TIGHTENING CURVES AND GRAPHS ON SURFACES

BY

HSIEN-CHIH CHANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Jeff Erickson, Chair
Professor Chandra Chekuri
Professor Nathan M. Dunfield
Professor David Eppstein
Associate Professor Alexandra Kolla

獻給　親愛的爸爸媽媽
謝謝你們的愛

Abstract

Any continuous deformation of closed curves on a surface can be decomposed into a finite sequence of local changes on the structure of the curves; we refer to such local operations as *homotopy moves*. *Tightening* is the process of deforming given curves into their minimum position; that is, those with minimum number of self-intersections. While such operations and the tightening process has been studied extensively, surprisingly little is known about the *quantitative* bounds on the number of homotopy moves required to tighten an arbitrary curve.

An unexpected connection exists between homotopy moves and a set of local operations on graphs called *electrical transformations*. Electrical transformations have been used to simplify electrical networks since the 19th century; later they have been used for solving various combinatorial problems on graphs, as well as applications in statistical mechanics, robotics, and quantum mechanics. Steinitz, in his study of 3-dimensional polytopes, looked at the electrical transformations through the lens of medial construction, and implicitly established the connection to homotopy moves; later the same observation has been discovered independently in the context of knots.

In this thesis, we study the process of tightening curves on surfaces using homotopy moves and their consequences on electrical transformations from a quantitative perspective. To derive upper and lower bounds we utilize tools like curve invariants, surface theory, combinatorial topology, and hyperbolic geometry. We develop several new tools to construct efficient algorithms on tightening curves and graphs, as well as to present examples where no efficient algorithm exists. We then argue that in order to study electrical transformations, intuitively it is most beneficial to work with *monotonic* homotopy moves instead, where no new crossings are created throughout the process; ideas and proof techniques that work for monotonic homotopy moves should transfer to those for electrical transformations. We present conjectures and partial evidence supporting the argument.

Acknowledgments

First I would like to express my most sincere gratitude towards my Ph.D. advisor, Jeff Erickson, for your guidance throughout the years. It is an honor to be your student. You always give me the most honest advice; every meeting with you is fun and fruitful. Most of all, your patience and listening ears lighten the burden when life was difficult and research was slow. I am forever grateful to know you as a colleague, and a friend.

It is always joyful to talk to Sariel Har-Peled, either about research or just random fun. Thank you for introducing me around on my first day in school, and all the ideas and advice (and jokes!) you shared with me. I want to thank Alexandra Kolla for your constant support and trust. Thank you for always giving me opportunities to travel and collaborate, even when no progress has been made for a long time.

The theory group in University of Illinois at Urbana-Champaign provides the most healthy research environment for me to grow as a young researcher. The collaborative atmosphere allows us to explore freely among the topics that interest us. I am fortunate to have so many great colleagues to work and share experience with, and sometimes just having fun together in the last few wonderful years: Shashank Agrawal, Matthew Bauer, Shant Boodaghians, Charles Carlson, Timothy Chan, Karthik Chandrasekaran, Chandra Chekuri, Kyle Fox, Shamoli Gupta, Mitchell Jones, Konstantinos Koiliaris, Nirman Kumar, Patrick Lin, Vivek Madan, Sahand Mozaffari, Kent Quanrud, Benjamin Raichel, Tasos Sidiropulos, Matus Telgarsky, Ross Vasko, Yipu Wang, Chao Xu, Ching-Hua Yu, Xilin Yu, and many more that I didn't have the chance to know better.

I am very fortunate to have the chance of working with great people all around the world: Paweł Gawrychowski, Naonori Kakimura, David Letscher, Arnaud de Mesmay, Shay Mozes, Saul Schleimer, Eric Sedgwick, Dylan Thurston, Stephan Tillmann, and Oren Weimann. Working together with similar minds on interesting projects is much more fun than working alone. Also, I want to express my thanks to Erin Chambers, Ho-Lin Chen, Kai-Min Chung, Joshua Grochow, Chung-Shou Liao, Michael Pelsmajer, and Marcus Schaefer for providing me opportunities to travel and discuss research with you.

Without the days working with friends in the theory group of National Taiwan University, I might not even pursue the career of being a theoretical computer scientist. Thanks to Hsueh-I Lu for your inspiring lectures, guidance, and advice on research that lead me to where I am.

最後，特別感謝妻子祈祈與昕恒；你們是我生命中的喜樂。與妳攜手同行使我知道，愛是最大的奧秘。 Finally, special thanks to Chichi and Luke; you are the joy of my life. Being with you made me realize that love is the greatest mystery.

Table of Contents

Chapter 1

Introduction and History

Say you're me and you're in math class, and your teacher's talking about ... Well, who knows what your teacher's talking about. Probably a good time to start doodling.

— Vi Hart, *Doodling in Math Class*

一角兩角三角形，四角五角六角半。
One-gon, two-gon, tri-angle; four-gon, five-gon, six-gon half.

— *Mandarin fingerplay*

Given an arbitrary closed curve on some 2-dimensional surface, it is natural to look for ways to modify or deform the curve continuously into its "simplest" form. The meaning of "simple" varies according to the applications. To fix the terminology, our goal is to *tighten* the curve via continuous deformation (known as *homotopy*) into another closed curve with minimum complexity. Common complexity measures include the *time* to minimize the length of the curve [23, 29, 30, 54, 63, 65, 82, 140, 162]; the *area* of the homotopy [39, 98, 183, 255]; the *height* of the homotopy [28, 35, 36, 38, 133]; the *width* of the homotopy [32, 34]; and other properties desired for the process, like simplicity and monotonicity [37, 40, 41, 42]. Most of the measures studied inherently require some existing *geometry* associated with the surface and the curve. However, in some instances of the curve tightening problem, the input curve is only supplied by—or differentiated up to—its combinatorial structure, and therefore a better complexity measure, preferably based only on the changes to the structure, is desired.

Consider the following scenario: Given two curves in the plane, we want to decide which curve is more complicated than the other. Various methods are known to measure the "curviness" of the drawings, which can be served as a way to decide the complexity of the curves. However, there are cases when "curviness" might not be the most suited measure. For example, when the input curves are hand-drawn symbols, the length and shape of the curves various drastically from one drawer to the other. What is invariant is the combinatorial structure of the hand drawing, that is, what "symbols" they really are. Naïve measure like counting the number of crossings in the symbols helps, but it does not solve the problem as the number of planar curves with a fixed number of crossings grows exponentially.

In this thesis we propose and study the following *topological* complexity measure—the number of local operations called *homotopy moves* that change the combinatorial structure—for tightening closed curves on arbitrary surfaces. Such local operations have been studied in topology since almost a hundred years ago [6, 7, 104, 105, 202, 239]; however, to the best of our knowledge, no previous work has tackled the problem from a quantitative perspective.

1.1 Homotopy Moves

Homotopy is the process of continuously deforming one curve to the other. For the sake of discreting the process, we assume throughout the rest of the thesis that all the curves are *generic*—every self-intersection is formed by exactly two subpaths crossing each other properly without tangency (that is, a *transverse double* intersection). In this case, one can summarize the changes to the combinatorial structure of the curve on a surface during the homotopy using the following set of local operations:

- 1↔0: Remove/add an empty *monogon*.
- 2↔0: Remove/add an empty *bigon*.
- 3→3: Flip an empty *triangle*; equivalently, move one strand across a self-intersection point.

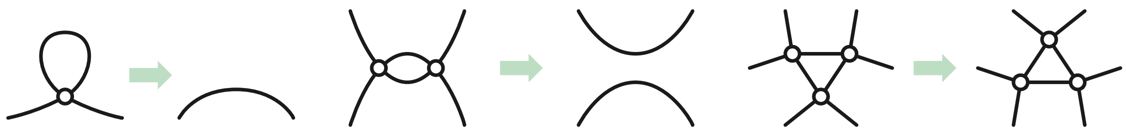


Figure 1.1. Homotopy moves 1→0, 2→0, and 3→3.

Each homotopy move is performed by continuously deforming the curve inside an open disk embedded on the surface, meeting γ as shown in Figure 1.1. Consequently, we call these operations ***homotopy moves***. Our notation is mnemonic; the numbers before and after each arrow indicate the number of local vertices before and after the move. (Similar notation has been used by Thurston [236].)

Homotopy moves are "shadows" of the classical Reidemeister moves used to manipulate knot and link diagrams [7, 202]. A compactness argument, first explicitly given by Titus [239] and Francis [104, 105] but implicit in earlier work of Alexander [6], Alexander and Briggs [7], and Reidemeister [202], implies that any continuous deformation between two generic closed curves on any surface is equivalent to—and therefore, any generic curve can be tightened by—a finite sequence of homotopy moves.

It is natural to ask *how many* homotopy moves are required to tighten a given closed curve on a surface to another curve with minimum number of self-intersections (known as the *geometric intersection number*). An algorithm to tighten any planar closed curve using at most $O(n^2)$ homotopy moves is implicit in Steinitz's proof that every 3-connected planar graph is the 1-skeleton of a convex polyhedron [230, 231]. (See Section 2.6 for a more detailed discussion on Steinitz's algorithm.) The $O(n^2)$ upper bound also follows from algorithms for *regular* homotopy, which forbids 0↔1 moves, by Francis [103], Vegter [251] (for polygonal curves), and Nowik [185].

On higher-genus orientable surfaces, a result of Hass and Scott [135] implies that every non-simple closed curve that is homotopic to a simple closed curve can be tightened using $O(n^2)$ moves, essentially by applying Steinitz's algorithm. Similar result for arbitrary curves on the torus can be derived and extracted from Hass and Scott [135]. De Graaf and Schrijver [125] proved that arbitrary curves on the annulus can be tightened using at most $O(n^2)$ moves.

When both the surface and the curve are unrestricted, Hass and Scott [136] and de Graff and Schrijver [125] independently proved that any closed curve on any surface can be tightened using a *finite* number of homotopy moves that never increase the number of self-intersections. Both results use discrete variants of curve-shortening flow. Grayson [126] and Angenent [9] provide similar results using differential curvature flow when the curves and surfaces are well-behaved. Later on Paterson proved the same result using a combinatorial algorithm [187]. None of these algorithms provide any bound on the number of homotopy moves performed as a function of the number of self-intersections. The monotonicity result, together with asymptotic bounds by Bender and Canfield [21] on the number of distinct (rooted) 4-regular maps with n vertices and genus g, immediately implies an upper bound of the form $n^{O(g)}2^{O(n)}$ on the number of homotopy moves required; this is the best upper bound previously known before our work.

1.2 Electrical Transformations

Let's change our focus from curves to graphs for a moment. Consider the following set of local operations defined on *plane graphs* (that is, planar graphs with embeddings on some surface), called ***electrical transformations*** (following Colin de Verdière *et al.* [68]), consisting of six operations in three dual pairs, as shown in Figure 1.2.

- *degree-1 reduction*: Contract the edge incident to a vertex of degree 1, or delete the edge incident to a face of degree 1
- *series-parallel reduction*: Contract either edge incident to a vertex of degree 2, or delete either edge incident to a face of degree 2
- ΔY *transformation*: Delete a vertex of degree 3 and connect its neighbors with three new edges, or delete the edges bounding a face of degree 3 and join the vertices of that face to a new vertex.

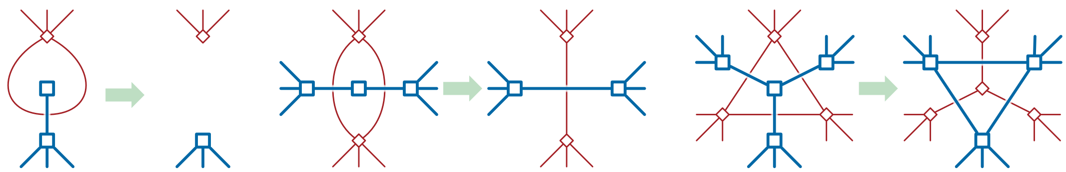


Figure 1.2. Facial electrical transformations in a plane graph G and its dual graph G^*.

It is natural to ask *how many* electrical transformations are required in the worst case. The earliest algorithm for reducing a plane graph to a single vertex again follows from Steinitz's proof of the convex polyhedron theorem [230, 231]. Later algorithms were given by Feo [99], Truemper [242], Feo and Provan [100], and others. Both Steinitz's algorithm and Feo and Provan's algorithm require at most $O(n^2)$ electrical transformations. (We will soon discuss Steinitz's algorithm in Section 2.6, and then Feo and Provan's algorithm later in Section 8.1.1.)

Even the special case of regular grids is interesting. Truemper [242, 244] describes a reduction from the problem of reducing general plane graphs to regular grids using graph minors, and show how to reduce the $p \times p$ grid in $O(p^3)$ steps. Poger and Sussmann [190] showed how to reduce the $(p+q) \times q$ grid in $O(pq^2 + q^3)$ steps. Nakahara and Takahashi [181] prove an upper bound of $O(\min\{pq^2, p^2q\})$ for the $p \times q$ cylindrical grid. Because every n-vertex plane graph is a minor of an $O(n) \times O(n)$ grid [233, 249], all of these results imply an $O(n^3)$ upper bound for arbitrary plane graphs (see Corollary 7.3). Both Gitler [115] and Feo and Provan [100] suspect the possibility that Truemper's algorithm actually performs only $O(n^2)$ electrical transformations. On the other hand, the smallest (cylindrical) grid containing every n-vertex plane graph as a minor has size $\Omega(n) \times \Omega(n)$ [249].

Most of these earlier algorithms actually solve a more difficult problem, considered by Akers [5] and Lehman [165], of reducing a planar graph with two special vertices called *terminals* to a single edge between the two. Epifanov [85] first proved that such reduction is always possible, using a nonconstructive argument; simpler constructive proofs were later given by Feo [99], Truemper [242, 244], Feo and Provan [100] (and Nakahara and Takahashi [181], whose algorithm is almost identical to Truemper's but performed on cylindrical grids instead). In fact, all existing algorithms that work for arbitrary plane graphs without terminals can be modified to work for the two-terminal case.

Feo-Provan Conjecture. Despite decades of prior work as we shown above, the complexity of the electrical reduction process is still poorly understood. Several authors have conjectured that the quadratic bound derived from Feo and Provan [100] can be improved. Without any restrictions on which transformations are permitted, the only known lower bound is the trivial $\Omega(n)$. Gitler [115] and Archdeacon *et al.* [12] asked whether the $O(n^{3/2})$ upper bound for square grids can be improved to near-linear. (We will show in Section 7.3.1 that turns out no improvements can be made.) As for arbitrary planar graphs, Feo and Provan [100] suggested that "there are compelling reasons to think that $O(|V|^{3/2})$ is the smallest possible order", possibly referring to earlier empirical results of Feo [99, Chapter 6]. Gitler [115] conjectured that a simple modification of Feo and Provan's algorithm requires only $O(n^{3/2})$ time.

1.3 Relation between Two Local Operations

Perhaps the most important and surprising connection we proposed in the thesis, is the existence of a quantitative relation between the electrical transformations and the homotopy moves. At the surface[1] such connection doesn't seem to make sense; after all, electrical transformations are performed on (embedded) graphs, whereas homotopy moves are performed on curves. We argue that, at an intuitive level, reduction using electrical transformations should be thought of as a variant of the *monotonic* homotopy process, where all 0→2 moves are forbidden.

Connections between graphs and planar curves can be traced back to Tait [234], when he came up the notion later known as the "Tait graph": Given a planar curve and the unique two-coloring of its regions in the plane, a graph can be constructed by taking one of the color classes as vertices, and two vertices are adjacent if the corresponding two regions share an intersection point of the curve. Notice that a Tait graph always comes with a planar embedding. The inverse operation to the Tait graph construction, now known as the *medial graph* construction, was discovered by Steinitz in his study of 3-dimensional convex polyhedron [230, 231], which he referred to as the "Θ-Prozeß". The medial graph $G^\times$ of an embedded graph G is constructed by taking the edges of G as vertices and connect two edges of G (with multiplicity) if they share both a vertex and a face in G. Every vertex in the medial graph has degree 4, and therefore one can naturally view the medial graph as a system of curves lying on the same surface as G, where each intersection point between two (possibly identical) constituent curves is transverse. We refer to the set of curves as the *medial curves*.

Through the lens of the medial construction, electrical transformations in any embedded graph G correspond to local operations in the medial graph $G^\times$ that bare extreme resemblance—perhaps almost identical to—homotopy moves. We refer to such local operations as ***medial electrical moves***.

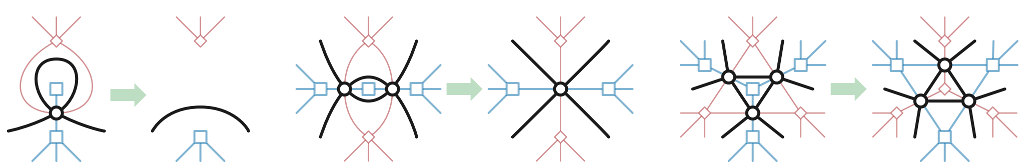


Figure 1.3. Electrical transformations and the corresponding medial electrical moves.

Now a natural bijection is established between graphs and systems of generic curves on a fixed surface; electrical transformations performed on the graph correspond to medial electrical moves performed on the medial curves. Many authors have observed and studied such correspondence, implicitly by Steinitz [230, 231] and Grünbaum [128], and explicitly by Yajima and Kinoshita [258], Goldman and Kauffman [117], and Nobel and Welsh [184].

The correspondence also provides another motivation to study electrical and homotopy moves on surfaces: When we perform electrical transformations on plane graphs, the terminals should not be involved in any local

[1] No pun intended.

operations. Under the medial graph construction, these terminals turn into punctures in the plane; no electrical moves will ever move the curves across a puncture. Therefore, by studying the relationship between electrical and homotopy moves on the punctured plane, we can bound the number of electrical transformations required to reduce plane graphs with terminals.

1.4 Results and Outline of the Thesis

The majority of the thesis is devoted to proving worst-case upper and lower bounds on the number of homotopy moves used for tightening curves and the number of electrical transformations required to reduce planar graphs.

We start with the preliminaries in Chapter 2, introducing the basic concepts used throughout the thesis and fixing the terminologies. Then in Chapter 3 we study the numerical curve invariant called *defect* introduced by Arnold [15, 16] and Aicardi [4]. Exact formulas of defect on specific families of curves are computed, along with several new properties of defect. The chapter finishes with some implications on *random knots*.

In Chapter 4 we derive lower bounds on the number of homotopy moves required to tighten curves on surfaces. We provide an $\Omega(n^{3/2})$ lower bound on tightening closed curves in the plane through the defect invariant. A natural generalization of defect to higher-genus surface gives a stronger $\Omega(n^2)$ bound for non-contractible curves on arbitrary orientable surfaces. The same $\Omega(n^2)$ bound can be proven and extended to arbitrary curves on any surface with non-positive Euler characteristic using a completely different potential function. In Chapter 5, a matching $O(n^{3/2})$ upper bound for planar curves is obtained using the *useful cycle technique*; we then extend the algorithm to arbitrary collection of closed curves in the plane.

In Chapter 6 we describe two methods to tighten curves on an arbitrary orientable surface by adapting Steinitz's algorithm: First, we present an $O((g+b)n^3)$-step algorithm for tightening curves on an arbitrary orientable genus-g surface with $b > 0$ boundary components. Next, we present an $O(gn^3 \log^2 n)$-step algorithm for tightening curves on an arbitrary orientable genus-g surface without boundary. We conclude the chapter with a discussion on monotonicity of the homotopy process.

In Chapter 7 we study the quantitative relation between electrical transformations and monotonic homotopy moves. After a brief discussion on some subtlety in the definition of electrical transformations, we will formally discuss the comparison between the two sets of operations, supported by some natural conjectures strengthening the relationship between electrical moves and homotopy moves. Evidence towards the conjectures and proofs for the special cases are provided in subsequent subsections; in passing, we will make Truemper's minor lemma [242] quantitative. We then apply the theory developed in previous chapters and sections to derive lower bounds on the number of electrical transformations required to reduce planar graphs, with or without terminals. One of the major theorem we rely on, based on arguments of Truemper [242] and Noble and Welsh [184], is that reducing a *unicursal* plane graph G—one whose medial graph is the image of a single closed curve—using electrical transformations requires at least as many steps as reducing the medial graph of G to a simple closed curve using homotopy moves.

1.5 History and Related Work

Applications of electrical transformations. Electrical transformations have been used since the end of the 19th century [155, 212] to analyze resistor networks and other electrical circuits, but many other applications have been discovered since. Akers [5] used the same transformations to compute shortest paths and maximum flows (see also Hobbs [141]). Lehman [165] used them to estimate network reliability; significant amount of work on such application follows [49, 131, 225, 240, 245] (see also [107, 129, 209, 210, 213, 224]). Further applications on solving combinatorial problems using electrical transformations have been found, including multicommodity flows [99]; counting spanning trees, perfect matchings, and cuts [33, 62]; evaluation of spin models in statistical mechanics [62, 146]; solving generalized Laplacian linear systems [127, 181]; kinematic analysis of robot manipulators [227]; flow estimation from noisy measurements [263]; constructing distance preservers [121]; and studying singularities in quantum field theory [200]. Lehman [164] gave a necessary condition on problems to which the electrical transformations applies. (See Chapter 7 and Appendix B of Gitler's thesis [115] for some discussion.)

Local operations related to homotopy moves. Tight bounds are known for two special cases where some homotopy moves are forbidden. First, Nowik [185] proved a tight $\Omega(n^2)$ lower bound for regular homotopy. Second, Khovanov [157] defined two curves to be *doodle equivalent* if one can be transformed into the other using $1{\leftrightarrow}0$ and $2{\leftrightarrow}0$ moves. Khovanov [157] and Ito and Takimura [144] independently proved that any planar curve can be transformed into its unique equivalent doodle with the smallest number of vertices, using only $1{\to}0$ and $2{\to}0$ moves. Thus, two doodle equivalent curves are connected by a sequence of $O(n)$ moves, which is obviously tight. It is not known which sets of curves are equivalent under $1{\leftrightarrow}0$ and $3{\to}3$ moves; indeed, Hagge and Yazinski only recently proved that this equivalence is nontrivial [132]; see also related results of Ito *et al.* [144, 145]. Looser bounds are also known for the minimum number of Reidemeister moves needed to reduce a diagram of the unknot [134, 159], to separate the components of a split link [138], or to move between two equivalent knot diagrams [70, 137].

Geometric intersection number. The *geometric intersection number* of a closed curve γ on a surface is the number of self-intersections of a tightening of γ. Several methods for characterizing and computing geometric intersection numbers are known [52, 53, 61, 119, 172]; however, none of these earlier results offers a full complexity analysis. Arettines [13] described a polynomial-time algorithm to compute geometric intersection number of a curve on an orientable surface with boundary, starting from the reduced crossing sequence of the curve with a system of arcs (defined in Section 6.2.1). Despré and Lazarus [77] described the first fully-analyzed polynomial-time algorithm to compute the geometric intersection number of arbitrary closed curves on an arbitrary orientable surface. Both of these algorithms follow a high-level strategy similar to ours, based on Hass and Scott's results about singular bigons, but neither algorithm computes an explicit sequence of homotopy moves. Instead, Arettines removes singular bigons by permuting their intersections along each arc, and Despré and Lazarus remove singular bigons by directly *smoothing* their endpoints. Further references can be found in Despré and Lazarus [77].

Beyond 2-terminal planar graphs. A vast amount of work has been done to extend the algorithms to planar graphs with more than two terminals. Gitler [115] and Gitler and Sagols [116] proved that any three-terminal planar graph can be reduced to a graph on the three terminals, confirming the speculation by Akers [5]. Poger [189] provided an alternative and efficient algorithm to reduce any three-terminal planar graph using only $O(n^2)$ steps. Archdeacon *et al.* [12] and Demasi and Mohar [76] characterized the four-terminal planar graphs that can be reduced to just four vertices. Gitler [68, 115] proved that for any integer k, any planar graph with k terminals on a common face can be reduced to a planar graph with $O(k^2)$ vertices. Gitler's results were significantly extended by Colin de Verdière *et al.* [66, 67, 68] and Curtis *et al.* [71, 72, 73] to the theory of circular planar networks; see also Postnikov [199] and Kenyon [156].

ΔY-reducible graphs. Gitler [115] proved that every K_5-minor-free or $K_{3,3}$-minor-free graph can be reduced to a single vertex; Wagner [253] proved similar results for almost-planar graphs and almost-graphic matroids, building on earlier matroid results of Truemper [243]; Truemper [242, Lemma 4] and several others [12, 115, 181, 184] proved that the class of ΔY-reducible graphs is closed under minor; Archdeacon *et al.* [12] extended the result to the class of terminal-reducible graphs, and characterized the class of ΔY-reducible projective-planar graphs; Yu [259, 260] showed there are at least 68 billion forbidden minors obstructions for the class of ΔY-reducible graphs, falling into 20 ΔY-equivalent classes.

The two obvious subclasses of ΔY-reducible graphs are the $\Delta\!\to\!Y$*-reducible* graphs and the $Y\!\to\!\Delta$*-reducible* graphs: graphs that are reducible under degree-1 reductions, series-parallel reductions, and exactly one of the two directions of ΔY transformation. These two classes of graphs are far more restrictive than the ΔY-reducible graphs; indeed, they are both subclasses of partial 4-trees [166]. The characterizations and recognition algorithms are known for both $\Delta\!\to\!Y$-reducible graphs [191, 192, 193] and $Y\!\to\!\Delta$-reducible graphs [14, 84, 194, 214].

Algebraic structures for curves on surfaces. The collection of multicurves forms a Lie bialgebra structure on the surface, first noticed by Goldman [118] and Turaev [247]. See Chas [50, 51] for a modern treatment of the topic. The electrical moves performed on curves is similar to the operation of the 0-Hecke monoid of the symmetric groups (also known as the Richardson–Springer monoid) [237], which can be viewed as electrical moves on flat braids. The main technical lemma in the monotonic tightening process for arbitrary annular curves of de Graaf and Schrijver [125, Theorem 4] can be extended to Weyl groups and more generally to Coxeter groups [110].

1.6 Acknowledgment

A major part of the thesis is based on (and extended from) the joint work [43, 44, 45, 46, 47, 48] with the following colleagues: Jeff Erickson, David Letscher, Arnaud de Mesmay, Saul Schleimer, Eric Sedgwick, Dylan Thurston, and Stephan Tillmann. The author is extremely grateful for the opportunity to work with them.

Sections 3.1, 6.3, 6.4, and 7.2 contain results that are either new, or have been improved from the corresponding versions in our earlier publications. Some of the other results in the thesis have also been rewritten to make the terminology and presentation consistent.

Chapter 2

Preliminaries

I don't like to define my music. To me, music is pure emotion. It's language that can communicate certain emotions and the rhythms cuts across genders, cultures and nationalities. All you need to do is close your eyes and feel those emotions.

— Yanni

I respectfully disagree.

— Laurel

We assume the readers are familiar with basic terminologies and definitions in graph theory and topo-logy. We refer the interested readers to the following references. For basic graph theory, see Diestel [80] and West [254]. For topology and manifolds, see Massey [176] and Lee [163]. For topological graph theory, see Mohar-Thomassen [179] and Lando-Zvonkin [160]. For combinatorial topology, see Stillwell [232].

2.1 Surfaces

Intuitively speaking, a ***2-dimensional manifold with boundary*** is a topological space where locally the neighborhood of any point in the interior of the space looks like an Euclidean plane, the neighborhood of any point on the boundary of the space looks like an Euclidean half-plane. A ***surface*** Σ is a 2-dimensional manifold, possibly with boundary. All surfaces are assumed to be connected unless stated otherwise. Every point $x \in \Sigma$ lies in an open neighborhood that is either homeomorphic to the plane $\mathbb{R}^2$ or homeomorphic to an open half-plane with x on its boundary. The points with half-plane neighborhoods form the ***boundary*** of Σ; the ***interior*** of Σ is the complement of its boundary.

The ***genus*** of an orientable surface is intuitively the number of *holes* the surface has. The ***Euler characteristic*** $\boldsymbol{\chi(\Sigma)}$ of a genus-g orientable surface Σ with b boundary components is equal to $2-2g-b$. Except for a few places (which we will mention explicitly), all the surfaces in the thesis are ***orientable***, which means that there is a consistent choice of the normal vectors everywhere on the surface. In other words, locally on the surface, words like "clockwise", "counter-clockwise", "left", and "right" are all well-defined. One of the most fundamental results in combinatorial topology is that all surfaces can be classified by their *genus*, *number of boundary components*, and *orientability* [25, 81, 106, 148, 177, 220]. (For an extended survey on the history, see Gallier and Xu [108].)

2.2 Curves and Graphs on Surfaces

2.2.1 Curves

Formally, a ***closed curve*** or a ***circle*** in a surface Σ is a continuous map γ from 1-dimensional circle S^1 to Σ, and a ***path*** in Σ is a continuous map $\eta\colon [0,1] \to \Sigma$. Depending on the context, we sometimes abuse the terminology and refer to a continuous map $\eta\colon (0,1) \to \Sigma$ as a *path* as well. We call the two points $\eta(0)$ and $\eta(1)$ as ***endpoints*** of η. A ***curve*** is either a closed curve or a path; its parametrization equips the curve with an orientation. We sometimes say the curve is ***directed*** when we want to emphasize its orientation. A curve is ***simple*** if it is injective. A ***subpath*** of a curve γ is the restriction of γ to an interval; again, a subpath is simple if the restriction is injective. We consider only ***generic*** curves, which are injective except at a finite number of self-intersections, each of which is a transverse double point (which means, no more than two subpaths intersect at the same point, and the tangent vectors are not a multiple of the other's). The double points avoid the boundary of Σ. (Some authors preferred the term *normal* [104, 105, 238, 257] or *stable* [186].) Unless specified otherwise, we do not distinguish between the function γ and its image. Sometimes we refer to closed curves in the plane as ***planar curves*** and closed curves in the annulus as ***annular curves***.

2.2.2 Graphs and Their Embeddings

A ***graph*** consists of some 0-dimensional points called ***vertices*** and a multiset containing pairs of vertices called ***edges***. An ***embedding*** of a graph G into an surface Σ maps the vertices of G to distinct points and the edges of G to simple interior-disjoint paths between those points. The ***faces*** of an embedding are the components of the complement of its image in Σ. We sometimes refer to graphs with embeddings as ***maps***. An embedding is ***cellular*** if every face is homeomorphic to an open disk. Any cellular embedding of G into an *orientable* surface can be encoded combinatorially by its ***rotation system***, which records the counterclockwise order of edges incident to each vertex of G. Two cellular embeddings of G are homeomorphic if and only if they have the same rotation system, up to reflections of the surface. An ***embedded graph*** is a graph G together with a cellular embedding of G into some surface Σ. A ***plane graph*** is a planar graph with some given cellular embedding in the plane.

The ***dual*** of a cellularly embedded graph G is another cellularly embedded graph G^* on the same surface, whose vertices, edges, and faces correspond to the faces, edges, and vertices of G, respectively. Specifically, the dual graph G^* has a vertex f^* for every face f of G, and two vertices of G^* are connected by an edge if and only if the corresponding faces of G are separated by an edge. The dual graph G^* inherits a cellular embedding into Σ from the embedding of G. The dual of the dual of a cellularly embedded graph G is (homeomorphic to) the original embedded graph G.

Let G be a graph cellularly embedded on surface Σ; each face of G being a disk implies that graph G must be connected. ***Euler's formula*** states that

$$n_v - n_e + n_f = \chi(\Sigma),$$

where n_v is the number of vertices, n_e is the number of edges, and n_f is the number of faces of G, respectively. In particular, any plane graph G has its number of vertices plus number of faces equals to its number of edges plus 2.

2.2.3 Curves as 4-regular Maps

The image of any non-simple closed curve γ has a natural structure as a 4-regular map, whose ***vertices*** are the self-intersections of γ, ***edges*** are maximal subpaths between vertices, and ***faces*** are components of $\Sigma \setminus \gamma$. We emphasize that the faces of γ are not necessarily disks. Every vertex x of γ has four ***corners*** adjacent to it; these are the four components of $D_x \setminus \gamma$ where D_x is a small disk neighborhood of x. Two curves γ and γ' are *isomorphic* if their images define combinatorially equivalent maps; we will not distinguish between isomorphic curves.

2.2.4 Jordan Curve Theorem

Given a *simple* closed curve σ on the sphere, the classical ***Jordan-Schönflies theorem*** [149, 150, 215, 250] states that σ separates the sphere into exactly two connected components, each of which is simply-connected. The weaker result, without the simply-connectedness conclusion, is often known as the ***Jordan curve theorem***. After projecting the curve σ into the plane, we refer to the two components of the complement of σ as the ***interior*** and the ***exterior*** of σ, depending on whether the component is bounded or not. Jordan-Schönflies theorem forms the basis of most of the arguments in the thesis. We will use the result implicitly without referring to its name. (A curious and tangled history regarding the proof(s) of the Jordan curve theorem(s) can be found in Jeff Erickson's notes on computational topology [88, Note 1].)

2.2.5 Multicurves

A ***multicurve*** on surface Σ is a collection of one or more closed curves in Σ; in particular, a ***k-curve*** is a collection of k circles. A multicurve is ***simple*** if it is injective, or equivalently, if it consists of pairwise disjoint simple closed curves. Again we only consider ***generic*** multicurves. The image of any multicurve is the disjoint union of simple closed curves and 4-regular maps. A ***component*** of a multicurve γ is any multicurve whose image is a connected component of the image of γ. We call the individual closed curves that comprise a multicurve its ***constituent curves***; see Figure 2.1. Most of the definitions on curves can be extended properly to multicurves.

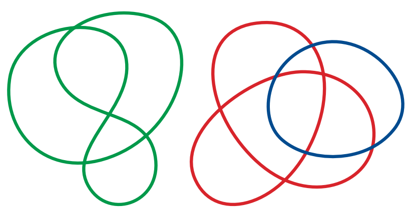


Figure 2.1. A multicurve with two components and three constituent curves, one of which is simple.

2.2.6 Tangles

A ***tangle*** $\oplus$[1] is a collection of boundary-to-boundary paths $\gamma_1, \gamma_2, \ldots, \gamma_s$ in a closed topological disk Σ, which (self-)intersect only pairwise, transversely, and away from the boundary of Σ. This terminology is borrowed from knot theory, where a tangle usually refers to the intersection of a knot or link with a closed 3-dimensional ball [57, 69]; our tangles are perhaps more properly called *flat tangles*, as they are images of tangles under appropriate projection. (Our tangles are unrelated to the obstructions to small branchwidth introduced by Robertson and Seymour [206].) Transforming a curve into a tangle is identical to (an inversion of) the *flarb* operation defined by Allen *et al.* [8].

We call each individual path γ_i a ***strand*** of the tangle. The ***boundary*** of a tangle $\oplus$ is the boundary of the disk Σ that defines $\oplus$; we usually denote the boundary by σ. By the Jordan-Schönflies theorem, we can assume without loss of generality that σ is actually a Euclidean circle. We can obtain a tangle from any closed curve γ by considering its restriction to any closed disk whose boundary σ intersects γ transversely away from its vertices; we call this restriction the ***interior tangle*** of σ.

The strands and boundary of any tangle define a plane graph whose boundary vertices each have degree 3 and whose interior vertices each have degree 4.

[1] Pronounced "Terra".

2.3 Homotopy

A ***homotopy*** between two closed curves γ and γ' on the same surface Σ is a continuous deformation from one curve to the other. Formally this is a continuous map $H: S^1 \times [0,1] \to \Sigma$ such that $H(\cdot, 0) = \gamma$ and $H(\cdot, 1) = \gamma'$. Similarly, a ***homotopy*** between two paths η and η' is a continuous deformation that keeps the endpoints fixed. Formally this is a continuous map $H: [0,1] \times [0,1] \to \Sigma$ such that $H(\cdot, 0) = \gamma$, and $H(\cdot, 1) = \gamma'$, and both $H(0, \cdot)$ and $H(1, \cdot)$ are constant functions. Two curves are ***homotopic***, or in the same ***homotopy class***, if there is a homotopy from one to the other. A closed curve γ is ***contractible*** if it is homotopic to a constant curve; intuitively, this says that γ can be continuously contracted to a single point. Otherwise we say γ is ***non-contractible***. The definition of homotopy extends naturally to multicurves.

A multicurve γ on a surface Σ can be ***tightened*** via homotopy to another multicurve γ' with minimum number of self-intersections. A multicurve is ***homotopically tight*** (***h-tight*** for short) if no homotopy leads to a multicurve with fewer vertices. As any contractible curve γ can be made simple through homotopy [135], we sometimes refer to the tightening process of a contractible curve γ as ***simplifying*** γ.

Similarly, a tangle is ***tight*** if no homotopy of the strands leads to another tangle with fewer vertices, or ***loose*** otherwise.

2.3.1 Covering Spaces and Fundamental Groups

A surface $\tilde{\Sigma}$ is a ***covering space*** of another surface Σ if there is a ***covering map*** from $\tilde{\Sigma}$ to Σ; that is, a continuous map $\pi: \tilde{\Sigma} \to \Sigma$ so that each point x on Σ has a neighborhood $U \subseteq \Sigma$ so that $\pi^{-1}(U)$ is a union of disjoint open sets $U_1 \cup U_2 \cup \cdots$, and, for any i, the restriction $\pi|_{U_i}: U_i \to U$ is a homeomorphism. The ***universal covering space*** $\hat{\Sigma}$ (or ***universal cover*** for short) is the unique simply-connected covering space of Σ.

The ***fundamental group*** $\boldsymbol{\pi_1(\Sigma)}$ of a surface Σ consists of all equivalence classes of closed curves passing through an arbitrary fixed basepoint on Σ up to homotopy, where the group operation comes from concatenating two curves at the fixed point. (For any path-connected space like surfaces, the result is independent to the choice of the basepoint up to isomorphism.)

There is a one-to-one correspondence between subgroups of $\pi_1(\Sigma)$ and covering spaces of Σ. To be precise, given any subgroup Γ of $\pi_1(\Sigma)$, each element α in group Γ acts on the universal cover $\hat{\Sigma}$ by moving the points according to the path that projects to the closed curve in Σ representing the element α. one can construct covering space $\tilde{\Sigma}_\Gamma$ of Σ as the *quotient space* $\hat{\Sigma}/\Gamma$, by identifying all the points in the same orbit under the action of Γ on the universal cover $\hat{\Sigma}$. For example, the trivial subgraph of $\pi_1(\Sigma)$ corresponds exactly to the universal cover of Σ.

2.3.2 Lifting

Let Σ be a surface and $\tilde{\Sigma}$ be a covering space of Σ with covering map π. A ***lift*** of a path η in Σ to $\tilde{\Sigma}$ is a path $\tilde{\eta}$ in $\tilde{\Sigma}$ such that $\eta = \pi \circ \tilde{\eta}$. A ***lift*** of a closed curve γ in Σ to $\tilde{\Sigma}$ is an infinite path $\tilde{\gamma}: \mathbb{R} \to \tilde{\Sigma}$ such that $\gamma(t \bmod 1) = \pi(\tilde{\gamma}(t))$. We sometimes view the closed curve γ as a path γ_x starting and ending at the same point x in Σ, and therefore abuse the terminology and refer to the lift of the path γ_x as ***lift*** of γ (at basepoint x) instead. Observe that the lift of γ at basepoint x is always a subpath of the lift of γ. A ***translate*** of a lift $\tilde{\alpha}$ is any other lift of α to the same covering space; equivalently, two paths $\tilde{\alpha}, \tilde{\beta}: [0,1] \to \tilde{\Sigma}$ are translates of each other if and only if $\pi \circ \tilde{\alpha} = \pi \circ \tilde{\beta}$.

The ***homotopy lifting property*** guarantees that any homotopy H from a curve γ to another curve γ' on Σ lifts to a homotopy $\tilde{H}$ from $\tilde{\gamma}$ to $\tilde{\gamma}'$ on the covering space $\tilde{\Sigma}$. If we decompose the homotopies H and $\tilde{H}$ into homotopy moves, any homotopy move in $\tilde{H}$ corresponds to a homotopy move in H by projection; however there might be additional homotopy moves in H where the strands involved are projected from different parts of the lift on $\tilde{\Sigma}$.

2.4 Combinatorial Properties of Curves

2.4.1 Monogons and Bigons

A ***monogon*** in a closed curve γ on surface Σ is a subpath of γ that begins and ends at some vertex x, intersects itself only at x, and bounds a disk in Σ containing exactly one of the four corners at x. A ***bigon*** in γ consists of two simple interior-disjoint subpaths of γ, sharing endpoints x and y, that together bound a disk in Σ containing exactly one corner at x and one at y. Since each subpath is simple, the vertices x and y are distinct.

We sometimes refer to the interior tangle of the boundary curve of the disk corresponding to the monogon (respectively, the bigon) as the ***interior tangle*** of the monogon (respectively, the bigon). A monogon or bigon is ***empty*** if its interior bigon does not intersect the rest of γ. A bigon β is ***minimal*** if its interior tangle $\oplus$ does not contain a smaller bigon, and no strand of $\oplus$ forms a bigon with β by intersecting either bounding path of β more than once.

2.4.2 Homotopy Moves

Consider the following set of local operations performed on any generic curve: **1→0** ***move*** removes an *empty* monogon, **2→0** ***move*** removes an *empty* bigon, and **3→3** ***move*** moves a subpath across a self-intersection. Collectively we call them (and their inverses) ***homotopy moves***.[2]

Each homotopy move can be executed by a homotopy inside an open disk embedded in Σ, meeting γ as shown in Figure 1.1. Conversely, Alexander's simplicial approximation theorem [6], together with combinatorial arguments of Alexander and Briggs [7] and Reidemeister [202], imply that any generic homotopy between two closed curves can be decomposed into a finite sequence of homotopy moves. The definition of homotopy and the decomposition of homotopies into homotopy moves extend naturally to multicurves and tangles.

2.4.3 Signs and Winding Numbers

We adopt a standard sign convention for vertices first used by Gauss [109]. Choose an arbitrary basepoint $\gamma(0)$ and orientation for the curve. For each vertex x, we define $\text{sgn}(x) = +1$ if the first traversal through the vertex crosses the second traversal from right to left, and $\text{sgn}(x) = -1$ otherwise. See Figure 2.2.

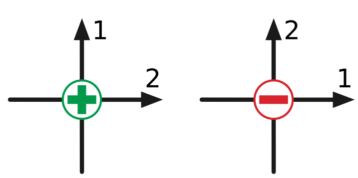


Figure 2.2. Gauss's sign convention.

[2]Unlike the situation for Reidemeister moves on knots, there is no consistent naming for the these local operations. Others called them Titus moves [104, 105]; shadow moves [246, 248]; perestroikas [15, 16]; Reidemeister-type moves [125, 196]; elementary moves [187]; basic moves [185]; and so on. Here we attempt to resolve the situation once and for all by proposing yet another name.

Let γ be a generic closed curve in the plane, and let p be any point not in the image of γ. Let ρ be any ray from p to infinity that intersects γ transversely. The ***winding number*** $\boldsymbol{wind(\gamma, p)}$ is the number of times γ crosses ρ from right to left, minus the number of times γ crosses ρ from left to right. The winding number does not depend on the particular choice of ray ρ. All points in the same face of γ have the same winding number; the winding numbers of two adjacent faces differ by 1, with the higher winding number on the left side of the edge. If p lies on the curve γ, we define $wind(\gamma, p)$ to be the average of the winding numbers of the faces incident to p with appropriate multiplicity—two faces if p lies on an edge, four if p is a vertex. The winding number of a vertex is always an integer.

The ***winding number*** of a directed closed curve γ in the annulus is the number of times any generic path ρ from one fixed boundary component to the other crosses γ from left to right, minus the number of times ρ crosses γ from right to left. Two directed closed curves in the annulus are homotopic if and only if their winding numbers are equal [142].

