polynomial invariant for knots and links via Von Neumann algebras. *Bull. Amer. Math. Soc.* 12:103–111.

This is the original announcement by Jones of his discovery of a new polynomial for knots and links.

Kanenobu, T. 1986. Infinitely many knots with the same polynomial. *Proc. Amer. Math. Soc.* 97:158–161.

A readable proof that there are infinitely many distinct knots with the same HOMFLY polynomial.

Kauffman, L. 1983. Combinatorics and knot theory. *Contemp. Math.* 20:181–200.

In this paper, the author introduces the concept of alternative links, a category that encompasses both alternating knots and torus knots. He then proves that the Seifert surface of an alternative link that comes from applying Seifert's algorithm to an alternative projection is a minimal genus Seifert surface, generalizing the result that was previously known for alternating and torus links.

———. 1988. New invariants in the theory of knots. *Amer. Math. Mon.* 195–242, March.

A nice survey article explaining the relationship between the bracket polynomial and the Jones polynomial, in addition to other interesting material.

Lickorish, W. B. R., and K. C. Millett. 1987. A polynomial invariant for oriented links. *Topology* 26:107–141.

This is the authors' proof of the existence of the HOMFLY polynomial, which was simultaneously discovered by the authors and by Freyd and Yetter; Ocneau; Hoste; and Przytycki and Traczyk.

———. 1988. The new polynomial invariants of knots and links. *Math. Mag.* 61(1):3–23, February.

This is a beautifully written introduction to the two-variable polynomials that generalize the Jones polynomial. This article is highly recommended reading.

Menasco, W., and M. Thistlethwaite. 1991. The Tait Flyping Conjecture. *Bull. Amer. Math. Soc.* 25(2):403–412, October.

Here is the paper where the authors announce their solution to this celebrated conjecture. It is a good paper to have a look at, since the first portion can be read without any special background.

Morton, H. 1986. Seifert circles and knot polynomials. *Proc. Cambridge Phil. Soc.* 99:107–109.

Relates braid index to the exponents of the HOMFLY polynomial. See also (Franks and Williams, 1987).

Murasugi, K. 1987. Jones polynomials and classical conjectures in knot theory. *Topology* 26:187–194.

Proves that the minimal crossing number of an alternating link occurs in a reduced alternating projection. See (Thistlethwaite, 1987) and (Kauffman, 1988) also. Also shows that alternating knots with odd crossing number must be chiral.

Thistlethwaite, M. 1987. A spanning tree expansion for the Jones polynomial. *Topology* 26:297–309.

See previous reference.

Chapter 7

Baxter, R. J. 1982. *Exactly Solved Models in Statistical Mechanics.* New York: Academic Press.

A seminal book in the development of the connections between knot theory and statistical mechanics. Here, the author solves the Ising model by utilizing the Yang-Baxter equation. A good technical reference for the detailed mathematics of the relevant statistical mechanics.

Dietrich-Buchecker, C., and J.-P. Sauvage. 1989. A synthetic molecular trefoil knot. *Angew. Chem.* 28(2):189–192.

This is the announcement of the first successful synthesis of a knotted molecule.

Flapan, E. 1990. Topological techniques to detect chirality. In *New Developments in Molecular Chirality,* edited by P. G. Mezey: 209–239. Amsterdam: Kluwer. 1990.

A nice readable survey on topological methods for determining chirality, written for an audience of nonmathematicians.

Jones, V. 1990. Knot theory and statistical mechanics. *Sci. Amer.* 263(5):98–103, November.

A readable account of the relationship between knots and statistical mechanics from the originator of the connection.

Jones, V. 1989. On knot invariants related to some statistical mechanical models. *Pac. J. Math.* 137(2):311–334.

This paper goes into some depth in explaining the connections between knot invariants and statistical mechanical models, including vertex models, Potts-type models, and IRF models. Although relatively technical, it fills in much of the material that we did not cover in Section 7.4.

King, R. B., and D. H. Rouvray Editors. 1987. *Graph Theory and Topology in Chemistry, Studies in Physical and Theoretical Chemistry 51.* New York: Elsevier.

A compendium of interesting articles. Particularly relevant articles include E. Flapan, Chirality of non-standardly embedded Mobius ladders; J. Simon, A topological approach to the stereochemistry of nonrigid molecules; D. W. Summers, Knots, macromolecules and chemical dynamics; and D. M. Walba, Topological stereochemistry: Knot theory of molecular graphs.

Pohl, W. F. 1980. DNA and differential geometry. *Math. Intell.* 3:20–27.

This paper discusses the results of White and Fuller on modeling DNA with ribbons.

Simon, J. 1986. Topological chirality of certain molecules. *Topology* 25(2):229–235.

Herein lies the proof that Möbius ladders with four or more rungs are always topologically chiral (they cannot be deformed to their mirror images).

Sumners, D. W. 1990. Untangling DNA. *Math. Intell.* 12(3):71–80.

A well-written introduction to knot theory as applied to DNA. Highly recommended.

