

## ROTATION DISTANCE, TRIANGULATIONS, AND HYPERBOLIC GEOMETRY

DANIEL D. SLEATOR, ROBERT E. TARJAN, AND WILLIAM P. THURSTON

### 1. INTRODUCTION

A *rotation* in a binary tree is a local restructuring of the tree that changes it into another tree. One can execute a rotation by collapsing an internal edge of the tree to a point, thereby obtaining a node with three children, and then re-expanding the node of order three in the alternative way into two nodes of order 2. The *rotation distance* between a pair of trees is the minimum number of rotations needed to convert one tree into the other. The problem addressed in this paper is: what is the maximum rotation distance between any pair of  $n$ -node binary trees? We show that for all  $n \geq 11$  this distance is at most  $2n - 6$  and that for all sufficiently large  $n$  this bound is tight. Culik and Wood [2] showed that the maximum rotation distance is at most  $2n - 2$ . Tom Leighton (private communication) showed that there exist trees whose rotation distance is at least  $7n/4 - O(1)$ . Pallo [7] proposed a heuristic search algorithm to compute the rotation distance between two given trees.

Our interest in this problem stems from our attempt to solve the dynamic optimality conjecture concerning the performance of splaying [8, 10]. Splaying is a heuristic for modifying the structure of a binary search tree in such a way that repeatedly accessing and updating the information in the tree is efficient. Although our solution to the problem of maximum rotation distance did not resolve the conjecture about splaying, the results in this paper are interesting for at least two other reasons. First, the combinatorial system of trees and their rotations is a fundamental one that is isomorphic to other natural combinatorial

---

Received by the editors January 29, 1988.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 51M20, 51M10, 05C05, 68P05.

This is a revised and expanded version of a paper that appeared in the 18th Annual ACM Symposium on Theory of Computing [9], Berkeley, CA, May 28–30, 1986.

Partial support provided for the first author by DARPA, ARPA order 4976, amendment 19, monitored by the Air Force Avionics Laboratory under contract F33615-87-C-1499, and by the National Science Foundation under grant CCR-8658139.

Partial support provided for the second author by the National Science Foundation under grant DCR-8605962.

Partial support provided for the third author by the National Science Foundation under grants DMR-8504984 and DCR-8505517.



systems. Results concerning this system are of interest from a purely mathematical point of view. Second, the method we use to solve the problem is novel and interesting in its own right and can potentially be applied to related problems.

A system that is isomorphic to binary trees related by rotations is that of triangulations of a polygon related by the *diagonal flip* operation. This is the operation that converts one triangulation of a polygon into another by removing a diagonal in the triangulation and adding the diagonal that subdivides the resulting quadrilateral in the opposite way. This type of move was studied by Wagner [14] in the context of arbitrary triangulated planar graphs and by Dewdney [3] in the case of graphs of genus one. They showed that any such graph can be transformed to any other by diagonal flips, but did not try to accurately estimate how many flips are necessary.

Our approach to solving the rotation distance problem is based on the observation that any sequence of diagonal flips converting one triangulation of a polygon into another gives a way to dissect (into tetrahedra) a polyhedron formed from the two triangulations. Using hyperbolic geometry, we construct polyhedra that require many tetrahedra to triangulate them. (Here and hereafter we use the word "triangulation" in a general sense meaning a dissection into simplices of appropriate dimension. A more rigorous definition is given in §2.4.) These polyhedra can be used to exhibit pairs of  $n$ -node trees (for all sufficiently large  $n$ ) such that the rotation distance between them is  $2n - 6$ .

In §2 we define the problem on trees, make the connection between trees and triangulations of a polygon, and show that sequences of diagonal flips are related to triangulations of polyhedra. In §3 we show how to use hyperbolic geometry to obtain a lower bound on the number of tetrahedra required to triangulate any polyhedron. We then construct particular polyhedra that require many tetrahedra to triangulate them. §4 contains remarks and some open problems.

## 2. DEFINITIONS AND EQUIVALENCES

**2.1. Binary trees.** A binary tree is a collection of nodes of two types, external and internal, and three relations among these nodes: parent, left child, and right child. Every node except a special one called the root has a parent, and every internal node has a left and a right child. External nodes have no children. A tree is said to be of size  $n$  if it has  $n$  internal nodes. A tree of size  $n$  has  $n + 1$  external nodes. (See [5,10] for a more complete description of binary trees and tree terminology.) The number of steps required to walk from the root of the tree to a node is the depth of that node. (Each step moves from a node to one of its children.)

A *symmetric order* traversal of the tree visits all of the nodes exactly once. This order can be described by a recursive algorithm as follows: If the node is an internal node, traverse its left subtree in symmetric order, visit the node itself, then traverse its right subtree in symmetric order. If the node is an external node, merely visit it. The order in which the nodes are visited is called the