2.5 Relating Graphs to Curves

2.5.1 Medial Construction

The ***medial graph*** of a graph G embedded on surface Σ, which we denote $\boldsymbol{G^\times}$, is another graph embedded on the same surface whose vertices correspond to the edges of G and whose edges correspond to incidences (with multiplicity) between vertices of G and faces of G. Two vertices of $G^\times$ are connected by an edge if and only if the corresponding edges in G are consecutive in cyclic order around some vertex, or equivalently, around some face in G. Every vertex in every medial graph has degree 4; thus, every medial graph is the image of a multicurve. Conversely, the image of every non-simple multicurve is the medial graph of some embedded graph on Σ. The medial graphs of any cellularly embedded graph G and its dual G^* are identical. To avoid trivial boundary cases, we define the medial graph of an isolated vertex to be a circle. We call an embedded graph G ***unicursal*** if its medial graph $G^\times$ is the image of a single closed curve.

The medial graph $G^\times$ of any 2-terminal plane graph G is properly considered as a multicurve in the annulus; the faces of $G^\times$ that correspond to the terminals are removed from the surface. In general, medial graph $G^\times$ of any k-terminal graph G embedded on surface Σ can be viewed as a multicurve on Σ with all faces of $G^\times$ representing terminals of G being removed.

2.5.2 Facial Electrical Moves

The ***facial electrical transformations*** consist of six operations in three dual pairs: ***degree-1 reduction***, ***series-parallel reduction***, and ***ΔY transformation***, as shown in Figure 1.2. (In Chapter 1 we simply called them *electrical transformations*; from this point on throughout the rest of the thesis, we reserve that name for the general set of transformations performed on arbitrary graphs without embeddings.) Facial electrical transformations on any graph G embedded in surface Σ correspond to local operations in the medial graph $G^\times$ on the same surface that closely resemble homotopy moves. Each degree-1 reduction in G corresponds to a 1→0 move in $G^\times$, and each ΔY transformation in G corresponds to a 3→3 move in $G^\times$. A series-parallel reduction in G contracts an empty bigon in $G^\times$ to a single vertex. Extending our earlier notation, we call this operation a ***2→1 move***. We collectively refer to these operations and their inverses as ***medial electrical moves***; see Figure 1.3.

A multicurve is ***electrically tight*** (***e-tight*** for short), if no sequence of medial electrical moves leads to another multicurve with fewer vertices. We use the terminology "tight" for both electrical and homotopic reductions. This is not a coincidence; we will justify its usage in Section 7.2.4.

2.5.3 Depths of Planar and Annular Multicurves

For any planar multicurve γ and any point p in the plane, let $\boldsymbol{depth(p,\gamma)}$ denote the minimum number of times a path from p to infinity crosses γ. Any two points in the same face of γ have the same depth, so each face f has a well-defined depth, which is its distance to the outer face in the dual graph of γ; see Figure 5.1. The depth of the multicurve, denoted $\boldsymbol{depth(\gamma)}$, is the maximum depth of the faces of γ; and the ***depth-sum potential*** $\boldsymbol{D\Sigma(\gamma)}$ (or just ***potential***) is the sum of depths of all the faces of γ. Euler's formula implies that any 4-regular plane graph with n vertices has exactly $n+2$ faces; thus, for any multicurve γ with n vertices, we have $n+1 \le D\Sigma(\gamma) \le (n+1)\cdot depth(\gamma)$.

Depths and potential of a tangle $\oplus$ are defined exactly the same as for planar curves: The depth of any face f of $\oplus$ is its distance to the outer face in the dual graph $\oplus^*$; the depth of the tangle is its maximum face depth; and the potential $D\Sigma(\oplus)$ of the tangle is the sum of all face depths.

The ***depth*** of any multicurve γ in the annulus is the minimum number of times a path from one boundary to the other crosses γ. Notice how this definition differs from the one for planar multicurves. If we embed the annulus in the punctured plane $\mathbb{R}^2 \setminus o$, the depth of the annular multicurve γ is in fact equivalent to $depth(o,\gamma)$. Just as the winding number around the boundaries is a complete homotopy invariant for annular curves, the depth turns out to be a complete invariant for facial electrical moves on annular multicurves. (See Section 7.2.3.) By definition the inequality $|wind(\gamma,o)| \le depth(o,\gamma)$ holds.

2.5.4 Smoothing

Suppose that γ is a generic closed curve and x is a vertex of γ. Let D_x be a small disk neighborhood of x. Then we may ***smooth*** the curve γ at x by removing $\gamma \cap D_x$ from γ and adding in two components of $\partial D_x \setminus \gamma$ to obtain another 4-regular map. Following Giller [113, 143], we refer to the resulting curve as a ***smoothing*** of γ.[3] There are two types of smoothings. One results in another closed curve, with the orientation of one subpath of γ reversed; the other breaks γ into a pair of closed curves, each retaining its original orientation. In the latter smoothing, let $\boldsymbol{\gamma_x^+}$ and $\boldsymbol{\gamma_x^-}$ respectively denote the closed curve locally to the left and to the right of x, as shown in Figure 2.3. For any vertex x and any other point p, we have $wind(\gamma,p) = wind(\gamma_x^+,p) + wind(\gamma_x^-,p)$. More generally, a ***smoothing*** of a multicurve γ is any multicurve obtained by smoothing a subset of its vertices. For any embedded graph G, the smoothings of the medial graph $G^\times$ are precisely the medial graphs of minors of G.

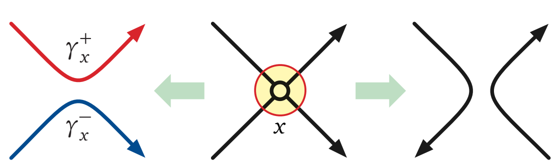


Figure 2.3. Smoothing a vertex. The left smoothing preserves orientation; the right smoothing preserves connectivity.

[3] The same operation is also known as a *split* or *splice* [152, 153]; an *opening* [219]; a *resolution* [170, 171]; or a *cut-and-paste* [182]. The word *smoothing* was later on picking up by Jones [147] and Kauffman [154].

2.6 Tightening Curves via Bigon Removal

As mentioned in the introduction, an algorithm to simplify any planar closed curve using at most $O(n^2)$ homotopy moves is implicit in Steinitz's proof that every 3-connected planar graph is the 1-skeleton of a convex polyhedron [230, 231]. Specifically, Steinitz proved that any non-simple planar multicurve or any loose tangle with no empty monogons contains a bigon. (It follows that a tangle is tight if every strand does not self-intersect, and every pair of strands intersects at most once.) Steinitz then proved that any *minimal* bigon with no empty monogons can be transformed into an empty bigon using a sequence of 3→3 moves, each removing one triangular face from the bigon. For the sake of completeness, we provide a succinct proof to the latter result here.

Lemma 2.1. *A minimal bigon that contains no empty monogons must have an empty triangle incident to either of the two bounding curves of the bigon; therefore such bigon can always be transformed into an empty bigon using a sequence of* 3→3 *moves.*

Proof: Let $\oplus$ be the interior tangle of the bigon. First we prove that all the strands of $\oplus$ are simple. Assume for contradiction that there is an inclusion-wise minimal monogon σ formed by some strand of $\oplus$; let's call the interior tangle of σ as $\oplus_\sigma$. Now all the strands of $\oplus_\sigma$ must be simple. However because σ is not empty, any strand of $\oplus_\sigma$ forms a bigon with σ, contradicting to the fact that the bigon itself is minimal.

Each pair of strands of $\oplus$ intersects at most once. Fixing one of the two curves λ forming the bigon, every vertex v inside the bigon (as the intersection point of two strands α and β) defines a closed region R_v formed by α, β, and λ. Now we argue the following: Any vertex v with inclusion-wise minimal R_v such that v has a neighbor w on λ (when viewed as a graph) must contain no other vertices besides v and the two intersections $(\alpha \cup \beta) \cap \lambda$. Assume for the contrary, without loss of generality that the neighbor w of v on λ is $\alpha \cap \lambda$. Consider the vertex y on β that is adjacent to $z := \beta \cap \lambda$; by our assumption y is not equal to v. Denote the other curve that intersects y as γ; by the above paragraph β does not self-intersect and thus $\gamma \neq \beta$. It is not hard to see now that R_y is contained in R_v, because γ must intersects λ on the subpath of λ between w and z as β and γ intersect at most once. This contradicts to the fact that R_v is inclusion-wise minimal. Therefore, there is an empty triangle incident to λ, and one 3→3 move will remove v from the bigon.

Recursively remove vertices from the bigon; the procedure will only stop when there are no vertices left. At this point all strands of $\oplus$ are simple and disjoint from each other; one can apply a sequence of 3→3 moves to remove all strands from $\oplus$, and thus making the bigon empty. □

Once the bigon is empty, it can be deleted with a single 2→0 or 2→1 move. Grünbaum [128] describes Steinitz's proof in more detail; indeed, Steinitz's proof is often incorrectly attributed to Grünbaum. See Gilmer and Litherland [114], Hass and Scott [136], Colin de Verdière *et al.* [68], or Nowik [185] for more modern treatments of Steinitz's technique.

Removing all the vertices inside the bigon takes as many 1→0 and 3→3 moves as the number of vertices, followed by another sequence of 3→3 moves that empties the bigon. This implies the following lemma, which will always be referred to as ***Steinitz's bigon removal algorithm***:

Lemma 2.2. *Any minimal bigon whose interior tangle contains n vertices and s strands can be removed using* $(n+s)$ 1→0 *and* 3→3 *moves followed by a single* 2→0 *or* 2→1 *move.*

Chapter 3

Curve Invariant — Defect

Porque una parte importante de la relación amorosa, se juega en esta posibilidad de reconocer los defectos del otro y preguntarse, sinceramente, si se puede ser feliz a pesar de ellos.

— Gabriel Rolón, *Historias de diván: ocho relatos de vida*

We consider a numerical invariant[1] of closed curves in the plane introduced by Arnold [15, 16] and Aicardi [4] called ***defect***. (We will only focus on planar curves in this chapter; later on we will discuss how to define defect invariant on higher genus surfaces in Section 4.2.) There are several equivalent definitions and closed-form formulas for defect and other closely related curve invariants [11, 56, 167, 169, 196, 222, 223, 252]; Polyak [195] proved that defect can be computed—or for our purposes, defined—as follows:

$$defect(\gamma) := -2 \sum_{x \between y} \operatorname{sgn}(x) \cdot \operatorname{sgn}(y).$$

Here the sum is taken over all *interleaved* pairs of vertices of γ: two vertices $x \neq y$ are ***interleaved***, denoted $\boldsymbol{x \between y}$, if they alternate in cyclic order—x, y, x, y—along γ. (The factor of -2 is a historical artifact, which we retain only to be consistent with Arnold's original definitions [15, 16].) Even though the signs of individual vertices depend on the basepoint and orientation of the curve, the defect of a curve is independent of those choices. Moreover, the defect of any curve is preserved by any homeomorphism from the plane (or the sphere) to itself, including reflection. Trivially, every simple closed curve has defect zero.

Arnold [15, 16] originally defined two related first-order curve invariants St ("strangeness") and J^+ by their changes under 2→0 and 3→3 moves, without giving explicit formulas. Aicardi [4] proved that the linear combination $2St + J^+$ is unchanged under 1→0 moves; Arnold dubbed this linear combination the "defect" of the curve [16]. Aicardi also described n-vertex curves with strangeness $-\lfloor n(n-1)/6 \rfloor$ and $n(n+1)/2$ for all n; Shumakovich [222, 223] later proved that all n-vertex curves have strangeness between these two extremes. (Nowik's $\Omega(n^2)$ lower bound for regular homotopy moves [185] follows immediately from Aicardi's analysis.) However, the curves with extremal strangeness actually have defect zero.

In Section 3.1, we compute the defect of the standard planar projection of any $p \times q$ torus knot where either $p \bmod q = 1$ or $q \bmod p = 1$, generalizing earlier results of Hayashi *et al.* [137, 139] and Even-Zohar *et al.* [95]. In particular, we show that the standard projection of the $p \times (p+1)$ torus knot, which has $p^2 - 1$ vertices, has defect $2\binom{p+1}{3}$.

Next, in Section 3.2, we prove that the defect of any generic closed curve γ with n vertices has absolute value at most $O(n^{3/2})$. Unlike most $O(n^{3/2})$ upper bounds involving planar graphs, our proof does *not* use the planar separator theorem [168]. First we prove that if the depth of the curve is $\Omega(\sqrt{n})$, there is a simple closed curve σ

[1] Here the invariance is maintained under curve *isotopy*, which preserves the combinatorial structure of the curve. For our purpose (and convention in computer science) it would be better suited to refer to it as a *potential function*.

that contains at least s^2 vertices of γ, where s is the number of strands in the interior tangle of σ. We establish an inclusion-exclusion relationship between the defects of the given curve γ, the curves obtained by tightening γ either inside or outside σ, and the curve obtained by tightening γ on both sides of σ. This relationship implies an unbalanced "divide-and-conquer" recurrence whose solution is $O(n^{3/2})$.

We prove the following surprising observation in Section 3.3: Although the medial graph of a plane graph G depends on the embedding of G, the defect of the medial graph of G does not. This result has some implications on lower bounds for electrical transformations in Section 7.3.3. The chapter ends with some discussion on models of random knots and its connection to defect bounds (Section 3.4).

3.1 Defect Lower Bound

3.1.1 Flat Torus Knots

For any relatively prime integers p and q, let $\boldsymbol{T(p,q)}$ denote the curve with the following parametrization, where θ runs from 0 to 2π:

$$T(p,q)(\theta) := ((\cos(q\theta)+2)\cos(p\theta),\ (\cos(q\theta)+2)\sin(p\theta)).$$

The curve $T(p,q)$ winds around the origin $|p|$ times, oscillates $|q|$ times between two concentric circles, and crosses itself exactly $(|p|-1)\cdot|q|$ times. We call these curves ***flat torus knots***.

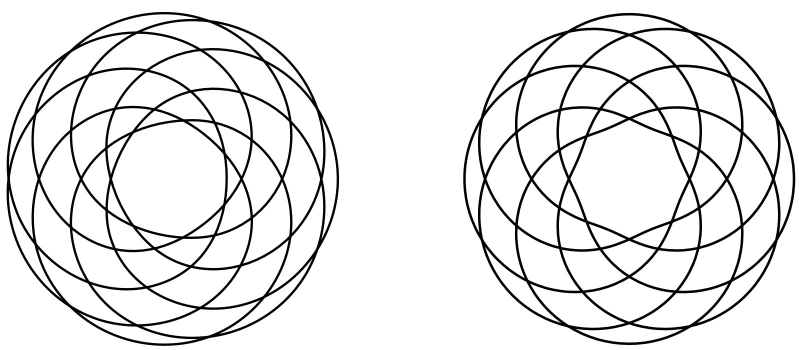


Figure 3.1. The flat torus knots $T(8,7)$ and $T(7,8)$.

Hayashi *et al.* [139, Proposition 3.1] proved that for any integer q, the flat torus knot $T(q+1,q)$ has defect $-2\binom{q}{3}$. Even-Zohar *et al.* [95] used a star-polygon representation of the curve $T(p,2p+1)$ as the basis for a universal model of random knots; in our notation, they proved that $\mathit{defect}(T(p,2p+1)) = 4\binom{p+1}{3}$ for any integer p. In this section we simplify and generalize both of these results to all flat torus knots $T(p,q)$ where either $q \bmod p = 1$ or $p \bmod q = 1$. For purposes of illustration, we cut $T(p,q)$ along a spiral path parallel to a portion of the curve, and then deform the p resulting subpaths, which we call *strands*, into a "flat braid" between two fixed diagonal lines. See Figure 3.2.

Lemma 3.1. *$\mathit{defect}(T(p,ap+1)) = 2a\binom{p+1}{3}$ for all integers $a \geq 0$ and $p \geq 1$.*

Proof: The curve $T(p,1)$ can be reduced to a simple closed curve using only 1→0 moves, so its defect is zero. For the rest of the proof, assume $a \geq 1$.

We define a *stripe* of $T(p,ap+1)$ to be a subpath from some innermost point to the next outermost point, or equivalently, a subpath of any strand from the bottom to the top in the flat braid representation. Each stripe

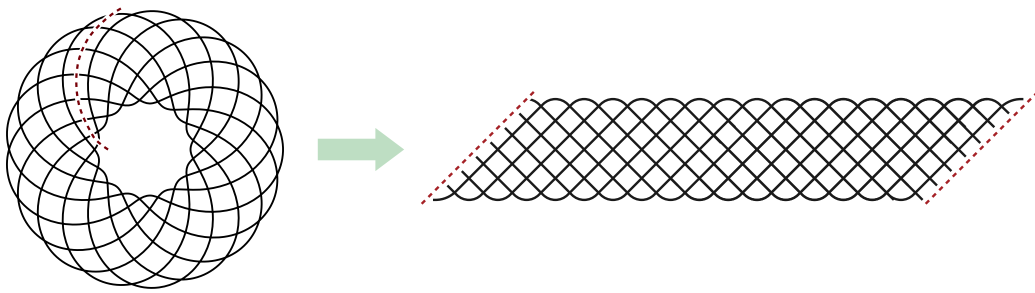


Figure 3.2. Transforming $T(8,17)$ into a flat braid.

contains exactly $p-1$ crossings. A *block* of $T(p, ap+1)$ consists of $p(p-1)$ crossings in p consecutive stripes; within any block, each pair of strands intersects exactly twice. We can reduce $T(p, ap+1)$ to $T(p, (a-1)p+1)$ by straightening any block one strand at a time. Straightening the bottom strand of the block requires the following $\binom{p}{2}$ moves, as shown in Figure 3.3.

- $\binom{p-1}{2}$ 3→3 moves pull the bottom strand downward over one intersection point of every other pair of strands. Just before each 3→3 move, exactly one of the three pairs of the three relevant vertices is interleaved, so each move decreases the defect by 2.
- $(p-1)$ 2→0 moves eliminate a pair of intersection points between the bottom strand and every other strand. Each of these moves also decreases the defect by 2.

Altogether, straightening one strand decreases the defect by $2\binom{p}{2}$. Proceeding similarly with the other strands, we conclude that $\mathit{defect}(T(p, ap+1)) = \mathit{defect}(T(p, (a-1)p+1)) + 2\binom{p+1}{3}$. The lemma follows immediately by induction. □

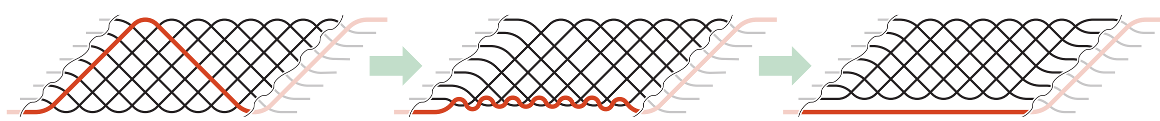


Figure 3.3. Straightening one strand in a block of $T(8, 8a+1)$.

Lemma 3.2. *$\mathit{defect}(T(aq-1, q)) = 2a\binom{q}{3}$ for all integers $a \geq 0$ and $q \geq 1$.*

Proof: The curve $T(q-1, q)$ is simple, so its defect is trivially zero. For any positive integer a, we can transform $T(aq-1, q)$ into $T((a-1)q-1, q)$ by incrementally removing the innermost q *loops*. We can remove the first loop using $\binom{q}{2}$ homotopy moves, as shown in Figure 3.4. (The first transition in Figure 3.4 just reconnects the top left and top right endpoints of the flat braid.)

- $\binom{q-1}{2}$ 3→3 moves pull the left side of the loop to the right, over the crossings inside the loop. Just before each 3→3 move, the three relevant vertices contain one interleaved pair, so each move *decreases* the defect by 2.
- $(q-1)$ 2→0 moves pull the loop over $q-1$ strands. The strands involved in each move are oriented in opposite directions, so these moves leave the defect unchanged.

- Finally, we can remove the loop with a single 1→0 move, which does not change the defect.

Altogether, removing one loop decreases the defect by $2\binom{q-1}{2}$. Proceeding similarly with the other loops, we conclude that $\mathit{defect}(T(aq-1,q)) = \mathit{defect}(T((a-1)q-1,q)) + 2\binom{q}{3}$. The lemma follows immediately by induction. □

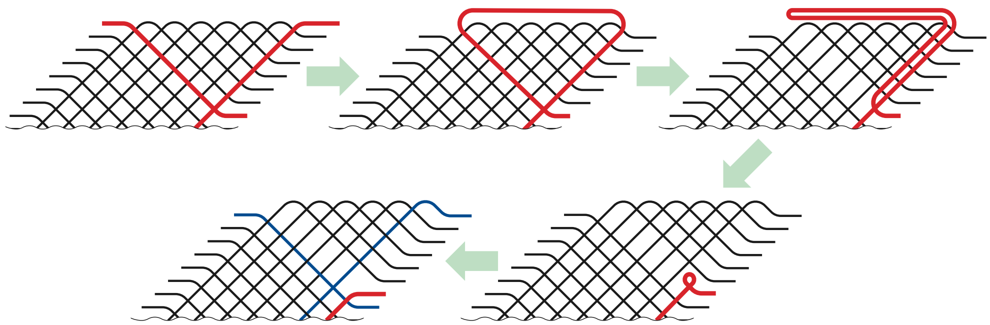


Figure 3.4. Removing one loop from the innermost block of $T(7a-1,7)$.

From Lemma 3.1 and Lemma 3.2 one concludes that the defect of planar curves can be of $\Omega(n^{3/2})$ in the worst case.

3.1.2 Defects of arbitrary flat torus knots

The argument in Lemma 3.1 and Lemma 3.2 can be used to compute the defect of *any* flat torus knot $T(p,q)$ using a process similar to Euclid's algorithm. The only subtlety is determining how many 3→3 moves increase or decrease the defect.

Let $[p]$ denote the set $\{0,1,\ldots,p-1\}$, and consider the permutation $\pi\colon [p]\to[p]$ defined by setting $\pi(i) := iq \bmod p$. We call a triple (i,j,k) of distinct indices in $[p]$ *positive* if $(\pi(i),\pi(j),\pi(k))$ is an even permutation of (i,j,k) and *negative* otherwise. Finally, let $\boldsymbol{\Delta(p,q)}$ denote the number of positive triples minus the number of negative triples. We easily observe that $\Delta(p,q) = \Delta(p, q \bmod p)$, and the proofs of Lemma 3.1 and Lemma 3.2 imply the recurrence

$$\mathit{defect}(T(p,q)) \;=\; \begin{cases} \mathit{defect}(T(p,q-p)) + 2\Delta(p,q) + 2\binom{p}{2} & \text{if } p<q,\\[1ex] \mathit{defect}(T(p-q,q)) - 2\Delta(q,p) & \text{if } p>q. \end{cases}$$

This recurrence immediately gives us an algorithm to compute $\mathit{defect}(T(p,q))$, similar to Euclid's algorithm. Indeed, we can express $\mathit{defect}(T(p,q))$ directly in terms of the continued fraction expansion of p/q as follows.

Let $r_0 := p$ and $r_1 := q$. For all $k\ge 1$ such that $r_k > 1$, define $a_k := \lfloor r_{k-1}/r_k \rfloor$ and $r_{k+1} := r_{k-1} \bmod r_k$. Then we have

$$\mathit{defect}(T(p,q)) \;=\; 2\sum_{k\ge 1} (-1)^k \cdot a_k \cdot \Delta(r_k, r_{k-1}) + 2\sum_{\substack{k\ge 1\\ k \text{ even}}} a_k \cdot \binom{r_k}{2}.$$

Using the above formula we can prove the *reciprocity formula* for the defect of flat torus knots.

Lemma 3.3. *For any positive integers p and q, $\mathit{defect}(T(p,q)) + \mathit{defect}(T(q,p)) = (p-1)(q-1)$.*

Proof: Let m be the smallest number such that $r_{m+1} = 1$. Then

$$\begin{aligned}
\mathit{defect}(T(p,q)) + \mathit{defect}(T(q,p)) &= 2\sum_{k=1}^{m} a_k \cdot \binom{r_k}{2} \\
&= \sum_{k=1}^{m} a_k \cdot r_k(r_k - 1) \\
&= \sum_{k=1}^{m} (r_{k-1} - r_{k+1})(r_k - 1) \\
&= (r_1 r_0 - r_m r_{m+1}) - (r_0 + r_1 - r_m - r_{m+1}) \\
&= (p-1)(q-1),
\end{aligned}$$

which proves the statement. □

One has the immediate corollary of Lemma 3.1, Lemma 3.2, and Lemma 3.3.

Corollary 3.1. *For all integers $a \geq 0$ and $p, q \geq 1$, we have*

$$\mathit{defect}(T(ap+1,p)) = -2a\binom{p}{3} \quad \text{and} \quad \mathit{defect}(T(p,ap-1)) = -2a\binom{p}{3} + 2a\binom{p}{2} - 2(p-1).$$

Fibonacci flat torus knots. An easy symmetry argument implies that the number of negative triples in π is exactly $\frac{p}{3} I(p,q)$, where $I(p,q)$ is the number of *inversions* in π. A classical theorem of Meyer [178] states that

$$I(p,q) = \frac{1}{2}\binom{p-1}{2} - 3p \cdot s(q,p).$$

Here $s(q,p)$ is the standard *Dedekind sum*

$$s(q,p) := \sum_{i=1}^{p-1} \left(\!\left(\frac{qi}{p}\right)\!\right)\left(\!\left(\frac{i}{p}\right)\!\right),$$

where $((\cdot))$ is the *sawtooth function*

$$((x)) := \begin{cases} 0 & \text{if } x \text{ is an integer,} \\ (x \bmod 1) - \frac{1}{2} & \text{otherwise.} \end{cases}$$

(For further background on Dedekind sums, including a self-contained proof of Meyer's theorem, see Rademacher and Grosswald [201].) It immediately follows that

$$\Delta(p,q) = 2p^2 \cdot s(q,p),$$

where $s(q,p)$ is the standard Dedekind sum. Dedekind [74] proved the following reciprocity formula when p and q are relatively prime:

$$s(p,q) + s(q,p) = -\frac{1}{4} + \frac{1}{12}\left(\frac{p}{q} + \frac{1}{pq} + \frac{q}{p}\right).$$

From this reciprocity formula and the easy identity $s(q,p) = s(q \bmod p, p)$, one can derive exact values for the Dedekind sum of consecutive Fibonacci numbers [10, p. 72],

$$s(F_{n+1}, F_n) = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \dfrac{F_n^2 + F_{n-1}^2 - 3F_nF_{n-1} + 1}{12F_nF_{n-1}} = \dfrac{-F_{n-2}}{6F_n} & \text{if } n \text{ is even} \end{cases}$$

and the exact values for defect on Fibonacci flat torus knots follow from careful calculations.

Lemma 3.4. *Let n be an odd number. We have the following.*

$$defect(T(F_{n+1}, F_n)) = \frac{1}{3}(F_n^2 - 1) - F_n + 1 \qquad defect(T(F_n, F_{n-1})) = \frac{1}{3}(F_n^2 - 1) - F_n + 1$$
$$defect(T(F_n, F_{n+1})) = -\frac{1}{3}(F_n^2 - 1) + F_nF_{n+1} - F_{n+1} \qquad defect(T(F_{n-1}, F_n)) = -\frac{1}{3}(F_n^2 - 1) + F_nF_{n-1} - F_{n-1}$$

As an immediate corollary, the absolute value of the defect of both flat torus knots $T(F_n, F_{n-1})$ and $T(F_{n-1}, F_n)$ are linear for any n.

Proof: First let us calculate $defect(T(F_n, F_{n-1}))$. Recall the defect formula using continued fraction of p/q:

$$defect(T(p,q)) \;=\; 2\sum_{k\geq 1}(-1)^k \cdot a_k \cdot \Delta(r_k, r_{k-1}) + 2\sum_{\substack{k\geq 1\\ k \text{ even}}} a_k \cdot \binom{r_k}{2}.$$

In the case of Fibonacci flat torus knots, $r_k = F_{n-k}$, and $a_k = 1$ for all $k \leq n-3$ (because $r_{n-2} = F_2 = 1$). With the assumption that n is an odd number, we have

$$\begin{aligned}
defect(T(F_n, F_{n-1})) &= 2\sum_{k\geq 1}^{n-3}(-1)^k \cdot \Delta(r_k, r_{k-1}) + 2\sum_{\substack{k\geq 1\\ k \text{ even}}}^{n-3} \binom{r_k}{2} \\
&= 2\sum_{k\geq 3}^{n-1}(-1)^k \cdot \Delta(F_k, F_{k+1}) + 2\sum_{\substack{k\geq 3\\ k \text{ even}}}^{n-1} \binom{F_k}{2} && \text{[replace } k \text{ with } n-k\text{]} \\
&= 2\sum_{\substack{k\geq 3\\ k \text{ even}}}^{n-1} 2F_k^2 \cdot \frac{-F_{k-2}}{6F_k} + 2\sum_{\substack{k\geq 3\\ k \text{ even}}}^{n-1} \binom{F_k}{2} && \text{[plug in } \Delta(F_k, F_{k+1})\text{]} \\
&= \frac{1}{3}\sum_{\substack{k\geq 3\\ k \text{ even}}}^{n-1} \left(3F_k^2 - 2F_kF_{k-2}\right) - \sum_{\substack{k\geq 3\\ k \text{ even}}}^{n-1} F_k && \text{[rearrange]} \\
&= \frac{1}{3}\sum_{\substack{k\geq 3\\ k \text{ even}}}^{n-1} \left(F_k^2 + 2F_kF_{k-1}\right) - \sum_{\substack{k\geq 3\\ k \text{ even}}}^{n-1} F_k && \text{[apply } F_k = F_{k-1} + F_{k-2}\text{]} \\
&= \frac{1}{3}\sum_{\substack{k\geq 3\\ k \text{ even}}}^{n-1} \left((F_k + F_{k-1})^2 - F_{k-1}^2\right) - F_n + 2 && \text{[apply } \textstyle\sum_{0\leq i<n} F_{2i} = F_{2n-1} - 1\text{]} \\
&= \frac{1}{3}(F_n^2 - 1) - F_n + 1. && \text{[telescope sum]}
\end{aligned}$$

Similarly we can calculate $defect(T(F_{n+1}, F_n))$. (In fact, the answer is exactly the same.) Applying the reciprocity formula (Lemma 3.3) gives us the last two equations. ☐

3.2 Defect Upper Bound

Polyak's formula [195] gives a straightforward quadratic upper bound on the defect of any closed curve in the plane. In this section, we prove an $O(n^{3/2})$ upper bound on the absolute value of the defect for any planar curves, using a recursive inclusion-exclusion argument. This bound matches the asymptotic worst case behavior of defect among all planar curves, as demonstrated in Section 3.1. Throughout this section, let γ be an arbitrary non-simple closed curve in the plane, and let n be the number of vertices of γ.

3.2.1 Winding Numbers and Diameter

First we derive an upper bound in terms of the depth of the curve. We parametrize γ as a function $\gamma\colon [0,1] \to \mathbb{R}^2$, where $\gamma(0) = \gamma(1)$ is an arbitrarily chosen basepoint. For each vertex x of γ, let $\boldsymbol{\gamma_x}$ denote the closed subpath of γ from the first occurrence of x to the second. More formally, if $x = \gamma(u) = \gamma(v)$ where $0 < u < v < 1$, then γ_x is the closed curve defined by setting $\gamma_x(t) := \gamma((1-t)u + tv)$ for all $0 \le t \le 1$.

Lemma 3.5. *For every vertex x, we have $\sum_{y \between x} \operatorname{sgn}(y) = 2\,\mathit{wind}(\gamma_x, x) - 2\,\mathit{wind}(\gamma_x, \gamma(0)) - \operatorname{sgn}(x)$.*

Proof: Our proof follows an argument of Titus [238, Theorem 1].

Fix a vertex $x = \gamma(u) = \gamma(v)$, where $0 < u < v < 1$. Let $\boldsymbol{\alpha_x}$ denote the subpath of γ from $\gamma(0)$ to $\gamma(u-\varepsilon)$, and let $\boldsymbol{\omega_x}$ denote the subpath of γ from $\gamma(v+\varepsilon)$ to $\gamma(1) = \gamma(0)$, for some sufficiently small $\varepsilon > 0$. Specifically, we choose ε such that there are no vertices $\gamma(t)$ where $u - \varepsilon \le t < u$ or $v < t \le v + \varepsilon$. (See Figure 3.5.) A vertex y interleaves with x if and only if y is an intersection point of γ_x with either α_x or ω_x, so

$$\sum_{y \between x} \operatorname{sgn}(y) \;=\; \sum_{y \in \alpha_x \cap \gamma_x} \operatorname{sgn}(y) \;+\; \sum_{y \in \gamma_x \cap \omega_x} \operatorname{sgn}(y).$$

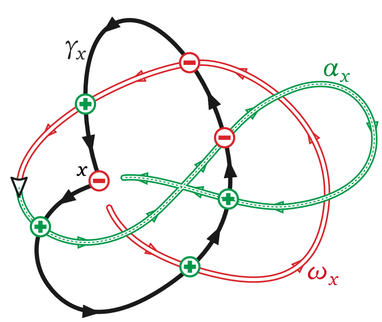


Figure 3.5. Proof of Lemma 3.5: $\mathit{wind}(\gamma_x, x) = +1 - 1 + 1 - \frac{1}{2} = \frac{1}{2}$

Now suppose we move a point p continuously along the path α_x, starting at the basepoint $\gamma(0)$. The winding number $\mathit{wind}(\gamma_x, p)$ changes by 1 each time this point γ_x. Each such crossing happens at a vertex of γ that lies on both α_x and γ_x; if this vertex is positive, $\mathit{wind}(\gamma_x, p)$ increases by 1, and if this vertex is negative, $\mathit{wind}(\gamma_x, p)$ decreases by 1. It follows that

$$\sum_{y \in \alpha_x \cap \gamma_x} \operatorname{sgn}(y) \;=\; \mathit{wind}(\gamma_x, \gamma(u-\varepsilon)) - \mathit{wind}(\gamma_x, \gamma(0)).$$

Symmetrically, if we move a point p *backward* along ω_x from the basepoint, the winding number $\mathit{wind}(\gamma_x, p)$ increases (resp. decreases) by 1 whenever $\gamma(t)$ passes through a positive (resp. negative) vertex in $\omega_x \cap \gamma_x$; see the red path in Figure 3.5. Thus,

$$\sum_{y \in \omega_x \cap \gamma_x} \operatorname{sgn}(y) \;=\; wind(\gamma_x, \gamma(v+\varepsilon)) - wind(\gamma_x, \gamma(0)).$$

Finally, our sign convention for vertices implies

$$wind(\gamma_x, \gamma(u-\varepsilon)) \;=\; wind(\gamma_x, \gamma(v+\varepsilon)) \;=\; wind(\gamma_x, x) - \operatorname{sgn}(x)/2,$$

which completes the proof. □

Lemma 3.6. *For any planar curve γ, we have $|defect(\gamma)| \le 2n \cdot depth(\gamma) + n$.*

Proof: Polyak's defect formula can be rewritten as

$$defect(\gamma) = -\sum_x \operatorname{sgn}(x) \left(\sum_{y \between x} \operatorname{sgn}(y) \right).$$

(This sum actually considers every pair of interleaved vertices twice, which is why the factor 2 is omitted.) Assume without loss of generality that the basepoint $\gamma(0)$ lies on the outer face of γ, so that $wind(\gamma_x, \gamma(0)) = 0$ for every vertex x. Then Lemma 3.5 implies

$$defect(\gamma) \;=\; \sum_x \operatorname{sgn}(x) \Big(\operatorname{sgn}(x) - 2\, wind(\gamma_x, x) \Big) \;=\; n - 2 \sum_x \operatorname{sgn}(x) \cdot wind(\gamma_x, x),$$

and therefore

$$|defect(\gamma)| \le n + 2 \sum_x |wind(\gamma_x, x)|\,.$$

We easily observe that $|wind(\gamma_x, x)| \le depth(x, \gamma_x) \le depth(x, \gamma)$ for every vertex x; the second inequality follows from the fact that no path crosses γ_x more times than it crosses γ. The lemma now follows immediately. □

The quantity $\sum_x \operatorname{sgn}(x) \cdot wind(\gamma_x, x)$ is equivalent, up to a factor of 4, to the curve invariant $\alpha(\gamma)$ introduced by Lin and Wang [167], which they defined as the limit of a certain integral (due to Bar-Natan [19]) over a smooth knot in $\mathbb{R}^3$ that projects to γ, as the knot approaches the plane of projection.

As we will see the upper bound $|defect(\gamma)| = O(n \cdot depth(\gamma))$ also follows from either our $O(n^{3/2})$ upper bound for homotopy moves (Lemma 5.1) which serves as an upper bound on defect (Lemma 4.1), or from the relation between number of medial electrical moves and defect (Theorem 7.2) and the electrical reduction algorithm of Feo and Provan [100].

3.2.2 Inclusion-Exclusion

Now let σ be a simple closed curve that intersects γ only transversely and away from its vertices. By the Jordan curve theorem, we can assume without loss of generality that σ is a Euclidean circle, the number of intersection points between γ and σ is even, and the intersection points are evenly spaced around σ. We arbitrarily refer to the two tangles defined by σ as the *interior* and *exterior* tangles of σ. Let $z_0, z_1, \dots, z_{s-1}$ be the points in $\sigma \cap \gamma$ in order along γ (*not* in order along σ). These intersection points decompose γ into a sequence of s subpaths $\gamma_1, \gamma_2, \dots, \gamma_s$; specifically, γ_i is the subpath of γ from z_{i-1} to $z_{i \bmod s}$, for each index i. Without loss of generality, every odd-indexed path γ_{2i+1} lies outside σ, and every even-indexed path γ_{2i} lies inside σ.

Let $\boldsymbol{\gamma \Cap \sigma}$ and $\boldsymbol{\gamma \Cup \sigma}$ denote the closed curves that result from tightening the interior and exterior tangles of σ, respectively.[2] To put it differently, let $\gamma \Cap \sigma$ denote a generic curve obtained from γ by continuously deforming all subpaths γ_i inside σ, keeping their endpoints fixed and never moving across σ, to minimize the number of intersections. (There may be several curves that satisfy the minimum-intersection condition; choose one arbitrarily.) Similarly, let $\gamma \Cup \sigma$ denote any generic curve obtained by continuously deforming the subpaths γ_i outside σ to minimize intersections. Finally, let $\boldsymbol{\gamma \circledcirc \sigma}$ denote the generic curve obtained by deforming *all* subpaths γ_i to minimize intersections; in other words, $\gamma \circledcirc \sigma := (\gamma \Cap \sigma) \Cup \sigma = (\gamma \Cup \sigma) \Cap \sigma$. See Figure 3.6.

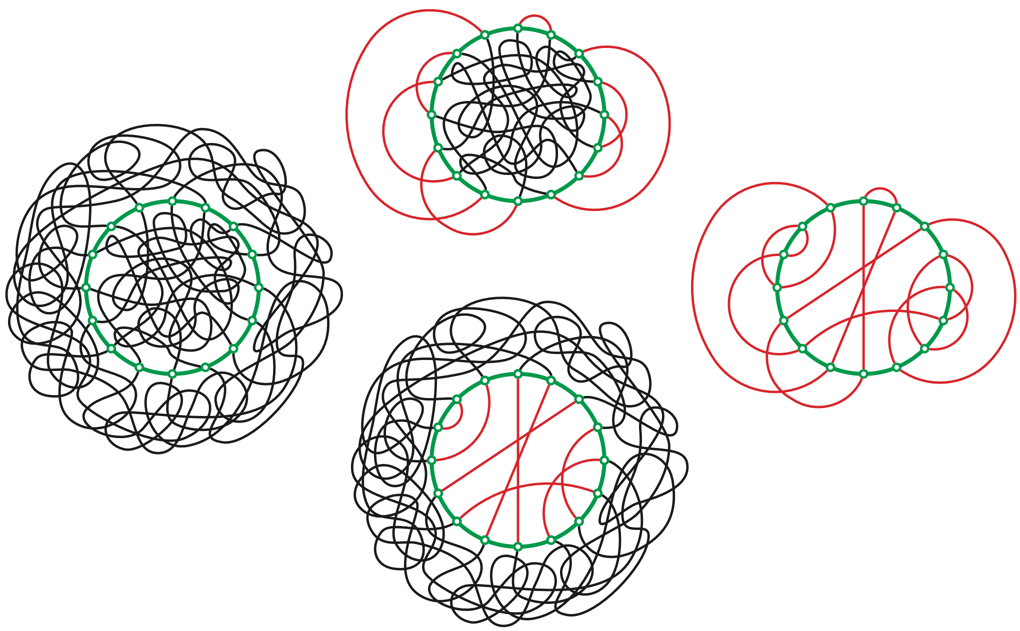


Figure 3.6. Clockwise from left: γ, $\gamma \Cup \sigma$, $\gamma \circledcirc \sigma$, and $\gamma \Cap \sigma$. The green circle in all four figures is σ.

To simplify notation, we define

$$\boldsymbol{defect(x,y)} \;:=\; [x \between y] \cdot \operatorname{sgn}(x) \cdot \operatorname{sgn}(y)$$

for any two vertices x and y, where $[x \between y] := 1$ if x and y are interleaved and $[x \between y] = 0$ otherwise. Then we can write the defect of γ as

$$\mathit{defect}(\gamma) \;=\; -2 \sum_{x,y} \mathit{defect}(x,y).$$

Every vertex of γ lies at the intersection of two (not necessarily distinct) subpaths. For any index i, let $\boldsymbol{X(i,i)}$ denote the set of self-intersection points of γ_i, and for any indices $i < j$, let $\boldsymbol{X(i,j)}$ be the set of points where γ_i intersects γ_j.

If two vertices $x \in X(i,k)$ and $y \in X(j,l)$ are interleaved, then we must have $i \le j \le k \le l$. Thus, we can express the defect of γ in terms of crossings between subpaths γ_i as follows.

$$\mathit{defect}(\gamma) \;=\; -2 \sum_{i \le j \le k \le l} \; \sum_{x \in X(i,k)} \; \sum_{y \in X(j,l)} \mathit{defect}(x,y)$$

[2]We recommend pronouncing $\Cap$ as "tightened inside" and $\Cup$ as "tightened outside"; note that the symbols $\Cap$ and $\Cup$ resemble the second letters of "inside" and "outside".

On the other hand, if $i < j < k < l$, then every vertex $x \in \gamma_i \cap \gamma_k$ is interleaved with every vertex of $y \in \gamma_j \cap \gamma_l$. Thus, we can express the contribution to the defect from pairs of vertices on four *distinct* subpaths as follows:

$$\boldsymbol{defect^{\#}(\gamma,\sigma)} \;:=\; -2 \sum_{i<j<k<l} \; \sum_{x \in X(i,k)} \; \sum_{y \in X(j,l)} \operatorname{sgn}(x) \cdot \operatorname{sgn}(y)$$

We can express this function more succinctly as

$$defect^{\#}(\gamma,\sigma) \;=\; -2 \sum_{i<j<k<l} defect(i,k) \cdot defect(j,l)$$

by defining

$$\boldsymbol{defect(i,j)} \;:=\; \sum_{x \in X(i,j)} \operatorname{sgn}(x)$$

for all indices $i < j$.

The following lemma implies that continuously deforming the subpaths γ_i without crossing σ leaves the value $defect^{\#}(\gamma,\sigma)$ unchanged, even though such a deformation may change the defect $defect(\gamma)$.

Lemma 3.7. *The value $defect(i,j)$ depends only on the parity of $i+j$ and the cyclic order of the endpoints of γ_i and γ_j around σ.*

Proof: There are only three cases to consider.