Sumners, D. W., Editor. 1993. *New Scientific Applications of Geometry and Topology. Proceedings of Symposia on Applied Mathematics*, Vol. 45 American Mathematical Society.

This is the proceedings from a special short course held at an American Mathematical Society Meeting in 1992. It contains fascinating articles by Nicholas Cozzarelli, Louis Kauffman, Jonathan Simon, Dewitt Sumners, James White, and Stuart Whittington.

Walba, D. M. 1985. Topological stereochemistry. *Tetrahedron* 41(16):3161–3212.

An encyclopedic article on the connections between knot theory and chemistry. An invaluable reference, leaning toward the chemists.

Wasserman, S., J. Dungan, and N. Cozzarelli. 1985. Discovery of a predicted DNA knot substantiates a model for site-specific recombination. *Science* 229:171–174, July.

Here are the first electron microscope pictures of knotted DNA.

White, J. 1969. Self-linking and the Gauss integral in higher dimensions. *Amer. J. Math.* XCI:693–728.

The geometry of ribbons in space.

Chapter 8

Appel, K., and W. Haken. 1977. See references for Chapter 1.

Conway, J. H., and C. McA. Gordon. 1983. Knots and links in spatial graphs. *J. Graph Theory* 7:445–453.

This article is the basis for Sections 8.1 and 8.2. This was the first article in which graphs were shown to be intrinsically knotted or intrinsically linked.

Howards, H., L. Klein, J. MacEachern, J. Mynttinen, J. Polito, and J. Terilla. 1991. Links in spatial embeddings of graphs. Preprint, September.

This article is an investigation into intrinsic linking by six undergraduates working with me in the summer of 1991. They prove numerous interesting results about intrinsic linking.

Kauffman, L. 1989. Invariants of graphs in three-space. *Trans. Amer. Math. Soc.* 311(2):697–710, February.

The author associates collections of knots and links to graphs in order to obtain invariants for the isotopy type of an embedded graph in three-space. Particularly readable.

———. 1983. Formal knot theory. In *Mathematical Notes,* 30. Princeton, N.J.: Princeton Univ. Press.

This book describes the Arf invariant very well. It also set the groundwork for Kauffman's subsequent work on the new polynomials, which appeared shortly after this book was published.

Negami, S. 1991. Ramsey theorems for knots, links and spatial graphs. *Trans. Amer. Math. Soc.* 324(2):527–541.

This paper proves that given any knot, there is a positive integer n such that ANY straight-line embedding of the complete graph K_n contains that knot. The author also looks at the relationship between the stick number of a knot and the crossing number of that knot.

Robertson, N., P. D. Seymour, and R. Thomas. 1993. Linkless embeddings of graphs in 3-space. *Bull. Amer. Math. Soc.* 28(1):84–89.

The authors announce the solution to the question of determining exactly which graphs have the property that they always contain a link, no matter how they are embedded in three-space.

Shimabara, M. 1988. Knots in certain spatial graphs. *Tokyo J. Math.*, 11(2):405–413.

This paper proves that $K_{5,5}$ is an intrinsically knotted graph.

Chapter 9

Cipra, B. A. 1988. *To have and have knot: when are two knots alike?, Science,* 241, 1291–1292.

A very readable account of the Gordon–Luecke result that distinct knots have distinct complements.

Gordon, C., and J. Luecke. 1989. Knots are determined by their complements. *J. Amer. Math. Soc.* 2(2):371–415.

The authors succeed in proving one of the big open questions (first posed by Tietze in 1908), namely, that two distinct knots cannot have the same complement. This is a hard technical paper.

Hempel, J. 1976. 3-manifolds. In *Annals of Mathematical Studies,* No. 86, Princeton, N.J.: Princeton Univ. Press.

The Bible of three-manifold theory. It assumes a strong background in algebraic topology and topology. Not for the novitiate.

Kauffman, L. 1991. *Knots and Physics.* Singapore: World Scientific.

This book is chock full of fascinating material. Besides containing excellent presentations of many of the connections between knots and physics, see Sections 16 and 17 for an advanced discussion on the three-manifold invariants that have come out of the new knot polynomials.

Rolfsen, D. 1976. See references for Chapter 1.

Weeks, J. 1985. See references for Chapter 5.

Chapter 10

Abbott, E. A. 1952. *Flatland.* New York: Dover.

This book, which was first published in 1884, is a mathematical classic. It tells of the adventures of A. Square, a creature that lives in the two-dimensional world of Flatland, and of his introduction to the three-dimensional world. A great book that broaches the idea of high dimensions.

Fox, R. H., 1962. See references for Chapter 1.

Lomonaco, S. J. Jr. 1983. Five dimensional knot theory. *Contemp. Math.* 20:249–270.

A description of how to visualize two-spheres in four-space and three-spheres in five-space.

Rucker, R. 1984. *The Fourth Dimension.* Boston: Houghton Mifflin.

A fun, relaxed discussion of how to think about higher dimensions. Lots of other great material about relativity, too.

Index