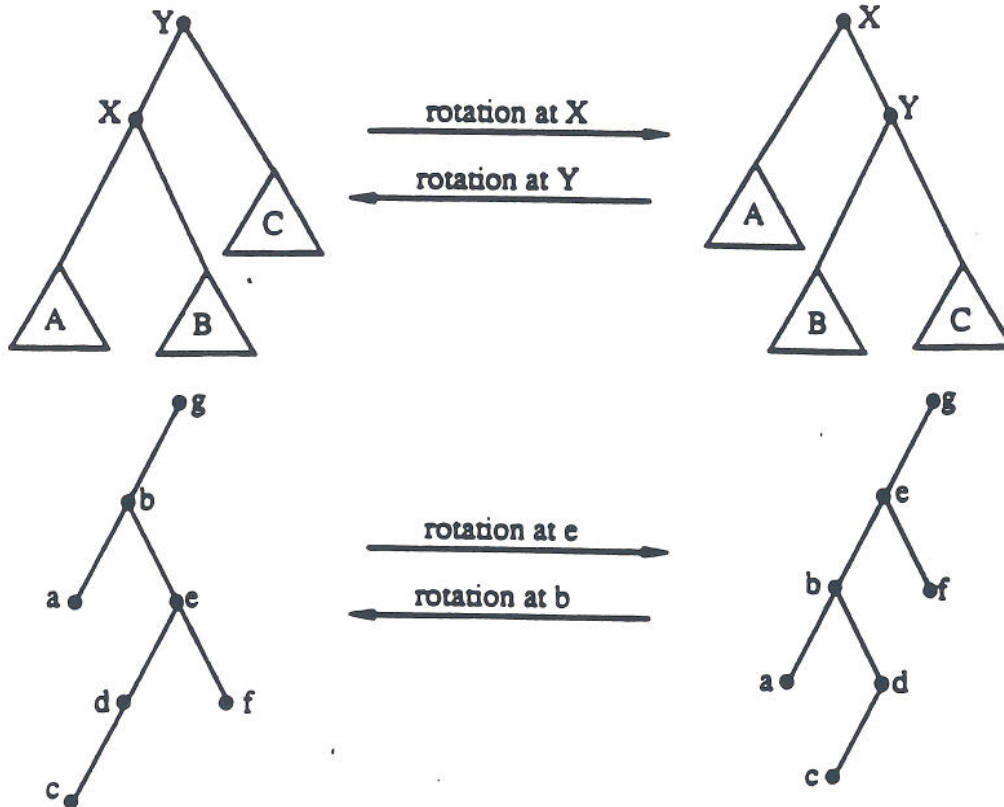


FIGURE 1. (a) The general definition of a rotation. Triangles denote subtrees. The tree shown could be part of a larger tree. (b) A rotation in a seven node tree. External nodes are not shown.

*symmetric order permutation* of the nodes (or simply the symmetric order of the nodes).

In a common computer-related application of binary trees the tree is used to store an ordered collection of pieces of information (called items). Each internal node of the tree is labeled with an item, and the order of the items is represented by the symmetric order of the nodes.

A *rotation* is an operation that changes one binary tree into another. In a tree of size  $n$  there are  $n - 1$  possible rotations, one corresponding to each nonroot internal node. Figure 1 shows the general rotation rule and the effect of a particular rotation on a particular tree. The rotation corresponding to a node changes the structure of the tree near that node, but leaves the structure elsewhere intact. A rotation maintains the symmetric order of the nodes, but changes the depths of some of them. Rotations are the primitives used by most schemes that maintain “balance” in binary trees [5, 10].

A rotation is an invertible operation; that is, if tree  $T$  can be changed into  $T'$  by a rotation, then  $T'$  can be changed back into  $T$  by a rotation. The *rotation graph* for trees of size  $n$  (denoted  $RG(n)$ ) is an undirected graph with



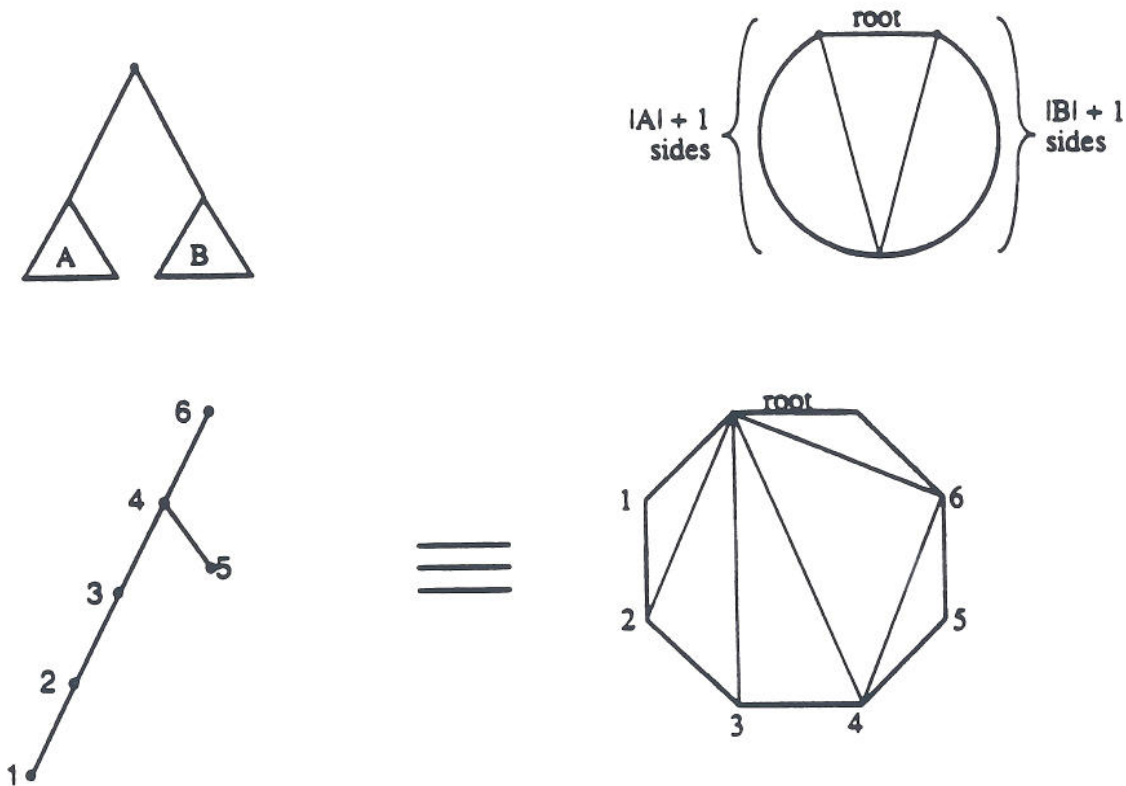


FIGURE 2. An example of a tree and its corresponding triangulation.