If $i+j$ is odd, then γ_i and γ_j lie on opposite sides of σ and therefore do not intersect, so $defect(i,j)=0$. For all other cases, $i+j$ is even, which implies without loss of generality that $j \geq i+2$.

Suppose the endpoints of γ_i and γ_j do not alternate in cyclic order around σ, or equivalently, that the corresponding subpaths of $\gamma \circledcirc \sigma$ are disjoint. The Jordan curve theorem implies that there must be equal numbers of positive and negative intersections between γ_i and γ_j, and therefore $defect(i,j)=0$.

Finally, suppose the endpoints of γ_i and γ_j alternate in cyclic order around σ, or equivalently, that the corresponding subpaths of $\gamma \circledcirc \sigma$ intersect exactly once. Then $defect(i,j)=1$ if the endpoints $z_i, z_j, z_{i-1}, z_{j-1}$ appear in clockwise order around σ and $defect(i,j)=-1$ otherwise. □

Now consider an interleaved pair of vertices $x \in X(i,k)$ and $y \in X(j,l)$ where at least two of the indices i,j,k,l are equal. Trivially, i and k have the same parity, and j and l also have the same parity. If $i=j$ or $i=l$ or $j=k$ or $j=l$, then all four indices have the same parity. If $i=k$, then we must also have $i=j$ or $i=l$ (or both), so again, all four indices have the same parity. We conclude that x and y are either both inside σ or both outside σ.

Lemma 3.8. *For any closed curve γ and any simple closed curve σ that intersects γ only transversely and away from its vertices, we have $defect(\gamma) = defect(\gamma \Cap \sigma) + defect(\gamma \Cup \sigma) - defect(\gamma \circledcirc \sigma)$.*

Proof: Let us write $defect(\gamma) = defect^{\#}(\gamma,\sigma) + defect^{\uparrow}(\gamma,\sigma) + defect^{\downarrow}(\gamma,\sigma)$, where

- $defect^{\#}(\gamma,\sigma)$ considers pairs of vertices on four different subpaths γ_i, as above,
- $defect^{\uparrow}(\gamma,\sigma)$ considers pairs of vertices inside σ on at most three different subpaths γ_i, and
- $defect^{\downarrow}(\gamma,\sigma)$ considers pairs of vertices outside σ on at most three different subpaths γ_i.

Lemma 3.7 implies that

$$defect^{\#}(\gamma,\sigma) = defect^{\#}(\gamma \Cap \sigma,\sigma) = defect^{\#}(\gamma \Cup \sigma,\sigma) = defect^{\#}(\gamma \circledcirc \sigma,\sigma).$$

The definitions of $\gamma \Cap \sigma$ and $\gamma \Cup \sigma$ immediately imply the following:

$$\begin{aligned} \mathit{defect}^{\uparrow}(\gamma \Cap \sigma, \sigma) &= \mathit{defect}^{\uparrow}(\gamma \circledcirc \sigma, \sigma) & \mathit{defect}^{\downarrow}(\gamma \Cap \sigma, \sigma) &= \mathit{defect}^{\downarrow}(\gamma, \sigma) \\ \mathit{defect}^{\uparrow}(\gamma \Cup \sigma, \sigma) &= \mathit{defect}^{\uparrow}(\gamma, \sigma) & \mathit{defect}^{\downarrow}(\gamma \Cup \sigma, \sigma) &= \mathit{defect}^{\downarrow}(\gamma \circledcirc \sigma, \sigma) \end{aligned}$$

The lemma now follows from straightforward substitution. □

Lemma 3.9. *For any closed curve γ and any simple closed curve σ that intersects γ only transversely and away from its vertices, we have $|\mathit{defect}(\gamma \circledcirc \sigma)| = O\left(|\gamma \cap \sigma|^3\right)$.*

Proof: Fix an arbitrary reference point $z \in \sigma \setminus \gamma$. For any point p in the plane, there is a path from p to z that crosses $\gamma \circledcirc \sigma$ at most $O(s)$ times. Specifically, move from p to the nearest point on $\gamma \circledcirc \sigma$, then follow $\gamma \circledcirc \sigma$ to σ, and finally follow σ to the reference point z. It follows that $\mathit{depth}(\gamma \circledcirc \sigma) = O(s)$. The curve $\gamma \circledcirc \sigma$ has at most $2\binom{s/2}{2} = O(s^2)$ vertices. The bound $|\mathit{defect}(\gamma \circledcirc \sigma)| = O(s^3)$ now immediately follows from Lemma 3.6. □

3.2.3 Divide and Conquer

We call a simple closed curve σ ***useful*** for γ if σ intersects γ transversely away from its vertices, and the interior tangle $\oplus$ of σ has at least s^2 vertices, where $s := |\sigma \cap \gamma|/2$ is the number of strands in $\oplus$.[3]

Lemma 3.10. *Let γ be an arbitrary non-simple closed curve in the plane with n vertices. Either there is a useful simple closed curve for γ whose interior tangle has depth $O(\sqrt{n})$, or the depth of γ is $O(\sqrt{n})$.*

Proof: To simplify notation, let $d := \mathit{depth}(\gamma)$. For each integer j between 1 and d, let R_j be the set of points p with $\mathit{depth}(p, \gamma) \geq d + 1 - j$, and let $\tilde{R}_j$ denote a small open neighborhood of the closure of $R_j \cup \tilde{R}_{j-1}$, where $\tilde{R}_0$ is the empty set. Each region $\tilde{R}_j$ is the disjoint union of closed disks, whose boundary cycles intersect γ transversely away from its vertices, if at all. In particular, $\tilde{R}_d$ is a disk containing the entire curve γ.

Fix a point z such that $\mathit{depth}(z, \gamma) = d$. For each integer j, let Σ_j be the unique component of $\tilde{R}_j$ that contains z, and let σ_j be the boundary of Σ_j. Then $\sigma_1, \sigma_2, \ldots, \sigma_d$ are disjoint, nested, simple closed curves; see Figure 3.7. Let n_j be the number of vertices and let $s_j := |\gamma \cap \sigma_j|/2$ be the number of strands of the interior tangle of σ_j. For notational convenience, we define $\Sigma_0 := \varnothing$ and thus $n_0 = s_0 = 0$. We ignore the outermost curve σ_d, because it contains the entire curve γ. The next outermost curve σ_{d-1} contains every vertex of γ, so $n_{d-1} = n$.

By construction, for each j, the interior tangle of σ_j has depth $j + 1$. Thus, to prove the lemma, it suffices to show by induction that if none of the curves $\sigma_1, \sigma_2, \ldots, \sigma_{d-1}$ is useful, then $d = O(\sqrt{n})$.

Fix an index j. Each edge of γ crosses σ_j at most twice. Any edge of γ that crosses σ_j has at least one endpoint in the annulus $\Sigma_j \setminus \Sigma_{j-1}$, and any edge that crosses σ_j twice has both endpoints in $\Sigma_j \setminus \Sigma_{j-1}$. Conversely, each vertex in Σ_j is incident to at most two edges that cross σ_j and no edges that cross σ_{j+1}. It follows that $|\sigma_j \cap \gamma| \leq 2(n_j - n_{j-1})$, and therefore $n_j \geq n_{j-1} + s_j$. Thus, by induction, we have

$$n_j \geq \sum_{i=1}^{j} s_i$$

for every index j.

[3] We could define σ to be useful if there are at least $\alpha \cdot s^2$ vertices in the interior tangle, and then optimize α to minimize the resulting upper bound.

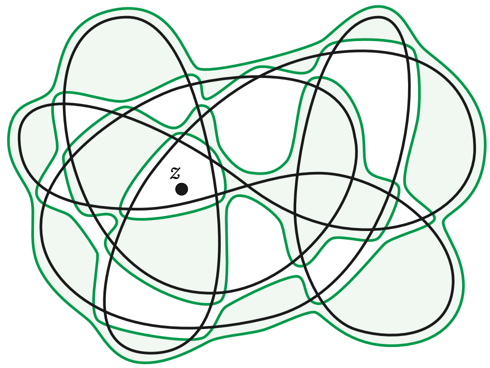


Figure 3.7. Nested depth cycles around a point of maximum depth.

Now suppose no curve σ_j with $1 \le j < d$ is useful. Then we must have $s_j^2 > n_j$ and therefore

$$s_j^2 > \sum_{i=1}^{j} s_i$$

for all $1 \le j < d$. Trivially, $s_1 \ge 1$, because γ is non-simple. A straightforward induction argument implies that $s_j \ge (j+1)/2$ and therefore

$$n \;=\; n_{d-1} \;\ge\; \sum_{i=1}^{d-1} \frac{i+1}{2} \;\ge\; \frac{1}{2}\binom{d+1}{2} \;>\; \frac{d^2}{4}.$$

We conclude that $d \le 2\sqrt{n}$, which completes the proof. □

We are now finally ready to prove our main upper bound.

Theorem 3.1. *$|defect(\gamma)| = O(n^{3/2})$ for every closed curve γ in the plane with n vertices.*

Proof: We prove by induction on n that $defect(\gamma) \le C \cdot n^{3/2}$ for any closed curve γ with n vertices, for some absolute constant C to be determined.

Let γ be an arbitrary closed curve with n vertices. Let σ be a simple closed curve that is useful for γ (that is, $m \ge s^2$) whose interior tangle has depth $O(\sqrt{n})$, as guaranteed by Lemma 3.10. (If there are no useful curves, then Lemma 3.6 implies that $|defect(\gamma)| = O(n^{3/2})$.) Let $s := |\gamma \cap \sigma|/2$. Lemma 3.8 implies

$$defect(\gamma) \;=\; defect(\gamma \Cap \sigma) + defect(\gamma \Cup \sigma) - defect(\gamma \circledcirc \sigma).$$

Suppose there are m vertices of γ lying in the interior of σ. Because the interior tangle of σ has depth $O(\sqrt{n})$, it follows that $depth(\gamma \Cup \sigma) = O(\sqrt{n} + s)$, and therefore by Lemma 3.6 and Lemma 3.9 this implies

$$|defect(\gamma \Cup \sigma)| + |defect(\gamma \circledcirc \sigma)| \;=\; O\big((\sqrt{n}+s)\cdot(m+s^2/2)\big) \;=\; O(m\sqrt{n}).$$

Because σ is useful for γ, $\gamma \Cap \sigma$ has at most $n - m + s^2/2 < n$ vertices. By the inductive hypothesis one has

$$|defect(\gamma)| \;\le\; C\left(n - m + s^2/2\right)^{3/2} + c \cdot m\sqrt{n}$$

for some constant c. The inequality $(x-y)^{3/2} \le (x-y)x^{1/2} = x^{3/2} - yx^{1/2}$ now implies

$$|defect(\gamma)| \;\le\; Cn^{3/2} - C(m - s^2/2)\sqrt{n} + c \cdot m\sqrt{n}.$$

Finally, again because σ is useful, we must have $m - s^2/2 \ge m/2$, which implies

$$\begin{aligned} |defect(\gamma)| \;&\le\; Cn^{3/2} - C(m/2)\sqrt{n} + c \cdot m\sqrt{n} \\ &=\; Cn^{3/2} - (C/2 - c)m\sqrt{n}. \end{aligned}$$

Provided $C/c \ge 2$, then $|defect(\gamma)| \le Cn^{3/2}$, as required. □

3.3 Medial Defect is Independent of Planar Embeddings

Recall that an embedded graph G *unicursal* if its medial graph $G^\times$ is the image of a single closed curve. The goal of the section is to prove that following surprising property about defect: The defect of the medial graph of an arbitrary unicursal planar graph G does not depend on its embedding.

Theorem 3.2. *Let G and H be planar embeddings of the same abstract planar graph. If G is unicursal, then H is unicursal and $defect(G^\times) = defect(H^\times)$.*

3.3.1 Navigating Between Planar Embeddings

A classical result of Adkisson [3] and Whitney [256] is that every 3-connected planar graph has an essentially unique planar embedding. Mac Lane [175] described how to count the planar embeddings of any biconnected planar graph, by decomposing it into its triconnected components. Stallmann [228, 229] and Cai [31] extended Mac Lane's algorithm to arbitrary planar graphs, by decomposing them into biconnected components. Mac Lane's decomposition is also the basis of the SPQR-tree data structure of Di Battista and Tamassia [78, 79], which encodes all planar embeddings of an arbitrary planar graph.

Mac Lane's structural results imply that any planar embedding of a 2-connected planar graph G can be transformed into any other embedding by a finite sequence of ***split reflections***, defined as follows. A ***split curve*** is a simple closed curve σ whose intersection with the embedding of G consists of two vertices x and y; without loss of generality, σ is a circle with x and y at opposite points. A split reflection modifies the embedding of G by reflecting the subgraph inside σ across the line through x and y.

Lemma 3.11. *Let G be an arbitrary 2-connected planar graph. Any planar embedding of G can be transformed into any other planar embedding of G by a finite sequence of split reflections.*

To navigate among the planar embeddings of *arbitrary* connected planar graphs, we need two additional operations. First, we allow split curves that intersect G at only a single cut vertex; a ***cut reflection*** modifies the embedding of G by reflecting the subgraph inside such a curve. More interestingly, we also allow degenerate split curves that pass through a cut vertex x of G *twice*, but are otherwise simple and disjoint from G. The interior of a degenerate split curve σ is an open topological disk. A ***cut eversion*** is a degenerate split reflection that everts the embedding of the subgraph of G inside such a curve, intuitively by mapping the interior of σ to an open circular disk (with two copies of x on its boundary), reflecting the interior subgraph, and then mapping the resulting embedding back to the interior of σ. Structural results of Stallman [228, 229] and Di Battista and Tamassia [79, Section 7] imply the following.

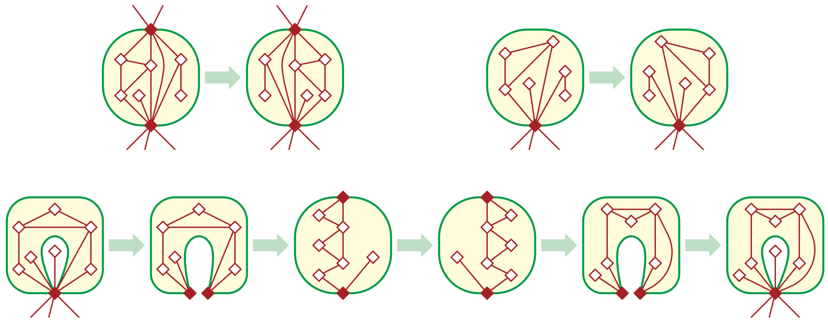


Figure 3.8. Top row: A regular split reflection and a cut reflection. Bottom row: a cut eversion.

Lemma 3.12. *Let G be an arbitrary connected planar graph. Any planar embedding of G can be transformed into any other planar embedding of G by a finite sequence of split reflections, cut reflections, and cut eversions.*

3.3.2 Tangle Flips

Now consider the effect of the operations stated in Lemma 3.12 on the medial graph $G^\times$. By assumption, G is unicursal so that $G^\times$ is a single closed curve. Let σ be any (possibly degenerate) split curve for G. Embed $G^\times$ so that every medial vertex lies on the corresponding edge in G, and every medial edge intersects σ at most once. Then σ intersects at most four edges of $G^\times$, so the tangle of $G^\times$ inside σ has at most two strands. Moreover, reflecting (or everting) the subgraph of G inside σ induces a *flip* of this tangle of $G^\times$. Any tangle can be ***flipped*** by reflecting the disk containing it, so that each strand endpoint maps to a different strand endpoint; see Figure 3.9. Straightforward case analysis implies that flipping any tangle of $G^\times$ with at most two strands transforms $G^\times$ into another closed curve.

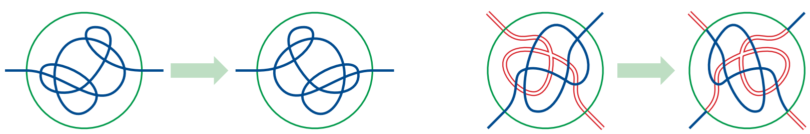


Figure 3.9. Flipping tangles with one and two strands.

The main result of this subsection is that the resulting curve has the same defect as $G^\times$.

Lemma 3.13. *Let γ be an arbitrary closed curve on the sphere. Flipping any tangle of γ with one strand yields another closed curve γ' with $defect(\gamma') = defect(\gamma)$.*

Proof: Let σ be a simple closed curve that crosses γ at exactly two points. These points decompose σ into two subpaths $\alpha \cdot \beta$, where α is the unique strand of the interior tangle and β is the unique strand of the exterior tangle. Let Σ denote the interior disk of σ, and let $\phi : \Sigma \to \Sigma$ denote the homeomorphism that flips the interior tangle. Flipping the interior tangle yields the closed curve $\gamma' := rev(\phi(\alpha)) \cdot \beta$, where *rev* denotes path reversal.

No vertex of α is interleaved with a vertex of β; thus, two vertices in γ' are interleaved if and only if the corresponding vertices in γ are interleaved. Every vertex of $rev(\phi(\alpha))$ has the same sign as the corresponding vertex of α, since both the orientation of the vertex and the order of traversals through the vertex changed. Thus, every vertex of γ' has the same sign as the corresponding vertex of γ. We conclude that $defect(\gamma') = defect(\gamma)$. □

Lemma 3.14. *Let γ be an arbitrary closed curve on the sphere. Flipping any tangle of γ with two strands yields another closed curve γ' with defect(γ') = defect(γ).*

Proof: Let σ be a simple closed curve that crosses γ at exactly four points. These four points naturally decompose γ into four subpaths $\alpha \cdot \delta \cdot \beta \cdot \varepsilon$, where α and β are the strands of the interior tangle of σ, and δ and ε are the strands of the exterior tangle. Flipping the interior tangle either exchanges α and β, reverses α and β, or both; see Figure 3.10. In every case, the result is a single closed curve γ'.

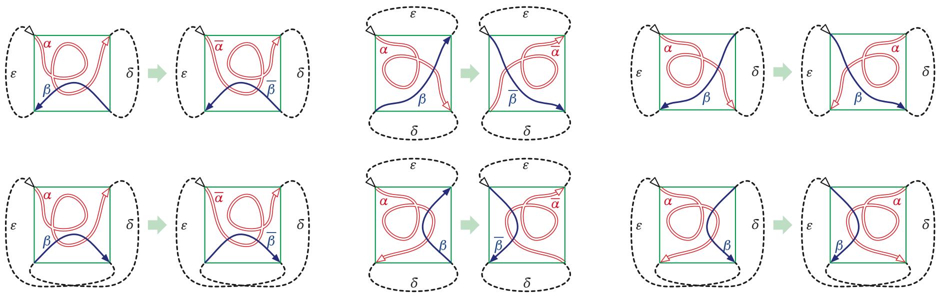


Figure 3.10. Flipping all six types of 2-strand tangle.

Let γ' be the result of flipping the interior tangle. The curve $\gamma' \Cup \sigma$ is just a reflection of $\gamma \Cup \sigma$, which implies that $\mathit{defect}(\gamma' \Cup \sigma) = \mathit{defect}(\gamma \Cup \sigma)$, and straightforward case analysis implies $\gamma' \Cap \sigma = \gamma \Cap \sigma$ and $\gamma' \odot \sigma = \gamma \odot \sigma$. We conclude by the inclusion-exclusion formula for defect (Lemma 3.8) that

$$\begin{aligned}\mathit{defect}(\gamma') &= \mathit{defect}(\gamma' \Cap \sigma) + \mathit{defect}(\gamma' \Cup \sigma) - \mathit{defect}(\gamma' \odot \sigma)\\ &= \mathit{defect}(\gamma \Cap \sigma) + \mathit{defect}(\gamma \Cup \sigma) - \mathit{defect}(\gamma \odot \sigma) \;=\; \mathit{defect}(\gamma).\end{aligned}$$

□

Now Theorem 3.2 follows immediately from Lemmas 3.12, 3.13, and 3.14.

3.4 Implications for Random Knots

Finally, we describe some interesting implications of our results on the expected behavior of random knots, following earlier results of Lin and Wang [167], Polyak [195], and new results of Even-Zohar, Hass, Linial, and Nowik [95, 96, 97]. We refer the reader to Burde and Zieschang [27] or Kauffman [151] for further background on knot theory, to Chmutov, Duzhin, and Mostovoy [57] for a detailed overview of finite-type knot invariants, and Even-Zohar [94] for a survey and some new results on the random knot models; we include only a few elementary definitions to keep the presentation self-contained.

A ***knot*** is (the image of) a continuous injective map from the circle into $\mathbb{R}^3$. Two knots are considered equivalent (more formally, *ambient isotopic*) if there is a continuous deformation of $\mathbb{R}^3$ that deforms one knot into the other. Knots are often represented by ***knot diagrams***, which are 4-regular plane graphs defined by a generic projection of the knot onto the plane, with an annotation at each vertex indicating which branch of the knot is "over" or "under" the other. Call any crossing x in a knot diagram *ascending* if the first branch through x after the basepoint passes over the second, and *descending* otherwise.

The ***Casson invariant*** c_2 is the simplest finite-type knot invariant; it is also equal to the second coefficient of the Conway polynomial [24, 198]. Polyak and Viro [197, 198] derived the following combinatorial formula for the

Casson invariant of a knot diagram κ:

$$c_2(\kappa) \;=\; -\sum_{\text{descending } x}\;\sum_{\text{ascending } y} [x \between y] \cdot \operatorname{sgn}(x) \cdot \operatorname{sgn}(y).$$

Like defect, the value of $c_2(\kappa)$ is independent of the choice of basepoint or orientation of the underlying curve γ; moreover, if the knots represented by diagrams κ and κ' are equivalent, then $c_2(\kappa) = c_2(\kappa')$.

Polyak [195, Theorem 7] observed that if a knot diagram κ is obtained from an arbitrary closed curve γ by independently resolving each crossing as ascending or descending with equal probability, then one can relate the expectation of Casson invariant $c_2(\kappa)$ and the defect of γ by

$$\mathrm{E}[c_2(\kappa)] = \mathit{defect}(\gamma)/8.$$

The same observation is implicit in earlier results of Lin and Wang [167]; and (for specific curves) in the later results of Even-Zohar *et al.* [95].

Even-Zohar *et al.* [95] studied the distribution of the Casson invariant for two models of random knots, the *Petaluma* model of Adams *et al.* [1, 2], which uses singular one-vertex diagrams consisting of $2p+1$ disjoint non-nested loops for some integer p, and the *star* model, which uses (a polygonal version of) the flat torus knot $T(p, 2p+1)$ for some integer p. Even-Zohar *et al.* prove that the expected value of the Casson invariant is $\binom{p}{2}/12$ in the Petaluma model and $\binom{p+1}{3}/2 \approx 0.03n^{3/2}$ in the star model. Later they studied the Petaluma model in further details [96]; in particular, the probability that Casson invariant of a random knot is equal to a given value decreases to zero as the number of petals grows.

Our defect analysis in Section 3.2 implies an upper bound on the Casson invariant for knot diagrams generated from *any* family of generic closed curves.

Corollary 3.2. *Let γ be any generic closed curve with n vertices, and let κ be a knot diagram obtained by resolving each vertex of γ independently and uniformly at random. Then $|\mathrm{E}[c_2(\kappa)]| = O(n^{3/2})$.*

Our results also imply that the distribution of the Casson invariant depends strongly on the precise parameters of the random model; even the sign and growth rate of $\mathrm{E}[c_2]$ depend on which curves are used to generate knot diagrams. For example:

- For random diagrams over the flat torus knot $T(p+1, p)$, we have $\mathrm{E}[c_2(\kappa)] = -\binom{p}{3}/4 = -n^{3/2}/24 + \Theta(n)$.
- For random diagrams over the Fibonacci flat torus knot $T(F_{k+1}, F_k)$, we have $\mathrm{E}[c_2(\kappa)] = \frac{1}{3}(F_k^2 - 1) - F_k + 1 = n/3\phi + \Theta(\sqrt{n})$, where $\phi := (\sqrt{5}+1)/2$ is the golden ratio.
- For random diagrams over the connected sum $T(p-1, p) \# T(p+1, p)$, we have $\mathrm{E}[c_2(\kappa)] = 0$.

Chapter 4

Lower Bounds for Tightening Curves

Lower bounds are hard. But that doesn't mean that no progress can be made. To get a lower bound, it is required [...] that you make a (possible restrictive) model of all algorithms or data-structures that can solve your problem.

— Discrete lizard, *Computer Science Stack Exchange q91156*

In this chapter we prove the first non-trivial lower bounds on number of homotopy moves required to tighten closed curves, both in the plane and on higher-genus surfaces. First, in Section 4.1, we derive an $\Omega(n^{3/2})$ lower bound on number of homotopy moves required to simplify any planar curve, using lower bound results on defect we have in Section 3.1, and the fact that each homotopy move changes the defect of a closed curve by at most 2. As for planar multicurves, using winding-number arguments we prove that in the worst case, simplifying an arrangement of k closed curves requires $\Omega(n^{3/2} + nk)$ homotopy moves, with an additional $\Omega(k^2)$ term if the target configuration is specified in advance.

In Section 4.2, we consider curves on surfaces of higher genus. Extending the notion of defect invariant, we prove that $\Omega(n^2)$ homotopy moves are required in the worst case to transform one non-contractible closed curve to another on the torus, and therefore on any orientable surface. Results of Hass and Scott [135] imply that this lower bound is tight if the non-contractible closed curve is homotopic to a simple closed curve.

We then construct an infinite family of contractible curves on the annulus that require at least $\Omega(n^2)$ moves to tighten in Section 4.3, using a complete different curve invariant than defect. Our new lower bound generalizes to any surface that has the annulus as a covering space—that is, any surface except for the sphere, the disk, or the projective plane.

4.1 Lower Bounds for Planar Curves

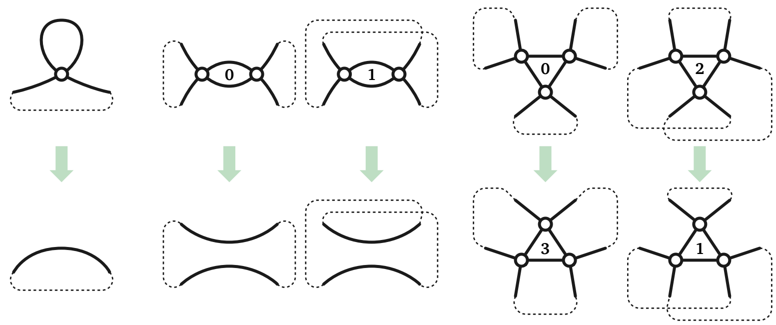
Now we prove our lower bounds for simplifying closed curves in the plane through the defect invariant. Straightforward case analysis [195] implies that any single homotopy move changes the defect of a curve by at most 2; the various cases are listed below and illustrated in Figure 4.1.

- A 1→0 move leaves the defect unchanged.
- A 2→0 move decreases the defect by 2 if the two disappearing vertices are interleaved, and leaves the defect unchanged otherwise.
- A 3→3 move increases the defect by 2 if the three vertices before the move contain an even number of interleaved pairs, and decreases the defect by 2 otherwise.

In light of this case analysis, the following lemma is trivial:

Lemma 4.1. *Simplifying any closed curve γ in the plane requires at least $|\mathit{defect}(\gamma)|/2$ homotopy moves.*

Move	1→0	2→0		3→3	



St	$-w$	0	0	+1	+1
J^+	$2w$	0	−2	0	0
defect	0	0	−2	+2	+2

Figure 4.1. Changes to Arnold's invariants: St, J^+, and defect incurred by homotopy moves. Numbers in each figure indicate how many pairs of vertices are interleaved; dashed lines indicate how the rest of the curve connects. The variable w shown in the 0→1 move column represents the winding number of the vertex.

As we have mentioned in Chapter 3, Arnold [15, 16] originally defined the curve invariants St and J^+ by their changes under 2→0 and 3→3 homotopy moves. Specifically, as shown in Figure 4.1, 3→3 moves change strangeness by ±1 but do not affect J^+; 2→0 moves change J^+ by either 0 or 2 but do not affect strangeness.

Defect bound from either Lemma 3.1, Lemma 3.2, or Corollary 3.1 implies the following lower bound on number of homotopy moves, which is also implicit in the work of Hayashi *et al.* [139] and Even-Zohar *et al.* [95].

Theorem 4.1. *For every positive integer n, there are closed curves with n vertices whose defects are $n^{3/2}/3 - O(n)$ and $-n^{3/2}/3 + O(n)$, and therefore require at least $n^{3/2}/6 - O(n)$ homotopy moves to reduce to a simple closed curve.*

Proof: The lower bound follows from Lemma 3.1, Lemma 3.2, or Corollary 3.1 by setting $a := 1$. If n is a prefect square, then the flat torus knot $T(\sqrt{n}+1, \sqrt{n})$ has n vertices and defect $-2\binom{\sqrt{n}}{3}$. If n is not a perfect square, we can achieve defect $-2\binom{\lfloor\sqrt{n}\rfloor}{3}$ by applying 0→1 moves to the curve $T(\lfloor\sqrt{n}\rfloor+1, \lfloor\sqrt{n}\rfloor)$. Similarly, we obtain an n-vertex curve with defect $2\binom{\lfloor\sqrt{n+1}\rfloor+1}{3}$ by adding monogons to the curve $T(\lfloor\sqrt{n+1}\rfloor, \lfloor\sqrt{n+1}\rfloor+1)$. Lemma 4.1 now immediately implies the lower bound on homotopy moves. □

4.1.1 Multicurves

Our previous results immediately imply that simplifying a multicurve with n vertices requires at least $\Omega(n^{3/2})$ homotopy moves; in this section we derive additional lower bounds in terms of the number of constituent curves. We distinguish between two natural variants of simplification: transforming a multicurve into an *arbitrary* set of disjoint simple closed curves, or into a *particular* set of disjoint simple closed curves.

Lemma 4.2. *Transforming a k-curve with n vertices in the plane into k arbitrary disjoint circles requires $\Omega(nk)$ homotopy moves in the worst case.*

Proof: For arbitrary positive integers n and k, we construct a multicurve with k disjoint constituent curves, all but one of which are simple, as follows. The first $k-1$ constituent curves $\gamma_1, \ldots, \gamma_{k-1}$ are disjoint circles inside the open unit disk centered at the origin. (The precise configuration of these circles is unimportant.) The remaining constituent curve γ_o is a spiral winding $n+1$ times around the closed unit disk centered at the origin, plus a line segment connecting the endpoints of the spiral; γ_o is the simplest possible curve with winding number $n+1$ around the origin. Let γ be the disjoint union of these k curves; we claim that $\Omega(nk)$ homotopy moves are required to simplify γ. See Figure 4.2.

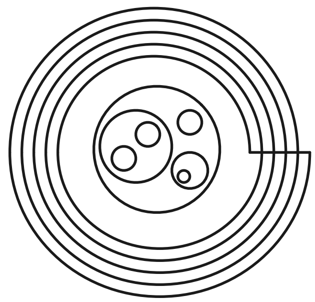


Figure 4.2. Simplifying this multicurve requires $\Omega(nk)$ homotopy moves.

Consider the faces of the outer curve γ_o during any homotopy of γ. Adjacent faces of γ_o have winding numbers that differ by 1, and the outer face has winding number 0. Thus, for any non-negative integer w, as long as the maximum absolute winding number $\left|\max_p wind(\gamma_o, p)\right|$ is at least w, the curve γ_o has at least $w+1$ faces (including the outer face) and therefore at least $w-1$ vertices, by Euler's formula. On the other hand, if any curve γ_i intersects a face of γ_o, no homotopy move can remove that face until the intersection between γ_i and γ_o is removed. Thus, before the simplification of γ_o is complete, each curve γ_i must intersect only faces with winding number 0, 1, or -1.

For each index i, let w_i denote the maximum absolute winding number of γ_o around any point of γ_i:

$$w_i := \max_\theta \left| wind\Big(\gamma_o, \gamma_i(\theta)\Big) \right|.$$

Let $W := \sum_i w_i$. Initially, $W = k(n+1)$, and when γ_o first becomes simple, we must have $W \le k$. Each homotopy move changes W by at most 1; specifically, at most one term w_i changes at all, and that term either increases or decreases by 1. The $\Omega(nk)$ lower bound now follows immediately. □

Theorem 4.2. *Transforming a k-curve with n vertices in the plane into an arbitrary set of k simple closed curves requires $\Omega(n^{3/2} + nk)$ homotopy moves in the worst case.*

We say that a collection of k disjoint simple closed curves is ***nested*** if some point lies in the interior of every curve, and ***unnested*** if the curves have disjoint interiors.

Lemma 4.3. *Transforming k nested circles in the plane into k unnested circles requires $\Omega(k^2)$ homotopy moves.*

Proof: Let γ and γ' be two nested circles, with γ' in the interior of γ and with γ directed counterclockwise. Suppose we apply an arbitrary homotopy to these two curves. If the curves remain disjoint during the entire homotopy, then γ' always lies inside a face of γ with winding number 1; in short, the two curves remain nested. Thus, any sequence of homotopy moves that takes γ and γ' to two non-nested simple closed curves contains at least one $0 \to 2$ move that makes the curves cross (and symmetrically at least one $2 \to 0$ move that makes them disjoint again).

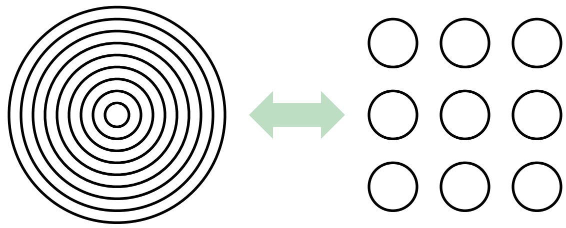


Figure 4.3. Nesting or unnesting k circles requires $\Omega(k^2)$ homotopy moves.

Consider a set of k nested circles. Each of the $\binom{k}{2}$ pairs of circles requires at least one 0→2 move and one 2→0 move to unnest. Because these moves involve distinct pairs of curves, at least $\binom{k}{2}$ 0→2 moves and $\binom{k}{2}$ 2→0 moves, and thus at least $k^2 - k$ moves altogether, are required to unnest every pair. □

Theorem 4.3. *Transforming a k-curve with n vertices in the plane into k nested (or unnested) circles requires $\Omega(n^{3/2} + nk + k^2)$ homotopy moves in the worst case.*

Corollary 4.1. *Transforming one k-curve with at most n vertices into another k-curve with at most n vertices requires $\Omega(n^{3/2} + nk + k^2)$ homotopy moves in the worst case.*

Although our lower bound examples consist of disjoint curves, all of these lower bounds apply without modification to *connected* multicurves, because any k-curve can be connected with at most $k-1$ 0→2 moves. On the other hand, any connected k-curve has at least $2k-2$ vertices, so the $\Omega(k^2)$ terms in Theorem 4.3 and Corollary 4.1 are redundant.

4.2 Quadratic Bound for Curves on Surfaces

In this section we consider the natural generalization of the defect invariant to closed curves on orientable surfaces of higher genus. Because these surfaces have non-trivial topology, not every closed curve is homotopic to a single point or even to a simple curve.

Although defect was originally defined as an invariant of *plane* curves, Polyak's formula

$$\mathit{defect}(\gamma) = -2 \sum_{x \between y} \operatorname{sgn}(x) \operatorname{sgn}(y)$$

extends naturally to closed curves on any orientable surface; homotopy moves change the invariant exactly as described in Figure 4.1. Thus, Lemma 4.1 immediately generalizes to any orientable surface as follows.

Lemma 4.4. *Let γ and γ' be arbitrary closed curves that are homotopic on an arbitrary orientable surface. Transforming γ into γ' requires at least $|\mathit{defect}(\gamma) - \mathit{defect}(\gamma')|/2$ homotopy moves.*

In contrast to Theorem 3.1, the following construction gives toroidal curves with quadratic defect, implying a quadratic lower bound for tightening non-contractible curves on orientable surfaces with positive genus.

Lemma 4.5. *For any positive integer n, there is a closed curve on the torus with n vertices and defect $\Omega(n^2)$ that is homotopic to a simple closed curve but not contractible.*

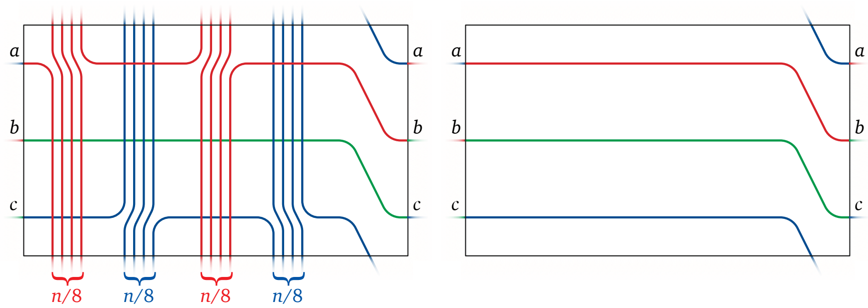


Figure 4.4. A curve γ on the torus with defect $\Omega(n^2)$ and a simple curve homotopic to γ.

Proof: Without loss of generality, suppose n is a multiple of 8. The curve γ is illustrated on the left in Figure 4.4. The torus is represented by a rectangle with opposite edges identified. We label three points a, b, and c on the vertical edge of the rectangle and decompose the curve into a simple red path from a to b, a simple green path from b to c, and a simple blue path from c to a. The red and blue paths each wind vertically around the torus, first $n/8$ times in one direction, and then $n/8$ times in the opposite direction.

As in previous proofs, we compute the defect of γ by describing a sequence of homotopy moves that tightens the curve, while carefully tracking the changes in the defect that these moves incur. We can unwind one turn of the red path by performing one 2→0 move, followed by $n/8$ 3→3 moves, followed by one 2→0 move, as illustrated in Figure 4.5. Repeating this sequence of homotopy moves $n/8$ times removes all intersections between the red and green paths, after which a sequence of $n/4$ 2→0 moves straightens the blue path, yielding the simple curve shown on the right in Figure 4.4. Altogether, we perform $n^2/64 + n/2$ homotopy moves, where each 3→3 move increases the defect of the curve by 2 and each 2→0 move decreases the defect of the curve by 2. We conclude that $\mathit{defect}(\gamma) = -n^2/32 + n$. □

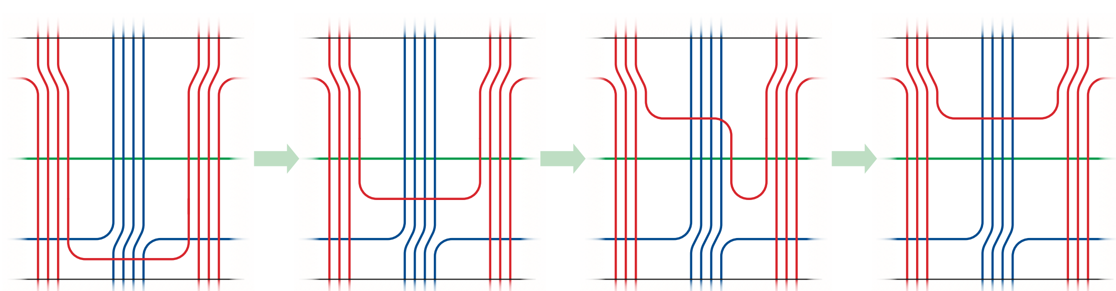


Figure 4.5. Unwinding one turn of the red path.

Theorem 4.4. *Tightening a closed curve with n crossings on a torus requires $\Omega(n^2)$ homotopy moves in the worst case, even if the curve is homotopic to a simple curve.*

Later in Section 4.3.2, we will describe a sequence of *contractible* closed curves on the annulus that requires $\Omega(n^2)$ homotopy moves to tighten through a different kind of curve invariant. Such curves must have defect $O(n^{3/2})$ by Theorem 3.1.

4.3 Quadratic Bound for Contractible Curves on Surfaces

We now prove a quadratic lower bound on the worst-case number of homotopy moves required to tighten closed curves in the annulus; we extend this lower bound to more complex surfaces in Section 4.3.3. Rather than considering the standard annulus $S^1 \times [0,1]$, it will be more convenient to work in the punctured plane $\mathbb{R}^2 \setminus \{o\}$, which is homeomorphic to the open annulus $S^1 \times (0,1)$; here o is an arbitrary point, which we call the ***origin***.

For any homotopy in the punctured plane, homotopy moves across the face containing o are forbidden. This makes the quadratic lower bound possible; without this restriction, any planar curve can be simplified using at most $O(n^{3/2})$ moves, as we will see in Section 5.1.

4.3.1 Traces and Types

To simplify the presentation, we identify the vertices before and after a 3→3 move as indicated in Figure 1.1. Each 3→3 move involves three subpaths of γ, which intersect in three vertices; intuitively, each of these vertices moves continuously across the opposite subpath. Thus, in any homotopy from one curve γ to another curve γ', each vertex of the evolving curve either starts as a vertex of γ or is created by a 0→1 or 0→2 move, moves continuously through a finite sequence of 3→3 moves, and either ends as a vertex of γ' or is destroyed by a 1→0 or 2→0 move.

Let H be a homotopy that transforms γ into γ', represented as a finite sequence of homotopy moves. We define a graph $\boldsymbol{\mathit{Trace}(H)}$, called the ***trace*** of H, whose nodes are the vertices of γ, the vertices of γ', and the 1↔0 and 2↔0 moves in H; each edge of $\mathit{Trace}(H)$ corresponds to the lifetime of a single vertex of the evolving curve. Every node of $\mathit{Trace}(H)$ has degree 1 or 2; thus, $\mathit{Trace}(H)$ is the disjoint union of paths and cycles.

Recall that curves γ_x^+ and γ_x^- are obtained from smoothing the curve γ at vertex x in a way that breaks γ into two closed curves, each respecting the orientation of the original. (See Figure 2.3.) We define the ***type*** of any vertex x of any annular curve γ as the winding number of the simpler curve γ_x^+ around the origin o (*not* around the vertex x); that is, we define $\boldsymbol{\mathit{type}(\gamma, x)} := \mathit{wind}(\gamma_x^+, o)$. Vertex x is ***irrelevant*** if either $\mathit{type}(\gamma, x) = 0$ or $\mathit{type}(\gamma, x) = \mathit{wind}(\gamma, o)$ and ***relevant*** otherwise. Two vertices x and y have ***complementary*** types if $\mathit{type}(\gamma, x) + \mathit{type}(\gamma, y) = \mathit{wind}(\gamma, o)$, or equivalently, if $\mathit{wind}(\gamma_x^+, o) = \mathit{wind}(\gamma_y^-, o)$. If two vertices have complementary types, then either both are relevant or both are irrelevant.

Lemma 4.6. *The following hold for any annular curve:*

(1) Each 1↔0 move creates or destroys an irrelevant vertex.

(2) Each 2↔0 move creates or destroys two vertices with complementary types and identical winding numbers.

(3) Each 3↔3 move changes the winding numbers of three vertices, each by exactly 1.

(4) Except as stated in (1), (2), and (3), homotopy moves do not change the type or winding number of any vertex.

Proof: Claim (1) is immediate. Up to symmetry, there are only two cases to consider to prove claim (2): The two sides of the empty bigon are oriented in the same direction or in opposite directions. In both cases, γ_x^+ and γ_y^- are homotopic and $\mathit{wind}(\gamma, x) = \mathit{wind}(\gamma, y)$, where x and y are the vertices of the bigon. See Figure 4.6. Claim (3) follows immediately from the observation that each vertex involved in a 3→3 move passes over the curve exactly once. Finally, claim (4) follows from the fact that winding number is a homotopy invariant; specifically, if there is a homotopy between two planar curves γ and γ' whose image does not include a point p, then $\mathit{wind}(\gamma, p) = \mathit{wind}(\gamma', p)$ [142]. □

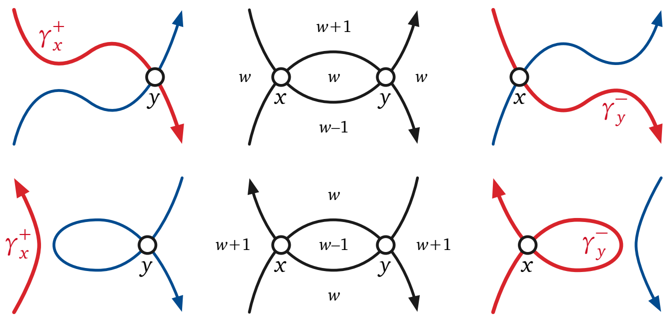


Figure 4.6. The vertices of empty bigons have complementary types and identical winding numbers.

Lemma 4.6 implies that no homotopy move transforms a relevant vertex into an irrelevant vertex or vice versa, and that relevant vertices are neither created by 0→1 moves nor destroyed by 1→0 moves. Let $\boldsymbol{Trace_*(H)}$ denote the subgraph of edges in the trace graph $Trace(H)$ that correspond to relevant vertices of the evolving curve. Again, $Trace_*(H)$ is the disjoint union of paths and cycles. Each path in $Trace_*(H)$ connects either two vertices of γ with complementary types, two vertices of γ' with complementary types, or a vertex of γ and a vertex of γ' with identical types. Intuitively, each path in $Trace_*(H)$ is the record of a single relevant vertex alternately moving forward and backward in time, reversing directions and types at every 0↔2 move. We say that the nodes at the end of each path in $Trace_*(H)$ are ***paired*** by the homotopy H. We emphasize that different homotopies may lead to different pairings.

Between 2↔0 moves, a relevant vertex can participate in any finite number of 3→3 moves. By Lemma 4.6(3), each 3→3 move changes the winding numbers of each of the three moving vertices by 1, and Lemma 4.6(4) implies that the winding number of a vertex changes only when it participates in a 3→3 move. Thus, the homotopy H must contain at least

$$\frac{1}{3}\sum_{x\sim y}\Big|wind(x) - wind(y)\Big|$$

3→3 moves, where the sum is over all pairs of paired vertices of $Trace_*(H)$, and the winding number of each vertex is defined with respect to the curve (γ or γ') that contains it.

4.3.2 A Bad Contractible Annular Curve

Theorem 4.5. *For any positive integer n, there is a contractible annular curve with n vertices that requires $\Omega(n^2)$ homotopy moves to tighten.*

Proof: For any pair of relatively prime integers p and q, the *flat torus knot* $T(p,q)$ described in Section 3.1.1 has exactly $(|p|-1)\cdot|q|$ vertices and winding number p around the origin. For any odd integer p, let $\mathbf{\Pi_p}$ denote the closed curve obtained by placing a scaled copy of $T(-p,1)$ inside the innermost face of $T(p,2)$ and attaching the two curves as shown in Figure 4.7. For purposes of illustration, we perform homotopy to move all crossings into a narrow horizontal rectangle to the right of the origin, which is also where we join the two curves. The resulting curve Π_p has winding number zero around the origin and thus is contractible, and it has $3(p-1)$ vertices. Within the rectangle (treated as a *tangle*), the curve consists of $2p$ simple strands; the endpoints of the strands are connected by disjoint parallel paths outside the rectangle. In the left half of the rectangle, strands are directed downward; in the right half, strands are directed upward. All but two strands connect the top and bottom of the rectangle; the only exceptions are the strands that connect the two flat torus knots.

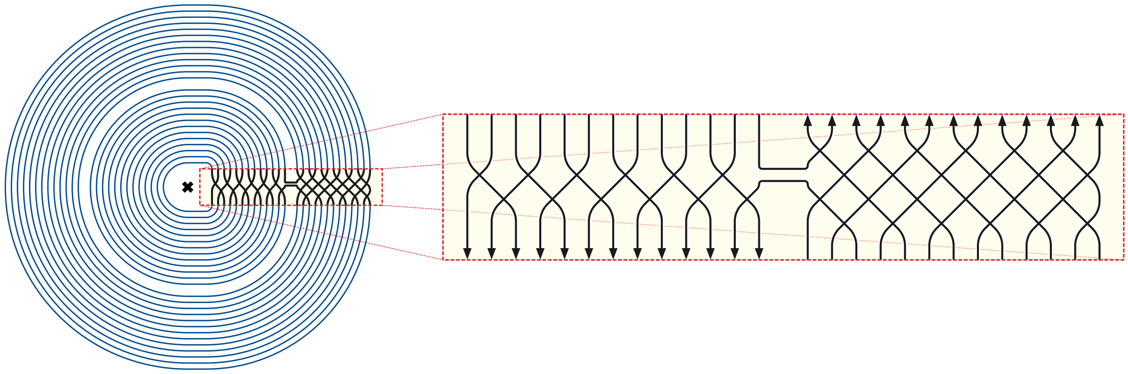


Figure 4.7. Our bad example curve Π_{13} in the punctured plane.

We catalog the vertices of Π_p as follows; see Figure 4.8. In the left half of the rectangle, Π_p has one vertex a_i with type i and winding number i, for each integer i from 1 to p. In the right half, Π_p has four vertices for each index i between 1 and $(p-1)/2$:

- two vertices x_i and x_i' with type $-i$ and winding number $2i$;
- one vertex y_i with type i and winding number $p-2i$; and
- one vertex z_i with type $i-p$ and winding number $p-2i$.

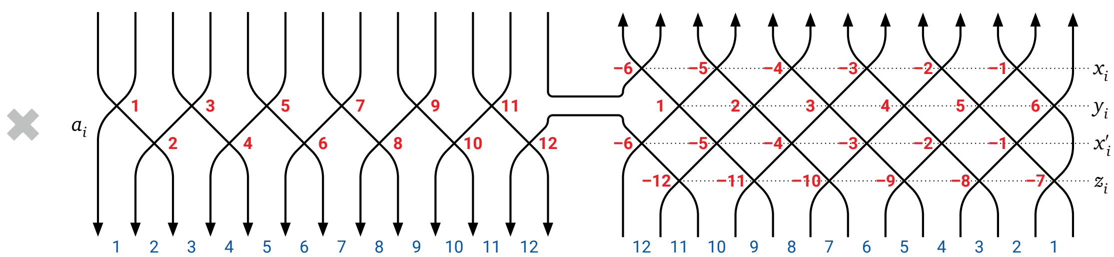


Figure 4.8. Vertices of Π_{13} annotated by type (bold red numbers next to each vertex) and winding number (thin blue numbers directly below each vertex).

Every homotopy from Π_p to a simple closed curve defines an essentially unique pairing of the vertices of Π_p; without loss of generality, a_i is paired with x_i', a_{p-i} is paired with z_i, and x_i is paired with y_i, for each integer i between 1 and $(p-1)/2$. Thus, the number of 3→3 moves in any homotopy that contracts Π_p is at least

$$\frac{1}{3}\sum_{i=1}^{(p-1)/2}\Big(|i-2i|+|(p-i)-(p-2i)|+|2i-(p-2i)|\Big)$$
$$=\frac{1}{3}\sum_{i=1}^{(p-1)/2}\Big(2i+|4i-p|\Big) = \frac{1}{3}\left(\sum_{i=1}^{(p-1)/2}2i+\sum_{j=1}^{(p-1)/2}(2j+1)\right) = \frac{p(p-1)}{6}.$$

This completes the proof. □

4.3.3 More Complicated Surfaces

We extend Theorem 4.5 to surfaces with more complex topology as follows. A closed curve in any surface Σ is ***two-sided*** if it has a neighborhood homeomorphic to the annulus. Let Σ be a compact surface, possibly with boundary or non-orientable, that contains a simple two-sided non-contractible cycle α; the only compact surfaces that do *not* contain such a cycle are the sphere, the disk, and the projective plane. To create a bad example curve for Σ, we simply embed our previous annular curve Π_p in an annular neighborhood A of α. The resulting curves are still contractible in Σ and, as we will shortly prove, still require $\Omega(n^2)$ homotopy moves to simplify.

However, winding numbers are not well-defined in surfaces of higher genus, so we need a more careful argument to prove the quadratic lower bound. Instead of reasoning directly about homotopy moves on Σ, we lift everything to a certain covering space of Σ previously considered by several authors [64, 87, 135, 161, 217].

Theorem 4.6. *Let Σ be a compact connected surface, possibly with boundary or non-orientable (but not the sphere, the disk, or the projective plane). For any positive integer n, there is a contractible curve with n vertices in Σ that requires $\Omega(n^2)$ homotopy moves to simplify.*

Proof: Let α be a simple two-sided non-contractible closed curve in Σ, that is, a non-contractible curve that lies in an open neighborhood A homeomorphic to the open annulus $S^1 \times (0,1)$. Every compact connected surface (other than the sphere, the disk, or the projective plane) contains such a curve.

The ***cyclic covering space*** $\hat{\Sigma}_\alpha$ of Σ with respect to α is the quotient of the universal covering space of Σ by the infinite-cyclic subgroup of the fundamental group $\pi_1(\Sigma)$ generated by α. Let $\pi\colon \hat{\Sigma}_\alpha \to \Sigma$ be the corresponding covering map. Standard covering space results imply that α has a unique lift $\hat{\alpha}$ to $\hat{\Sigma}_\alpha$ that is a simple closed curve. Also, $\hat{\alpha}$ has an open annular neighborhood $\hat{A}$ with non-contractible boundary components in $\hat{\Sigma}$. Moreover, we may assume that the restriction of the covering map π to $\hat{A}$ is a homeomorphism to A.

Let $\hat{\gamma}$ be an arbitrary contractible curve in $\hat{A}$, and let γ be the projection of $\hat{\gamma}$ to A. The two curves γ and $\hat{\gamma}$ have the same number of vertices and edges. By homotopy lifting property, any homotopy $H\colon S^1 \times [0,1] \to \Sigma$ from γ to a point lifts to a homotopy $\hat{H}\colon S^1 \times [0,1] \to \hat{\Sigma}_\alpha$ from $\hat{\gamma}$ to a point. Each homotopy move in $\hat{H}$ projects to a homotopy move in H, but H may include additional homotopy moves, where the strands involved are projected from different parts of the covering space. It follows that simplifying γ in Σ requires at least as many homotopy moves as simplifying $\hat{\gamma}$ in $\hat{\Sigma}_\alpha$.

Standard covering space results imply that the interior of $\hat{\Sigma}_\alpha$ is homeomorphic to an open annulus, and therefore to the punctured plane $\mathbb{R}^2 \setminus \{o\}$. (See, for example, Schrijver [217, Proposition 2].) The lower bound now follows directly from Theorem 4.5, by setting $\hat{\gamma} := \Pi_p$ for some $p = \Theta(n)$, as defined in Section 4.3.2. If Σ has non-empty boundary, then $\hat{\Sigma}_\alpha$ also has non-empty boundary, but without loss of generality, any homotopy that contracts $\hat{\gamma}$ avoids the boundary of $\hat{\Sigma}_\alpha$. □

Theorem 4.6 strengthens the $\Omega(n^2)$ lower bound in Section 4.2 for tightening non-contractible curves in orientable surfaces. Results of Hass and Scott [135, Theorem 2.7] imply that our lower bound is tight for the Möbius band, the Klein bottle, and any orientable surface except the sphere or the disk; any contractible curve on these surfaces can be simplified using at most $O(n^2)$ homotopy moves.

The only missing case is the projective plane; see Section 8.2 for a discussion.

Chapter 5

Tightening Planar Curves

I used to love to untangle chains when I was a child. I had thin, busy fingers, and I never gave up. Perhaps there was a psychiatric component to my concentration but like much of my psychic damage, this worked to everyone's advantage.

— Anne Lamott, *Plan B: Further Thoughts on Faith*

We develop a new algorithm to simplify any closed curve in the plane in $O(n^{3/2})$ homotopy moves in Section 5.1. First we describe an algorithm that uses $O(D\Sigma)$ moves, where $D\Sigma$ is the sum of the face depths of the input curve. At a high level, our algorithm can be viewed as a variant of Steinitz's algorithm that empties and removes *monogons* instead of bigons. We then extend our algorithm to *tangles*: collections of boundary-to-boundary paths in a closed disk. Our algorithm tightens a tangle in $O(D\Sigma + ns)$ moves, where $D\Sigma$ is the sum of the depths of the tangle's faces, s is the number of strands, and n is the number of intersection points. Using the result from Section 3.2.3, we can find a simple closed curve whose interior tangle has m vertices, at most $\sqrt{m}$ strands, and maximum face depth $O(\sqrt{n})$. Tightening this tangle and then recursively simplifying the resulting curve requires a total of $O(n^{3/2})$ moves. We show that this simplifying sequence of homotopy moves can be computed in $O(1)$ amortized time per move, assuming the curve is presented in an appropriate graph data structure. We conclude this chapter by proving that any arrangement of k closed curves can be simplified in $O(n^{3/2} + nk)$ homotopy moves, or in $O(n^{3/2} + nk + k^2)$ homotopy moves if the target configuration is specified in advance, precisely matching our lower bounds for all values of n and k.

5.1 Planar Curves

5.1.1 Contracting Simple Loops

Lemma 5.1. *Every closed curve γ in the plane can be simplified using at most $3\,D\Sigma(\gamma) - 3$ homotopy moves.*

Proof: We prove the statement by induction on the number of vertices in γ. The lemma is trivial if γ is already simple, so assume otherwise. Let $x := \gamma(\theta) = \gamma(\theta')$ be the first vertex to be visited twice by γ after the (arbitrarily chosen) basepoint $\gamma(0)$. Let α denote the subpath of γ from $\gamma(\theta)$ to $\gamma(\theta')$; our choice of x implies that α is a simple monogon. Let m and s denote the number of vertices and strands in the interior tangle of α, respectively.

Finally, let γ' denote the closed curve obtained from γ by removing α. The first stage of our algorithm transforms γ into γ' by contracting the monogon α via homotopy moves.

We remove the vertices and edges from the interior of α one at a time as follows; see Figure 5.2. If we can perform a 2→0 move to remove one edge of γ from the interior of α and decrease s, we do so. Otherwise, either α is empty, or some vertex of γ lies inside α. In the latter case, at least one vertex x inside α has a neighbor that lies on α. We move x outside α with a 0→2 move (which increases s by 1) followed by a 3→3 move (which decreases m by 1). Once α is an empty monogon, we remove it with a single 1→0 move. Altogether, our algorithm

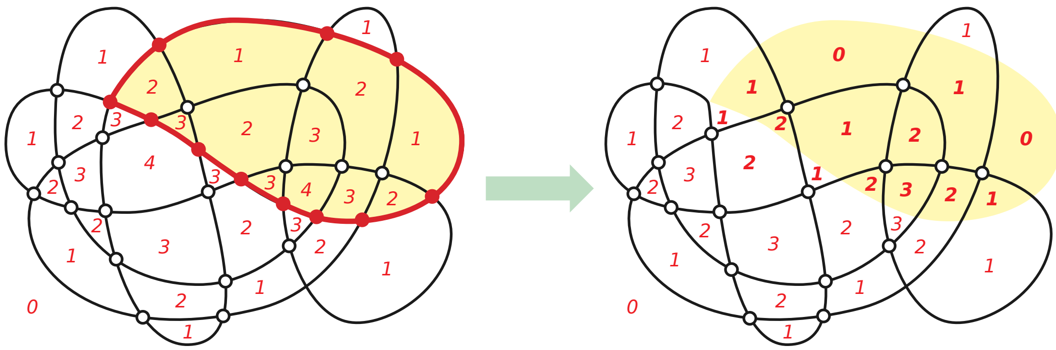


Figure 5.1. Transforming γ into γ' by contracting a simple monogon. Numbers are face depths.

transforms γ into γ' using at most $3m + s + 1$ homotopy moves. Let M denote the actual number of homotopy moves used.

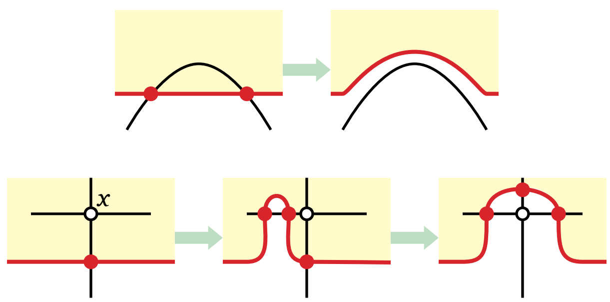


Figure 5.2. Moving a monogon over an interior empty bigon or an interior vertex.

Euler's formula implies that α contains exactly $m + s + 1$ faces of γ. The Jordan curve theorem implies that $depth(p, \gamma') \leq depth(p, \gamma) - 1$ for any point p inside α, and trivially $depth(p, \gamma') \leq depth(p, \gamma)$ for any point p outside α. It follows that $D\Sigma(\gamma') \leq D\Sigma(\gamma) - (m + s + 1) \leq D\Sigma(\gamma) - M/3$, and therefore $M \leq 3D\Sigma(\gamma) - 3D\Sigma(\gamma')$. The induction hypothesis implies that we can recursively simplify γ' using at most $3D\Sigma(\gamma') - 3$ moves. The lemma now follows immediately. □

Our upper bound is a factor of 3 larger than Feo and Provan's [100]; however our algorithm has the advantage that it extends to *tangles*, as described in the next subsection.

5.1.2 Tangles

Recall that a tangle is *tight* if every pair of strands intersects at most once and *loose* otherwise. Every loose tangle contains either an empty monogon or a (not necessarily empty) bigon. Thus, any tangle with n vertices can be transformed into a tight tangle—or less formally, *tightened*—in $O(n^2)$ homotopy moves using Steinitz's algorithm. On the other hand, there are infinite classes of loose tangles for which no homotopy move that decreases the potential, so we cannot directly apply Feo and Provan's algorithm to this setting. (See Section 8.1.1).

We describe a two-phase algorithm to tighten any tangle. First, we remove any self-intersections in the individual strands, by contracting monogons as in the proof of Lemma 5.1. Once each strand is simple, we move the strands so that each pair intersects at most once. See Figure 5.3.

Lemma 5.2. *Every n-vertex tangle $\oplus$ with s simple strands can be tightened using at most $3ns$ homotopy moves.*

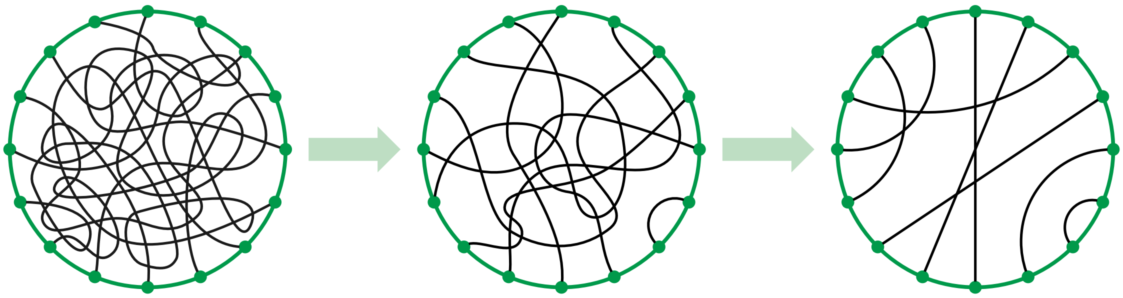


Figure 5.3. Tightening a tangle in two phases: First simplifying the individual strands, then removing excess crossings between pairs of strands.

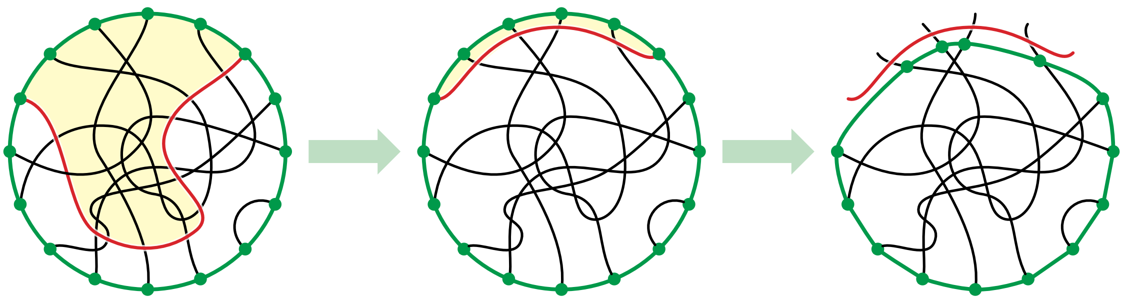


Figure 5.4. Moving one strand out of the way and shrinking the tangle boundary.

Proof: We prove the lemma by induction on s. The base case when $s = 1$ is trivial, so assume $s \geq 2$.

Fix an arbitrary reference point on the boundary circle σ that is not an endpoint of a strand. For each index i, let σ_i be the arc of σ between the endpoints of γ_i that does not contain the reference point. A strand γ_i is *extremal* if the corresponding arc σ_i does not contain any other arc σ_j.

Choose an arbitrary extremal strand γ_i. Let m_i denote the number of tangle vertices in the interior of the disk bounded by γ_i and the boundary arc σ_i; call this disk Σ_i. Let s_i denote the number of intersections between γ_i and other strands. Finally, let γ_i' be a path inside the disk Σ defining tangle $\oplus$, with the same endpoints as γ_i, that intersects each other strand in $\oplus$ at most once, such that the disk bounded by σ_i and γ_i' has no tangle vertices inside its interior. (See Figure 5.4 for an example; the red strand in the left tangle is γ_i, the red strand in the middle tangle is γ_i', and the shaded disk is Σ_i.)

We can deform γ_i into γ_i' using essentially the algorithm from Lemma 5.1; the disk Σ_i is contracted along with γ_i in the process. If Σ_i contains an empty bigon with one side in γ_i, remove it with a 2→0 move (which decreases s_i by 1). If Σ_i has an interior vertex with a neighbor on γ_i, remove it using at most two homotopy moves (which increases s_i by 1 and decreases m_i by 1). Altogether, this deformation requires at most $3m_i + s_i \leq 3n$ homotopy moves.

After deforming γ_i to γ_i', we redefine the tangle by shrinking its boundary curve slightly to exclude γ_i', without creating or removing any vertices in the tangle or endpoints on the boundary; see the right of Figure 5.4. We emphasize that shrinking the boundary does not modify the strands and therefore does not require any homotopy moves. The resulting smaller tangle has exactly $s - 1$ strands, each of which is simple. Thus, the induction hypothesis implies that we can recursively tighten this smaller tangle using at most $3n(s-1)$ homotopy moves. □

Corollary 5.1. *Every n-vertex s-strand tangle $\oplus$ can be tightened using at most $3\,D\Sigma(\oplus) + 3ns$ homotopy moves.*

Proof: As long as $\oplus$ contains at least one non-simple strand, we identify a simple monogon α in that strand and remove it as described in the proof of Lemma 5.1. Suppose there are m vertices and t strands in the interior tangle of α, and let M be the number of homotopy moves required to remove α. The algorithm in the proof of Lemma 5.1 implies that $M \leq 3m+t+1$, and Euler's formula implies that α contains $m+t+1 \geq M/3$ faces. Removing α decreases the depth of each of these faces by at least 1 and therefore decreases the potential of the tangle by at least $M/3$.

Let $\oplus'$ be the remaining tangle after all such monogons are removed. Our potential analysis for a single monogon implies inductively that transforming $\oplus$ into $\oplus'$ requires at most $3D\Sigma(\oplus) - 3D\Sigma(\oplus') \leq 3D\Sigma(\oplus)$ homotopy moves. Because $\oplus'$ still has s strands and at most n vertices, Lemma 5.2 implies that we can tighten $\oplus'$ with at most $3ns$ additional homotopy moves. □

5.1.3 Main Algorithm

Our main algorithm repeatedly finds a useful closed curve using Lemma 3.10 whose interior tangle has depth $O(\sqrt{n})$, and tightens its interior tangle; if there are no useful closed curves, then we fall back to the monogon-contraction algorithm of Lemma 5.1.

Theorem 5.1. *Every closed curve in the plane with n vertices can be simplified in $O(n^{3/2})$ homotopy moves.*

Proof: Let γ be an arbitrary closed curve in the plane with n vertices. If γ has depth $O(\sqrt{n})$, Lemma 5.1 and the trivial upper bound $D\Sigma(\gamma) \leq (n+1) \cdot \mathit{depth}(\gamma)$ imply that we can simplify γ in $O(n^{3/2})$ homotopy moves. For purposes of analysis, we charge $O(\sqrt{n})$ of these moves to each vertex of γ.

Otherwise, let σ be an arbitrary useful closed curve chosen according to Lemma 3.10. Suppose the interior tangle of σ has m vertices, s strands, and depth d. Lemma 3.10 implies that $d = O(\sqrt{n})$, and the definition of useful implies that $s \leq \sqrt{m}$, which is $O(\sqrt{n})$. Thus, by Corollary 5.1, we can tighten the interior tangle of σ in $O(md + ms) = O(m\sqrt{n})$ moves. This simplification removes at least $m - s^2/2 \geq \Omega(m)$ vertices from γ, as the resulting tight tangle has at most $s^2/2$ vertices. Again, for purposes of analysis, we charge $O(\sqrt{n})$ moves to each deleted vertex. We then recursively simplify the resulting closed curve.

In either case, each vertex of γ is charged $O(\sqrt{n})$ moves as it is deleted. Thus, simplification requires at most $O(n^{3/2})$ homotopy moves in total. □

The bound in Theorem 5.1 is asymptotically optimal as it matches the lower bound in Theorem 4.1 up to constant factors. As an immediate corollary of Theorem 5.1 and Theorem 7.2, we obtain an alternative proof to the subquadratic defect upper bound in Section 3.2.

5.1.4 Efficient Implementation

Here we describe how to implement our curve-simplification algorithm to run in $O(n^{3/2})$ time; in fact, our implementation spends only constant amortized time per homotopy move. We assume that the input curve is given in a data structure that allows fast exploration and modification of plane graphs, such as a quad-edge data structure [130] or a doubly-connected edge list [22]. If the curve is presented as a polygon with m edges, an appropriate graph representation can be constructed in $O(m \log m + n)$ time using classical geometric algorithms [55, 60, 180]; more recent algorithms can be used for piecewise-algebraic curves [83].

Theorem 5.2. *Given a simple closed curve γ in the plane with n vertices, we can compute a sequence of $O(n^{3/2})$ homotopy moves that simplifies γ in amortized constant time per move.*

Proof: We begin by labeling each face of γ with its depth, using a breadth-first search of the dual graph in $O(n)$ time. Then we construct the depth contours of γ—the boundaries of the regions $\tilde{R}_j$ from the proof of Lemma 3.10—and organize them into a *contour tree* in $O(n)$ time by brute force. Another $O(n)$-time breadth-first traversal computes the number of strands and the number of interior vertices of every contour's interior tangle; in particular, we identify which depth contours are useful. To complete the preprocessing phase, we place all the leafmost useful contours into a queue. We can charge the overall $O(n)$ preprocessing time to the $\Omega(n)$ homotopy moves needed to simplify the curve.

As long as the queue of leafmost useful contours is non-empty, we extract one contour σ from this queue and simplify its interior tangle $\oplus$ as follows. Suppose $\oplus$ has m interior vertices.

Following the proof of Theorem 5.1, we first simplify every monogon in each strand of $\oplus$. We identify monogons by traversing the strand from one endpoint to the other, marking the vertices as we go; the first time we visit a vertex that has already been marked, we have found a monogon α. We can perform each of the homotopy moves required to shrink α in $O(1)$ time, because each such move modifies only a constant-radius boundary of a vertex on α. After the monogon is shrunk, we continue walking along the strand starting at the most recently marked vertex.

The second phase of the tangle-simplification algorithm proceeds similarly. We walk around the boundary of $\oplus$, marking vertices as we go. As soon as we see the second endpoint of any strand γ_i, we pause the walk to straighten γ_i. As before, we can execute each homotopy move used to move γ_i to γ_i' in $O(1)$ time. We then move the boundary of the tangle over the vertices of γ_i', and remove the endpoints of γ_i' from the boundary curve, in $O(1)$ time per vertex.

The only portions of the running time that we have not already charged to homotopy moves are the time spent marking the vertices on each strand and the time to update the tangle boundary after moving a strand aside. Altogether, the uncharged time is $O(m)$, which is less than the number of moves used to tighten $\oplus$, because the contour σ is useful. Thus, tightening the interior tangle of a useful contour requires $O(1)$ amortized time per homotopy move.

Once the tangle is tight, we must update the queue of useful contours. The original contour σ is still a depth contour in the modified curve, and tightening $\oplus$ only changes the depths of faces that intersect $\oplus$. Thus, we could update the contour tree in $O(m)$ time, which we could charge to the moves used to tighten $\oplus$; but in fact, this update is unnecessary, because no contour in the interior of σ is useful. We then walk up the contour tree from σ, updating the number of interior vertices until we find a useful ancestor contour. The total time spent traversing the contour tree for new useful contours is $O(n)$; we can charge this time to the $\Omega(n)$ moves needed to simplify the curve. □

5.2 Planar Multicurves

Finally, we describe how to extend our $O(n^{3/2})$ upper bound to multicurves in the plane. Just as in Section 4.1.1, we distinguish between two variants, depending on whether the target of the simplification is an *arbitrary* set of disjoint cycles or a *particular* set of disjoint cycles. In both cases, our upper bounds match the lower bounds proved in Section 4.1.1.

First we extend our monogon-contraction algorithm from Lemma 5.1 to the multicurve setting. Recall that a *component* of a multicurve γ is any multicurve whose image is a component of the image of γ, and the individual closed curves that comprise γ are its *constituent curves*. The main difficulty is that one component of the multicurve might lie inside a face of another component, making progress on the larger component impossible. To handle this potential obstacle, we simplify the *innermost* components of the multicurve first, and we move isolated simple closed curves toward the outer face as quickly as possible. Figure 5.5 sketches the basic steps of our algorithm when the input multicurve is connected.

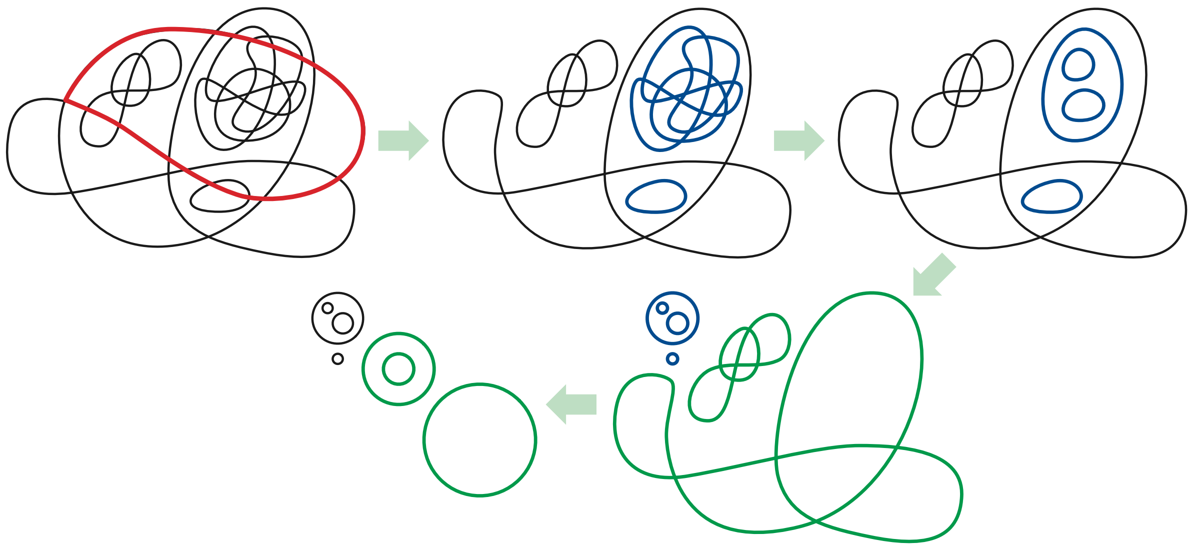


Figure 5.5. Simplifying a connected multicurve: shrink an arbitrary simple monogon or cycle, recursively simplify any inner components, translate inner circle clusters to the outer face, and recursively simplify the remaining non-simple components.

Lemma 5.3. *Every n-vertex k-curve γ in the plane can be transformed into k disjoint simple closed curves using at most $3D\Sigma(\gamma) + 4nk$ homotopy moves.*

Proof: Let γ be an arbitrary k-curve with n vertices. If γ is connected, we either contract and delete a monogon, exactly as in Lemma 5.1, or we contract a simple constituent curve to an isolated circle, using essentially the same algorithm. In either case, the number of moves performed is at most $3D\Sigma(\gamma) - 3D\Sigma(\gamma')$, where γ' is the multicurve after the contraction. The lemma now follows immediately by induction.

We call a component of γ an ***outer component*** if it is incident to the unbounded outer face of γ, and an ***inner component*** otherwise. If γ has more than one outer component, we partition γ into subpaths, each consisting of one outer component γ_o and all inner components located inside faces of γ_o, and we recursively simplify each subpath independently; the lemma follows by induction. If any outer component is simple, we ignore that component and simplify the rest of γ recursively; again, the lemma follows by induction.

Thus, we can assume without loss of generality that our multicurve γ is disconnected but has only one outer component γ_o, which is non-simple. For each face f of γ_o, let γ_f denote the union of all components inside f. Let n_f and k_f respectively denote the number of vertices and constituent curves of γ_f. Similarly, let n_o and k_o respectively denote the number of vertices and constituent curves of the outer component γ_o.

We first recursively simplify each subpath γ_f; let κ_f denote the resulting *cluster* of k_f simple closed curves. By the induction hypothesis, this simplification requires at most $3D\Sigma(\gamma_f) + 4n_f k_f$ homotopy moves. We *translate* each cluster κ_f to the outer face of γ_o by shrinking κ_f to a small ε-ball and then moving the entire cluster along a shortest path in the dual graph of γ_o. This translation requires at most $4n_o k_f$ homotopy moves; each circle in κ_f uses one 2→0 move and one 0→2 move to cross any edge of γ_o, and in the worst case, the cluster might cross all $2n_0$ edges of γ_o. After all circle clusters are in the outer face, we recursively simplify γ_o using at most $3D\Sigma(\gamma_o) + 4n_o k_o$ homotopy moves.

The total number of homotopy moves used in this case is

$$\sum_f 3D\Sigma(\gamma_f) + 3D\Sigma(\gamma_o) \ + \ \sum_f 4n_f k_f + \sum_f 4n_o k_f + 4n_o k_o.$$

Each face of γ_o has the same depth as the corresponding face of γ, and for each face f of γ_o, each face of the subpath γ_f has lesser depth than the corresponding face of γ. It follows that

$$\sum_f D\Sigma(\gamma_f) + D\Sigma(\gamma_o) \le D\Sigma(\gamma).$$

Similarly, $\sum_f n_f + n_o = n$ and $\sum_f k_f + k_o = k$. The lemma now follows immediately. □

To reduce the leading term to $O(n^{3/2})$, we extend the definition of a tangle to the intersection of a multicurve γ with a closed disk whose boundary intersects the multicurve transversely away from its vertices, or not at all. Such a tangle can be decomposed into boundary-to-boundary paths, called ***open*** strands, and closed curves that do not touch the tangle boundary, called ***closed*** strands. Each closed strand is a constituent curve of γ. A tangle is ***tight*** if every strand is simple, every pair of open strands intersects at most once, and otherwise all strands are disjoint.

Theorem 5.3. *Every k-curve in the plane with n vertices can be transformed into a set of k disjoint simple closed curves using $O(n^{3/2} + nk)$ homotopy moves.*

Proof: Let γ be an arbitrary k-curve with n vertices. Following the proof of Lemma 5.3, we can assume without loss of generality that γ has a single outer component γ_o, which is non-simple.

When γ is disconnected, we follow the strategy in the previous proof. Let γ_f denote the union of all components inside any face f of γ_o. For each face f, we recursively simplify γ_f and translate the resulting cluster of disjoint circles to the outer face; when all faces are empty, we recursively simplify γ_o. The theorem now follows by induction.

When γ is non-simple and connected, we follow the useful closed curve strategy from Theorem 5.1. We define a closed curve σ to be useful for γ if the interior tangle of σ has its number of vertices at least the square of the number of *open* strands; then the proof of Lemma 3.10 applies to connected multicurves with no modifications. So let $\oplus$ be a tangle with m vertices, $s \le \sqrt{m}$ open strands, ℓ closed strands, and depth $d = O(\sqrt{n})$. We straighten $\oplus$ in two phases, almost exactly as in Section 5.1.2, contracting monogons and simple closed strands in the first phase, and straightening open strands in the second phase.

In the first phase, contracting one monogon or simple closed strand uses at most $3D\Sigma(\oplus) - 3D\Sigma(\oplus')$ homotopy moves, where $\oplus'$ is the tangle after contraction. After each contraction, if $\oplus'$ is disconnected—in particular, if we just contracted a simple closed strand—we simplify and extract any isolated components as follows. Let $\oplus'_o$ denote the component of $\oplus'$ that includes the boundary cycle, and for each face f of $\oplus'_o$, let γ_f denote the union of all components of $\oplus'$ inside f. We simplify each multicurve γ_f using the algorithm from Lemma 5.3—*not recursively!*—and then translate the resulting cluster of disjoint circles *to the outer face of* γ. See Figure 5.6. Altogether, simplifying and translating these subpaths requires at most $3D\Sigma(\oplus') - 3D\Sigma(\oplus'') + 4n\sum_f k_f$ homotopy moves, where $\oplus''$ is the resulting tangle.

The total number of moves performed in the first phase is at most $3D\Sigma(\oplus) + 4m\ell = O(m\sqrt{n} + n\ell)$. The first phase ends when the tangle consists entirely of simple open strands. Thus, the second phase straightens

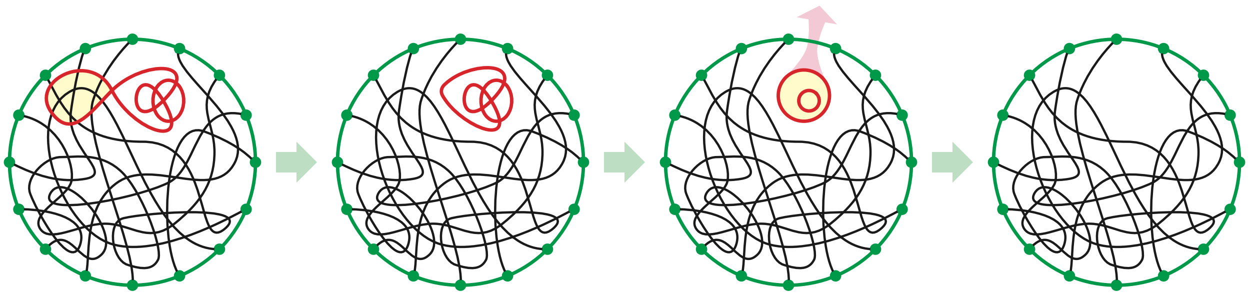


Figure 5.6. Whenever shrinking a monogon or simple closed strand disconnects the tangle, simplify each isolated component and translate the resulting cluster of circles to the outer face of the entire multicurve.

the remaining open strands exactly as in the proof of Lemma 5.2; the total number of moves in this phase is $O(ms) = O(m\sqrt{n})$. We charge $O(\sqrt{n})$ time to each deleted vertex and $O(n)$ time to each constituent curve that was simplified and translated outward. We then recursively simplify the remaining multicurve, ignoring any outer circle clusters.

Altogether, each vertex of γ is charged $O(\sqrt{n})$ time as it is deleted, and each constituent curve of γ is charged $O(n)$ time as it is translated outward. □

With $O(k^2)$ additional homotopy moves, we can transform the resulting set of k disjoint circles into k nested or unnested circles.

Theorem 5.4. *Any k-curve with n vertices in the plane can be transformed into k nested (or unnested) simple closed curves using $O(n^{3/2} + nk + k^2)$ homotopy moves.*

Corollary 5.2. *Any k-curve with at most n vertices in the plane can be transformed into any other k-curve with at most n vertices using $O(n^{3/2} + nk + k^2)$ homotopy moves.*

Theorems 4.2 and 4.3 and Corollary 4.1 imply that these upper bounds are tight in the worst case for all possible values of n and k. As in the lower bounds, the $O(k^2)$ terms are redundant for connected multicurves.

More careful analysis implies that any k-curve with n vertices and depth d can be simplified in $O(n \min\{d, n^{1/2}\} + k \min\{d, n\})$ homotopy moves, or transformed into k unnested circles using $O(n \min\{d, n^{1/2}\} + k \min\{d, n\} + k \min\{d, k\})$ homotopy moves. Moreover, these upper bounds are tight, up to constant factors, for all possible values of n, k, and d. We leave the details of this extension as an exercise for the reader.

Chapter 6

Tightening Curves on Surfaces

Let bigons be bygones.

— Anna, *The Geometric Supposer: What is it a case of?*

In this chapter we prove that any n-vertex closed curve on an arbitrary orientable surface of negative Euler characteristic can be tightened in polynomially many homotopy moves. Throughout the chapter we assume the reader is familiar with fundamentals of combinatorial topology on surfaces. We refer the readers to Massey [176], Stillwell [232], and Giblin [112] for comprehensive introductions to the topic.

Our main technical contribution is to extend Steinitz's bigon removal algorithm to *singular bigons*—bigons that wrap around the surface and overlap themselves but nevertheless have well-defined disjoint bounding paths—whose existence is guaranteed by a theorem of Hass and Scott [135, Theorem 2.7]. (A formal definition of the singular bigon can be found in Section 6.1.) To work with singular bigons, it is conceptually easier to look at a lift of the bigon in the universal cover. Unlike the case when the bigon is embedded, moving the two bounding paths of the bigon now also moves all their *translates* in the universal cover, which potentially changes the structure inside the lifted bigon. We overcome this difficulty by carefully subdividing the homotopy into phases, each performed inside a subset of the universal cover that maps injectively onto the original surface.

We provide two algorithms to remove singular bigons: one for orientable surface with boundary and one for those without. We consider surfaces with boundary first, not only because the bound obtained is stronger, but also because the proof is simpler and provides important intuition for the more difficult proof of the boundary-less case. The benefit of working on surface with boundary is that the fundamental group of such surface is *free*; intuitively one can always find a way to decrease the complexity of the bigon wrapping around the surface.

Our proof for surface without boundary uses a discrete analog of the classical *isoperimetric inequality* in the hyperbolic plane to bound the number of vertices inside the lifted bigon (area) in terms of the number of vertices on its boundary (perimeter). To make the presentation self-contained, we provide an elementary proof of this inequality using the combinatorial Gauss-Bonnet theorem [18, 93, 173, 188]. The second algorithm is surprisingly complex and subtle, with multiple components and tools drawn from discrete and computational topology.

6.1 Singular Bigons and Singular Monogons

Here we generalize the Steinitz's bigon removal algorithm to any closed curves on arbitrary orientable surfaces. Following Hass and Scott [135], a ***singular bigon*** in γ consists of two subpaths of γ that are disjoint in the domain, and the two subpaths are homotopic to each other in Σ. Similarly, a ***singular monogon*** is a subpath of γ whose two endpoints are identical in Σ, and that forms a null-homotopic closed curve in Σ.

Our algorithms rely on the following simple property of singular monogons and bigons, which follows immediately from their definition.

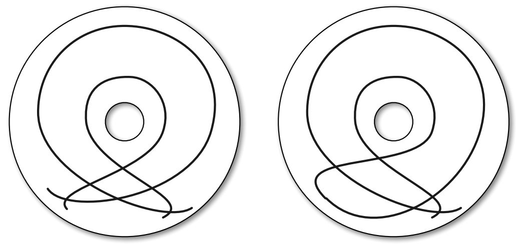


Figure 6.1. A basic singular bigon and a basic singular monogon in the annulus.