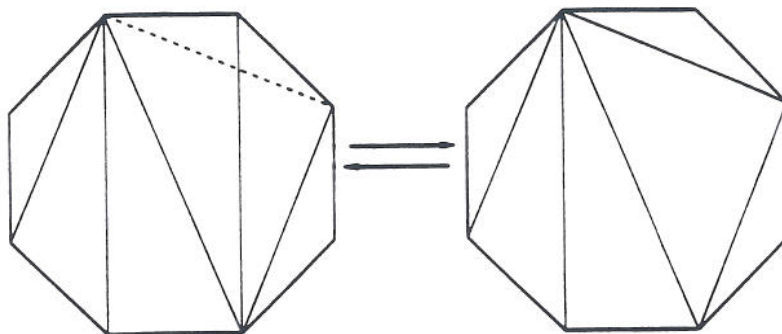


FIGURE 3. A diagonal flip in a triangulation of an octagon.

one vertex for each tree of size  $n$  and an edge between vertices  $T$  and  $T'$  if there is a rotation that changes  $T$  into  $T'$ .

Any binary tree of size  $n$  can be converted into any other by performing an appropriate sequence of rotations. Therefore the rotation graph is connected. We can define the *rotation distance* between two trees as the length of the shortest path in the rotation graph between the two trees, i.e. the minimum number of rotations required to convert one tree into the other. The main problem we address in this paper is that of estimating the diameter of  $RG(n)$ , i.e., the maximum rotation distance between any two  $n$ -node binary trees.



**2.2. Polygon triangulations.** Problems concerning rotation distance can be formulated with respect to a different system of combinatorial objects and their transformations. This alternative formulation is perhaps more natural and also seems to supply more insight.

Suppose we are given a binary tree  $T$  of size  $n$ . Take a collection of triangles indexed by the internal nodes of  $T$ . Now glue the triangles together along their edges according to the pattern of the tree, i.e., according to the parent-child relation. The resulting surface is homeomorphic to a disk. In fact, we can choose a standard convex  $(n+2)$ -gon and choose one of its edges to be the "root" edge. Label the  $n$  vertices of the  $(n+2)$ -gon that are not endpoints of the root edge by the  $n$  internal nodes of  $T$ , counterclockwise in symmetric order. Now the triangles can be inductively mapped into the  $(n+2)$ -gon by gluing one edge of each triangle to the appropriate edge of its parent triangle and sending the remaining vertex to the vertex labeled by its node. As a special case, the root triangle is attached to the root edge and the root vertex. See Figure 2.

In this way, we obtain a 1-1 correspondence between binary trees and triangulations of the  $(n+2)$ -gon with no interior vertices. We refer to the  $n+2$  sides of the polygon as *edges* and the chords that divide it into triangles as *diagonals*. Any triangulation of the  $(n+2)$ -gon has  $n-1$  diagonals and  $n$  triangles. We regard the polygon as having a distinguished edge and orientation.

A *diagonal flip* is an operation that transforms one triangulation of a polygon into another. The effect of a diagonal flip is shown in Figure 3 and can be described as follows: A diagonal inside the polygon is removed, creating a face with four sides. The opposite diagonal of this quadrilateral is inserted in place of the one removed, restoring the diagram to a triangulation of the polygon.

Let  $TG(n+2)$  be a graph with one node for each triangulation of an  $(n+2)$ -gon and an edge between two nodes if the two nodes are related by a diagonal flip. We see that

**Lemma 1.** *The graph  $TG(n+2)$  is isomorphic to the rotation graph  $RG(n)$ .*

The proof of this lemma is a straightforward application of the correspondence described above. A more detailed discussion of the relationships between trees, triangulations, and Catalan numbers can be found in Chapter 20 of [4].

**2.3. Results of polygon triangulations.** As we saw in §2.2, a study of the rotation distance between trees can be formulated as a study of the distance between triangulations under the diagonal flip operation. Let  $d(\tau_1, \tau_2)$  be the minimum number of diagonal flips needed to transform triangulation  $\tau_1$  into triangulation  $\tau_2$ . For convenience, we shall now change our use of the variable " $n$ ." We consider triangulations of an  $n$ -gon and let  $d(n)$  be the maximum distance between any pair of such triangulations. That is,  $d(n)$  is the diameter of  $TG(n)$  or equivalently of  $RG(n-2)$ . Figure 4 shows  $TG(6)$ , whose vertices are the fourteen triangulations of a hexagon. The greatest distance between a pair of triangulations is four: there are several pairs that achieve this distance.