Lemma 6.1. *The bounding paths of any singular monogon or bigon in γ cross γ at most $2n$ times.*

Proof: Any point traversing the entire curve γ passes through each of the n self-intersection points twice, and the bounding paths of a singular bigon are disjoint in the domain by definition. □

An important subtlety of Hass and Scott's definition is that a lift of a singular bigon to the universal cover is not necessarily an *embedded* bigon. First, the lifted boundary paths of the bigon need not be simple or disjoint. More subtly, the endpoints of the lifted bigon might not enclose single corners: an embedded bigon looks like a lens ◊, but a lift of a singular bigon might resemble a heart ♡ or a butt ⊂⊃. Similarly, a lift of a singular monogon is not necessarily an *embedded* monogon; the lifted subpath might self-intersect way from its endpoint, and it may not enclose a single corner at its endpoint.

We define a singular monogon or singular bigon to be ***basic*** if any of its lifts on the universal cover is an *embedded* monogon or bigon, respectively. Hass and Scott proved that any closed curve with excess intersections on an arbitrary orientable surface, with or without boundary, must contain a singular monogon or a singular bigon [135, Theorem 4.2]. However, a close reading of their proof reveals that the singular monogon or singular bigon they find is in fact basic. We thus restate their result without repeating the proof.

Lemma 6.2 (Hass and Scott [135]). *Let γ be a closed curve on an arbitrary orientable surface. If γ has excess intersections, then there is a basic singular monogon or a basic singular bigon in γ.*

In their paper Hass and Scott also demonstrated a multicurve with excess intersections that *does not* contain any singular monogons or bigons. Therefore our algorithms do not generalize to multicurves directly. In fact, such question is still open.

Conjecture 6.1. *Any multicurve on an arbitrary orientable surface can be tightened using polynomially many homotopy moves.*

6.2 Surfaces with Boundary

In this section, we consider the case of surfaces with boundary.

Theorem 6.1. *On an oriented surface of genus g with $b > 0$ boundary components, a closed curve with n self-intersections can be tightened using at most $O((g+b)n^3)$ homotopy moves.*

Later in Section 6.3 we will describe a similar algorithm for closed curves on an arbitrary orientable surface without boundary. The reader is encouraged to follow the order of the presentation and get an intuitive sense of how the bigon removal algorithm operates in this simpler setting.

Removing singular bigons, as guaranteed by Lemma 6.2, is the foundation of our upper bound proofs. Given a curve γ with n vertices that is not already tightened, we decrease the number of vertices of γ as follows. If γ contains an *embedded* monogon or bigon, we delete it following Steinitz's algorithm (Lemma 2.2), using $O(n)$ homotopy moves. Otherwise, if γ contains a basic singular bigon, we attempt to remove it, essentially by swapping the two bounding curves; however, if at any point γ has only $n-2$ vertices, we immediately abort the bigon removal and recurse. Finally, if γ contains no basic singular bigons, Lemma 6.2 implies that γ must contain a basic singular monogon; we perform a single 0→1 move to transform it into a basic singular bigon (as shown in Figure 6.2) and then defer to the previous case.

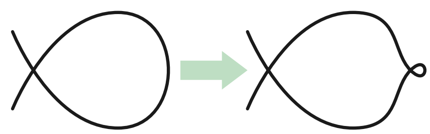


Figure 6.2. A single 0→1 move transforms a basic singular monogon into a basic singular bigon.

The curve γ is tightened after repeating the previous reduction process at most n times. Thus, Theorem 6.1 follows immediately from the following lemma, which we prove in the remainder of this section.

Lemma 6.3. *Let Σ be an orientable surface of genus g with $b > 0$ boundary components, and let γ be a closed curve in Σ with n vertices that contains a basic singular bigon, but no embedded monogons or bigons. The number of vertices of γ can be decreased by 2 using $O((g+b)n^2)$ homotopy moves.*

6.2.1 Removing a Basic Singular Bigon

Fix a surface Σ and a closed curve γ with n vertices, satisfying the conditions of Lemma 6.3. A ***system of arcs*** $\boldsymbol{\Delta}$ on the surface Σ is a collection of simple disjoint boundary-to-boundary paths that cuts the surface Σ open into one single polygon. Euler's formula implies that every system of arcs contains exactly $2g+b-1$ arcs. Cutting the surface along these arcs leaves a topological disk P whose boundary alternates between arcs (each arc in Δ appears twice) and subpaths of the boundary. We refer to P as the ***fundamental polygon*** of Σ with respect to Δ.

For any closed curve γ on any orientable surface Σ with boundary, there is a system of arcs Δ satisfying the following ***crossing property***: Each arc in Δ intersects each edge of γ at most twice, and only transversely. (For examples of such a construction, see Colin de Verdière and Erickson [64, Section 6.1] or Erickson and Nayyeri [92, Section 3].) The fundamental polygon induces a tiling of the universal cover of Σ; we call each lift of the fundamental polygon a ***tile***.

Any basic singular bigon β of γ in Σ lifts to a bigon $\hat\beta$ in the universal cover of Σ, with two bounding subpaths λ and ρ that are disjoint in the domain of γ except possibly at their endpoints. Since $\hat\beta$ bounds a disk in the universal cover, any lift of any arc of Δ intersects $\hat\beta$ an even number of times. The intersection of a tile with $\hat\beta$ may have several components; we call each component a ***block***. A block is ***transverse*** if it is adjacent to both λ and ρ, and ***extremal*** otherwise. The *transverse* blocks have a natural linear ordering $B_1, \dots, B_k$ along either λ or ρ.

Our process for removing the bigon $\hat\beta$ has three stages: (1) Sweep inward over the extremal blocks, (2) sweep across the sequence of transverse blocks, and finally (3) remove one small empty bigon at a corner of $\hat\beta$. The first two stages are illustrated in Figure 6.3. This homotopy projects to a homotopy on Σ. We will prove that at the end of this bigon removal process, γ has exactly $n-2$ vertices.

To simplify our algorithm, we actually abort the bigon-removal process immediately as soon as γ has $n-2$ vertices; however, for purposes of analysis, we conservatively assume that the removal process runs to completion.

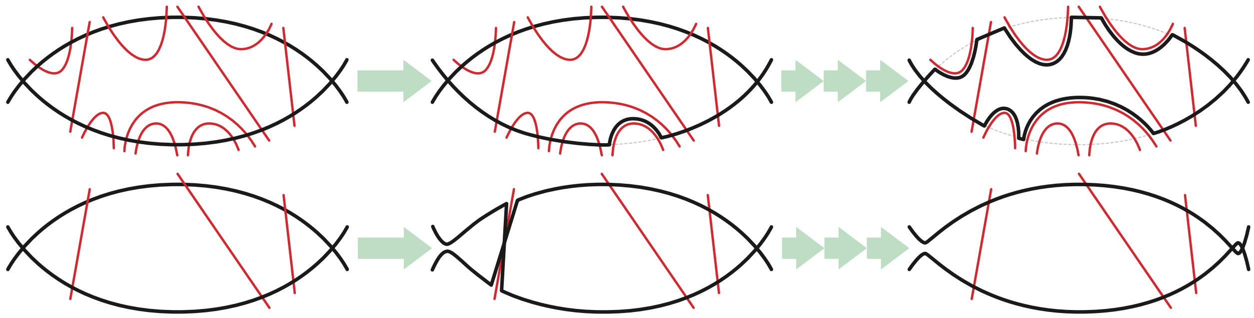


Figure 6.3. Removing a basic singular bigon on a surface with boundary. Top: Sweeping extremal blocks. Bottom: Sweeping transverse blocks.

We separately analyze stage (1) and stage (2) next.

Lemma 6.4. *All extremal blocks can be removed from $\hat{\beta}$ using $O((g+b)n^2)$ moves, without changing the number of vertices of γ.*

Proof: We actually describe how to remove every embedded bigon formed by a subpath of γ and a subpath of any arc in Δ using at most $O((g+b)n^2)$ homotopy moves, each of which is a 3→3 move. Every extremal block in $\hat{\beta}$ projects to such an embedded bigon, because tiles (and a fortiori blocks) project injectively into the surface Σ.

We proceed inductively as follows. Suppose γ and Δ bound an embedded bigon, since otherwise there is nothing to prove. Let B be a *minimal* embedded bigon with respect to containment, bounded by a subpath δ of an arc in Δ, and a subpath α of the curve γ. Because γ has no embedded monogons or bigons, every subpath of γ inside B is simple, and every pair of such subpaths intersects at most once. Moreover, every such subpath has one endpoint on α and the other endpoint on δ. Thus, the number of intersections between δ and γ is equal to number of intersections between α and $\gamma \setminus \alpha$.

To remove B, we apply the following homotopy process similar to Steinitz's algorithm (Lemma 2.2); the only difference here is that δ is not part of the curve γ, and therefore no actual homotopy move is required if some subpaths of δ participate in a move. We first sweep the subpath α across B until the bigon defined by α and δ has no vertices in its interior, and then sweep α across δ without performing any additional homotopy moves, as shown in Figure 6.4. Because the number of intersections between δ and γ is equal to number of intersections between α and $\gamma \setminus \alpha$, this sweep does not change the number of vertices of γ.

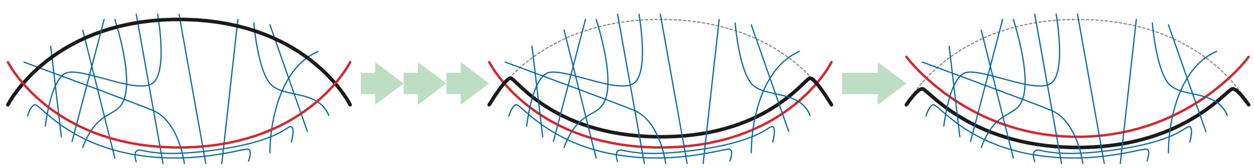


Figure 6.4. Sweeping a minimal embedded bigon bounded by a subpath of γ (black) and a subpath of Δ (red). Thin (blue) lines are other subpaths of γ.

To implement the sweep, Steinitz's lemma (Lemma 2.1) implies that if the interior of B contains any vertices of γ, then some triangular face of γ lies inside B and adjacent to an edge of α. Thus, we can reduce the number of interior vertices of B with a single 3→3 move. It follows inductively that the number of moves required to sweep over B is equal to the number of vertices in the interior of B, which is trivially at most n.

Removing a minimal embedded bigon between γ and Δ takes at most n moves and decreases the number of intersections between γ and Δ by 2. Each of the $O(g+b)$ arcs in Δ intersects each of the $O(n)$ edges of γ

at most twice by the crossing property of Δ, so the total number of such intersections is at most $O((g+b)n)$. Finally, because every move is a 3→3 move, we *never* change the number of vertices of γ. The lemma follows immediately. □

Now let $B_1, B_2, \ldots, B_k$ denote the sequence of transverse blocks of $\hat{\beta}$, and let $\boldsymbol{\delta_i}$ denote the common boundary B_i and B_{i+1} for each index i. Each path δ_i is a subpath of a lift of some arc in Δ. For notational convenience, let $x := \delta_0$ and $y := \delta_k$ denote the endpoints of $\hat{\beta}$, so that each block B_i has paths δ_{i-1} and δ_i on its boundary.

Recall that λ and ρ denote the bounding subpaths of $\hat{\beta}$. To sweep over the transverse blocks, we *intuitively* maintain a path ϕ from a point on λ to a point on ρ, which we call the ***frontier***. The frontier starts as a trivial path at the endpoint δ_0. Then we repeatedly sweep the frontier over B_i from δ_{i-1} to δ_i, as i goes from 1 to k. After these k iterations, the frontier lies at the endpoint δ_k.

Our actual homotopy modifies the bounding curves λ and ρ as shown in the bottom of Figure 6.3. Intuitively, the prefixes of λ and ρ "behind" ϕ are swapped; the frontier itself is actually an arbitrarily close pair of crossing subpaths connecting the swapped prefixes of λ and ρ with the unswapped suffixes "ahead" of the frontier. Replacing the single path ϕ with a close pair of crossing paths increases the number of homotopy moves to perform the sweep by only a constant factor.

Lemma 6.5. *Sweeping ϕ over one transverse block requires at most $O(n)$ homotopy moves.*

Proof: Consider a sweep over B_i, from δ_{i-1} to δ_i. We start by moving the frontier just inside B_i, without performing any homotopy moves. The main sweep passes ϕ over every vertex in B_i, including the vertices on the bounding paths λ and ρ, stopping ϕ just before it reaches δ_i. Finally, we move the frontier onto δ_i without performing any homotopy moves. Because the interior of each block projects injectively onto the surface, no other translate of ϕ intersects B_i during the sweep.

Up to constant factors, the number of homotopy moves required to sweep B_i is bounded by the number of vertices of γ inside B_i, plus the number of intersections between γ and the bounding subpaths δ_{i-1} or δ_i. There are trivially at most n vertices in B_i, and the crossing property of the system of arcs Δ implies that each arc in Δ intersects γ at most $O(n)$ times. □

With the two previous lemmas in hand, we are finally ready to prove Lemma 6.3. Let γ be a closed curve in Σ with a basic singular bigon β, let $\hat{\beta}$ be a lift of β to the universal cover of Σ, and let λ be one of the bounding paths of $\hat{\beta}$.

The definition of singular bigon implies that λ contains at most $2n$ edges of γ by Lemma 6.1. Each of these edges crosses each arc of Δ at most twice, and there are $O(g+b)$ arcs in Δ, so λ crosses Δ at most $O((g+b)n)$ times. Each transverse block B_i except the last can be charged to the unique intersection point $\delta_i \cap \lambda$. We conclude that $\hat{\beta}$ contains $O((g+b)n)$ transverse blocks.

Sweeping inward over all extremal blocks in $\hat{\beta}$ requires $O((g+b)n^2)$ homotopy moves and does not change the number of vertices of γ by Lemma 6.4. Sweeping over all $O((g+b)n)$ transverse blocks requires a total of $O((g+b)n^2)$ homotopy moves by Lemma 6.5. Sweeping the transverse blocks has the same effect as smoothing one endpoint of the bigon and doubling the other endpoint, as shown on the bottom right of Figure 6.3, which implies that γ still has n vertices. Removing the final empty bigon with a single 2→0 move reduces the number of vertices to $n-2$.

This completes the proof of Lemma 6.3, and therefore the proof of Theorem 6.1.

6.3 Surfaces Without Boundary

In this section, we prove our upper bound for closed curves on surfaces without boundary. The following theorem improves over the $O(n^4)$ bound given by the earlier conference version [47] when the genus g is at most $n/\log^2 n$.

Theorem 6.2. *On an oriented surface without boundary, a closed curve with n self-intersections can be tightened using at most $O(gn^3 \log^2 n)$ homotopy moves.*

We follow the same high-level strategy described in Section 6.2; consequently, it suffices to prove that a basic singular bigon can be removed using $O(gn^2 \log^2 n)$ homotopy moves.

Instead of a system of arcs, we decompose the surface using a *reduced cut graph*; this cut graph induces a regular hyperbolic tiling in the universal cover of the surface. In Section 6.3.1 we describe how to compute a cut graph whose induced tiling intersects the bounding paths of any basic singular bigon at most $O(n)$ times. In Section 6.3.2, we apply Dehn's isoperimetric inequality for regular hyperbolic tilings [75] to bound the number of tiles lying in the interior of the bigon. Then we describe our process for removing a singular bigon at two levels of detail. First, in Section 6.3.3, we provide a coarse description of the homotopy as a sequence of moves in the *bigon graph*, which is the decomposition of the lifted bigon by the tiling. We process the regions in this decomposition in a particular order to keep the number of *chords* created by translates of the moving path under control. Finally in Section 6.3.4 we obtain the actual sequence of homotopy moves by carefully perturbing the curves in the previous homotopy into general position; bounding the intersections between perturbed chords is the most delicate portion of our analysis.

6.3.1 Dual Reduced Cut Graphs

A ***tree-cotree decomposition*** of a cellularly embedded graph G is a partition (T, L, C) of the edges of G into three disjoint subsets: a spanning tree T of G, the edges C corresponding to a spanning tree of the dual graph G^*, and exactly $2g$ leftover edges $L := E(G) \setminus (T \cup C)$, where g is the genus of the underlying surface [86].

Let γ be a closed curve on Σ; we temporarily view γ as a 4-regular graph with some given embedding. However, the embedding of γ is not necessarily cellular; let G be a cellular refinement of γ obtained by triangulating every face. A ***dual reduced cut graph X*** (hereafter, just ***cut graph***) is a cellularly embedded graph obtained from a tree-cotree decomposition (T, L, C) of G as follows: Start with the subgraph of G^* induced by the dual spanning tree C^* and the leftover edges L^*, repeatedly delete vertices with degree one, and finally perform series reductions on all vertices with degree two [91].

The cut graph X inherits a cellular embedding into Σ from the embedding of G^*; by construction, this embedding has exactly one face. Because every vertex of X has degree 3, Euler's formula implies that X has exactly $4g - 2$ vertices and $6g - 3$ edges. To be consistent with the terminology in Section 6.2.1, we call the edges of X ***arcs***. Cutting the surface Σ along X yields a polygon with $12g - 6$ sides, which we call the ***fundamental polygon*** of X. The cut graph induces a regular tiling $\hat{X}$ of the universal cover $\hat{\Sigma}$ of Σ; we refer to each lift of the fundamental polygon of X as a ***tile***.

By construction, the cut graph X satisfies the following ***crossing property***: *Each edge of the curve γ crosses X at most once*. We emphasize that this crossing property might no longer hold when we start moving the curve γ. Compared with the system of arcs we used in Section 6.2 which satisfies a weaker crossing property (that each edge of γ crosses *each* arc at most $O(1)$ times), the cut graph gives us an improved upper bound on the number of tiles intersecting the bounding paths of an embedded bigon in the universal cover of Σ.

Bigon graph. The tiling of the universal cover of Σ induced by the cut graph X decomposes the disk bounded by $\hat{\beta}$ into pieces; we call this decomposition the ***bigon graph***. More formally, we define the bigon graph G as follows. By construction, $\hat{\beta}$ intersects the tiling $\hat{X}$ transversely. The vertices of G are the two endpoints of $\hat{\beta}$, the intersections of λ and ρ with arcs of the tiling, and the vertices of the tiling in the interior of $\hat{\beta}$. The arcs of G are subpaths of $\lambda \cup \rho$ and subpaths of tiling arcs bounded by these vertices. Finally, the bounded faces of G are the components of the intersection of each tile with the interior of $\hat{\beta}$. We emphasize that the intersection of a single tile with the interior of $\hat{\beta}$ may have several components.

6.3.2 Isoperimetric Inequality

Consider an embedded bigon $\hat{\beta}$ in the universal cover of surface Σ, which is a lift of a basic singular bigon in the curve γ on Σ. Unlike the case of surface with boundary in Section 6.2, there may be tiles lying completely in the *interior* of the bigon $\hat{\beta}$, without intersecting the two bounding paths. We bound the number of such interior tiles using a discrete isoperimetric inequality, which is a consequence of Dehn's seminal observation that the graph metric defined by a regular tiling of the hyperbolic plane is a good approximation of the continuous hyperbolic metric [75]. We provide a self-contained proof of this inequality, using a combinatorial version of the Gauss-Bonnet theorem described at varying levels of generality by Banchoff [18], Lyndon and Schupp [174], and Gersten and Short [111].

Let G be a graph with a cellular embedding onto surface Σ, and let $\chi(\Sigma)$ be the Euler characteristic of Σ, defined as the number of vertices and faces in G minus the number of edges in G, which is equal to $\chi(\Sigma) = 2 - 2g - b$, where g is the genus of Σ and b is the number of boundary components of Σ. One can view the definition of the Euler characteristic of Σ through a different lens. Assign an arbitrary real "interior angle" $\angle c$ to each corner c of Σ. Define the ***curvature*** $\boldsymbol{\kappa(v)}$ of a vertex v of G as $1 - \frac{1}{2}\deg v + \sum_{c \in v}(\frac{1}{2} - \angle c)$, and the ***curvature*** $\boldsymbol{\kappa(f)}$ of a face f of G as $1 - \sum_{c \in f}(\frac{1}{2} - \angle c)$. The following equality, which is an immediate consequence of Euler's formula, is known as the ***combinatorial Gauss-Bonnet theorem***:

$$\sum_{v} \kappa(v) + \sum_{f} \kappa(f) = \chi(\Sigma).$$

Now we are ready to bound the number of faces of the bigon graph G. The ***perimeter*** $L(G)$ of the bigon graph G is the number of intersections between the two bounding paths of the bigon and arcs of $\hat{X}$.

Lemma 6.6. *Let Σ be a closed surface of genus $g > 1$, let γ be a closed curve on Σ, let X be the cut graph of γ on Σ, and let G be a bigon graph of some embedded bigon $\hat{\beta}$ in $\hat{\Sigma}$. Then the number of faces in the bigon graph G is at most $O(L(G))$.*

Proof: Let I denote the union of all tiles in $\hat{X}$ that lie entirely in the interior of $\hat{\beta}$ (that is, the union of all faces of G that are actually complete tiles). The region I may be empty or disconnected; however, every component of I is a closed disk. First we connect the number of tiles in I with the number of vertices on the boundary of I.

Let D be an arbitrary component of I. Let A denote the number of tiles in D, and let L denote the number of vertices on the boundary of D. Every boundary vertex is either incident to one interior tile and has degree 2 (convex) or incident to two interior tiles and has degree 3 (concave). Let L^+ and L^- respectively denote the number of convex and concave vertices on the boundary of D. Assign angle $1/3$ to each corner of D, so that

- every interior vertex has curvature 0,

- every convex vertex has curvature $1/6$,
- every concave vertex has curvature $-1/6$, and
- every face has curvature $2-2g$.

The combinatorial Gauss-Bonnet theorem now implies that $(L^+ - L^-)/6 + (2-2g)A = 1$, and therefore $L^+ - L^- = (12g-12)A+6$. (In particular, $L^+ \geq L^-$.) Thus, some face f is incident to at least $12g-11$ convex vertices, and therefore at least $12g-10$ arcs on the boundary of D. Deleting f from D removes at least $12g-10$ boundary arcs and exposes at most 4 interior arcs. The isoperimetric inequality $A \leq L/(12g-14)$ now follows immediately by induction.

Now consider the embedded bigon $\hat{\beta}$. Because each vertex in $\hat{X}$ has degree 3, every convex vertex of I is either incident to an arc intersecting $\hat{\beta}$, or incident to another convex vertex of I, in which case the two convex vertices belongs to different components of I. The number of components of I having no convex vertices incident to $\hat{\beta}$ is strictly less than the number of those do, and therefore by an easy charging argument, there are at most $O(L(G))$ convex vertices on the boundary of I. Using the deduced inequality $L^+ \geq L^-$ from the previous paragraph, we have now showed that I contains at most $O(L(G))$ vertices and thus at most $O(L(G)/g)$ tiles. In the mean while, at most $O(L(G))$ faces are incident to the boundary of $\hat{\beta}$. Thus, the total number of faces of G is at most $O(L(G))$, as claimed. □

Lemma 6.1 and the crossing property of the cut graph X imply that at most $O(n)$ tiles of Σ intersect the two bounding paths λ and ρ of $\hat{\beta}$. Thus Lemma 6.6 implies that the bigon graph G has at most $O(n)$ faces, and therefore $O(n)$ vertices and arcs by Euler's formula.

As a corollary, one can derive a logarithmic bound on the maximum distance from any vertex of $\hat{X}$ inside the bigon to one of the two bounding paths.

Lemma 6.7. *Let Σ be an orientable surface of genus $g > 1$ and γ be a closed curve on Σ. Let X be the cut graph of γ on Σ and G be the corresponding bigon graph of some embedded bigon $\hat{\beta}$ in $\hat{\Sigma}$. Denote n the number of vertices in G. Then the maximum distance from a vertex of G to either bounding path of $\hat{\beta}$ is at most $O(\log n)$.*

Proof: Consider the set $S_{\leq k}$ of vertices of G with distance at most k to some fixed vertex v of G. As k grows, the set $S_{\leq k}$ grows exponentially in size since $\hat{X}$ is a hyperbolic tiling. This implies that for some distance $k = O(\log n)$ the set $S_{\leq k}$ has non-empty intersection with the given bounding path of $\hat{\beta}$. □

6.3.3 Coarse Homotopy

Let β be a basic singular bigon in γ, let $\hat{\beta}$ be its lift to the universal cover, and let λ and ρ be the bounding paths of $\hat{\beta}$. Our goal is to remove this bigon by swapping the bounding paths λ and ρ, which has the same effect as smoothing the two endpoints of β, reducing the number of vertices of γ by 2. See Figure 6.5. In this section, we construct a homotopy from λ to ρ, not as a sequence of individual homotopy moves, but as a coarser sequence of moves in the bigon graph of $\hat{\beta}$. Applying the reversal of this sequence of moves to ρ moves it to the original position of λ, completing the exchange of the two bounding paths.

Discrete homotopy. We construct a *discrete homotopy* [35, 36, 133] through the bigon graph G that transforms one bounding path λ of the bigon into the other bounding path ρ. This discrete homotopy is a sequence of walks in G—which may traverse the same arc in G more than once—rather than a sequence of generic curves. In the

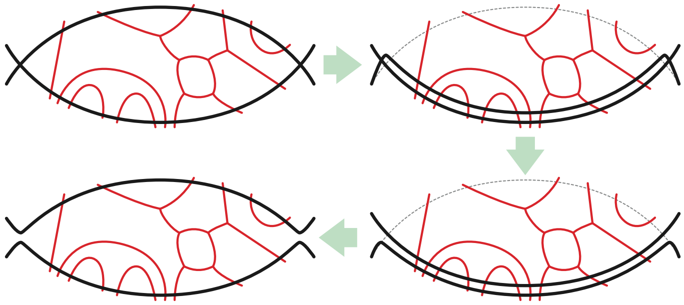


Figure 6.5. Swapping the two bounding paths of a bigon.

next section, we will carefully perturb these walks into generic curves, and implement each step of the discrete homotopy as a finite sequence of homotopy moves.

Let W be a walk on the bigon graph G from one endpoint of the bigon to the other. A ***spike*** in W is an arc of G followed immediately by the same arc in the opposite direction. We define two local operations for modifying W; see Figure 6.6.

- **Face move:** Replace a single arc e in W with the complementary boundary walk around some face f of G that is incident to e.
- **Spike move:** Delete a spike from W and decrease the length of W by two.

We emphasize that after a face move across face f, the frontier walk W may traverse some arcs of f more than once; moreover, these multiple traversals may or may not be spikes. Because every face f is a disk, the arc e and its complementary boundary walk around f share endpoints, and thus any face move can be implemented by a homotopy across f. Similarly, a spike move can be implemented by a homotopy in the arc containing the spike. A discrete homotopy in G is a finite sequence of face moves and spike moves. We refer to the current walk W at any stage of this homotopy as the ***frontier walk***.

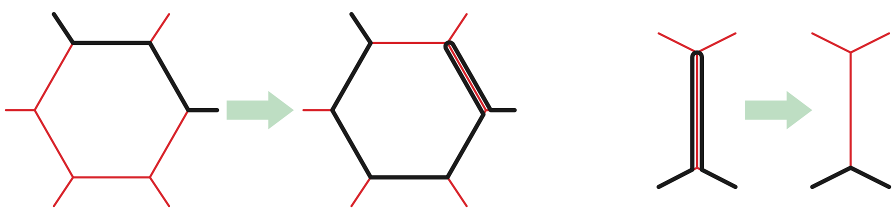


Figure 6.6. A face move and a spike move.

The following result by Har-Peled *et al.* [133] guarantees the existence of some discrete homotopy whose frontier walk is short at all times. The original result assumes that the underlying graph is triangulated; however the proof still works on regular tilings.

Lemma 6.8 (Har-Peled *et al.* [133, Theorem 1]). *Given an n-vertex arc-weighted bigon graph G with two bounding paths λ and ρ, there is a discrete homotopy from λ to ρ whose frontier walk has (weighted) length at most*

$$O(|\lambda| + |\rho| + f_* \cdot (d_* + w_*) \cdot \log n),$$

where f_* is the maximum size of the faces, d_* is the maximum distance between a vertex in G and a vertex on λ, and w_* is the maximum arc weight over all arcs. Furthermore, each arc of G is traversed at most twice by any *frontier walk in the discrete homotopy.*

Set the weight of each arc e in the bigon graph G to 0 if e is on the bounding path λ or ρ, and set the weight to 1 otherwise. By Lemma 6.7, the maximum distance between a vertex in G and a vertex on λ is at most $O(\log n)$. Now apply Lemma 6.8 to G, one obtains a discrete homotopy from λ to ρ where all the frontier walks have at most $O(g \log^2 n)$ arcs of $\hat{X}$ not on λ or ρ, and each arc of G is traversed at most twice by any frontier walk in the discrete homotopy. By crossing property of the cut graph X, there is at most $O(n)$ crossings between $\hat{\gamma}$ and X, and therefore together with Lemma 6.1 all the frontier walks have at most $O(gn \log^2 n)$ crossings with $\hat{\gamma}$. We refer these as the ***frontier property*** of the coarse homotopy, summarized as follow: At every stage of the discrete homotopy,

(a) the frontier walk W passes through at most $O(g \log^2 n)$ arcs of $\hat{X}$ not on λ or ρ,

(b) the frontier walk W intersects (the original) $\hat{\gamma}$ at most $O(gn \log^2 n)$ times, and

(c) each arc of G is traversed at most twice by any frontier walk in the discrete homotopy.

6.3.4 Fine Homotopy

Interactions between the moving frontier and the original curve present a significant subtlety in our algorithm. We refine the discrete homotopy in the previous section, first by perturbing the moving frontier walk so that after every graph move γ is a generic curve, and then by decomposing the perturbed graph moves into a finite sequence of homotopy moves.

Perturbing the frontier. First, given the frontier walk $\hat{W}$ at any stage of the coarse homotopy, perturb $\hat{W}$ into a simple path in the universal cover $\hat{\Sigma}$. Based on the frontier property (c) of the coarse homotopy described in Section 6.3.3, combinatorially there is only one such perturbation. We will denote the *perturbed frontier walk* in $\hat{\Sigma}$ by $\hat{\omega}$. Project the perturbed frontier walk $\hat{\omega}$ back to the surface Σ to obtain the ***frontier curve*** $\boldsymbol{\omega}$. Notice that the number of self-intersections of ω near any vertex of (the original) γ is at most 4 (locally it looks like the symbol #). The frontier curve ω is not necessarily generic, as subpaths of ω near the cut graph X could overlap each other in unspecified ways.

To specify the perturbation near the cut graph X, we define a convenient family $\mathcal{O}$ of open sets, which we call ***bubbles***, that covers the cut graph X and its complement face in Σ, following a construction of Babson and Chan [17]. (See also Erickson [89].) Each bubble in $\mathcal{O}$ is either a vertex bubble, an arc bubble, or a face bubble.

The vertex bubbles are disjoint open balls around the vertices of X. The arc bubbles are disjoint open neighborhoods of the portions of the arcs of X away from the vertices. Finally, the face bubble is an open neighborhoods of the portions of $\Sigma \setminus X$ away from the vertices and the arcs; there is only one face bubble in $\mathcal{O}$. The intersection of all pairs of two bubbles of different types is the disjoint union of open disks, one for each incidence between the corresponding vertex and arc, vertex and face, or arc and face of X. See Figure 6.7.

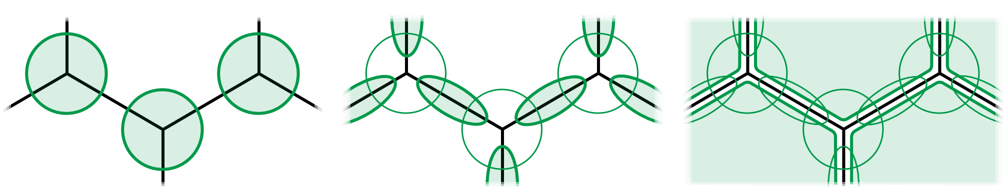


Figure 6.7. Vertex bubbles, arc bubbles, and face bubbles.

We now describe how to draw the frontier curve ω near the cut graph X so that the complexity inside each bubble of X is controlled. We model each arc bubble as a *Euclidean* rectangle containing several straight segments parallel to the arc, which we call ***tracks***, arranged so that each track in the arc bubble of e intersects γ transversely. (The metric is merely a convenience, so that we can write "straight" and "parallel"; the tracks can be defined purely combinatorially.)

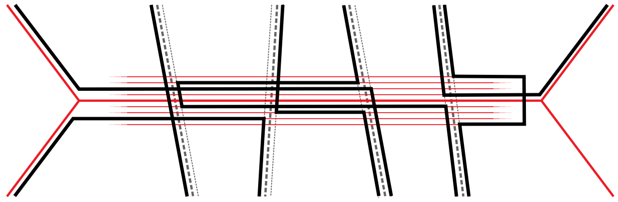


Figure 6.8. Closeup of an arc bubble of some arc e in X, showing subpaths of the frontier ω sticking to subpaths in e, including the perturbations of two spikes.

Now consider a frontier curve ω; each maximal subpath along some arc e of X is drawn on a unique track in the arc bubble of e. Moreover, when ω switches from arc e to another arc e' (including at the tip of a spike at a vertex of X, which we view as a zero-length walk), there is a corner at the intersection of those two tracks. The part of ω that follows the bounding paths λ and ρ stays unchanged.

Thus, every subpath of ω inside the arc bubble of some arc e alternates between tracks parallel to e and either (1) tracks parallel to other arcs or (2) parallel to the bounding paths λ and ρ. Intuitively, we say that a subpath of ω ***sticks to*** an arc of X if the subpath lies on some track in the corresponding arc bubble. Similarly, we say that a subpath of ω ***sticks to*** the bounding paths λ and ρ if the corresponding subpath of $\hat{W}$ traverses arcs of G on $\hat{\lambda}$ and $\hat{\rho}$ in the universal cover $\hat{\Sigma}$. See Figure 6.8.

Graph moves revisited. In our perturbed homotopy, we require every face move to be performed entirely within the corresponding face bubble, and every spike move to be performed entirely within the corresponding arc bubble, while maintaining the track structure of the perturbed frontier ω. To this end, we introduce two additional graph moves for modifying the frontier curve ω.

- **Arc move**: Move a maximal subpath sticking to an arc e of X into an incident face bubble, within the arc bubble of e implemented by switching tracks.
- **Vertex move**: Move the curve across a vertex v of X within the corresponding vertex bubble.

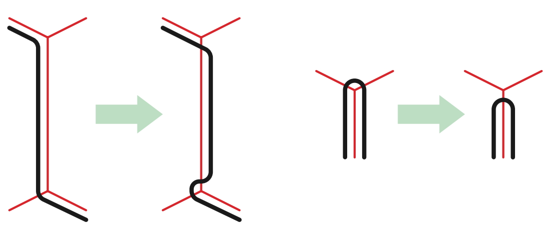


Figure 6.9. An arc move and a vertex move.

The arc moves and vertex moves can be seen as preprocessing steps to ensure that a subpath of ω lies in the proper face or arc bubble before performing a face or a spike move. Thus, our perturbed coarse homotopy still follows the outline given in Section 6.3.3, but now each face move might be prefaced by a single arc move, and each spike move might be prefaced by a single vertex move, as shown in Figure 6.10.

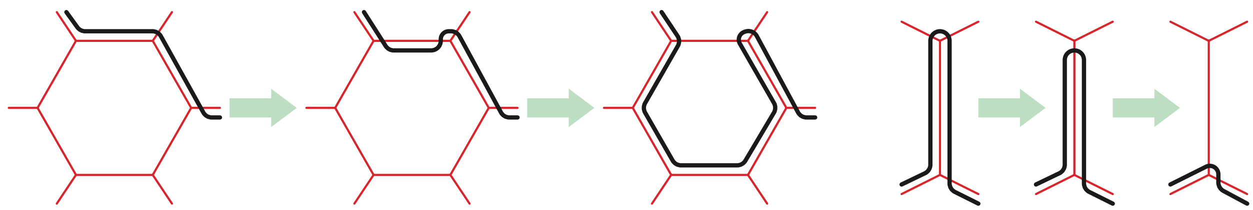


Figure 6.10. Top: An arc move followed by a face move. Bottom: A vertex move followed by a spike move.

We emphasize that every face move is performed entirely within a face bubble, every arc move and spike move is performed entirely within an arc bubble, and every vertex move is performed entirely within a vertex bubble. Therefore each graph move can be implemented solely on the original surface Σ.

The final homotopy. Finally, we construct a sequence of homotopy moves moving one bounding path λ to the other bounding path ρ by decomposing the perturbed graph moves.

Lemma 6.9. *Let Σ be an orientable surface without boundary, and let γ be a closed curve with n vertices on Σ that contains a basic singular bigon β, but no embedded monogons or bigons. Then β can be removed using $O(gn^2 \log^2 n)$ homotopy moves, without changing the rest of γ.*

Proof: Let $\hat{\beta}$ be the lift of β to the universal cover $\hat{\Sigma}$; let $\hat{\lambda}$ and $\hat{\rho}$ be the bounding curves of $\hat{\beta}$ in $\hat{\Sigma}$; let G be the corresponding bigon graph. Our earlier analysis implies that G has at most $O(n)$ vertices, arcs, and faces. Thus, moving $\hat{\lambda}$ to $\hat{\rho}$ requires at most $O(n)$ graph moves.

Each of these graph moves is performed within a bubble in $\mathcal{O}$ that embeds in Σ, and therefore can be realized using $O(m)$ homotopy moves, where m is the number of vertices of γ within that bubble before the graph move begins. It remains only to prove the following claim:

> Between any two graph moves, the number of vertices of γ inside any bubble is at most $O(gn \log^2 n)$.

The proof of this claim is surprisingly delicate. All the properties we mentioned in each of the previous subsections contribute to avoid the danger of increasing the number of vertices in γ uncontrollably during the process: (1) dividing the homotopy into graph moves (Section 6.3.1 and Section 6.3.2), (2) the order of face moves (Section 6.3.3), and (3) the way the perturbed frontier curve lies near the cut graph (Section 6.3.4). For the rest of the proof we refer to subpaths of the frontier curve ω within a bubble simply as ***chords***.

The maximum number of vertices of curve γ inside a bubble between two graph moves is at most the sum of the number of vertices of γ before the homotopy, the number of intersections between the original γ and the chords, and the number of intersections between pairs of chords. The first term is at most n by definition; the frontier property (b) of the coarse homotopy implies that the second term is at most $O(gn \log^2 n)$. To bound the last term, we separately consider each type of bubble:

- *Face bubble:* Because vertices and arcs of the tiling do not lie inside a face bubble, every chord in a face bubble sticks to the bounding paths λ or ρ. The way we construct the perturbed frontier curve ω ensures that at most two chords stick to the same subpath of λ or ρ in the bubble. Since both λ and ρ are subpaths of γ, we can charge the intersections between chords to the corresponding vertices in (the original) γ. There are at most n vertices of γ in the face bubble, and therefore at most $O(n)$ vertices are created by intersecting chords within the bubble.

- *Arc bubbles:* Our construction ensures that the chords within each arc bubble are polygonal curves, and the number of intersections between two such chords does not exceed the sum of the number of segments in each of them. The frontier property (c) of the coarse homotopy implies that each arc of X is traversed by $\hat{W}$ at most twice, and therefore each chord sticking to e consists of at most $O(1)$ segments. It follows that any pair of chords that stick to the arc intersect $O(1)$ times. The frontier property (a) of the coarse homotopy implies that $\hat{W}$ always traverse at most $O(g \log^2 n)$ arcs not on the bounding paths λ and ρ. This in turn implies that there are at most $O(g \log^2 n)$ chords inside any arc bubble that stick to the arc, and at most $O(g \log^2 n)$ tracks are needed for any arc bubble. By crossing property of the cut graph, at most $O(n)$ chords stick to λ and ρ, We conclude that at most $O(gn \log^2 n)$ vertices are created by intersecting chords within any arc bubble.

- *Vertex bubbles:* Being subpaths of γ, bounding paths λ and ρ will never intersect the vertex bubbles. Thus, each chord in a vertex bubble sticks to a walk on the arcs of X incident to the vertex, which must have length 2. Our construction of the perturbed frontier ω ensures that each pair of these chords intersects at most $O(1)$ times. Similar to the case of the arc bubble, there are at most $O(\log^2 n)$ chords inside any vertex bubble. We conclude that at most $O(\log^4 n)$ vertices are created by chords intersecting within any vertex bubble.

This concludes the proof. □

Summary. We conclude by summarizing our proof of Theorem 6.2. Let γ be a closed curve on an orientable surface without boundary. If γ is not yet tightened, Lemma 6.2 implies that after at most one $0 \to 1$ move (see Figure 6.2), γ contains at least one basic singular bigon. By Lemma 6.9, we can decrease the number of vertices of γ by two by removing one basic singular bigon in $O(gn^2 \log^2 n)$ homotopy moves. After $O(n)$ such bigon removals, all the excess intersections of γ must have been removed. We conclude that γ can be tightened using at most $O(gn^3 \log^2 n)$ homotopy moves.

6.4 Tightening Curves Using Monotonic Homotopy Moves

The proofs of Hass and Scott [136] and de Graaf and Schrijver [125] have the additional benefit that the number of vertices of the curve never increases during the homotopy process.[1] It would be much preferable if our efficient homotopy processes in Section 6.2 and Section 6.3 are monotonic as well; in other words, we are looking for a sequence of polynomially many *monotonic* homotopy moves to tighten the given multicurve.

We made partial progress towards such a goal. In particular, we show that it is sufficient to assume the surface has boundary. Let γ be a multicurve on Σ, and let $\boldsymbol{\gamma_*}$ be the unique ***geodesic*** of γ on Σ—a multicurve consisting of each shortest representative among the homotopy class of the constituent curves in γ. Let the ***ε-neighborhood*** of a curve γ be the union of all ε-disks centered at some point of γ. We say the curve γ is ***ε-close*** to the geodesic γ_* if the lift of γ in the universal cover lies in an ε-neighborhood of the lift of γ_*.

Lemma 6.10. *Let γ be an n-vertex noncontractible curve on a genus-g orientable surface Σ and let γ_* be the unique geodesic of γ on Σ. Curve γ can be made ε-close to γ_* using $O(n^5 \log^3 g / g^2)$ monotonic homotopy moves for some $\varepsilon = O(g/(n \log g))$; furthermore, the ε-neighborhood of γ_* does not cover the whole surface Σ.*

[1]De Graaf and Schrijver's result requires a fourth type of homotopy move, which moves an isolated simple contractible constituent curve from one face of the rest of the multicurve to another. However, since this move can only be applied to disconnected multicurves, it does not affect our argument.

The proof of Lemma 6.10 can be viewed as an efficient implementation of the first step of the algorithm by de Graaf and Schrijver, moving the curve close to the unique geodesic of its homotopy class. Our proof relies heavily on hyperbolic trigonometry; for a clean introduction to the topic see Traver [241].

6.4.1 Moving Curves Close to Geodesics

In this subsection we prove Lemma 6.10. Let $\oplus$ be a tangle whose disk is endowed with a hyperbolic metric. Tangle $\oplus$ is ***straightened*** if all the strands of $\oplus$ are geodesics with respect to d_H. We emphasize the difference between *straightened* and *tightened*; a straightened tangle must be tightened, but *not* vice versa. We will make use of the following quantitative version of Ringel's homotopy theorem [203, 204] (see also [120, 125, 136, 211]).

Lemma 6.11 (Hass and Scott [136, Lemma 1.6]). *Any n-vertex tangle can be straightened (with respect to a hyperbolic metric) monotonically using $O(n^2)$ homotopy moves.*

Construct hyperbolic metric. First we modify γ by straightening all the strands of within the open disk $\Sigma \setminus X$ using $O(n^2)$ moves by Lemma 6.11. Now we construct a hyperbolic metric on surface Σ such that

(1) the length of the (modified) curve γ is at most $O(n \log g)$, and

(2) the length of the shortest non-contractible cycle (known as the *systole*) is at least 1.

The construction is similar to the argument in Dehn's seminal result [75] that the graph distance on a regular tiling of the universal cover $\hat{\Sigma}$ approximates the hyperbolic metric on $\hat{\Sigma}$. Construct the cut graph X from curve γ such that every edge of γ crosses X at most $O(1)$ times, like we described in Section 6.3.1. Lift the cut graph X to the universal cover endowed with the unique hyperbolic metric, such that each corner has angle $1/3$ circles; this implies that each side of the fundamental polygon has length at least 1.[2] One can project the metric back to the original surface; denote the hyperbolic metric constructed as $\boldsymbol{d_H}$.

To prove that the hyperbolic metric d_H defined on surface Σ satisfies property (1), consider the modified curve γ where all strands within the open disk $\Sigma \setminus X$ are straightened. Note that any geodesic path not intersecting X has length at most $O(\log g)$ (which is the diameter of the fundamental polygon with respect to d_H). By Lemma 6.1 this implies that the length of the modified γ is at most $O(n \log g)$, thus the hyperbolic metric d_H satisfies property (1).

As for property (2), consider any non-contractible cycle σ on surface Σ; without loss of generality assume σ to be a geodesic. If we lift σ to the universal cover $\hat{\Sigma}$ such that the lift $\hat{\sigma}$ starts and ends on the lift $\hat{X}$ of the cut graph X, because σ is non-contractible, the two arcs of $\hat{X}$ where $\hat{\sigma}$ starts and ends respectively are two different translates of the same arc in X. Consider the sequence of arcs $a_0, \ldots, a_k$ in $\hat{X}$ intersected by $\hat{\sigma}$. Because σ is a geodesic and every vertex in $\hat{X}$ has degree 3, one has $a_i \neq a_{i+1}$ and no a_i is incident to a_{i+2} for all i. If for some i the two arcs a_i and a_{i+1} are not incident to each other (that is, a_i and a_{i+1} do not share a vertex in $\hat{X}$), then by hyperbolic trigonometry the length of the subpath of $\hat{\sigma}$ connecting a_i to a_{i+1} is at least the length of the side of the polygon, which is at least 1. Otherwise, if a_i is incident to a_{i+1} and a_{i+1} is incident to a_{i+2}, as a_i is not incident to a_{i+2}, by reflecting the subpath of $\hat{\sigma}$ from a_{i+1} to a_{i+2} to the tile that contains a_i and a_{i+1} we again have the length of the subpath of $\hat{\sigma}$ lower-bounded by the length of a_{i+1}. This proves that d_H satisfies property (2).

Tortuosity. Let $\gamma : [0,1] \to \Sigma$ be a curve. Denote $\boldsymbol{D(x,r)}$ the disk centered at point x with radius r (with respect the metric d_H). Let I_t be the maximal interval of $[0,1]$ containing t such that $\gamma(I_t)$ lies in the disk $D(\gamma(t), 1/2)$.

[2] To be accurate, the length of the side is equal to $2\cosh^{-1}(\sin(2\pi/6) \cdot \cos(2\pi/(24g-12)))$.

The ***tortuosity*** [125] of curve γ at point t, denoted as ***tort(γ, t)***, is the difference between the length of the subpath of γ lying in the disk of radius $1/2$ centered at $\gamma(t)$ and the geodesic distance between the two endpoints of the subpath. Formally,

$$tort(\gamma, t) \coloneqq |\gamma(I_t)| - d_H(\gamma(I_t(0)), \gamma(I_t(1)))\,.$$

Practically speaking, the tortuosity of γ at point t is equal the improvement one will make after straightening the disk $D(\gamma(t), 1/2)$. The ***tortuosity*** of curve γ is the supremum of $tort(\gamma, t)$ where t ranges over $[0, 1]$. The goal of the following lemma is to prove that when the tortuosity of the curve is small, then the whole curve is ε-close to its unique geodesic. In other words, as long as the curve γ has points that are at least ε away from the geodesic, we can always find a disk centered at some point of γ whose straightening will decrease the length of γ by at least fixed amount, depending only on ε.

Lemma 6.12. *For any small $\varepsilon > 0$, if the tortuosity of γ is at most $O(\varepsilon^2)$, then γ is ε-close to the geodesic γ_*.*

Proof: We will prove the contrapositive statement using hyperbolic trigonometry. For the sake of generality we temporarily treat r as a variable; at the end of the calculation one just plugin $r \coloneqq 1/2$. Here we list two identities that will be used in our proof.

(1) For any real number x, $\sinh(2x) = 2 \sinh x \cosh x$ and $(\cosh(x))^2 - (\sinh(x))^2 = 1$.

(2) Given an arbitrary *Saccheri quadrilateral* with the lengths of the legs, base, and top as a, b, and c respectively, then

$$\sinh \frac{c}{2} = \cosh a \cdot \sinh \frac{b}{2}.$$

Lift both γ and γ_* to the universal cover $\hat{\Sigma}$; denote the resulting paths as $\hat{\gamma}$ and $\hat{\gamma}_*$ accordingly. Let t be a point in $[0, 1]$ such that $\hat{\gamma}(t)$ has maximum distance to $\hat{\gamma}_*$. Refer to point $\hat{\gamma}(t)$ as p and the maximum distance as δ; by assumption δ is at least ε. Our goal is to prove that the tortuosity of γ at t is at least $\Omega(\varepsilon^2)$. Denote the two endpoints of the maximal subpath of $\hat{\gamma}$ in $D(p, r)$ containing p as x and y, and the maximal subpath itself as $\hat{\gamma}[x, y]$. One has

$$tort(\gamma, t) = |\hat{\gamma}[x, y]| - d_H(x, y) \geq 2r - d_H(x, y).$$

Here without loss of generality we will assume that x and y are both at distance exactly δ to $\hat{\gamma}_*$. The reason one can make such an assumption is because, as one moves x and y perpendicularly along the geodesics away from $\hat{\gamma}_*$, $d_H(x, y)$ increases and therefore the tortuosity when both x and y are at distance δ is a lower bound to the original tortuosity.

What is left is to upper bound $d_H(x, y)$. Let x^*, p^*, and y^* be the points on $\hat{\gamma}_*$ that have minimum distance to x, p, and y respectively. By identity (2) one has

$$\sinh\left(d_H(x, y)/2\right) = \cosh \delta \cdot \sinh\left(d_H(x^*, y^*)/2\right)$$

and

$$\sinh(r/2) = \cosh \delta \cdot \sinh\left(d_H(x^*, y^*)/4\right).$$

The second equality gives us

$$d_H(x^*, y^*)/2 = 2 \sinh^{-1}\left(\frac{\sinh(r/2)}{\cosh \delta}\right),$$

plug back to the first equation one has

$$\sinh\left(d_H(x,y)/2\right) = \cosh\delta \cdot \sinh\left(2\sinh^{-1}\left(\frac{\sinh(r/2)}{\cosh\delta}\right)\right).$$

Apply identity (1) on the first hyperbolic sine, one has

$$\begin{aligned}
\sinh\left(d_H(x,y)/2\right) &= \cosh\delta \cdot 2 \cdot \sinh\left(\sinh^{-1}\left(\frac{\sinh(r/2)}{\cosh\delta}\right)\right) \cdot \cosh\left(\sinh^{-1}\left(\frac{\sinh(r/2)}{\cosh\delta}\right)\right) \\
&= \cosh\delta \cdot 2 \cdot \left(\frac{\sinh(r/2)}{\cosh\delta}\right) \cdot \cosh\left(\sinh^{-1}\left(\frac{\sinh(r/2)}{\cosh\delta}\right)\right) \\
&= 2 \cdot \sinh(r/2) \cdot \left(1 + \left(\sinh\left(\sinh^{-1}\left(\frac{\sinh(r/2)}{\cosh\delta}\right)\right)\right)^2\right)^{1/2} \\
&= 2 \cdot \sinh(r/2) \cdot \left(1 + \left(\frac{\sinh(r/2)}{\cosh\delta}\right)^2\right)^{1/2}.
\end{aligned}$$

This shows that

$$d_H(x,y) = 2 \cdot \sinh^{-1}\left(2 \cdot \sinh(r/2) \cdot \left(1 + \left(\frac{\sinh(r/2)}{\cosh\delta}\right)^2\right)^{1/2}\right)$$

by identity (2). Taylor expand $d_H(x,y)$ around $\delta = 0$ gives us

$$d_H(x,y) = 2r - \frac{(\sinh(r/2))^3}{\cosh(r/2)\cdot\cosh(r)}\delta^2 + O(\delta^4),$$

and therefore $tort(\gamma, t) \geq \Omega(\delta^2) \geq \Omega(\varepsilon^2)$. □

Exposing points outside the neighborhood. Now we proceed to bound ε so that the ε-neighborhood of the geodesic γ_* does not cover the whole surface Σ.

Lemma 6.13. *Let γ be an n-vertex curve on Σ. Then the ε-neighborhood of γ_* does not cover the whole surface Σ if ε is at most $O(g/(n\log g))$.*

Proof: Given any curve γ with the corresponding unique (close) geodesic γ_* on surface Σ with the constructed hyperbolic metric d_H, the length of γ_* is at most $O(n\log g)$ by property (1). For small enough ε, the area of the ε-neighborhood of a curve with length ℓ is at most $O(\varepsilon\ell)$.[3] The area of the surface is precisely $(4g-4)\pi$. (This follows directly from Gauss-Bonnet theorem which is independent to the hyperbolic metric up to scaling.[4]) This implies that for the ε-neighborhood of γ_* to cover the whole surface Σ, the following holds:

$$\varepsilon \;\geq\; \frac{(4g-4)\pi}{O(n\log g)} \;\geq\; \Omega\left(\frac{g}{n\log g}\right)$$

In other words, if we set $\varepsilon \leq O(g/(n\log g))$, then the ε-neighborhood of γ_* cannot cover the whole surface Σ, thus proving the lemma. □

[3] To see this, cover the neighborhood with kite-like *Lambert quadrilaterals* with length of the short sides as ε. The only acute angle α of the quadrilateral is equal to $\arccos((\sinh\varepsilon)^2)$. The area of the quadrilateral is equal to the angle deficit, which is $\pi/2-\alpha$. In combine the area of the quadrilateral is at most $O(\varepsilon^2)$, and thus the total area of the ε-neighborhood on Σ is at most $O(\varepsilon^2\cdot\ell/\varepsilon) = O(\varepsilon\ell)$.

[4] Alternatively, one can derive the area directly: divide the fundamental polygon into $12g-6$ triangles by drawing straight-lines from the center of the polygon to all vertices, and use the area formula for triangles.

Basmajian, Parlier, and Souto [20] showed that for any fixed genus g, the $O(1/n)$ bound in Lemma 6.13 is tight up to logarithmic factors.

Putting it together. Now we are ready to prove Lemma 6.10. Consider the set of disks centered at each point on the curve with radius 1/2, which is smaller than half the systole by property (2); therefore all such disks are embedded in Σ. Straighten any disk using Lemma 6.11 if the tortuosity of the center point is at least ε^2. Once every point on γ has tortuosity less than ε^2, by Lemma 6.12 the curve γ now lies in the ε-neighborhood of γ_*.

Straighten a disk takes $O(n^2)$ moves using Lemma 6.11. The tortuosity at a center of each disk is a lower bound on the difference between the lengths of the curve γ before and after straightening. From property (1) of the hyperbolic metric d_H the length of γ is at most $O(n \log g)$. Every time a disk is straightened the length of the curve γ will drop by at least ε^2. Since γ is noncontractible, the length of any curve homotopic to γ is at least the systole, which is $\Omega(1)$ by property (2). Therefore at most $O(n \log g/\varepsilon^2)$ disks will be straighten before every point has tortuosity less than ε^2. In total at most $O(n^3 \log g/\varepsilon^2)$ homotopy moves are performed. From Lemma 6.13, setting $\varepsilon \coloneqq O(g/(n \log g))$ concludes the proof of Lemma 6.10.

Chapter 7

Electrical Transformations

I believe in love at first sight.
You want that connection, and then you want some problems.

— Keanu Reeves

In this section we explore the close relationship between electrical transformations for graphs and homotopy moves for curves on arbitrary surfaces. We start with some possible different definitions of electrical transformations performed on graphs with embeddings. Then we focus on the most restrictive version—the *facial* electrical transformations—and work with them at the level of medial multicurves. Quantitative connections between such transformations and the homotopy moves are the main focus of the rest of the section. We conclude our discussion of the connection with an application, by providing tight lower bounds on the number of electrical transformations required to reduce plane graphs with or without terminals using results we derived in previous chapters.

7.1 Types of Electrical Transformations

Electrical transformations defined on general graphs consist of the following set of local operations performed on any graph:

- *Leaf contraction*: Contract the edge incident to a vertex of degree 1.
- *Loop deletion*: Delete the edge of a loop.
- *Series reduction*: Contract either edge incident to a vertex of degree 2.
- *Parallel reduction*: Delete one of a pair of parallel edges.
- *$Y\to\Delta$ transformation*: Delete a vertex of degree 3 and connect its neighbors with three new edges.
- *$\Delta\to Y$ transformation*: Delete the edges of a 3-cycle and join the vertices of the cycle to a new vertex.

We distinguish between three increasingly general types of electrical transformations on graphs embedded on a surface: *facial*, *crossing-free*, and *arbitrary*.

An electrical transformation in a graph G embedded on a surface Σ is ***facial*** if any deleted cycle is a face of G. All leaf contractions, series reductions, and $Y\to\Delta$ transformations are facial, but loop deletions, parallel reductions, and $\Delta\to Y$ transformations may not be facial. As we have seen in the introduction and preliminaries (Sections 1.2 and 2.5.2), facial electrical transformations form three dual pairs, as shown in Figure 1.2; for example, any series reduction in G is equivalent to a parallel reduction in the dual graph G^*.

An electrical transformation in G is ***crossing-free*** if it preserves the embeddablility of the underlying graph into the same surface. Equivalently, an electrical transformation is crossing-free if the vertices of the cycle deleted by the transformation are all incident to a common face (in the given embedding) of G. All facial electrical transformations are trivially crossing-free, as are all loop deletions and parallel reductions. If the graph embeds in the plane then crossing-free electrical transformations are also called ***planar***. (For ease of presentation, we assume throughout this chapter that plane graphs are actually embedded on the *sphere* instead of the plane.) The only non-crossing-free electrical transformation is a $\Delta\to Y$ transformation whose three vertices are *not* incident to a common face; any such transformation introduces a $K_{3,3}$-minor into the graph, connecting the three vertices of the Δ to an interior vertex, an exterior vertex, and the new Y vertex.

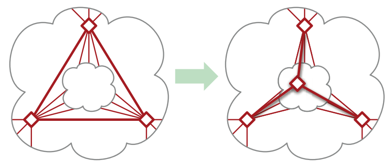


Figure 7.1. A non-planar $\Delta\to Y$ transformation.

7.2 Connection Between Electrical and Homotopy Moves

Recall that facial electrical transformations in any plane graph G correspond to local operations in the medial graph $G^\times$ known as the *medial electrical moves*; we refer them as ***electrical moves*** for short in this chapter (see Figure 1.3).

For any connected multicurve (or 4-regular graph) γ on surface Σ,

- let $\boldsymbol{X(\gamma)}$ denote the minimum number of electrical moves required to tighten γ on Σ,
- let $\boldsymbol{H^{\downarrow}(\gamma)}$ denote the minimum number of homotopy moves required to tighten γ on Σ, without ever increase the number of vertices. In other words, no $0\to1$ and $0\to2$ moves are allowed.
- let $\boldsymbol{H(\gamma)}$ denote the minimum number of homotopy moves required to tighten γ on Σ.

As we mentioned in Section 6.4, de Graaf and Schrijver [125] proved that any multicurve γ can be tightened using monotonic homotopy moves, which implies that $H^{\downarrow}(\gamma) = 0$ if and only if $H(\gamma) = 0$. In other words, standard homotopy moves and monotonic homotopy moves share the same set of target multicurves with minimum number of vertices. Now by definition one has $H^{\downarrow}(\gamma) \geq H(\gamma)$ for any multicurve γ on surface Σ.

Tightening curves using electrical moves is a more difficult problem than tightening curves using homotopy moves. Modulo some conjectures we will discuss shortly, in the following subsections we argue that the number of electrical moves required is polynomially-related to the number of *monotonic* homotopy moves required.

As initial evidence, both Steinitz's algorithm and Feo-Provan's algorithm can easily be adapted to simplify planar curves monotonically, simply by replacing each $2\to1$ move encountered with a $2\to0$ move and recursing. A subtlety here is that we do not know *a priori* whether tightening a multicurve using electrical moves will result in the same multicurve as tightening using homotopy moves (or whether the two tightened multicurves even have the same number of vertices). Notice that we don't have such a problem in the plane as all planar multicurves can be tightened to simple curves using either electrical or homotopy moves. One direction follows from de Graaf and Schrijver [125].

Lemma 7.1. *Let γ be a connected multicurve on an arbitrary surface Σ. If γ is electrically tight, then γ is homotopically tight.*

Proof: Let γ be a connected multicurve in some arbitrary surface, and suppose γ is not homotopically tight. Result of de Graaf and Schrijver [125] implies that γ can be tightened by a finite sequence of homotopy moves that never increases the number of vertices. In particular, applying some finite sequence of 3→3 moves to γ creates either an empty monogon, which can be removed by a 1→0 move, or an empty bigon, which can be removed by either a 2→0 move or a 2→1 move. Thus, γ is not e-tight. □

The main obstacle in showing the opposite direction is that we don't have a similar monotonicity result like de Graaf and Schrijver [125] for electrical moves on arbitrary surfaces. In Sections 7.2.2 and 7.2.3 the monotonicity results are established for both planar and annular multicurves, which implies that the two types of tightness are indeed equivalent for those multicurves. We conjecture that the same holds for arbitrary multicurve on any surface.

Conjecture 7.1. *Any multicurve on any surface Σ is electrically tight if and only if it is homotopically tight.*

Assume Conjecture 7.1 holds, we can formally compare the number of electrical moves to the number of homotopy moves required to tighten a multicurve. The following lemma demonstrates that monotonic homotopy moves are indeed closely related to electrical moves.

Lemma 7.2. *Assume Conjecture 7.1 holds. Fix an arbitrary surface Σ. Let $f(n)$ be a non-decreasing function. If $H^{\downarrow}(\gamma) \le f(n)$ holds for all multicurves γ on Σ with n vertices, then $X(\gamma) \le n \cdot f(n)$ also holds for all γ.*

Proof: Given a minimum-length sequence of monotonic homotopy moves that tightens γ. If $H^{\downarrow}(\gamma) = 0$, assuming Conjecture 7.1 one has $X(\gamma) = 0$ as well and thus the statement trivially holds. Otherwise, consider the first move in the sequence that decreases the number of vertices in γ (that is, either a 1→0 or 2→0 move). Replace the 2→0 move with a 2→1 if needed, one arrives at a curve γ' that has strictly less vertices than γ. The number of homotopy moves in the sequence from the original γ to γ' is at most $H^{\downarrow}(\gamma)$. Now by induction on the number of vertices,

$$
\begin{aligned}
X(\gamma) &\le X(\gamma') + H^{\downarrow}(\gamma) \\
&\le (n-1) \cdot H^{\downarrow}(\gamma') + H^{\downarrow}(\gamma) \\
&\le (n-1) \cdot f(n-1) + f(n) \\
&\le n \cdot f(n),
\end{aligned}
$$

which proves the lemma. □

After presenting all the necessary terminologies, in Section 7.2.4 we will introduce the *strong smoothing conjecture* (Conjecture 7.3) which implies both Conjecture 7.1 (and thus Lemma 7.2 without the assumption), and the opposite direction of the inequality between $H^{\downarrow}(\gamma)$ and $X(\gamma)$ (see Lemma 7.16). We discuss other consequences and partial attempts towards proving Conjecture 7.1 in the same section.

Before that, we provide evidence to the conjecture(s) in Section 7.2.2 and Section 7.2.3 by showing that for arbitrary planar and annular curves, both Conjecture 7.1 and the inequality $X(\gamma) + O(n) \ge H^{\downarrow}(\gamma)$ holds. This demonstrates that $X(\gamma)$ and $H^{\downarrow}(\gamma)$ are at most a linear factor away from each other for planar or annular curve γ.

7.2.1 Smoothing Lemma—Inductive case

The following key lemma follows from close reading of proofs by Truemper [242, Lemma 4] and several others [12, 115, 181, 184] that every minor of a ΔY-reducible graph is also ΔY-reducible. Our proof most closely resembles an argument of Gitler [115, Lemma 2.3.3], but restated in terms of electrical moves on multicurves to simplify the case analysis. In his PhD thesis [122, Proposition 5.1], de Graaf provided a proof to some special case of the lemma at the level of medial curves.

Lemma 7.3. *Let γ be any connected multicurve on surface Σ, and let $\check{\gamma}$ be a connected smoothing of γ. Applying any sequence of electrical moves to γ to obtain γ'; let x be the number of electrical moves in the sequence. Then one can apply a similar sequence of electrical moves of length at most x to $\check{\gamma}$ to obtain a (possibly trivial) connected smoothing $\check{\gamma}'$ of γ'.*

Proof: We prove the statement by induction on the number of electrical moves in the sequence and the number of smoothed vertices. If $\check{\gamma} = \gamma$ then the statement trivially holds. Otherwise, we first consider the special case where $\check{\gamma}$ is obtained from γ by smoothing a single vertex x. Without loss of generality let γ' be the result of the first electrical move. There are two nontrivial cases to consider.

First, suppose the move from γ to γ' does not involve the smoothed vertex x. Then we can apply the same move to $\check{\gamma}$ to obtain a new multicurve $\check{\gamma}'$; the same multicurve can also be obtained from γ' by smoothing x.

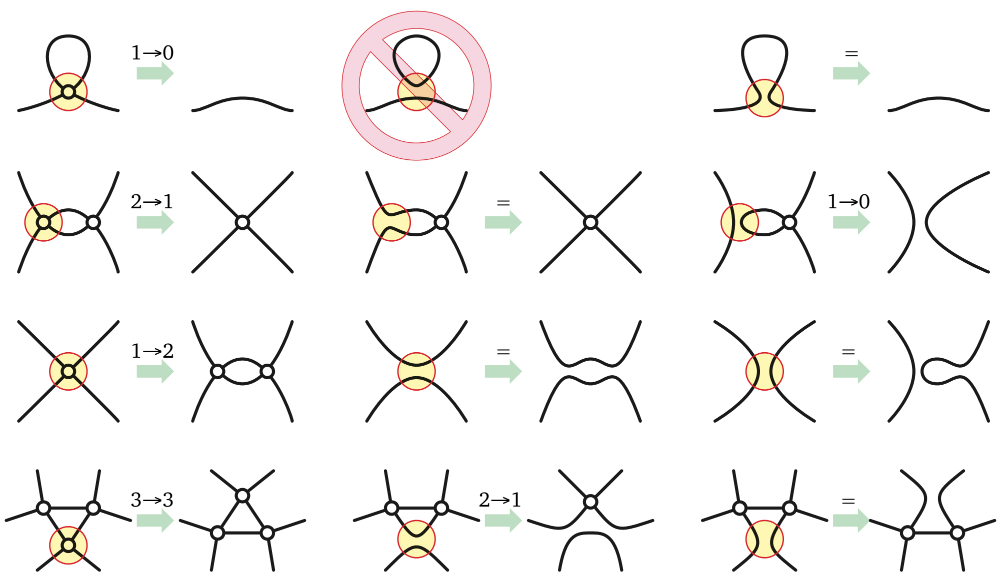


Figure 7.2. Cases for the proof of the Lemma 7.3; the circled vertex is x.

Now suppose the first move does involve x. In this case, we can apply at most one electrical move to $\check{\gamma}$ to obtain a (possibly trivial) smoothing $\check{\gamma}'$ of γ'. There are eight subcases to consider, shown in Figure 7.2. One subcase for the 0→1 move is impossible, because $\check{\gamma}$ is connected. In the remaining 0→1 subcase and one 2→1 subcase, the curves $\check{\gamma}$, $\check{\gamma}'$, and γ' are all isomorphic. In all remaining subcases, $\check{\gamma}'$ is a connected proper smoothing of γ'.

Finally, we consider the more general case where $\check{\gamma}$ is obtained from γ by smoothing more than one vertex. Let $\tilde{\gamma}$ be any intermediate curve, obtained from γ by smoothing just one of the vertices that were smoothed to obtain $\check{\gamma}$. As $\check{\gamma}$ is a connected smoothing of $\tilde{\gamma}$, the curve $\tilde{\gamma}$ itself must be connected too. Our earlier argument implies that there is a sequence of electrical moves that changes $\tilde{\gamma}$ to a smoothing $\tilde{\gamma}'$ of γ'. The inductive hypothesis implies that there is a sequence of electrical moves that changes $\check{\gamma}$ to a smoothing $\check{\gamma}'$ of $\tilde{\gamma}'$, which is itself a smoothing of γ'. This completes the proof. □

As a remark, using a similar argument one can recover a result by Newmann-Coto [182]: any homotopy from multicurve γ to another multicurve γ' that never removes vertices can be turned into a homotopy from a smoothing of γ to a smoothing of γ'. Chambers and Liokumovich [40] studied a similar problem where one wants to convert a homotopy between two *simple* curves on surface into an *isotopy*, without increasing the length of any intermediate curve by too much. They showed that the desired isotopy can be obtained from a clever Euler-tour argument on the graph of all possible complete smoothings of the intermediate curves.

7.2.2 In the Plane

The main result of this subsection is that the number of *homotopy* moves required to simplify a closed curve in the plane is a lower bound on the number of *electrical moves* required to simplify the same closed curve. Our result makes explicit the quantitative bound implicit in the work of Noble and Welsh [184], and most of our proofs closely follow theirs.

We also establish two other results on the fly—the function $X(\cdot)$ never increases under smoothings, and the monotonicity of electrical moves—which are interesting in their own right. The fact that every planar curve can be simplified using either electrical or homotopy moves makes the proofs in this subsection slightly easier comparing to the annular case (see Section 7.2.3).

Lemma 7.4. *$X(\check{\gamma}) \le X(\gamma)$ for every connected smoothing $\check{\gamma}$ of every connected multicurve γ in the plane.*

Proof: Let γ be a connected multicurve, and let $\check{\gamma}$ be a connected smoothing of γ. If γ is already simple, the lemma is vacuously true. Otherwise, applying a minimum-length sequence of electrical moves that simplifies γ. By Lemma 7.3 there is another sequence of electrical moves of length at most $X(\gamma)$ that simplifies $\check{\gamma}$. We immediately have $X(\check{\gamma}) \le X(\gamma)$ and the lemma is proved. □

Lemma 7.5. *For every connected multicurve γ, there is a minimum-length sequence of electrical moves that simplifies γ to a simple closed curve that does not contain $0{\to}1$ or $1{\to}2$ moves.*

Proof: Consider a minimum-length sequence of electrical moves that simplifies an arbitrary connected multicurve γ to a simple closed curve. For any integer $i \ge 0$, let γ_i denote the result of the first i moves in this sequence; in particular, $\gamma_0 = \gamma$ and $\gamma_{X(\gamma)}$ is a simple closed curve. Minimality of the simplification sequence implies that $X(\gamma_i) = X(\gamma) - i$ for all i; in particular, $X(\gamma_i)$ decreases as i grows. Now let i be an arbitrary index such that γ_i has one more vertex than γ_{i-1}. Then γ_{i-1} is a connected proper smoothing of γ_i, so Lemma 7.4 implies that $X(\gamma_{i-1}) \le X(\gamma_i)$, giving us a contradiction. □

Lemma 7.6. *$X(\gamma) \ge H^{\downarrow}(\gamma) \ge H(\gamma)$ for every closed curve γ in the plane.*

Proof: The second inequality is straightforward. The proof of the first inequality proceeds by induction on $X(\gamma)$.

Let γ be a closed curve. If $X(\gamma) = 0$, then γ is already simple, so $H^{\downarrow}(\gamma) = 0$. Otherwise, consider a minimum-length sequence of electrical moves that simplifies γ to a simple closed curve. Lemma 7.5 implies that we can assume that the first move in the sequence is neither $0{\to}1$ nor $1{\to}2$. If the first move is $1{\to}0$ or $3{\to}3$, the theorem immediately follows by induction.

The only interesting first move is $2{\to}1$. Let γ' be the result of this $2{\to}1$ move, and let γ° be the result of the corresponding $2{\to}0$ move. The minimality of the sequence implies that $X(\gamma) = X(\gamma') + 1$, and we trivially have $H^{\downarrow}(\gamma) \le H^{\downarrow}(\gamma^{\circ}) + 1$. Because γ consists of *one* single curve, γ° is also a single curve and is therefore connected. The curve γ° is also a proper smoothing of γ', so the Lemma 7.4 implies $X(\gamma^\circ) \le X(\gamma') < X(\gamma)$. Finally, the inductive hypothesis implies that $X(\gamma^\circ) \ge H^\downarrow(\gamma^\circ)$, and therefore

$$H^\downarrow(\gamma) - 1 \le H^\downarrow(\gamma^\circ) \le X(\gamma^\circ) \le X(\gamma') = X(\gamma) - 1$$

which completes the proof. □

7.2.3 In the Annulus

Tight curves on the annulus. To prove similar results in the annulus, first we have to prove Conjecture 7.1 for annular multicurves. Recall that the *depth* of any annular multicurve γ is the minimum number of times a path from one boundary to the other crosses γ. In many ways, depth can be viewed an unsigned version of winding number. Just as the winding number around the boundaries is a complete homotopy invariant for curves in the annulus, the depth turns out to be a complete invariant for electrical moves on the annular multicurve.

Lemma 7.7. *Electrical moves do not change the depth of any connected multicurve in the annulus.*

Proof: Let γ be a connected multicurve in the annulus. For any face of γ that could be deleted by a electrical move, exhaustive case analysis implies that there is a shortest path in the dual of γ between the two boundary faces of γ that avoids that face. □

For any integer $d > 0$, let $\boldsymbol{\alpha_d}$ denote the unique closed curve in the annulus with $d-1$ vertices and winding number d. Up to isotopy, this curve can be parametrized in the plane as

$$\alpha_d(\theta) \;:=\; \big((\cos(\theta)+2)\cos(d\theta),\; (\cos(\theta)+2)\sin(d\theta)\big).$$

In the notation of Section 3.1.1, α_d is the *flat torus knot* $T(d, 1)$.

Lemma 7.8. *For any integer $d > 0$, the curve α_d is both h-tight and e-tight.*

Proof: Every connected multicurve in the annulus with either winding number d or depth d has at least $d+1$ faces (including the faces containing the boundaries of the annulus) and therefore, by Euler's formula, has at least $d-1$ vertices. □

Lemma 7.9. *If γ is an h-tight connected multicurve in the annulus, then $\gamma = \alpha_d$ for some integer d.*

Proof: A multicurve in the annulus is h-tight if and only if its constituent curves are h-tight *and disjoint*. Thus, any *connected* h-tight multicurve is actually a single closed curve. Any two curves in the annulus with the same winding number are homotopic [142]. Finally, up to isotopy, α_d is the only closed curve in the annulus with winding number d and $d-1$ vertices [135, Lemma 1.12]. □

The following corollaries are now immediate by Lemma 7.1.

Corollary 7.1. *A connected multicurve γ in the annulus is e-tight if and only if $\gamma = \alpha_{depth(\gamma)}$; therefore, any multicurve γ is e-tight if and only if γ is h-tight.*

Corollary 7.2. *Let γ and γ' be two connected multicurves in the annulus. Then γ can be transformed into γ' by electrical moves if and only if $depth(\gamma) = depth(\gamma')$.*

Equipped with the understanding of tight annular curves, we are ready to extend the results in Section 7.2.2 to the annulus.

Lemma 7.10. *For any connected smoothing $\check{\gamma}$ of any connected multicurve γ in the annulus, we have $X(\check{\gamma}) + \frac{1}{2}\,depth(\check{\gamma}) \le X(\gamma) + \frac{1}{2}\,depth(\gamma)$.*

Proof: Let γ be an arbitrary connected multicurve in the annulus, and let $\check{\gamma}$ be an arbitrary connected smoothing of γ. Without loss of generality, we can assume that γ is non-simple, since otherwise the lemma is vacuous.

If γ is already e-tight, then $\gamma = \alpha_d$ for some integer $d \ge 2$ by Corollary 7.1. (The curves α_0 and α_1 are simple.) First, suppose $\check{\gamma}$ is a connected smoothing of γ obtained by smoothing a single vertex x. The smoothed curve $\check{\gamma}$ contains a single monogon if x is the innermost or outermost vertex of γ, or a single bigon otherwise. Applying one 1→0 or 2→0 move transforms $\check{\gamma}$ into the curve α_{d-2}, which is e-tight by Lemma 7.8. Thus we have $X(\check{\gamma}) = 1$ and $depth(\check{\gamma}) = d - 2$, which implies $X(\check{\gamma}) + \frac{1}{2}\,depth(\check{\gamma}) = X(\gamma) + \frac{1}{2}\,depth(\gamma)$. As for the general case when $\check{\gamma}$ is obtained from γ by smoothing more than one vertices, the statement follows from the previous case by induction on the number of smoothed vertices.

If γ is not e-tight, applying a minimum-length sequence of electrical moves that tightens γ into some curve γ'. By Lemma 7.3 there is another sequence of electrical moves of length at most $X(\gamma)$ that tightens $\check{\gamma}$ to some connected smoothing $\check{\gamma}'$ of γ', which can be further tightened electrically to an e-tight curve using arguments in the previous paragraph because γ' is e-tight. This implies that $X(\check{\gamma}) \le X(\gamma) + \frac{1}{2}(depth(\gamma') - depth(\check{\gamma}'))$. By Lemma 7.7, γ and γ' have the same depth, and $\check{\gamma}$ and $\check{\gamma}'$ have the same depth. Therefore $X(\check{\gamma}) + \frac{1}{2}\,depth(\check{\gamma}) \le X(\gamma) + \frac{1}{2}\,depth(\gamma)$ and the lemma is proved. □

Lemma 7.11. *For every connected multicurve γ in the annulus, there is a minimum-length sequence of electrical moves that tightens γ to $\alpha_{depth(\gamma)}$ without 0→1 or 1→2 moves.*

Proof: Consider a minimum-length sequence of electrical moves that tightens an arbitrary connected multicurve γ in the annulus. For any integer $i \ge 0$, let γ_i denote the result of the first i moves in this sequence. Suppose γ_i has one more vertex than γ_{i-1} for some index i. Then γ_{i-1} is a connected proper smoothing of γ_i, and $depth(\gamma_i) = depth(\gamma_{i-1})$ by Lemma 7.7; so Lemma 7.10 implies that $X(\gamma_{i-1}) \le X(\gamma_i)$, contradicting our assumption that the reduction sequence has minimum length. □

Lemma 7.12. *$X(\gamma) + \frac{1}{2}\,depth(\gamma) \ge H^{\downarrow}(\gamma) \ge H(\gamma)$ for every closed curve γ in the annulus.*

Proof: Again the second inequality is straightforward, as explained at the start of the section. Let γ be a closed curve in the annulus. If γ is already e-tight, then $X(\gamma) = H^{\downarrow}(\gamma) = 0$ by Lemma 7.1, so the lemma is trivial. Otherwise, consider a minimum-length sequence of electrical moves that tightens γ. By Lemma 7.11, we can assume that the first move in the sequence is neither 0→1 nor 1→2. If the first move is 1→0 or 3→3, the theorem immediately follows by induction on $X(\gamma)$, since by Lemma 7.7 neither of these moves changes the depth of the curve.

The only interesting first move is 2→1. Let γ' be the result of this 2→1 move, and let γ° be the result if we perform the 2→0 move on the same empty bigon instead. The minimality of the sequence implies $X(\gamma) = X(\gamma') + 1$, and we trivially have $H^{\downarrow}(\gamma) \le H^{\downarrow}(\gamma^{\circ}) + 1$. Because γ is a single curve, γ° is also a single curve and therefore a connected proper smoothing of γ'. Thus, Lemma 7.7, Lemma 7.10, and induction on the number of vertices imply

$$X(\gamma) + \frac{1}{2}\,depth(\gamma) \;=\; X(\gamma') + \frac{1}{2}\,depth(\gamma') + 1$$

$$\begin{aligned} &\geq X(\gamma^\circ) + \frac{1}{2}\,depth(\gamma^\circ) + 1 \\ &\geq H^{\downarrow}(\gamma^\circ) + 1 \\ &\geq H^{\downarrow}(\gamma), \end{aligned}$$

which completes the proof. □

7.2.4 Towards Connection between Electrical and Monotonic Homotopy Moves

In this subsection we discuss some attempts to establish a formal connection between electrical and monotonic homotopy moves. In particular, we formulate two versions of *smoothing conjecture* that imply both Conjecture 7.1 and the relation between functions X and $H^{\downarrow}$.

A closed curve γ is ***primitive*** if γ is not homotopic to a proper multiple of some other closed curve. A multicurve is *primitive* if all its constituent curves are primitive. We show equivalence between the following concepts on primitive multicurves. Let γ be a multicurve on an orientable surface Σ such that each constituent curve of γ is primitive. Define the ***μ-function*** as

$$\boldsymbol{\mu(\gamma, \sigma)} \;\coloneqq\; \min_{\substack{\sigma' \sim \sigma \\ \sigma' \pitchfork \gamma}} \operatorname{cr}(\gamma, \sigma'),$$

where $\operatorname{cr}(\gamma, \sigma')$ is the number of crossing between γ and σ', and the minimum is ranging over all closed curve σ' homotopic to the given closed curve σ on Σ, intersecting γ transversely.[1] Denote $\boldsymbol{\mu_\gamma}$ as the single-variable function $\mu(\gamma, \cdot)$. The notion of μ-function is deeply related to the *representativity* or *facewidth* of a graph studied in topological graph theory [205, 208, 235]. The μ-function is invariant under electrical moves and isotopy of γ.

The μ-function is a higher-genus analogue to the *depth* function defined in the annulus. The following result that μ is invariant under electrical moves can be found in Robertson and Vitray [208]; we sketch a proof for sake of completeness.

Lemma 7.13 (Robertson and Vitray [208, Proposition 14.4]). *Electrical moves do not change μ_γ for any multicurve γ on surface Σ.*

Proof: For any face of γ intersected by some closed curve σ that could be deleted after an electrical move, exhaustive case analysis implies that there is another closed curve σ' that avoids that face. □

Multicurve γ satisfies ***simplicity conditions*** [217] if (1) any lifting of γ_i in the universal cover $\hat{\Sigma}$ does not self-intersect for any constituent curve γ_i of γ, and (2) any distinct liftings of γ_i and γ_j in $\hat{\Sigma}$ intersect each other at most once for any pair of (possibly identical) constituent curves γ_i and γ_j of γ. Multicurve γ is ***minimally crossing*** [217, 219] if each constituent curve of γ has minimum number of self-intersections in its homotopy class, and every pair of constituent curves has minimum intersections with each other, in their own homotopy classes. In notation, one has

$$\operatorname{cr}(\gamma_i) \;=\; \min_{\gamma_i' \sim \gamma_i} \operatorname{cr}(\gamma_i') \qquad \text{and} \qquad \operatorname{cr}(\gamma_i, \gamma_j) \;=\; \min_{\substack{\gamma_i' \sim \gamma_i \\ \gamma_j' \sim \gamma_j}} \operatorname{cr}(\gamma_i', \gamma_j')$$

for all constituent curves γ_i and γ_j of γ; $\operatorname{cr}(\gamma_i)$ denotes the number of self-intersections of curve γ_i. Multicurve γ is ***crossing-tight*** [217, 219] if $\mu_\gamma \neq \mu_{\check{\gamma}}$ for any proper smoothing $\check{\gamma}$ of γ.

[1] In Schrijver [219], the μ-function is defined with respect to the graph corresponding to γ through medial construction; the function defined here is denoted as μ' in his paper.

Our proof of equivalence relies on machineries developed extensively in the sequence of work by de Graaf and Schrijver [123, 124, 125, 216, 217, 218, 219] who did all the weight-lifting. However the original work does not address the problem of relating electrical and homotopy moves.

Theorem 7.1. *Let γ be a multicurve on an orientable surface whose constituent curves are all primitive. The following statements are equivalent: (1) Multicurve γ satisfies simplicity conditions, (2) γ is minimally crossing, (3) γ is crossing-tight, (4) γ is e-tight, and (5) γ is h-tight.*

Proof (sketch): (1) $\Longleftrightarrow$ (2) $\Longleftrightarrow$ (3): Schrijver [217, Proposition 12] showed that γ satisfies simplicity conditions if and only if γ is minimally crossing and each constituent curve is primitive. Later in the same paper [217, Theorem 5] he also showed that γ is minimally crossing and each constituent curve is primitive if and only if γ is crossing-tight. An alternative proof using the monotonicity of homotopy process can be found in de Graaf's thesis [122].

(3) $\Rightarrow$ (4): In another paper Schrijver [219, Theorem 2] showed that two crossing-tight multicurves γ and γ' can be transformed into each other using only 3→3 moves if (and only if) $\mu_\gamma = \mu_{\check{\gamma}}$. This result implies that if multicurve γ is crossing-tight then γ is e-tight, as electrical moves preserves the μ-function by Lemma 7.13.

(4) $\Rightarrow$ (5): Any e-tight multicurve must be h-tight by de Graaf and Schrijver [125] (see Lemma 7.1).

(5) $\Rightarrow$ (3): If γ is h-tight and primitive, then by Hass and Scott [135, Lemma 3.4] multicurve γ satisfies simplicity conditions. To elaborate, assume for contradiction that γ violates the simplicity conditions. As γ is h-tight one can push each constituent curve of γ close to its unique geodesic on the surface without even decreases the number of vertices, similar to the algorithm of de Graaf and Schrijver [125]. Therefore all the intersections between lifts of constituent curves of γ remains after the push. The primitiveness of the curve γ guarantees that each lift of any constituent curve does not self-intersect, and two different lifts of the same constituent curve intersects at most once on $\hat{\Sigma}$. Between the lifts of two distinct geodesics there is at most one intersection in the universal cover, and thus the same holds for the lifts of two distinct constituent curves of γ.

This concludes the proof. □

Unfortunately Theorem 7.1 does not imply immediately a relation between number of electrical versus homotopy moves required to tighten a multicurve on surface, because primitive multicurves can have non-primitive smoothings. Still, one would hope that some forms of the smoothing lemma hold on general orientable surface, possibly with assumptions on the applicable smoothings.