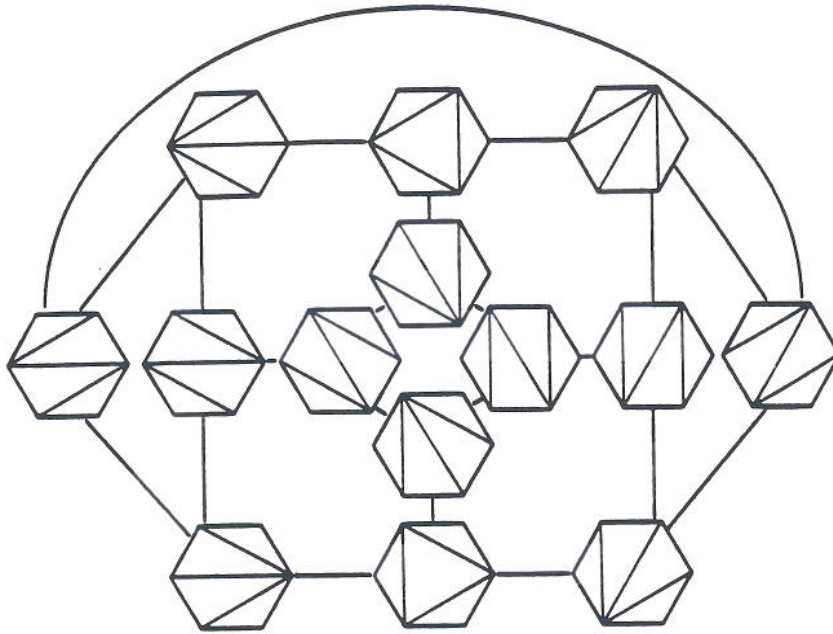


FIGURE 4. The rotation graph of a hexagon,  $RG(6)$ .

The added symmetry revealed in the triangulation system that is hidden in the binary tree system enables us to improve Culik and Wood's upper bound on  $d(n)$  from  $2n - 6$  to  $2n - 10$ .

**Lemma 2.**  $d(n) \leq 2n - 10$  for all  $n > 12$ .

*Proof.* Any triangulation of an  $n$ -gon has  $n - 3$  diagonals. Given any vertex  $x$  of degree  $\deg(x) < n - 3$ , we can increase  $\deg(x)$  by one by a suitable flip. Thus in  $n - 3 - \deg(x)$  flips we can produce the unique triangulation all of whose diagonals have one end at  $x$ . It follows that given any two triangulations  $\tau_1$  and  $\tau_2$  we can convert  $\tau_1$  into  $\tau_2$  in  $2n - 6 - \deg_1(x) - \deg_2(x)$  flips, where  $x$  is any vertex and the degree of  $x$  is  $\deg_1(x)$  in  $\tau_1$  and  $\deg_2(x)$  in  $\tau_2$ . The average over vertices  $x$  of  $\deg_1(x)$  is  $2 - 6/n$ , and of  $\deg_1(x) + \deg_2(x)$  is  $4 - 12/n$ . It follows that if  $n > 12$ , there is a vertex  $x$  such that  $\deg_1(x) + \deg_2(x) \geq 4$ .

□

The following lemma about sequences of diagonal flips shows that in some situations it is easy to find the first flip in an optimal sequence of flips.

**Lemma 3.** (a) If it is possible to flip one diagonal of  $\tau_1$  creating  $\tau'_1$  so that  $\tau'_1$  has one more diagonal in common with  $\tau_2$  than does  $\tau_1$ , then there exists a shortest path from  $\tau_1$  to  $\tau_2$  in which the first flip creates  $\tau'_1$ . (b) If  $\tau_1$  and  $\tau_2$  have a diagonal in common, then a shortest path from  $\tau_1$  to  $\tau_2$  never flips this diagonal. In fact, any path that flips this diagonal is at least two flips longer than a shortest path.

*Proof.* Let  $S$  be a sequence of adjacent triangulations connecting  $\tau_1$  to  $\tau_2$ .

$$S = t_0 (= \tau_1), t_1, t_2, \dots, t_k (= \tau_2).$$



Assume that  $t_1 \neq \tau'_1$ . We shall construct a new sequence of adjacent triangulations  $S'$  also connecting  $\tau_1$  and  $\tau_2$  whose length is no longer than the length of  $S$  and in which the first flip creates  $\tau'_1$ . This will suffice to prove (a) of the lemma.

Let  $l$  and  $r$  be the endpoints of the diagonal that  $\tau'_1$  and  $\tau_2$  have in common but  $\tau_1$  and  $\tau_2$  do not have in common. Any triangulation  $\tau$  can be *normalized* with respect to the diagonal  $(l, r)$  to create a new triangulation  $N(\tau)$ . The diagonals of  $N(\tau)$  are of three types: (1)  $N(\tau)$  contains the diagonal  $(l, r)$ , (2)  $N(\tau)$  contains every diagonal of  $\tau$  that does not cross the diagonal  $(l, r)$  (two diagonals with an endpoint in common are not said to cross), (3) if  $\tau$  contains a diagonal  $(a, b)$  that crosses the diagonal  $(l, r)$  then  $N(\tau)$  contains the diagonals  $(a, r)$  and  $(b, r)$ . (See Figure 5.)