Conjecture 7.2. *Let γ be any connected multicurve on surface Σ, and let $\check{\gamma}$ be a connected smoothing of γ, satisfying $\mu_{\check{\gamma}} = \mu_\gamma$. Then $X(\check{\gamma}) \le X(\gamma)$ holds.*

Lemma 7.14. *Assume Conjecture 7.2 holds. For every connected multicurve γ, there is a minimum-length sequence of electrical moves that tightens γ and does not contain* 0→1 *or* 1→2 *moves.*

Proof: Consider a minimum-length sequence of electrical moves that reduces an arbitrary connected multicurve γ to a simple closed curve. For any integer $i \ge 0$, let γ_i denote the result of the first i moves in this sequence; in particular, $\gamma_0 = \gamma$ and $\gamma_{X(\gamma)}$ is a simple closed curve. Minimality of the reduction sequence implies that $X(\gamma_i) = X(\gamma) - i$ for all i; in particular, $X(\gamma_i)$ strictly decreases as i increases. Now let i be an arbitrary index such that γ_i has one more vertex than γ_{i-1} after applying either a 0→1 or 1→2 move. Then γ_{i-1} is a connected proper smoothing of γ_i satisfying $\mu_{\gamma_i} = \mu_{\gamma_{i-1}}$; so Lemma 7.13 and Conjecture 7.2 imply that $X(\gamma_{i-1}) \le X(\gamma_i)$, giving us a contradiction. □

Using Lemma 7.14, we can show that the two different notions of tightness are indeed equivalent, thus proving Conjecture 7.1.

Lemma 7.15. *Conjecture 7.2 implies Conjecture 7.1.*

Proof: The only if direction follows directly from Lemma 7.1. Conversely, suppose γ is not e-tight. Lemma 7.14 implies that γ can be tightened by a finite sequence of electrical moves that never increases the number of vertices. In particular, some finite sequence of 3→3 moves to γ reveals either an empty monogon or an empty bigon. Thus, γ is not h-tight. □

Strong smoothing conjecture. We don't have the result corresponding to Lemma 7.6 in general surfaces, because that requires us to prove the following stronger version of the smoothing lemma.

Conjecture 7.3. *Let γ be any connected multicurve on surface Σ, and let $\check{\gamma}$ be a connected smoothing of γ. Then*

$$X(\check{\gamma}) + C \cdot \sum_{\sigma \in \Gamma_0} \mu_{\check{\gamma}}(\sigma) \leq X(\gamma) + C \cdot \sum_{\sigma \in \Gamma_0} \mu_{\gamma}(\sigma),$$

for some absolute constant C, where Γ_0 is some finite collection of simple curves on surface Σ.

It is immediate that Conjecture 7.3 implies Conjecture 7.2. Using the strong smoothing conjecture we can prove the analogous result to Lemma 7.6.

Lemma 7.16. *Assume Conjecture 7.3 holds, then $X(\gamma) + C \cdot \sum_{\sigma \in \Gamma_0} \mu_{\gamma}(\sigma) \geq H^{\downarrow}(\gamma) \geq H(\gamma)$ for any closed curve γ.*

Proof: The second inequality is straightforward, as explained in the start of the section. Let γ be a closed curve. If γ is e-tight, then γ is h-tight as well by Lemma 7.1 so the inequality trivially holds. Otherwise, consider a minimum-length sequence of electrical moves that tightens γ. Conjecture 7.3 implies Conjecture 7.2, so by Lemma 7.14 we can assume that the first move in the sequence is neither 0→1 nor 1→2. If the first move is 1→0 or 3→3, the theorem immediately follows by induction.

The only interesting first move is 2→1. Let γ' be the result of this 2→1 move, and let γ° be the result of the corresponding 2→0 homotopy move. The minimality of the sequence implies that $X(\gamma) = X(\gamma') + 1$, and we trivially have $H(\gamma) \leq H(\gamma^{\circ}) + 1$. Because γ consists of *one* single curve, γ° is also a single curve and is therefore connected. The curve γ° is also a proper smoothing of γ'. Thus, Lemma 7.13, Conjecture 7.3, and induction on number of vertices imply

$$\begin{aligned}
X(\gamma) + C \cdot \sum_{\sigma \in \Gamma_0} \mu_{\gamma}(\sigma) &= X(\gamma') + C \cdot \sum_{\sigma \in \Gamma_0} \mu_{\gamma'}(\sigma) + 1 \\
&\geq X(\gamma^{\circ}) + C \cdot \sum_{\sigma \in \Gamma_0} \mu_{\gamma^{\circ}}(\sigma) + 1 \\
&\geq H(\gamma^{\circ}) + 1 \\
&\geq H(\gamma),
\end{aligned}$$

which completes the proof. □

7.3 Lower Bounds on Electrical Transformations

7.3.1 Plane Graphs

Lemma 7.4 immediately implies the following corollary through the medial graph construction; we state the corollary explicitly as it generalizes the result of Truemper's [242] that any minor of a ΔY-reducible plane graph is also ΔY-reducible.

Corollary 7.3. *For any connected plane graph G, reducing any connected proper minor of G to a single vertex requires strictly fewer facial electrical transformations than reducing G to a single vertex.*

Recall a plane graph G is *unicursal* if its medial graph $G^\times$ is the image of a single closed curve.

Theorem 7.2. *For every connected plane graph G and every unicursal minor H of G, reducing G to a single vertex requires at least $|\mathit{defect}(H^\times)|/2$ facial electrical transformations.*

Proof: Either H equals G, or Corollary 7.3 states that reducing a proper minor H of G to a single vertex requires strictly fewer facial electrical transformations than reducing G to a single vertex. Note that facial electrical transformations performed on H corresponds precisely to electrical moves performed on $H^\times$. Now because $\gamma := H^\times$ is unicursal, Lemma 4.1 and Lemma 7.6 implies that $X(\gamma) \geq H(\gamma) \geq |\mathit{defect}(\gamma)|/2$. □

We can also derive explicit lower bounds for the number of facial electrical transformations required to reduce any plane graph of treewidth t to a single vertex. For any positive integers p and q, we define two cylindrical grid graphs; see Figure 7.3.

- $\boldsymbol{C(p,q)}$ is the Cartesian product of a cycle of length q and a path of length $p-1$. If q is odd, then the medial graph of $C(p,q)$ is the flat torus knot $T(2p,q)$.
- $\boldsymbol{C'(p,q)}$ is obtained by connecting a new vertex to the vertices of one of the q-gonal faces of $C(p,q)$, or equivalently, by contracting one of the q-gonal faces of $C(p+1,q)$ to a single vertex. If q is even, then the medial graph of $C'(p,q)$ is the flat torus knot $T(2p+1,q)$.

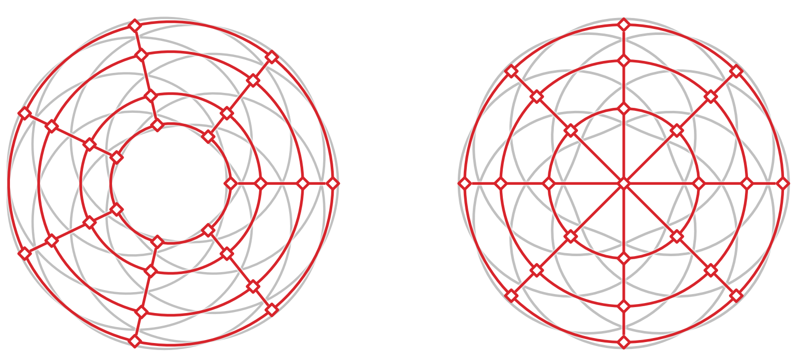


Figure 7.3. The cylindrical grid graphs $C(4,7)$ and $C'(3,8)$ and (in light gray) their medial graphs $T(8,7)$ and $T(7,8)$.

Corollary 7.4. *For all positive integers p and q, the cylindrical grid $C(p,q)$ requires $\Omega(\min\{p^2q, pq^2\})$ facial electrical transformations to reduce to a single vertex.*

Proof: First suppose $p \le q$. Because $C(p-1,q)$ is a minor of $C(p,q)$, we can assume without loss of generality that p is even and $p < q$. Let H denote the cylindrical grid $C(p/2, ap+1)$, where $a := \lfloor (q-1)/p \rfloor \ge 1$. H is a minor of $C(p,q)$ (because $ap+1 \le q$), and the medial graph of H is the flat torus knot $T(p, ap+1)$. Lemma 3.1 implies

$$\mathit{defect}\Big(T(p, ap+1)\Big) = 2a\binom{p+1}{3} = \Omega(ap^3) = \Omega(p^2 q).$$

Theorem 7.2 now implies that reducing $C(p,q)$ requires at least $\Omega(p^2 q)$ facial electrical transformations.

The symmetric case $p > q$ is similar. We can assume without loss of generality that q is odd. Let H denote the cylindrical grid $C'(aq, q)$, where $a := \lfloor (p-1)/q \rfloor \ge 1$. H is a proper minor of $C(p,q)$ (because $aq < p$), and the medial graph of H is the flat torus knot $T(2aq+1, q)$. Corollary 3.1 implies

$$\Big|\mathit{defect}\Big(T(2aq+1, q)\Big)\Big| = 4a\binom{q}{3} = \Omega(aq^3) = \Omega(pq^2).$$

Theorem 7.2 now implies that reducing $C(p,q)$ requires at least $\Omega(pq^2)$ facial electrical transformations. □

In particular, reducing any $\Theta(\sqrt{n}) \times \Theta(\sqrt{n})$ cylindrical grid requires at least $\Omega(n^{3/2})$ facial electrical transformations. Our lower bound matches an $O(\min\{pq^2, p^2 q\})$ upper bound by Nakahara and Takahashi [181]. Because every $p \times q$ rectangular grid contains $C(\lfloor p/3 \rfloor, \lfloor q/3 \rfloor)$ as a minor, the same $\Omega(\min\{p^2 q, pq^2\})$ lower bound applies to rectangular grids. In particular, Truemper's $O(p^3) = O(n^{3/2})$ upper bound for the $p \times p$ square grid [242] is tight. Finally, because every plane graph with treewidth t contains an $\Omega(t) \times \Omega(t)$ grid minor [207], reducing *any* n-vertex plane graph with treewidth t requires at least $\Omega(t^3 + n)$ facial electrical transformations. Therefore, our result answers the question by Gitler [115] and Archdeacon *et al.* [12] negatively.

An interesting open question is to determine the asymptotically bound to reduce any plane graph of treewidth t. We ambitiously conjecture that the correct answer is in fact $\Theta(nt)$. Of course proving this conjecture would be hard because it implies the Feo-Provan conjecture that any plane graph can be reduced using $O(n^{3/2})$ facial electrical transformations. However even a tight lower bound seems to be non-trivial as there are n-vertex planar graphs of treewidth t that do not contain any $\Omega(n) \times \Omega(t)$ grid minors.

Conjecture 7.4. *Any n-vertex plane graph of treewidth t can be reduced to a single vertex using $O(nt)$ facial electrical transformations, and the bound is tight in the worst case.*

7.3.2 Two-Terminal Plane Graphs

Most applications of electrical reductions, starting with Kennelly's classical computation of effective resistance [155], designate two vertices of the input graph as *terminals* and require a reduction to a single edge between those terminals. In this context, electrical transformations that delete either of the terminals are forbidden: specifically, leaf contractions when the leaf is a terminal, series reductions when the degree-2 vertex is a terminal, and $Y{\to}\Delta$ transformations when the degree-3 vertex is a terminal.

Epifanov [85] was the first to prove that any 2-terminal planar graph can be reduced to a single edge between the terminals using a finite number of electrical transformations, roughly 50 years after Steinitz proved the corresponding result for planar graphs without terminals [230, 231]. Epifanov's proof is non-constructive; algorithms for reducing 2-terminal planar graphs were later described by Feo [99], Truemper [242], and Feo and Provan [100]. (An algorithm in the spirit of Steinitz's reduction proof can also be derived from results of de Graaf and Schrijver [125].)

An important subtlety that complicates both Epifanov's proof and its algorithmic descendants is that not every 2-terminal planar graph can be reduced to a single edge using only *facial* electrical transformations. The simplest bad example is the three-vertex graph shown in Figure 7.4; the solid vertices are the terminals. Although this graph has more than one edge, it has no reducible leaves, empty loops, cycles of length 2 or 3, or vertices with degree 2 or 3. We will soon see that this graph cannot be reduced to an edge even if we allow "backward" facial electrical transformations that make the graph more complicated.

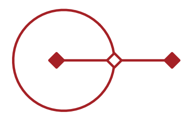


Figure 7.4. A facially irreducible 2-terminal plane graph.

Existing algorithms for reducing an arbitrary 2-terminal plane graphs to a single edge rely on an additional operation which we call a *terminal-leaf contraction*, in addition to facial electrical transformations. We discuss this subtlety in more detail in Section 7.3.4.

Bullseyes. The graph in Figure 7.4 is just one example of an infinite family of irreducible 2-terminal plane graphs. For any $k > 0$, let $\boldsymbol{B_k}$ denote the 2-terminal plane graph that consists of a path of length k between the terminals, with a loop attached to each of the $k-1$ interior vertices, embedded so that collectively they form concentric circles that separate the terminals. We call each graph B_k a ***bullseye***. For example, B_1 is just a single edge; B_2 is shown in Figure 7.4; and B_4 is shown on the left in Figure 7.5. The medial graph $B_k^\times$ of the kth bullseye is the curve α_{2k}, as we have seen in Section 7.2.3. Because different bullseyes have different medial depths, Lemma 7.7 implies that no bullseye can be transformed into any other bullseye by facial electrical transformations.

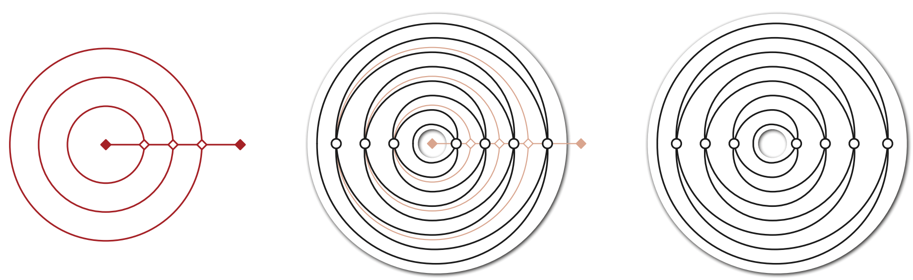


Figure 7.5. The bullseye graph B_4 and its medial graph α_8.

The following corollaries are now immediate from results in Section 7.2.3.

Corollary 7.5. *Let G be an arbitrary 2-terminal plane graph. Graph G can be reduced to the bullseye B_k using a finite sequence of facial electrical transformations if and only if $depth(G^\times) = 2k$.*

Corollary 7.6. *Let G and H be arbitrary 2-terminal plane graphs. Graph G can be transformed to graph H using a finite sequence of facial electrical transformations if and only if $depth(G^\times) = depth(H^\times)$.*

Theorem 7.3. *Let G be an arbitrary 2-terminal plane graph, and let γ be any unicursal smoothing of $G^\times$. Reducing G to a bullseye requires at least $H(\gamma) - \frac{1}{2}\,depth(\gamma)$ facial electrical transformations.*

In Section 4.3.2, we describe an infinite family of contractible curves in the annulus that require $\Omega(n^2)$ homotopy moves to simplify. Because these curves are contractible, they have even depth, and thus are the medial graphs of

2-terminal plane graphs. Euler's formula implies that every n-vertex curve in the annulus has exactly $n+2$ faces (including the boundary faces) and therefore has depth at most $n+1$.

Corollary 7.7. *Reducing a 2-terminal plane graph to a bullseye requires $\Omega(n^2)$ facial electrical transformations in the worst case.*

7.3.3 Planar Electrical Transformations

We extend our earlier $\Omega(n^{3/2})$ lower bound in Section 7.3.1 for reducing plane graphs *without* terminals using only facial electrical transformations to the larger class of planar electrical transformations. Our extension to non-facial electrical transformations is based on the following surprising observation, shown in Section 3.3: Although the medial graph of G depends on its embedding, the *defect* of the medial graph of G does not.

Each planar electrical transformation in a planar graph G induces the same change in the medial graph $G^\times$ as a finite sequence of 1- and 2-strand tangle flips (hereafter simply called "tangle flips") followed by a single electrical move. (See Section 3.3.2 for the definition of tangle flips.) For an arbitrary connected multicurve γ on the sphere, let $\boldsymbol{\bar{X}(\gamma)}$ denote the minimum number of electrical moves in a mixed sequence of electrical moves and tangle flips that simplifies γ. Similarly, let $\boldsymbol{\bar{H}(\gamma)}$ denote the minimum number of homotopy moves in a mixed sequence of homotopy moves and tangle flips that simplifies γ. We emphasize that tangle flips are "free" and do not contribute to either $\bar{X}(\gamma)$ or $\bar{H}(\gamma)$.

Our lower bound on planar electrical moves follows our earlier lower bound proof for facial electrical moves almost verbatim; the only subtlety is that the embedding of the graph can effectively change at every step of the reduction. We repeat the arguments here to keep the presentation self-contained.

Lemma 7.17. *$\bar{X}(\check{\gamma}) \le \bar{X}(\gamma)$ for every connected proper smoothing $\check{\gamma}$ of every connected multicurve γ on the sphere.*

Proof: Let γ be a connected multicurve, and let $\check{\gamma}$ be a connected proper smoothing of γ. The proof proceeds by induction on $\bar{X}(\gamma)$. If $\bar{X}(\gamma)=0$, then γ is already simple, so the lemma is vacuously true.

First, suppose $\check{\gamma}$ is obtained from γ by smoothing a single vertex x. Consider an optimal mixed sequence of tangle flips and electrical moves that simplifies γ. This sequence starts with zero or more tangle flips, followed by a electrical move. Let γ' be the multicurve that results from the initial sequence of tangle flips; by definition, we have $\bar{X}(\gamma)=\bar{X}(\gamma')$. Moreover, applying the same sequence of tangle flips to $\check{\gamma}$ yields a connected multicurve $\check{\gamma}'$ such that $\bar{X}(\check{\gamma})=\bar{X}(\check{\gamma}')$. Thus, we can assume without loss of generality that the first operation in the sequence is a electrical move.

Now let γ' be the result of this move; by definition, we have $\bar{X}(\gamma)=\bar{X}(\gamma')+1$. As in the proof of Lemma 7.4, there are several subcases to consider, depending on whether the move from γ to γ' involves the smoothed vertex x, and if so, the specific type of move; see Figure 7.2. In every subcase, by Lemma 7.3 we can apply at most one electrical move to $\check{\gamma}$ to obtain a (possibly trivial) smoothing $\check{\gamma}'$ of γ', and then apply the inductive hypothesis on γ' and $\check{\gamma}'$ to prove the statement. We omit the straightforward details.

Finally, if $\check{\gamma}$ is obtained from γ by smoothing more than one vertex, the lemma follows immediately by induction from the previous analysis. □

Lemma 7.18. *For every connected multicurve γ, there is an intermixed sequence of electrical moves and tangle flips that simplifies γ to a simple closed curve, contains exactly $\bar{X}(\gamma)$ electrical moves, and does not contain $0{\to}1$ or $1{\to}2$ moves.*

Proof: Consider an optimal sequence of electrical moves and tangle flips that simplifies γ, and let γ_i denote the result of the first i moves in this sequence. If any γ_i has more vertices than its predecessor γ_{i-1}, then γ_{i-1} is a connected proper smoothing of γ_i, and Lemma 7.17 implies a contradiction. □

Lemma 7.19. *$\bar{X}(\gamma) \ge \bar{H}(\gamma)$ for every closed curve γ on the sphere.*

Proof: Let γ be a planar closed curve. The proof proceeds by induction on $\bar{X}(\gamma)$. If $\bar{X}(\gamma) = 0$, then γ is simple and thus $\bar{H}(\gamma) = 0$, so assume otherwise.

Consider an optimal sequence of electrical moves and tangle flips that simplifies γ, and let γ_i be the curve obtained by applying a prefix of the sequence up to and including the first electrical move. The minimality of the sequence implies that $\bar{X}(\gamma) = \bar{X}(\gamma') + 1$. By Lemma 7.18, we can assume without loss of generality that the first electrical move in the sequence is neither 0→1 nor 1→2, and if this first electrical move is 1→0 or 3→3, the theorem immediately follows by induction.

The only remaining move to consider is 2→1. Let γ° denote the result of applying the same sequence of tangle flips to γ, but replacing the final 2→1 move with a 2→0 move, or equivalently, smoothing the vertex of γ' left by the final 2→1 move. We immediately have $\bar{H}(\gamma) \le \bar{H}(\gamma^\circ) + 1$. Because γ° is a connected proper smoothing of γ', Lemma 7.17 implies $\bar{X}(\gamma^\circ) < \bar{X}(\gamma') = \bar{X}(\gamma) - 1$. Finally, the inductive hypothesis implies that $\bar{X}(\gamma^\circ) \ge \bar{H}(\gamma^\circ)$, which completes the proof. □

Lemma 7.20. *$\bar{H}(\gamma) \ge |defect(\gamma)|/2$ for every closed curve γ on the sphere.*

Proof: Each homotopy move decreases $|defect(\gamma)|$ by at most 2, and Lemmas 3.13 and 3.14 imply that tangle flips do not change $|defect(\gamma)|$ at all. Every simple curve has defect 0. □

Theorem 7.4. *Let G be an arbitrary planar graph, and let γ be any unicursal smoothing of $G^\times$ (defined with respect to any planar embedding of G). Reducing G to a single vertex requires at least $|defect(\gamma)|/2$ planar electrical transformations.*

Proof: The minimum number of planar electrical transformations required to reduce G is at least $\bar{X}(G^\times)$. Because γ is a single curve, it must be connected, so Lemma 7.17 implies that $\bar{X}(G^\times) \ge \bar{X}(\gamma)$. The theorem now follows immediately from Lemmas 7.19 and 7.20. □

The following corollary is now immediate from either Lemma 3.1, Lemma 3.2, or Corollary 3.1.

Corollary 7.8. *Reducing any n-vertex planar graph to a single vertex requires $\Omega(n^{3/2})$ planar electrical transformations in the worst case.*

7.3.4 Terminal-Leaf Contractions

The electrical reduction algorithms of Feo [99], Truemper [242], and Feo and Provan [100] rely exclusively on facial electrical transformations, plus one additional operation.

- *Terminal-leaf contraction*: Contract the edge incident to a *terminal* vertex with degree 1. The neighbor of the deleted terminal becomes a new terminal.

Terminal-leaf contractions are also called *FP-assignments*, after Feo and Provan [76, 115, 116]. Later algorithms for reducing plane graphs with three or four terminals [12, 76, 116] also use only facial electrical transformations and terminal-leaf contractions.

Formally, terminal-leaf contractions are *not* electrical transformations, as they can change the target value one wants to compute in application. For example, if the edges in the graph shown in Figure 7.4 represent 1Ω resistors, a terminal-leaf contraction changes the effective resistance between the terminals from 2Ω to 1Ω. However, both Gilter [115] and Feo and Provan [100] observed that any sequence of facial electrical transformations and terminal-leaf contractions can be simulated on the fly by a sequence of *planar* electrical transformations. Specifically, we simulate the first leaf contraction at either terminal by simply marking that terminal and proceeding as if its unique neighbor were a terminal. Later electrical transformations involving the neighbor of a marked terminal may no longer be facial, but they will still be planar; terminal-leaf contractions at the unique neighbor of a marked terminal become series reductions. At the end of the sequence of transformations, we perform a final series reduction at the unique neighbor of each marked terminal.

Unfortunately, terminal-leaf contractions change both the depth of the medial graph and the curve invariants that imply the quadratic homotopy lower bound. As a result, our quadratic lower bound proof breaks down if we allow terminal-leaf contractions. Indeed, we conjecture that any 2-terminal plane graph can be reduced to a single edge using only $O(n^{3/2})$ facial electrical transformations and terminal-leaf contractions, matching the lower bound proved in Section 7.3.3. (See Section 8.1.)

Chapter 8

Conclusions and Open Problems

Qui rogat, non errat.
— Latin proverb

Let us conclude the thesis with a list of conjectures along with some discussion.

8.1 Feo-Provan Conjecture

Perhaps the most compelling, and the primary motivation for our work, is to decide whether $\Theta(n^2)$ is indeed the best possible bound on the number of electrical transformations required to reduce any planar graph without terminals to a single vertex. Like Feo and Provan [100], Gitler [115], and Archdeacon *et al.* [12], we conjecture that $O(n^{3/2})$ facial electrical transformations suffice. However, perhaps we are less certain in light of the quadratic lower bound on reducing 2-terminal plane graphs from Section 7.3.2. Similarly, it is an open question whether any 2-terminal plane graph can be reduced to a single edge using $O(n^{3/2})$ facial electrical transformations and terminal-leaf contractions, as mentioned in Section 7.3.4. Proving these conjectures appears to be challenging.

Conjecture 8.1. *Any n-vertex plane graph can be reduced to a single vertex using at most $O(n^{3/2})$ facial electrical transformations. Any n-vertex plane graph with two terminals can be reduced to an edge using at most $O(n^{3/2})$ facial electrical transformations and terminal-leaf contractions.*

Once we go beyond facial and planar electrical transformations, none of our lower bound techniques apply, and we do not have any results about non-planar electrical transformations or electrical reduction of non-planar graphs. Indeed, the only lower bound known in the most general setting, for *any* family of electrically reducible graphs, is the trivial $\Omega(n)$. It seems unlikely that planar graphs can be reduced more quickly by using non-planar electrical transformations, but we can't prove anything. Any non-trivial lower bound for this problem would be interesting.

One way to prove the Feo-Provan conjecture is to extend Theorem 5.1 to the medial electrical setting. To do so it is sufficient to provide a way to tighten any tangle of depth $O(\sqrt{n})$ using $O(n^{3/2})$ electrical moves, similar to Lemma 5.1. One subtle difference between the two types of local operations is that a 2→1 move cannot be realized by homotopy of curves, and therefore the strategy for proving Lemma 5.1 and Lemma 5.1 by contracting monogons and tightening strands no longer works. Lemma 5.1 can be substituted by the algorithm of Feo and Provan [100] because the input is a closed curve; however as we are about to see, their algorithm does not work on tangles.

8.1.1 Feo-Provan's Algorithm

Call an electrical or a homotopy move *positive* if it decreases the sum of the face depths; in particular, every 1→0, 2→0, and 2→1 move is positive. A key technical lemma of Feo and Provan implies that every non-simple curve in the plane admits a positive homotopy move [100, Theorem 1]. Therefore, Feo and Provan's algorithm requires at most $O(D\Sigma)$ moves, where $D\Sigma$ is the sum of face depths of the input curve. Euler's formula implies that every curve with n crossings has $O(n)$ faces, and each of these faces has depth $O(n)$ in the worst case. Thus, the quadratic upper bound on simplifying planar curves using homotopy moves follows from algorithm of Feo and Provan as well.

One major benefit to view Feo and Provan's algorithm through the lens of medial construction is that the consistency of labeling scheme comes for free once we interpret the labels as depths of the faces in the medial graph. Unfortunately, all the existing proofs of positive-move lemma [100, 189] are quite long and complicated. Indeed, like we mentioned in Section 5.1.2, there are infinite classes of loose tangles that do not admit an positive moves. This suggests that any proof to the lemma needs to utilize the fact that the given (multi-)curve is indeed closed in the plane. On top of that, the proof by Feo and Provan is presented at the graph level which complicates the presentation. Here we raise the question in search of a better proof using the language of (multi-)curves.

Gitler [115] conjectured that a variant of Feo and Provan's algorithm that always makes the *deepest* positive move requires only $O(n^{3/2})$ moves. This conjecture is supported by the empirical results of Feo [99, Chapter 6]. Song [226] observed that if Feo and Provan's algorithm always chooses the *shallowest* positive move, it can be forced to make $\Omega(n^2)$ moves even when the input curve can be simplified using only $O(n)$ moves.

8.1.2 Steinitz's Algorithm

Another possible approach is an efficient implementation of Steinitz's bigon removal algorithm. In general removing a minimal bigon takes $\Theta(n)$ steps, so only a quadratic upper bound follows on tightening an arbitrary tangle.

It is natural to ask whether Steinitz's algorithm can be improved, either by carefully choosing which bigon to remove in each stage, by more carefully choosing how to empty each bigon, and/or by more refined analysis. (It is not hard to show that Steinitz's algorithm can be forced to perform $\Omega(n^2)$ moves if the bigons are chosen adversarially.) For example, one might repeatedly reduce the bigon containing the smallest number of faces. As we will see in Section 8.1.3, we cannot always hope for a bigon with sublinear number of faces inside. However, we can prove that a bigon with small *perimeter* does always exist.

As electrical moves in general do not preserve the number of strands of a multicurve or a tangle, we need to generalize the definition of tangle in this situation. In this subsection a ***tangle*** is a collection of boundary-to-boundary paths $\gamma_1, \gamma_2, \ldots, \gamma_s$ and a collection of *closed curves* $\kappa_1, \kappa_2, \ldots, \kappa_t$ in a closed topological disk Σ, which (self-)intersect only pairwise, transversely, and away from the boundary of Σ. We call each individual path γ_i an ***open strand*** and each closed curve κ_j a ***closed strand***; collectively we refer to them as ***strands***. A closed strand κ is ***lingering*** if κ does not intersect any other strands in the tangle. Throughout the subsection we assume that our tangle does not have lingering closed strands. A tangle is ***tight*** if every strand is simple, every pair of open strands intersects at most once, and otherwise all strands are disjoint; otherwise the tangle is ***loose***.

Let $\oplus$ be a tangle and let β be a bigon in the tangle. Let $\#\boldsymbol{on}(\boldsymbol{\beta})$ denote the number of intersections between the tangle and the boundary of an ε-neighborhood of the bigon β, for some small enough ε such that the boundary of ε-neighborhood only intersects the two curves that forms the bigon and the extension of the strands of the bigon. Also let $\#\boldsymbol{in}(\boldsymbol{\beta})$ denote the number of vertices inside the ε-neighborhood of bigon β.

Lemma 8.1. *Let n be a fixed integer. Let $\oplus$ be any loose tangle with at most n vertices, at most $3\sqrt{n}$ open strands, and no lingering closed strands. Then there is either an empty monogon in $\oplus$, or a minimal bigon β with $\#\text{on}(\beta) \le 8\sqrt{n}$.*[1]

Proof: Let the *length* of a (not necessarily simple) subpath η of a planar curve γ, denoted $|\eta|$, defined to be the number of *vertices* of γ on η (counted with multiplicity). We will prove the statement by induction on the number of vertices inside the tangle. First we argue that whenever we found a monogon of length at most $4\sqrt{n}$ then we are done. Because in this case either the monogon is empty, or we can apply the lemma recursively on an ε-neighborhood of the monogon excluding the double point (as a tangle of at most $3\sqrt{n}$ open strands); such a tangle cannot be tight.

Consider the following three cases. First, if all strands of tangle $\oplus$ has length at most $4\sqrt{n}$, the any minimal bigon in $\oplus$ will have $\#\text{on}(\beta) \le 8\sqrt{n}$.

The second case is when there is a closed strand κ of length less than $4\sqrt{n}$. Now either there is a monogon formed by κ and we are done, or κ is simple. Since all the closed strands are not lingering, there must be another strand of $\oplus$ intersecting κ at least twice. In this case we recurse on the interior tangle $\oplus'$ formed by curve κ. If $\oplus'$ is not tight then we are done. If $\oplus'$ is tight, let n' denote the number of vertices in $\oplus'$ and s' denote the number of strands of $\oplus'$. As all the strands of $\oplus'$ intersects each other at most once, we have $n' \le \binom{s'}{2}$ and there is a strand η of $\oplus'$ satisfying

$$|\eta| \le \frac{2n'}{s'} \le \frac{2}{s'} \cdot \frac{s'(s'-1)}{2} = s'-1 \le 2\sqrt{n},$$

as the length of κ is at most $4\sqrt{n}$. Again if η has a monogon then we are done. Otherwise, either the tangle $\oplus''$ is tight thus the bigon is minimal, or we can recurse on the tangle $\oplus''$ formed by an ε-neighborhood of the bigon σ formed by κ and η excluding the two double points, which has at most $(2+2)\sqrt{n}/2 \le 3\sqrt{n}$ open strands (since $|\kappa| \le 4\sqrt{n}$ and η separates κ into two arcs, one of the arcs has length at most $2\sqrt{n}$). In either case the statement is proved.

The third case is when all closed strands in $\oplus$ has length at least $4\sqrt{n}$ (there might be no closed strands at all), and one of the (open or closed) strands α in $\oplus$ has length at least $4\sqrt{n}$. As all the closed strands in $\oplus$ has length at least $4\sqrt{n}$, there are at most $0.5\sqrt{n}$ closed strands in $\oplus$. Take an arbitrary subpath of α of length $4\sqrt{n}$ and call it η. We refer to $\alpha \setminus \eta$ as a *semi-strand* of $\oplus$. Now either there is an monogon in η of length at most $4\sqrt{n}$ (in which case we are done); or η is a simple curve and there is another (semi-)strand of $\oplus$ that intersects η with at least two vertices by pigeonhole principle, as there are in total at most $(3+0.5)\sqrt{n}$ strands in $\oplus$, either open or closed.

Now one can prove that there must be a (semi-)strand λ of $\oplus$ such that

$$|\lambda \cap \eta| \ge \frac{|\lambda|}{\sqrt{n}-1} + 1.$$

Assume the contrary, we consider the sum over all (semi-)strands of $\oplus$:

$$\begin{aligned} |\eta| = \sum_{\lambda} |\lambda \cap \eta| &\le \frac{1}{\sqrt{n}-1} \sum_{\lambda} |\lambda| \\ &\le \frac{1}{\sqrt{n}-1} \left(2n - 4\sqrt{n}\right) \\ &\le 2\sqrt{n}, \end{aligned}$$

[1]The constants here are not optimal. In general, an upper bound $c\sqrt{n}$ on the number of open strands for some constant $c \ge \frac{4}{\sqrt{6}}$ will imply $\#\text{on}(\beta) \le (c + (c^2+8)^{1/2})\sqrt{n} \le (2c + \frac{4}{c})\sqrt{n}$ for some minimal bigon β.

which is a contradiction as $|\eta| = 4\sqrt{n}$.

Take two points x and y on $\lambda \cap \eta$, such that a subpath λ' of λ from x to y (including both endpoints) has length at most

$$\left\lfloor \frac{|\lambda|-1}{|\lambda \cap \eta|-1} \right\rfloor + 1 \leq \sqrt{n}.$$

(The assumption that $|\lambda \cap \eta| > 1$ is from the pigeonhole principle.) If λ' contains a monogon then we are done. Otherwise λ' is simple and there is a *quasi-bigon* formed by λ' and η. As η has length $4\sqrt{n}$, we again apply the lemma recursively on an ε-neighborhood of such a quasi-bigon excluding the two double points, as a tangle of at most $(4+1)\sqrt{n}/2 \leq 3\sqrt{n}$ open strands. □

8.1.3 Curves where All Bigons are Large

Now we introduce an infinite family of multicurves built on Fibonacci lattices, which we call *Fibonacci cubes*, in which every bigon and monogon contains $\Omega(n)$ faces. Prior work and applications on Fibonacci lattice include discrepancy and numerical integration [261, 262]; image processing and memory layout [58, 59, 101, 102]; data structures and lower bounds [26, 158].

For each integer k, the ***kth Fibonacci cube*** $\mathcal{F}_k$ is constructed from six identical tilted square lattices on the faces of a cube. Specifically, let $\mathcal{L}$ denote the dual of the standard integer lattice, with vertices $(x + 1/2, y + 1/2)$ for all integers x and y, and with edges between horizontal and vertical neighbors. Let $\mathcal{L}_k$ denote the two-dimensional Fibonacci lattice generated by the orthogonal integer vectors (F_{k-1}, F_k) and $(F_k, -F_{k-1})$. Each face of $\mathcal{F}_k$ contains the restriction of $\mathcal{L}$ with the square induced by lattice $\mathcal{L}_k$ with vertices $(0,0)$, (F_{k-1}, F_k), $(F_k, -F_{k-1})$, and (F_{k+1}, F_{k-2}), where F_i denotes the ith Fibonacci number. The graph $\mathcal{F}_k$ has exactly $n_k := 6 \cdot F_{2k-1}$ vertices and thus exactly $6 \cdot F_{2k-1} + 2$ faces.

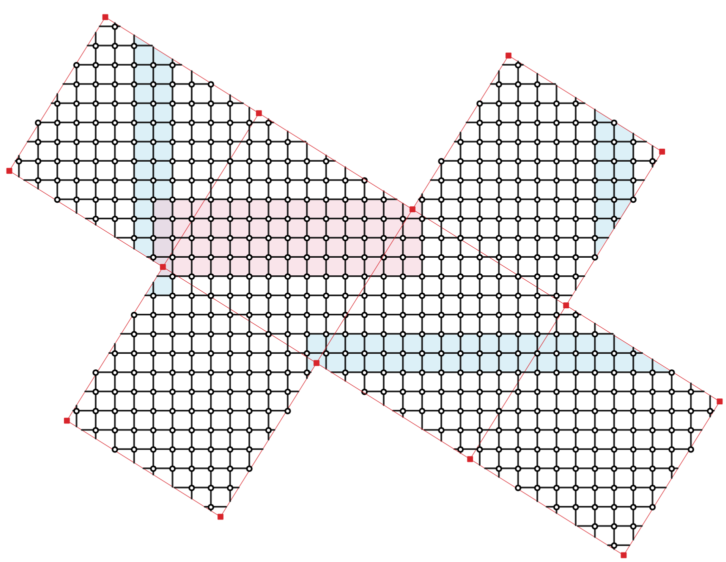


Figure 8.1. An unfolded Fibonacci cube $\mathcal{F}_6$ with two minimal bigons shaded.

Discrete Gauss-Bonnet theorem implies that every bigon in $\mathcal{F}_k$ contains exactly two triangular faces, which must lie on the boundary by Steinitz lemma on minimal bigons (see the proof of Lemma 2.1). Any minimal bigon—unfolded into the plane—looks like a rectangle with two opposite corners clipped off (to make the triangular faces). The other two opposite corners of the rectangle are the vertices of the bigon. Thus, the number of vertices in the interior of any minimal bigon is equal to the area of an axis-aligned rectangle in the plane with two opposite corners in the Fibonacci lattice $\mathcal{L}_k$. (See Figure 8.1.)

We use the following elementary but crucial discrepancy property of the Fibonacci lattice [58, Lemma 7] [59, Lemma 5].

Lemma 8.2. *Any axis-aligned rectangle containing more than one point of the Fibonacci lattice $\mathcal{L}_k$ has area at least $(F_{k-1}+1)(F_k+1) \geq F_{2k-1}/\sqrt{5} \geq n_k/(6\sqrt{5})$.*

Theorem 8.1. *Every bigon in $\mathcal{F}_k$ contains $\Theta(n_k)$ vertices and therefore $\Theta(n_k)$ faces.*

Erickson [90] conjectured that any Fibonacci cube $\mathcal{F}_k$ has a constant number of constituent curves, and any constituent curve γ of $\mathcal{F}_k$ (which is a single closed curve) satisfies the following property: each face of γ is the union of a constant number of faces of $\mathcal{F}_k$, and the number of vertices of γ is a constant fraction of n_k. This implies that any bigon in γ contains a constant fraction of the faces of γ, and therefore also have linear size.

8.2 Homotopy Moves on Low-genus Surfaces

As we have seen in Section 4.3, Theorem 4.6 implies an $\Omega(n^2)$ lower bound for tightening curves on any surface except for the sphere, the disk, and the projective plane. Our result in Section 5.1 shows that any planar curve and can be simplified in $O(n^{3/2})$ moves. Now two cases remain.

8.2.1 Tangles

If we only consider closed curves on the disk, this is no different than the planar case as our tightening algorithm does not make use of homotopy moves performed on the infinite face. (Although it is not hard to construct examples where the optimal number of moves required depends on whether the curve lies in the sphere or the disk.)

In many ways, tangles can be view as curve systems on the disk. (In general, when one talks about curves on surface without boundary, it makes sense to include all the boundary-to-boundary paths.) Our algorithm for simplifying planar curves (Theorem 5.1) generalizes directly to tangles; besides some minor details (say one should remove all the strands without intersections ahead of time), the only missing part is the lemma that proves the existence of useful cycles in tangles, analogous to Lemma 3.10. If one looks closely at the proof, there are no places where we use the assumption that the outermost contour contains the whole curve. Therefore we summarize the result without repeating its proof.

Theorem 8.2. *Every n-vertex tangle can be tightened in $O(n^{3/2})$ homotopy moves.*

If in addition we want to enforce monotonicity by disallowing 0→2 moves, the problem becomes open. In light of the close relation between electrical reduction and monotonic homotopy reduction process we have seen in Section 6.4, we believe that proving the following conjecture is as hard as its electrical counterpart, the Feo-Provan conjecture (see Section 8.1).

Conjecture 8.2. *Any n-vertex tangle can be tightened monotonically using $O(n^{3/2})$ homotopy moves.*

8.2.2 Projective Plane

The only missing case is the projective plane. Using the fact that the oriented double-cover of the projective plane is the sphere, an argument similar to the proof of Theorem 4.6 implies an $\Omega(n^{3/2})$ lower bound on homotopy moves, by plugging in the lower bound for the planar case (Theorem 4.1).

We left the task of finding a matching upper bound as an open question to the readers. One would expect a solution follows from extending the useful cycle technique to the projective planar setting.

Conjecture 8.3. *Any curves on the projective plane can be tightened using at most $O(n^{3/2})$ homotopy moves.*

8.3 Monotonic Homotopy Moves on Arbitrary Surfaces

Finally, in light of Theorem 6.2 and Lemma 6.10, we conjecture that any multicurve on an arbitrary surface can be tightened monotonically using polynomially many homotopy moves. This conjecture, if true, will generalize both Theorem 6.2 and Conjecture 6.1.

Conjecture 8.4. *Any multicurve on an arbitrary surface can be tightened monotonically using polynomially many homotopy moves.*

References

[1] Colin Adams. Triple crossing number of knots and links. *J. Knot Theory Ramif.* 22(2):1350006 (17 pages), 2013. arXiv:1207.7332.

[2] Colin Adams, Thomas Crawford, Benjamin DeMeo, Michael Landry, Alex Tong Lin, MurphyKate Montee, Seojung Park, Saraswathi Venkatesh, and Farrah Yhee. Knot projections with a single multi-crossing. *J. Knot Theory Ramif.* 24(3):1550011 (30 pages), 2015. arXiv:1208.5742.

[3] Virgil W. Adkisson. Cyclicly connected continuous curves whose complementary domain boundaries are homeomorphic, preserving branch points. *C. R. Séances Soc. Sci. Lett. Varsovie III* 23:164–193, 1930.

[4] Francesca Aicardi. Tree-like curves. *Singularities and Bifurcations*, 1–31, 1994. Advances in Soviet Mathematics 21, Amer. Math. Soc.

[5] Sheldon B. Akers, Jr. The use of wye-delta transformations in network simplification. *Oper. Res.* 8(3):311–323, 1960.

[6] James W. Alexander. Combinatorial analysis situs. *Trans. Amer. Math. Soc.* 28(2):301–326, 1926.

[7] James W. Alexander and G. B. Briggs. On types of knotted curves. *Ann. Math.* 28(1/4):562–586, 1926–1927.

[8] Sarah R. Allen, Luis Barba, John Iacono, and Stefan Langerman. Incremental Voronoi diagrams. *Proc. 32nd Int. Symp. Comput. Geom.*, 15:1–15:16, 2016. Leibniz International Proceedings in Informatics 51. ⟨http://drops.dagstuhl.de/opus/volltexte/2016/5907⟩. arXiv:1603.08485.

[9] Sigurd Angenent. Parabolic equations for curves on surfaces: Part II. Intersections, blow-up and generalized solutions. *Ann. Math.* 133(1):171–215, 1991.

[10] Tom M. Apostol. *Modular Functions and Dirichlet Series in Number Theory*, 2nd edition. Graduate Texts in Mathematics 41. Springer-Verlag, 1990.

[11] Hideyo Arakawa and Tetsuya Ozawa. A generalization of Arnold's strangeness invariant. *J. Knot Theory Ramif.* 8(5):551–567, 1999.