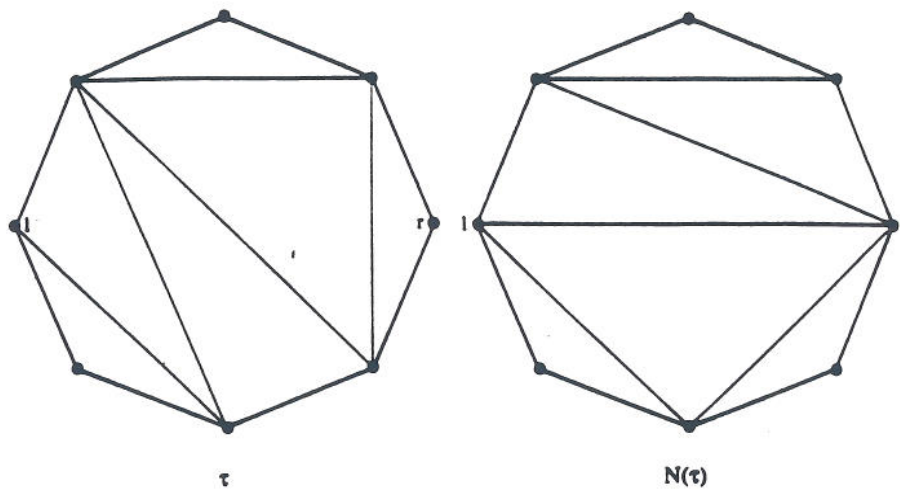


FIGURE 5. A triangulation  $\tau$  and its normalized version  $N(\tau)$ .

Consider the sequence of triangulations

$$N = t_0, N(t_0), N(t_1), \dots, N(t_k).$$

A straightforward case analysis shows that successive triangulations of this sequence are either identical or adjacent. Eliminating all but one of each group of identical consecutive triangulations in this sequence gives the desired sequence  $S'$ . A priori  $S'$  might contain  $k+2$  triangulations, but this cannot be the case for the following reason. Consider the triangulations  $t_i$  and  $t_{i+1}$  in  $S$  with the property that  $t_i$  does not contain diagonal  $(l, r)$  and  $t_{i+1}$  does. (There must be such a pair since the final triangulation contains the diagonal  $(l, r)$  and the initial one does not.) It is easily verified that the triangulations  $N(t_i)$  and  $N(t_{i+1})$  must be equal, and therefore occur only once in  $S'$ . Thus  $S'$  contains at most  $k+1$  triangulations. Verifying that  $S'$  starts and ends with  $\tau_1$  and  $\tau_2$ , and that its second triangulation is  $\tau'_1$ , is straightforward. This completes the proof of (a) of the lemma.



The same technique serves to prove (b) of the lemma. Let  $S = t_0 (= \tau_1)$ ,  $t_1, \dots, t_k (= \tau_2)$  be a sequence that transforms  $\tau_1$  into  $\tau_2$  in which the first move is one that flips a diagonal  $(l, r)$  common to both  $\tau_1$  and  $\tau_2$ . Normalize this sequence with respect to the diagonal  $(l, r)$  and eliminate redundancies to create a sequence  $S'$ . Then  $S'$  transforms  $\tau_1$  into  $\tau_2$  in two fewer flips than does  $S$ . The reason is that neither the first flip of  $S$ , which misaligns  $(l, r)$ , nor a later flip that aligns  $(l, r)$ , occurs in  $S'$ .  $\square$

A refinement of the lower bound proof in Lemma 2 for small values of  $n$  and a computer search have produced the exact values of  $d(n)$  for  $n \leq 18$ , which appear below.

$n$	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$d(n)$	0	1	2	4	5	7	9	11	12	15	16	18	20	22	24	26

**2.4. Simplicial complexes.** This section is a digression into the notation and terminology of simplicial complexes. Although we will apply these concepts in at most three dimensions, we have developed them in full  $n$ -dimensional generality. The sole purpose of this and the next section is to develop tools required to prove Lemma 7. At the end of §2.5 is an example that illustrates many of the concepts about to be introduced.

An  $n$ -simplex is the  $n$ -dimensional generalization of a triangle. It has  $n + 1$  vertices and  $n + 1$  faces each of which is an  $n - 1$ -simplex. The simplex can be thought of as a list of the names of its vertices (in sorted order by vertex name).

A *simplicial complex*  $K$  is the union of a collection of simplices of assorted dimensions, called *cells*. Whenever  $\alpha$  is an  $n$ -simplex in the collection and  $\beta$  is a  $k$ -face of  $\alpha$ , then  $\beta$  is also in the collection. The intersection of any two simplices in the collection is also a simplex in the collection. The *dimension* of  $K$  is the maximum dimension of a simplex of  $K$ . The  $k$ -skeleton  $K_k$  of  $K$  is the union of the simplices of dimension not exceeding  $k$ , with its inherited structure as a simplicial complex.

A *triangulation* of a space  $X$  is a simplicial complex  $K$  and a map  $f$  from simplices onto  $X$  such that the images of the simplices in  $X$  intersect exactly as they do in  $K$ .

An *orientation* for a  $k$ -simplex is an ordering of its vertices. An oriented simplex thus consists of a sorted list of the names of its vertices, in addition to a permutation of its vertex names. The orientation is said to be positive if the permutation of the vertices is an even permutation, and negative otherwise.

Associated with a simplicial complex  $K$  of dimension  $n$  is a collection of vector spaces  $C_k(K)$ , for  $0 \leq k \leq n$ , where  $C_k$  is defined to consist of finite formal linear combinations of oriented  $k$ -simplices of  $K$  with real coefficients, subject to the relation that two oriented  $k$ -simplices whose orientations differ by a single inversion are negatives of each other. Elements of  $C_k$  are called



$k$ -chains of  $K$ . Using the above relation we can reduce a  $k$ -chain to an equivalent one in which each simplex of  $K$  occurs at most once.

The *boundary map*

$$\partial_k: C_k \rightarrow C_{k-1},$$

is a linear map from  $k$ -chains to  $(k - 1)$ -chains. Because it is linear, its value on an arbitrary  $k$ -chain is a linear combination of its values on each simplex of the given  $k$ -chain. The boundary map is defined on an oriented simplex to be its oriented boundary: to determine the orientation on a  $(k - 1)$ -face  $\beta$  of a  $k$ -simplex  $\alpha$ , choose an ordering of the vertices of  $\alpha$  compatible with the orientation and such that the vertices of  $\beta$  come last. The induced ordering of the vertices of  $\beta$  is the correct orientation. As a special case, we define  $\partial_0$  as a trivial map to a trivial vector space consisting of 0 alone.

Elements of the kernel of  $\partial_k$  (that is, chains whose boundary is 0) are called  $k$ -cycles. The image of  $\partial_{k+1}$  automatically is contained in the kernel of  $\partial_k$  because the boundary of a simplex is a cycle. (Elements of this image are called  $k$ -boundaries. The quotient space of  $k$ -cycles by  $k$ -boundaries is called the  $k$ th homology of  $K$ ,  $H_k(K; R)$ . This does not depend on the structure of  $K$  as a simplicial complex, but only on the underlying space of  $K$ .)

A  $k$ -simplex is said to be *geodesic* if it is contained in an affine  $k$ -dimensional subspace of  $R^n$  (but not in any  $(k - 1)$ -dimensional subspace of  $R^n$ ), and all of its  $(k - 1)$ -faces are also geodesic. A map  $f$  is said to be *affine* on a  $k$ -simplex if the image of the simplex under  $f$  is geodesic.

Let  $f$  be an affine map that maps an  $n$ -simplex  $s$  into an  $n$ -simplex  $s'$  in  $R^n$ . We may assume that the domain of  $f$  is simplex  $s$ , placed in  $R^n$  in the canonical fashion. (The canonical placement of an  $n$ -simplex in  $R^n$  is obtained by taking the convex body formed by convex combinations of the origin and the  $n$  unit vectors. To label the canonical placement, place the labels in sorted order on the origin, then on each of the unit vectors of  $R^n$  respectively.) Each vertex of  $s'$  may be labeled according to which vertex of  $s$  is mapped to it by  $f$ . The map  $f$  is said to be *orientation preserving* if the canonical placement of  $s$  in  $R^n$  can be transformed to  $s'$  via a continuous family of affine embeddings of  $s$  in  $R^n$  so that the labels on the vertices match. If this is not possible, then  $f$  is said to be *orientation inverting*.

**2.5. The volume of an  $(n - 1)$ -cycle.** We shall now develop the tools necessary to define the volume of an  $(n - 1)$ -cycle embedded in  $R^n$ .

Suppose that we have a simplicial complex  $K$ , an  $n$ -chain  $c$  in  $C_n(K)$ , and a map  $f$  of  $K$  into  $R^n$  that is affine on each simplex. Choose a basis  $\alpha_1, \alpha_2, \dots$  of  $C_n(K)$  such that the orientation of  $\alpha_i$  is positive if  $f$  is orientation preserving on  $\alpha_i$ , and negative otherwise. (This definition of orientation has the property that all the basis simplices have the "same" orientation. For example, if we are working in two dimensions, then the orientation of each triangle is either clockwise or counterclockwise, and in this case our basis has the property that all the basis triangles are counterclockwise.)



This data determines a certain step function  $\text{degree}(c, f)$  on  $R^n$ , as follows. Write  $c = \sum a_i \alpha_i$ . First suppose that  $x$  is any point in  $R^n$  which is not in the image of the  $n - 1$  skeleton of  $K$ . To compute the degree at  $x$ , we add the coefficients of the simplices whose image contains the point  $x$ . If  $x$  is in the image of the  $n - 1$  skeleton, we define  $\text{degree}(c, f)(x)$  to be the maximum value of the degree that occurs in arbitrarily small neighborhoods of  $x$ .

The degree map we have defined satisfies the following lemma, which makes it useful in defining volumes.

**Lemma 4.** *If  $c$  is an  $n$ -cycle in a simplicial complex  $K$ , then  $\text{degree}(c, f) = 0$ . Furthermore if  $d$  and  $e$  are any chains such that  $\partial d = \partial e$ , then  $\text{degree}(d, f) = \text{degree}(e, f)$ .*

*Proof.* The degree function  $\text{degree}(c, f)$  is constant everywhere except possibly in the image of the union of the boundaries of the  $\alpha_i$ . We shall first show that crossing these boundaries also leaves this function constant. Let  $\alpha$  and  $\alpha'$  be two simplices from the basis that have a boundary simplex  $\beta$  in common. If  $\alpha$  and  $\alpha'$  are mapped by  $f$  to the same side of  $\beta$  then the coefficient of  $\beta$  in  $\partial\alpha$  is the same as the coefficient of  $\beta$  in  $\partial\alpha'$ . If  $\alpha$  and  $\alpha'$  are mapped to opposite sides of  $\beta$ , then the coefficients are negatives of each other. Since  $c$  is a cycle, the boundary vanishes, so the coefficient of  $\beta$  in the boundary is zero. Therefore the sum of the coefficients in  $c$  over those  $n$ -simplices that map to one side of  $\beta$  equals the sum of the coefficients in  $c$  over those that map to the other. This shows that the degree is the same on both sides of  $\beta$ .