[12] Dan Archdeacon, Charles J. Colbourn, Isidoro Gitler, and J. Scott Provan. Four-terminal reducibility and projective-planar wye-delta-wye-reducible graphs. *J. Graph Theory* 33(2):83–93, 2000.

[13] Chris Arettines. A combinatorial algorithm for visualizing representatives with minimal self-intersection. *J. Knot Theory Ramif.* 24(11):1550058–1–1550058–17, 2015. arXiv:1101.5658.

[14] Stefan Arnborg, Andrzej Proskurowski, and Derek G. Corneil. Forbidden minors characterization of partial 3-trees. *Discrete Math.* 810:1–19, 1990.

[15] Vladimir I. Arnold. Plane curves, their invariants, perestroikas and classifications. *Singularities and Bifurcations*, 33–91, 1994. Adv. Soviet Math. 21, Amer. Math. Soc.

[16] Vladimir I. Arnold. *Topological Invariants of Plane Curves and Caustics*. University Lecture Series 5. Amer. Math. Soc., 1994.

[17] Eric K. Babson and Clara S. Chan. Counting faces of cubical spheres modulo two. *Discrete Math.* 212(3):169–183, 2000. arXiv:9811085v1.

[18] Thomas Banchoff. Critial points and curvature for embedded polyhedra. *J. Diff. Geom.* 1:245–256, 1967.

[19] Dror Bar-Natan. On the Vassiliev knot invariants. *Topology* 34:423–472, 1996.

[20] Ara Basmajian, Hugo Parlier, and Juan Souto. Geometric filling curves on surfaces. *Bulletin of the London Mathematical Society* 49(4):660–669. Wiley Online Library, 2017.

[21] Edward A. Bender and E. Rodney Canfield. The asymptotic number of rooted maps on a surface. *J. Comb. Theory Ser. A* 43(2):244–257, 1986.

[22] Mark de Berg, Otfried Cheong, Marc van Kreveld, and Mark Overmars. *Computational Geometry: Algorithms and Applications*, 3rd edition. Springer-Verlag, 2008.

[23] Sergei Bespamyatnikh. Computing homotopic shortest paths in the plane. *J. Algorithms* 49(2):284–303, 2003.

[24] Joan S. Birman and Xiao-Song Lin. Knot polynomials and Vassiliev's invariants. *Invent. Math.* 111:225–270, 1993.

[25] Henry R. Brahana. Systems of circuits on two-dlmensional manifolds. *Ann. Math.* 23(2):144–168, 1922.

[26] Gerth Stølting Brodal, Pooya Davoodi, Moshe Lewenstein, Rajeev Raman, and Satti Srinivasa Rao. Two dimensional range minimum queries and Fibonacci lattices. *Proc. 20th Ann. European Symp. Algorithms*, 217–228, 2012. Lecture Notes Comput. Sci. 7501, Springer.

[27] Gerhard Burde and Neiher Zieschang. *Knots*, 2nd revised and extended edition. de Gruyter Studies in Mathematics 5. Walter de Gruyter, 2003.

[28] Benjamin Burton, Erin Chambers, Marc van Kreveld, Wouter Meulemans, Tim Ophelders, and Bettina Speckmann. Computing optimal homotopies over a spiked plane with polygonal boundary. *LIPIcs-Leibniz International Proceedings in Informatics*, vol. 87, 2017.

[29] Sergio Cabello, Matt DeVos, Jeff Erickson, and Bojan Mohar. Finding one tight cycle. *Proc. 19th Ann. ACM-SIAM Symp. Discrete Algorithms*, 527–531, 2008.

[30] Sergio Cabello, Yuanxin Liu, Andrea Mantler, and Jack Snoeyink. Testing homotopy for paths in the plane. *Discrete Comput. Geom.* 31:61–81, 2004.

[31] Jaizhen Cai. Counting embeddings of planar graphs using DFS trees. *SIAM J. Discrete Math.* 6(3):335–352, 1993.

[32] Allan Calder and Jerrold Siegel. On the width of homotopies. *Topology* 19(3):209–220. Elsevier, 1980.

[33] Gabriel D. Carroll and David Speyer. The cube recurrence. *Elec. J. Combin.* 11:#R73, 2004.

[34] Erin W. Chambers, Éric Colin de Verdière, Jeff Erickson, Sylvain Lazard, Francis Lazarus, and Shripad Thite. Homotopic Fréchet distance between curves or, walking your dog in the woods in polynomial time. *Comput. Geom. Theory Appl.* 43(3):295–311, 2010.

[35] Erin W. Chambers and David Letscher. On the height of a homotopy. *Proc. 21st Canadian Conference on Computational Geometry*, vol. 9, 103–106, 2009.

[36] Erin W. Chambers and David Letscher. Erratum for on the height of a homotopy, 2010. ⟨http://mathcs.slu.edu/~chambers/papers/hherratum.pdf⟩.

[37] Erin Wolf Chambers, Gregory R Chambers, Arnaud de Mesmay, Tim Ophelders, and Regina Rotman. Constructing monotone homotopies and sweepouts. Preprint, August 2017. arXiv:1704.06175.

[38] Erin Wolf Chambers, Arnaud de Mesmay, and Tim Ophelders. On the complexity of optimal homotopies. *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1121–1134, 2018.

[39] Erin Wolf Chambers and Yusu Wang. Measuring similarity between curves on 2-manifolds via homotopy area. *Proceedings of the twenty-ninth annual symposium on Computational geometry*, 425–434, 2013.

[40] Gregory R. Chambers and Yevgeny Liokumovich. Converting homotopies to isotopies and dividing homotopies in half in an effective way. *Geometric and Functional Analysis* 24(4):1080–1100. Springer, 2014.

[41] Gregory R Chambers and Yevgeny Liokumovich. Optimal sweepouts of a Riemannian 2-sphere. Preprint, June 2016. arXiv:1411.6349.

[42] Gregory R Chambers and Regina Rotman. Monotone homotopies and contracting discs on Riemannian surfaces. *Journal of Topology and Analysis* 1–32. World Scientific, 2016.

[43] Hsien-Chih Chang and Jeff Erickson. Electrical reduction, homotopy moves, and defect. Preprint, October 2015. arXiv:1510.00571.

[44] Hsien-Chih Chang and Jeff Erickson. Untangling planar curves. *Proc. 32nd Int. Symp. Comput. Geom.*, 29:1–29:15, 2016. Leibniz International Proceedings in Informatics 51. ⟨http://drops.dagstuhl.de/opus/volltexte/2016/5921⟩.

[45] Hsien-Chih Chang and Jeff Erickson. Lower bounds for planar electrical reduction. Submitted, 2017.

[46] Hsien-Chih Chang and Jeff Erickson. Unwinding annular curves and electrically reducing planar networks. Accepted to Computational Geometry: Young Researchers Forum, Proc. 33rd Int. Symp. Comput. Geom., 2017.

[47] Hsien-Chih Chang, Jeff Erickson, Arnaud de Mesmay, David Letscher, Saul Schleimer, Eric Sedgwick, Dylan Thurston, and Stephan Tillmann. Untangling curves on surfaces via local moves. *Proc. 29th Annual ACM-SIAM Symposium on Discrete Algorithms*, 121–135, 2018.

[48] Hsien-Chih Chang and Arnaud de Mesmay. Tightening curves on surfaces: Better and faster. Manuscript, 2018.

[49] Manoj K. Chari, Thomas A. Feo, and J. Scott Provan. The delta-wye approximation procedure for two-terminal reliability. *Oper. Res.* 44(5):745–757, 1996.

[50] Moira Chas. Minimal intersection of curves on surfaces. *Geometriae Dedicata* 144(1):25–60. Springer, 2010.

[51] Moira Chas. The Goldman bracket and the intersection of curves on surfaces. *Geometry, Groups and Dynamics: ICTS Program "Groups, Geometry and Dynamics", December 3–16, 2012, Almora, India*, 73–83, 2015. Contemporary Mathematics 639, American Mathematical Soc.

[52] Moira Chas and Fabiana Krongold. An algebraic characterization of simple closed curves on surfaces with boudnary. *J. Topol. Anal.* 2(3):395–417, 2010.

[53] Moira Chas and Fabiana Krongold. Algebraic characteriation of simple closed curves via Turaev's cobracket. *J. Topology* 9(1):91–104, 2015.

[54] Bernard Chazelle. A theorem on polygon cutting with applications. *Proc. 23rd Ann. IEEE Symp. Found. Comput. Sci.*, 339–349, 1982.

[55] Bernard Chazelle and Herbert Edelsbrunner. An optimal algorithm for intersecting line segments in the plane. *J. ACM* 39(1):1–54, 1992.

[56] Sergei Chmutov and Sergei Duzhin. Explicit formulas for Arnold's generic curve invariants. *Arnold-Gelfand Mathematical Seminars: Geometry and Singularity Theory*, 123–138, 1997. Birkhäuser.

[57] Sergei Chmutov, Sergei Duzhin, and Jacob Mostovoy. *Introduction to Vassiliev knot invariants*. Cambridge Univ. Press, 2012. ⟨http://www.pdmi.ras.ru/∼duzhin/papers/cdbook⟩. arXiv:1103.5628.

[58] Benny Chor, Charles E. Leiserson, and Ronald L. Rivest. An application of number theory to the organization of raster-graphics memory (extended abstract). *Proc. 23rd Ann. IEEE Symp. Found. Comput. Sci.*, 92–99, 1982.

[59] Benny Chor, Charles E. Leiserson, Ronald L. Rivest, and James B. Shearer. An application of number theory to the organization of raster-graphics memory. *J. ACM* 33(1):86–104, 1986.

[60] Kenneth L. Clarkson and Peter W. Shor. Applications of random sampling in computational geometry, II. *Discrete Comput. Geom.* 4:387–421, 1989.

[61] Marshall Cohen and Martin Lustig. Paths of geodesics and geometric intersection numbers: I. *Combinatorial Group Theory and Topology*, 479–500, 1984. Annals of Math. Studies 111, Princeton Univ. Press.

[62] Charles J. Colbourn, J. Scott Provan, and Dirk Vertigan. A new approach to solving three combinatorial enumeration problems on planar graphs. *Discrete Appl. Math.* 60:119–129, 1995.

[63] Éric Colin de Verdière and Jeff Erickson. Tightening non-simple paths and cycles on surfaces. *Proc. 17th Ann. ACM-SIAM Symp. Discrete Algorithms*, 192–201, 2006.

[64] Éric Colin de Verdière and Jeff Erickson. Tightening non-simple paths and cycles on surfaces. *SIAM J. Comput.* 39(8):3784–3813, 2010.

[65] Éric Colin de Verdière and Francis Lazarus. Optimal pants decompositions and shortest homotopic cycles on an orientable surface. *J. ACM* 54(4), 2007.

[66] Yves Colin de Verdière. Reseaux électriques planaires. *Prepublications de l'Institut Fourier* 225:1–20, 1992.

[67] Yves Colin de Verdière. Réseaux électriques planaires I. *Comment. Math. Helvetici* 69:351–374, 1994.

[68] Yves Colin de Verdière, Isidoro Gitler, and Dirk Vertigan. Réseaux électriques planaires II. *Comment. Math. Helvetici* 71:144–167, 1996.

[69] John H. Conway. An enumeration of knots and links, and some of their algebraic properties. *Computational Problems in Abstract Algebra*, 329–358, 1970. Pergamon Press.

[70] Alexander Coward and Marc Lackenby. An upper bound on Reidemeister moves. *Amer. J. Math.* 136(4):1023–1066, 2014. arXiv:1104.1882.

[71] Edward B. Curtis, David Ingerman, and James A. Morrow. Circular planar graphs and resistor networks. *Linear Alg. Appl.* 283(1–3):115–150, 1998.

[72] Edward B. Curtis, Edith Mooers, and James Morrow. Finding the conductors in circular networks from boundary measurements. *Math. Mod. Num. Anal.* 28(7):781–813, 1994.

[73] Edward B. Curtis and James A. Morrow. *Inverse Problems for Electrical Networks*. World Scientific, 2000.

[74] Richard Dedekind. Erläuterungen zu den Fragmenten XXVIII. *Berhard Riemann's Gesammelte Mathematische Werke und wissenschaftlicher Nachlass*, 2nd edition, 466–478, 1892. Teubner.

[75] Max Dehn. Über unendliche diskontinuierliche Gruppen. *Math. Ann.* 71(1):116–144, 1911.

[76] Lino Demasi and Bojan Mohar. Four terminal planar Delta-Wye reducibility via rooted $K_{2,4}$ minors. *Proc. 26th Ann. ACM-SIAM Symp. Discrete Algorithms*, 1728–1742, 2015.

[77] Vincent Despré and Francis Lazarus. Computing the geometric intersection number of curves. *Proc. 33rd Int. Symp. Comput. Geom.*, 35:1–35:15, 2017. Leibniz Int. Proc. Informatics 77. arXiv:1511.09327.

[78] Giuseppe Di Battista and Roberto Tamassia. Incremental planarity testing. *Proc. 30th Ann. IEEE Symp. Foundations Comput. Sci.*, 436–441, 1989.

[79] Giuseppe Di Battista and Roberto Tamassia. On-line planarity testing. *SIAM J. Comput.* 25(5):956–997, 1996.

[80] Reinhard Diestel. *Graph Theory*, 5th edition. Springer Publishing Company, Incorporated, 2017.

[81] Walther Dyck. Beiträge zur Analysis situs I. Aufsatz. Ein- und zweidimensionale Mannigfaltigkeiten. *Math. Ann.* 32(4):457–512, 1888.

[82] Alon Efrat, Stephen G. Kobourov, and Anna Lubiw. Computing homotopic shortest paths efficiently. *Comput. Geom. Theory Appl.* 35(3):162–172, 2006.

[83] Arno Eigenwillig and Michael Kerber. Exact and efficient 2D-arrangements of arbitrary algebraic curves. *Proc. 19th Ann. ACM-SIAM Symp. Discrete Algorithms*, 122–131, 2008.

[84] Ehab S. El-Mallah and Charles J. Colbourn. On two dual classes of planar graphs. *Discrete mathematics* 80(1):21–40. Elsevier, 1990.

[85] G. V. Epifanov. Reduction of a plane graph to an edge by a star-triangle transformation. *Dokl. Akad. Nauk SSSR* 166:19–22, 1966. In Russian. English translation in *Soviet Math. Dokl.* 7:13–17, 1966.

[86] David Eppstein. Dynamic generators of topologically embedded graphs. *Proc. 14th Ann. ACM-SIAM Symp. Discrete Algorithms*, 599–608, 2003. arXiv:cs.DS/0207082.

[87] David B. A. Epstein. Curves on 2-manifolds and isotopies. *Acta Mathematica* 115:83–107, 1966.

[88] Jeff Erickson. Simple polygons. Preliminary draft, 2013. ⟨http://jeffe.cs.illinois.edu/teaching/comptop/schedule.html⟩.

[89] Jeff Erickson. Efficiently hex-meshing things with topology. *Discrete Comput. Geom.* 52(3):427–449. Springer, 2014.

[90] Jeff Erickson. Personal communication, May 2018.

[91] Jeff Erickson and Sariel Har-Peled. Optimally cutting a surface into a disk. *Discrete Comput. Geom.* 31(1):37–59, 2004.

[92] Jeff Erickson and Amir Nayyeri. Minimum cuts and shortest non-separating cycles via homology covers. *Proc. 22nd Ann. ACM-SIAM Symp. Discrete Algorithms*, 1166–1176, 2011.

[93] Jeff Erickson and Kim Whittlesey. Transforming curves on surfaces redux. *Proc. 24th Ann. ACM-SIAM Symp. Discrete Algorithms*, 1646–1655, 2013.

[94] Chaim Even-Zohar. Models of random knots. *Journal of Applied and Computational Topology* 1(2):263–296. Springer, 2017.

[95] Chaim Even-Zohar, Joel Hass, Nati Linial, and Tahl Nowik. Invariants of random knots and links. *Discrete & Computational Geometry* 56(2):274–314, 2016. arXiv:1411.3308.

[96] Chaim Even-Zohar, Joel Hass, Nati Linial, and Tahl Nowik. The distribution of knots in the petaluma model. Preprint, April 2017. arXiv:1706.06571.

[97] Chaim Even-Zohar, Joel Hass, Nati Linial, and Tahl Nowik. Universal knot diagrams. Preprint, April 2018. arXiv:1804.09860.

[98] Brittany Terese Fasy, Selcuk Karakoc, and Carola Wenk. On minimum area homotopies of normal curves in the plane. Preprint, July 2017. arXiv:1707.02251.

[99] Thomas A. Feo. *I. A Lagrangian Relaxation Method for Testing The Infeasibility of Certain VLSI Routing Problems. II. Efficient Reduction of Planar Networks For Solving Certain Combinatorial Problems*. Ph.D. thesis, Univ. California Berkeley, 1985. ⟨http://search.proquest.com/docview/303364161⟩.

[100] Thomas A. Feo and J. Scott Provan. Delta-wye transformations and the efficient reduction of two-terminal planar graphs. *Oper. Res.* 41(3):572–582, 1993.

[101] Amos Fiat and Adi Shamir. Polymorphic arrays: A novel VLSI layout for systolic computers. *J. Comput. System Sci.* 33(1):47–65, 1986.

[102] Amos Fiat and Adi Shamir. How to find a battleship. *Networks* 19(3):361–371, 1989.

[103] George K. Francis. The folded ribbon theorem: A contribution to the study of immersed circles. *Trans. Amer. Math. Soc.* 141:271–303, 1969.

[104] George K. Francis. Titus' homotopies of normal curves. *Proc. Amer. Math. Soc.* 30:511–518, 1971.

[105] George K. Francis. Generic homotopies of immersions. *Indiana Univ Math. J.* 21(12):1101–1112, 1971/72.

[106] George K. Francis and Jeff Weeks. Conway's ZIP proof. *Amer. Math. Monthly* 106(5):393–399, 1999.

[107] J. P. Gadani. System effectiveness evaluation using star and delta transformations. *IEEE Trans. Reliability* R-30(1):43–47, 1981.

[108] Jean Gallier and Dianna Xu. *A guide to the classification theorem for compact surfaces*. Springer Science & Business Media, 2013.

[109] Carl Friedrich Gauß. Nachlass. I. Zur Geometria situs. *Werke*, vol. 8, 271–281, 1900. Teubner. Originally written between 1823 and 1840.

[110] Meinolf Geck and Gotz Pfeiffer. On the irreducible characters of hecke algebras. *Advances in Mathematics* 102(1):79–94, 1993.

[111] Steve M. Gersten and Hamish B. Short. Small cancellation theory and automatic groups. *Invent. Math.* 102:305–334, 1990.

[112] Peter Giblin. *Graphs, Surfaces and Homology*, 3rd edition. Cambridge Univ. Press, 2010.

[113] Cole A. Giller. A family of links and the Conway calculus. *Transactions of the American Mathematical Society* 270(1):75–109, 1982.

[114] Patrick M. Gilmer and Richard A. Litherland. The duality conjecture in formal knot theory. *Osaka Journal of Mathematics* 23(1):229–247. Osaka University and Osaka City University, Departments of Mathematics, 1986.

[115] Isidoro Gitler. *Delta-wye-delta Transformations: Algorithms and Applications*. Ph.D. dissertation, University of Waterloo, 1991.

[116] Isidoro Gitler and Feliú Sagols. On terminal delta-wye reducibility of planar graphs. *Networks* 57(2):174–186, 2011.

[117] Jay R. Goldman and Louis H. Kauffman. Knots, tangles, and electrical networks. *Adv. Appl. Math.* 14:267–306, 1993.

[118] William M. Goldman. Invariant functions on Lie groups and hamiltonian flows of surface group representations. *Inventiones mathematicae* 85(2):263–302. Springer, 1986.

[119] Daciberg L. Gonçalves, Elena Kudryavtseva, and Heiner Zieschang. An algorithm for minimal number of intersection points of curves on surfaces. *Proc. Seminar on Vector and Tensor Analysis* 26(139–167), 2005.

[120] Jacob E. Goodman and Stefan Felsner. Pseudoline arrangements. *Handbook of discrete and computational geometry*, 3rd edition, chapter 5, 125–157, 2017. Chapman and Hall/CRC.

[121] Gramoz Goranci, Monika Henzinger, and Pan Peng. Improved guarantees for vertex sparsification in planar graphs. Preprint, December 2017. arXiv:1702.01136.

[122] Maurits de Graaf. *Graphs and curves on surfaces*. Ph.D. dissertation, Universiteit van Amsterdam, 1994.

[123] Maurits de Graaf and Alexander Schrijver. Characterizing homotopy of systems of curves on a compact surface by crossing numbers. *Linear Alg. Appl.* 226–228:519–528, 1995.

[124] Maurits de Graaf and Alexander Schrijver. Decomposition of graphs on surfaces. *J. Comb. Theory Ser. B* 70:157–165, 1997.

[125] Maurits de Graaf and Alexander Schrijver. Making curves minimally crossing by Reidemeister moves. *J. Comb. Theory Ser. B* 70(1):134–156, 1997.

[126] Matthew A. Grayson. Shortening embedded curves. *Ann. Math.* 129(1):71–111, 1989.

[127] Keith Gremban. *Combinatorial Preconditioners for Sparse, Symmetric, Diagonally Dominant Linear Systems*. Ph.D. thesis, Carnegie Mellon University, 1996. ⟨https://www.cs.cmu.edu/~glmiller/Publications/b2hd-GrembanPHD.html⟩. Tech. Rep. CMU-CS-96-123.

[128] Branko Grünbaum. *Convex Polytopes*. Monographs in Pure and Applied Mathematics XVI. John Wiley & Sons, 1967.

[129] Doris Lloyd Grosh. Comments on the delta-star problem. *IEEE Trans. Reliability* R-32(4):391–394, 1983.

[130] Leonidas J. Guibas and Jorge Stolfi. Primitives for the manipulation of general subdivisions and the computation of Voronoi diagrams. *ACM Trans. Graphics* 4(2):75–123, 1985.

[131] Hariom Gupta and Jaydev Sharma. A delta-star transformation approach for reliability estimation. *IEEE Trans. Reliability* R-27(3):212–214, 1978.

[132] Tobias J. Hagge and Jonathan T. Yazinski. On the necessity of Reidemeister move 2 for simplifying immersed planar curves. *Knots in Poland III. Part III*, 101–110, 2014. Banach Center Publ. 103, Inst. Math., Polish Acad. Sci. arXiv:0812.1241.

[133] Sariel Har-Peled, Amir Nayyeri, Mohammad Salavatipour, and Anastasios Sidiropoulos. How to walk your dog in the mountains with no magic leash. *Discrete Comput. Geom.* 55(1):39–73, 2016.

[134] Joel Hass and Tal Nowik. Unknot diagrams requiring a quadratic number of Reidemeister moves to untangle. *Discrete Comput. Geom.* 44(1):91–95, 2010.

[135] Joel Hass and Peter Scott. Intersections of curves on surfaces. *Israel J. Math.* 51(1–2):90–120, 1985.

[136] Joel Hass and Peter Scott. Shortening curves on surfaces. *Topology* 33(1):25–43, 1994.

[137] Chuichiro Hayashi and Miwa Hayashi. Minimal sequences of Reidemeister moves on diagrams of torus knots. *Proc. Amer. Math. Soc.* 139:2605–2614, 2011. arXiv:1003.1349.

[138] Chuichiro Hayashi, Miwa Hayashi, and Tahl Nowik. Unknotting number and number of Reidemeister moves needed for unlinking. *Topology Appl.* 159:1467–1474, 2012. arXiv:1012.4131.

[139] Chuichiro Hayashi, Miwa Hayashi, Minori Sawada, and Sayaka Yamada. Minimal unknotting sequences of Reidemeister moves containing unmatched RII moves. *J. Knot Theory Ramif.* 21(10):1250099 (13 pages), 2012. arXiv:1011.3963.

[140] John Hershberger and Jack Snoeyink. Computing minimum length paths of a given homotopy class. *Comput. Geom. Theory Appl.* 4:63–98, 1994.

[141] Arthur M. Hobbs. Letter to the editor: Remarks on network simplification. *Oper. Res.* 15(3):548–551, 1967.

[142] Heinz Hopf. Über die Drehung der Tangenten und Sehnen ebener Kurven. *Compositio Math.* 2:50–62, 1935.

[143] Jim Hoste. The Arf invariant of a totally proper link. *Topology and its Applications* 18(2-3):163–177. Elsevier, 1984.

[144] Noburo Ito and Yusuke Takimura. (1,2) and weak (1,3) homotopies on knot projections. *J. Knot Theory Ramif.* 22(14):1350085 (14 pages), 2013. Addendum in *J. Knot Theory Ramif.* 23(8):1491001 (2 pages), 2014.

[145] Noburo Ito, Yusuke Takimura, and Kouki Taniyama. Strong and weak (1, 3) homotopies on knot projections. *Osaka J. Math* 52(3):617–647, 2015.

[146] François Jaeger. On spin models, triply regular association schemes, and duality. *J. Alg. Comb.* 4:103–144, 1995.

[147] Vaughan F. R. Jones. On knot invariants related to some statistical mechanical models. *Pacific journal of mathematics* 137(2):311–334. University of California, Department of Mathematics, 1989.

[148] Camille Jordan. Sur la déformation des surfaces. *J. Math. Pures Appl. (Série 2)* 11:105–109, 1866.

[149] Camille Jordan. Courbes continues. *Cours d'Analyse de l'École Polytechnique*, 1st edition, vol. 3, 587–594, 1887.

[150] Camille Jordan. Lignes continues. *Cours d'Analyse de l'École Polytechnique*, 2nd edition, vol. 1, 90–99, 1893.

[151] Louis H. Kauffman. *On Knots*. Princeton Univ. Press, 1987.

[152] Louis H. Kauffman. State models and the Jones polynomial. *Topology* 26(3):395–407. Elsevier, 1987.

[153] Louis H. Kauffman. New invariants in the theory of knots. *Amer. Math. Monthly* 95(3):195–242, 1988.

[154] Louis H. Kauffman. Gauss codes, quantum groups and ribbon Hopf algebras. *Reviews in Mathematical Physics* 5(4):735–773. World Scientific, 1993.

[155] Arthur Edwin Kennelly. Equivalence of triangles and three-pointed stars in conducting networks. *Electrical World and Engineer* 34(12):413–414, 1899.

[156] Richard W. Kenyon. The Laplacian on planar graphs and graphs on surfaces. *Current Developments in Mathematics*, 2011. Int. Press. arXiv:1203.1256.

[157] Mikhail Khovanov. Doodle groups. *Trans. Amer. Math. Soc.* 349(6):2297–2315, 1997.

[158] Elias Koutsoupias and David Scot Taylor. Tight bounds for 2-dimensional indexing schemes. *Proc. 17th Ann. ACM Symp. Principles Database Syst.* 52–58, 1998.

[159] Marc Lackenby. A polynomial upper bound on Reidemeister moves. *Ann. Math.* 182(2):491–564, 2015. arXiv:1302.0180.

[160] Sergei K. Lando and Alexander K. Zvonkin. *Graphs on Surfaces and Their Applications*. Low-Dimensional Topology II. Springer-Verlag, 2004.

[161] Francis Lazarus and Julien Rivaud. On the homotopy test on surfaces. *Proc. 53rd Ann. IEEE Symp. Foundations Comput. Sci.*, to appear, 2012. arXiv:1110.4573.

[162] Der-Tsai Lee and Franco P. Preparata. Euclidean shortest paths in the presence of rectilinear barriers. *Networks* 14:393–410, 1984.

[163] John M. Lee. *Introduction to Topological Manifolds*. Graduate Texts in Mathematics 202. Springer, 2000.

[164] Alfred Lehman. A necessary condition for wye-delta transformation. MRC Technical Summary Report 383, Math. Res. Center, Univ, Wisconsin, March 1963. ⟨www.dtic.mil/dtic/tr/fulltext/u2/403836.pdf⟩.

[165] Alfred Lehman. Wye-delta transformations in probabilistic network. *J. Soc. Indust. Appl. Math.* 11:773–805, 1963.

[166] Gerard T. Lingner, Themistocles Politof, and A. Satyanarayana. A forbidden minor characterization and reliability of a class of partial 4-trees. *Networks* 25(3):139–146. Wiley Online Library, 1995.

[167] Xiao-Song Lin and Zhenghan Wang. Integral geometry of plane curves and knot invariants. *J. Diff. Geom.* 44:74–95, 1996.

[168] Richard J. Lipton and Robert E. Tarjan. A separator theorem for planar graphs. *SIAM J. Applied Math.* 36(2):177–189, 1979.

[169] Chenghui Luo. *Numerical Invariants and Classification of Smooth and Polygonal Plane Curves*. Ph.D. thesis, Brown Univ., 1997.

[170] Feng Luo. Simple loops on surfaces and their intersection numbers. Preprint, January 1998. arXiv:math/9801018.

[171] Feng Luo. Simple loops on surfaces and their intersection numbers. *Journal of Differential Geometry* 85(1):73–116. Lehigh University, 2010.

[172] Martin Lustig. Paths of geodesics and geometric intersection numbers: II. *Combinatorial Group Theory and Topology*, 501–544, 1987. Annals of Math. Studies 111, Princeton Univ. Press.

[173] Roger C. Lyndon. On Dehn's algorithm. *Math. Ann.* 166:208–228, 1966.

[174] Roger C. Lyndon and Paul E. Schupp. *Combinatorial Group Theory*. Springer-Verlag, 1977.

[175] Saunders Mac Lane. A structural characterization of planar combinatorial graphs. *Duke Math. J.* 3(3):460–472, 1937.

[176] William S. Massey. *A basic course in algebraic topology*. Springer-Verlag, 1991.

[177] August F. Möbius. Theorie der elementaren Verwandtschaft. *Ber. Sächs. Akad. Wiss. Leipzig, Math.-Phys. Kl.* 17:18–57, 1863. *Gesammelte Werke* 2:433–471, Liepzig, 1886.

[178] Curt Meyer. Über einige Anwendungen Dedekindscher Summen. *J. Reine Angew. Math.* 198:143–203, 1957.

[179] Bojan Mohar and Carsten Thomassen. *Graphs on Surfaces*. Johns Hopkins Univ. Press, 2001.

[180] Ketan Mulmuley. A fast planar partition algorithm, I. *J. Symbolic Comput.* 10(3–4):253–280, 1990.

[181] Hiroyuki Nakahara and Hiromitsu Takahashi. An algorithm for the solution of a linear system by Δ-Y transformations. *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences* E79-A(7):1079–1088, 1996. Special Section on Multi-dimensional Mobile Information Network.

[182] Max Neumann-Coto. A characterization of shortest geodesics on surfaces. *Algebraic & Geometric Topology* 1:349–368, 2001.

[183] Zipei Nie. On the minimum area of null homotopies of curves traced twice. Preprint, December 2014. arXiv:1412.0101.

[184] Steven D. Noble and Dominic J. A. Welsh. Knot graphs. *J. Graph Theory* 34(1):100–111, 2000.

[185] Tahl Nowik. Complexity of planar and spherical curves. *Duke J. Math.* 148(1):107–118, 2009.

[186] Tahl Nowik. Order one invariants of planar curves. *Adv. Math.* 220:427–440, 2009. arXiv:0712.4071.

[187] Jane M. Paterson. A combinatorial algorithm for immersed loops in surfaces. *Topology Appl.* 123(2):205–234, 2002.

[188] George Pólya. An elementary analogue to the Gauss-Bonnet theorem. *Amer. Math. Monthly* 61(9):601–603, 1954.

[189] Sofya Poger. *Some New Results on Three-Terminal Planar Graph Reducibility*. Ph.D. dissertation, Stevens Inst. Tech., 2001.

[190] Sofya Poger and Yoram J. Sussman. A Δ-wye-Δ reduction for planar grid graphs in subquadratic time. *Algorithms and Complexity in Durham 2006: Proceedings of the Second ACiD Workshop*, 119–130, 2006.

[191] Themistocles Politof. *A Characterization and Efficient Reliability Computation of Delta-Y Reducible Networks*. Ph.D. dissertation, University of California, Berkeley, 1983.

[192] Themistocles Politof and A. Satyanarayana. Network reliability and inner-four-cycle-free graphs. *Mathematics of operations research* 11(3):484–505. INFORMS, 1986.

[193] Themistocles Politof and A. Satyanarayana. A linear-time algorithm to compute the reliability of planar cube-free networks. *IEEE Transactions on Reliability* 39(5):557–563. IEEE, 1990.

[194] Themistocles Politof, A. Satyanarayana, and L. Tung. An $O(n \cdot \log(n))$ algorithm to compute the all-terminal reliability of $(K_5, K_{2,2,2})$ free networks. *IEEE transactions on reliability* 41(4):512–517. IEEE, 1992.

[195] Michael Polyak. Invariants of curves and fronts via Gauss diagrams. *Topology* 37(5):989–1009, 1998.

[196] Michael Polyak. New Whitney-type formulae for plane curves. *Differential and symplectic topology of knots and curves*, 103–111, 1999. Amer. Math. Soc. Translations 190, Amer. Math. Soc.

[197] Michael Polyak and Oleg Viro. Gauss diagram formulas for Vassiliev invariants. *Int. Math. Res. Notices* 11:445–453, 1994.

[198] Michael Polyak and Oleg Viro. On the Casson knot invariant. *J. Knot Theory Ramif.* 10(5):711–738, 2001. arXiv:math/9903158.

[199] Alexander Postnikov. Total positivity, grassmannians, and networks. Preprint, Sep 2006. arXiv:math/0609764.

[200] Igor Prlina, Marcus Spradlin, and Stefan Stanojevic. All-loop singularities of scattering amplitudes in massless planar theories. Preprint, May 2018. arXiv:1805.11617.

[201] Hans Rademacher and Emil Grosswald. *Dedekind Sums*. Carus Math. Monographs 16. Math. Assoc. America, 1972.

[202] Kurt Reidemeister. Elementare Begründung der Knotentheorie. *Abh. Math. Sem. Hamburg* 5:24–32, 1927.

[203] Gerhard Ringel. Teilungen der Ebene durch Geraden oder topologische Geraden. *Math. Z.* 64(1):79–102, 1956.

[204] Gerhard Ringel. Über geraden in allgemeiner lage. *Elemente der Mathematik* 12:75–82, 1957.

[205] Neil Robertson and Paul D. Seymour. Graph minors. VII. Disjoint paths on a surface. *J. Comb. Theory Ser. B* 45(2):212–254, 1988.

[206] Neil Robertson and Paul D. Seymour. Graph minors. X. Obstructions to tree-decomposition. *J. Comb. Theory Ser. B* 52(2):153–190, 1991.

[207] Neil Robertson, Paul D. Seymour, and Robin Thomas. Quickly excluding a planar graph. *J. Comb. Theory Ser. B* 62(2):232–348, 1994.

[208] Neil Robertson and Richard Vitray. Representativity of surface embeddings. *Paths, Flows, and VLSI-Layout*, 293–328, 1990. Algorithms and Combinatorics 9, Springer-Verlag.

[209] Arnie Rosenthal. Note on 'Closed form solutions for delta-star and star-delta conversion of reliability networks'. *IEEE Trans. Reliability* R-27(2):110–111, 1978.

[210] Arnie Rosenthal and David E. Frisque. Transformations for simplifying network reliability calculations. *Networks* 7(2):97–111, 1977. Erratum in *Networks* 7(4):382, 1977.

[211] Jean-Pierre Roudneff. Tverberg-type theorems for pseudoconfigurations of points in the plane. *European Journal of Combinatorics* 9(2):189–198. Elsevier, 1988.

[212] Alexander Russell. The method of duality. *A Treatise on the Theory of Alternating Currents*, chapter XVII, 380–399, 1904. Cambridge Univ. Press.

[213] A. Satyanarayana and R. Tindell. Efficient algorithms for the evaluation of planar network reliability. Tech. rep., Department of Electrical Engineering and Computer Science, Stevens Institute of Technology, March 1993.

[214] A. Satyanarayana and L. Tung. A characterization of partial 3-trees. *Networks* 20(3):299–322. Wiley Online Library, 1990.

[215] Arthur Schönflies. Über einen Satz aus der Analysis situs. *Nachr. Ges. Wiss. Göttingen* 79–89, 1896.

[216] Alexander Schrijver. Homotopy and crossing of systems of curves on a surface. *Linear Alg. Appl.* 114–115:157–167, 1989.

[217] Alexander Schrijver. Decomposition of graphs on surfaces and a homotopic circulation theorem. *J. Comb. Theory Ser. B* 51(2):161–210, 1991.

[218] Alexander Schrijver. Circuits in graphs embedded on the torus. *Discrete Math.* 106/107:415–433, 1992.

[219] Alexander Schrijver. On the uniqueness of kernels. *J. Comb. Theory Ser. B* 55:146–160, 1992.

[220] Herbert Seifert and William Threlfall. *Lehrbook der Topologie*. Teubner, Leipzig, 1934. Reprinted by AMS Chelsea, 2003. English translation in [221].

[221] Herbert Seifert and William Threlfall. *A Textbook of Topology*. Pure and Applied Mathematics 89. Academic Press, New York, 1980. Edited by Joan S. Birman and Julian Eisner. Translated from [220] by Michael A. Goldman.

[222] Aleksandr Shumakovich. Explicit formulas for the strangeness of plane curves. *Algebra i Analiz* 7(3):165–199, 1995. Corrections in *Algebra i Analiz* 7(5):252–254, 1995. In Russian; English translation in [223].

[223] Aleksandr Shumakovich. Explicit formulas for the strangeness of a plane curve. *St. Petersburg Math. J.* 7(3):445–472, 1996. English translation of [222].

[224] C. Singh and S. Asgarpoor. Comments on "Closed form solutions for delta-star and star-delta conversion of reliability networks". *IEEE Trans. Reliability* R-25(5):336–339, 1976.

[225] C. Singh and S. Asgarpoor. Reliability evaluation of flow networks using delta-star transformations. *IEEE Trans. Reliability* R-35(4):472–477, 1986.

[226] Xiaohuan Song. Implementation issues for Feo and Provan's delta-wye-delta reduction algorithm. M.Sc. Thesis, University of Victoria, 2001.

[227] Ernesto Staffelli and Federico Thomas. Analytic formulation of the kinestatis of robot manipulators with arbitrary topology. *Proc. 2002 IEEE Conf. Robotics and Automation*, 2848–2855, 2002.

[228] Matthias F. M. Stallmann. Using PQ-trees for planar embedding problems. Tech. Rep. NCSU-CSC TR-85-24, Dept. Comput. Sci., NC State Univ., December 1985. ⟨https://people.engr.ncsu.edu/mfms/Publications/1985-TR_NCSU_CSC-PQ_Trees.pdf⟩.

[229] Matthias F. M. Stallmann. On counting planar embeddings. *Discrete Math.* 122:385–392, 1993.

[230] Ernst Steinitz. Polyeder und Raumeinteilungen. *Enzyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* III.AB(12):1–139, 1916.

[231] Ernst Steinitz and Hans Rademacher. *Vorlesungen über die Theorie der Polyeder: unter Einschluß der Elemente der Topologie*. Grundlehren der mathematischen Wissenschaften 41. Springer-Verlag, 1934. Reprinted 1976.

[232] John Stillwell. *Classical Topology and Combinatorial Group Theory*, 2nd edition. Graduate Texts in Mathematics 72. Springer-Verlag, 1993.

[233] James A. Storer. On minimal node-cost planar embeddings. *Networks* 14(2):181–212, 1984.

[234] Peter Guthrie Tait. On knots I. *Proc. Royal Soc. Edinburgh* 28(1):145–190, 1876–7.

[235] Carsten Thomassen. Embeddings of graphs with no short noncontractible cycles. *J. Comb. Theory Ser. B* 48(2):155–177, 1990.

[236] Dylan P. Thurston. Geometric intersection of curves on surfaces. Preprint, August 2008. ⟨http://www.math.columbia.edu/~dpt/DehnCoordinates.ps⟩.

[237] Alexander Tiskin. Semi-local string comparison: Algorithmic techniques and applications. Preprint, November 2013. arXiv:0707.3619.

[238] Charles J. Titus. A theory of normal curves and some applications. *Pacific J. Math.* 10:1083–1096, 1960.

[239] Charles J. Titus. The combinatorial topology of analytic functions on the boundary of a disk. *Acta Mathematica* 106:45–64, 1961.

[240] Lorenzo Traldi. On the star-delta transformation in network reliability. *Networks* 23(3):151–157, 1993.

[241] Tiffani Traver. Trigonometry in the hyperbolic plane. Manuscript, May 2014.

[242] Klaus Truemper. On the delta-wye reduction for planar graphs. *J. Graph Theory* 13(2):141–148, 1989.

[243] Klaus Truemper. A decomposition theory for matroids. VI. Almost regular matroids. *J. Comb. Theory Ser. B* 55:253–301, 1992.

[244] Klaus Truemper. *Matroid Decomposition*. Academic Press, 1992.

[245] Klaus Truemper. A note on delta-wye-delta reductions of plane graphs. *Congr. Numer.* 158:213–220, 2002.

[246] Vladimir G. Turaev. Quantum invariants of 3-manifolds and a glimpse of shadow topology. *C. R. Acad. Sci. Paris I* 313:395–398, 1991.

[247] Vladimir G. Turaev. Skein quantization of poisson algebras of loops on surfaces. *Annales scientifiques de l'Ecole normale supérieure*, vol. 24, 635–704, 1991.

[248] Vladimir G. Turaev. Shadow links and face models of statistical mechanics. *J. Diff. Geom.* 36:35–74, 1992.

[249] Leslie S. Valiant. Universality considerations in VLSI circuits. *IEEE Trans. Comput.* C-30(2):135–140, 1981.

[250] Oswald Veblen. Theory on plane curves in non-metrical analysis situs. *Trans. Amer. Math. Soc.* 6:83–98, 1905.

[251] Gert Vegter. Kink-free deformation of polygons. *Proceedings of the 5th Annual Symposium on Computational Geometry*, 61–68, 1989.

[252] Oleg Viro. Generic immersions of circle to surfaces and complex topology of real algebraic curves. *Topology of Real Algebraic Varieties and Related Topics*, 231–252, 1995. Amer. Math. Soc. Translations 173, Amer. Math. Soc.

[253] Donald Wagner. Delta-wye reduction of almost-planar graphs. *Discrete Appl. Math.* 180:158–167, 2015.

[254] Douglas Brent West. *Introduction to graph theory*, 2nd edition. Prentice hall Upper Saddle River, 2001.

[255] Brian White. Mappings that minimize area in their homotopy classes. *Journal of Differential Geometry* 20(2):433–446. Lehigh University, 1984.

[256] Hassler Whitney. Congruent graphs and the connectivity of graphs. *Amer. J. Math.* 54(1):150–168, 1932.

[257] Hassler Whitney. On regular closed curves in the plane. *Compositio Math.* 4:276–284, 1937.

[258] Takeshi Yajima and Shin'ichi Kinoshita. On the graphs of knots. *Osaka Mathematical Journal* 9(2):155–163. Department of Mathematics, Osaka University, 1957.

[259] Yaming Yu. Forbidden minors for wyw-delta-wye reducibility. *J. Graph Theory* 47(4):317–321, 2004.

[260] Yaming Yu. More forbidden minors for wye-delta-wye reducibility. *Elec. J. Combin.* 13:#R7, 2006.

[261] Stanisław K. Zaremba. Good lattice points, discrepancy, and numerical integration. *Annali di Matematica Pura ed Applicata* 73(1):293–317, 1966.

[262] Stanisław K. Zaremba. A remarkable lattice generated by Fibonacci numbers. *Fibonacci Quarterly* 8(2):185–198, 1970. ⟨http://www.fq.math.ca/8-2.html⟩.

[263] Ron Zohar and Dan Gieger. Estimation of flows in flow networks. *Europ. J. Oper. Res.* 176:691–706, 2007.