If  $f$  is a generic map (where  $f$  does not have degeneracies, such as mapping several  $(n - 1)$ -simplices to the same place) this is enough to show that  $\text{degree}(c, f)$  is continuous everywhere (and therefore constant). This is because in this case one can get from any generic point to any other generic point only crossing the image of a single  $(n - 1)$ -face at a time. Since the degree is zero near infinity, the degree must be zero everywhere. In the general case, when  $f$  might not be generic,  $f$  can be perturbed a little, without changing the degree at most points. The degree for the perturbations must be identically zero, so  $\text{degree}(c, f) = 0$  in this case as well.

As for the second assertion of the lemma, if  $d$  and  $e$  are chains such that  $\partial d = \partial e$ , then  $d - e$  is a cycle, so by the first part of the lemma  $\text{degree}(d - e, f) = 0$ . Because the degree operator is linear,  $\text{degree}(d, f) - \text{degree}(e, f) = \text{degree}(d - e, f)$ , which gives the result.  $\square$

If  $z$  is an  $n - 1$  cycle in a simplicial complex  $K$ , and if  $f$  is a map of  $K$  into  $R^n$ , it is always possible to enlarge  $K$  to a simplicial complex  $K'$  in which there are chains  $c$  for which  $z = \partial c$ . The easiest way is to define  $K'$  as the cone on  $K$ ; that is,  $K'$  has one more vertex  $v$  than  $K$ , and for each simplex  $\alpha$  of dimension  $k$  there is an additional  $(k + 1)$ -simplex with one face on  $\alpha$  and its extra vertex at  $v$ . The map of  $K$  into  $R^n$  easily extends to a map of  $K'$  determined by choosing where to send  $v$ . Now consider the



map  $\text{degree}(c, f)$ . Lemma 4 tells us that this step function only depends on  $z$ , not on the way  $z$  is expressed as a boundary. Thus we can rewrite it as  $w(z, f)(x) = \text{degree}(c, f)(x)$  and call it the *wrapping number* of  $z$  about  $x$ . The *algebraic volume* enclosed by  $f(z)$  is defined to be the integral of the wrapping number.

A special case that will become important later is the volume of a three-dimensional polyhedron. A triangulation  $\sigma$  of the sphere is a way to dissect the sphere into curvilinear triangles. Each such  $\sigma$  has a *fundamental 2-cycle*, defined to be the sum of the triangles of  $\sigma$  with orientation coming from the counterclockwise ordering of the vertices (when looking at the sphere from the outside). If  $f$  maps the sphere into  $R^3$  in such a way that each simplex of  $\sigma$  is geodesic, then we define the volume enclosed by  $f$  to be the volume enclosed by the fundamental 2-cycle of  $\sigma$ .

Figure 6 illustrates a simplicial complex  $K$  with 15 simplices:  $\langle 123 \rangle$ ,  $\langle 234 \rangle$ ,  $\langle 235 \rangle$ ,  $\langle 12 \rangle$ ,  $\langle 13 \rangle$ ,  $\langle 23 \rangle$ ,  $\langle 24 \rangle$ ,  $\langle 34 \rangle$ ,  $\langle 35 \rangle$ ,  $\langle 25 \rangle$ ,  $\langle 1 \rangle$ ,  $\langle 2 \rangle$ ,  $\langle 3 \rangle$ ,  $\langle 4 \rangle$ ,  $\langle 5 \rangle$ . (Here the numbers between brackets are the vertices of the simplex in sorted order.) The way these are mapped into  $R^2$  by a map  $f$  is shown in (a) of the figure. The canonical imbedding of simplices  $\langle 123 \rangle$ ,  $\langle 234 \rangle$ , and  $\langle 235 \rangle$  into  $R^2$  is shown in (b). The map  $f$  is orientation inverting on simplices  $\langle 123 \rangle$  and  $\langle 234 \rangle$ , and orientation preserving on simplex  $\langle 235 \rangle$ .

To define the degree of  $f$ , we could use the following basis for  $C_2(K)$ :  $\alpha_1 = \langle 123 \rangle(132)$ ,  $\alpha_2 = \langle 234 \rangle(243)$ ,  $\alpha_3 = \langle 235 \rangle(235)$ . (The permutation for an oriented simplex is shown in parentheses after the sorted vertex list of the simplex.) Other bases are possible; for example, we could have chosen  $\alpha_1$  to be  $\langle 123 \rangle(321)$ .

Let  $c$  be the chain  $-\langle 123 \rangle(132) + \langle 234 \rangle(243) + 2\langle 235 \rangle(235)$ . Part (c) of Figure 6 shows the value of  $\text{degree}(c, f)$  in  $R^2$ . Note that

$$\partial \langle 123 \rangle(132) = \langle 12 \rangle(21) + \langle 13 \rangle(13) + \langle 23 \rangle(32),$$

and

$$\partial c = \langle 12 \rangle(21) + \langle 13 \rangle(13) + \langle 34 \rangle(34) + \langle 24 \rangle(42) + 2\langle 23 \rangle(23) + 2\langle 35 \rangle(35) + 2\langle 25 \rangle(52).$$

**2.6. Triangulations of the sphere and the ball.** In this section we show that the quantity  $d(n)$  is related to the number of tetrahedra that are required to triangulate certain polyhedra.

Let  $\sigma$  be a triangulation of the sphere and let  $z$  be a fundamental 2-cycle of  $\sigma$ . Then  $T$  is an *exposed triangulation of the ball extending  $\sigma$*  if (1)  $T$  is a triangulation of the ball, (2) there exists a 3-chain  $c$  in  $C_3(T)$  such that  $z = \partial c$  and all the coefficients of  $c$  are  $\pm 1$ , and (3) all of the vertices of  $T$  occur as vertices in  $\sigma$ . The 3-chain  $c$  satisfying this definition is unique.

An exposed triangulation of the ball extending  $\sigma$  is the three dimensional analogue of a triangulation of an  $n$ -gon described in §2.2. In contrast to the situation in two dimensions, there are generally triangulations extending  $\sigma$  containing different numbers of tetrahedra. Although it happens to be true that



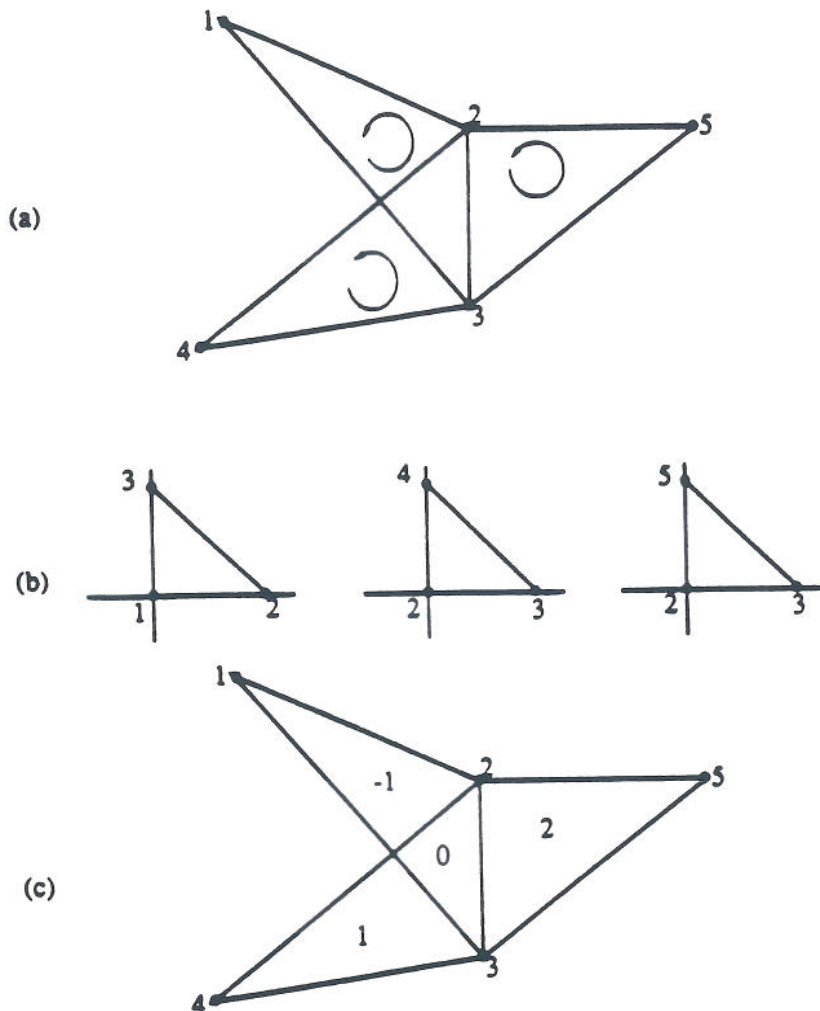


FIGURE 6. (a) shows the way a simplicial complex with fifteen simplices is mapped by a map  $f$  into  $R^2$ . (b) shows the canonical imbedding of simplices  $\langle 123 \rangle$ ,  $\langle 234 \rangle$ , and  $\langle 235 \rangle$  into  $R^2$ . (The map  $f$  is orientation inverting on simplices  $\langle 123 \rangle$  and  $\langle 234 \rangle$ , and orientation preserving on simplex  $\langle 235 \rangle$ .) (c) shows the value of degree  $(c, f)$  in  $R^2$ .

every triangulation of  $S^2$  is homeomorphic to a triangulation with geodesic faces, the situation is quite different from exposed triangulations of the ball: it is even possible to construct examples with knotted edges, so that they are forcibly curvilinear.

The union of two triangulations  $\tau_1$  and  $\tau_2$  of an  $n$ -gon, glued along their boundary, usually gives a triangulation of the sphere, which we denote by  $U(\tau_1, \tau_2)$ . (There are certain degenerate cases in which the union does not satisfy the definition of a simplicial complex. This happens if  $\tau_1$  and  $\tau_2$  have



a diagonal in common. Nonetheless,  $U(\tau_1, \tau_2)$  describes a triangulation of a slightly more general sort, which we shall not discuss.)

For any  $\sigma$  there is an exposed triangulation of the ball extending  $\sigma$ . This allows us to define  $t(\sigma)$  to be the minimum number of tetrahedra in any exposed triangulation of the ball extending  $\sigma$ . The following lemma relates triangulations to rotation distance.

**Lemma 5.** *If  $\tau_1$  and  $\tau_2$  have no diagonal in common, then*

$$t(U(\tau_1, \tau_2)) \leq d(\tau_1, \tau_2).$$

*Proof.* There exists a sequence of  $d(\tau_1, \tau_2)$  diagonal flips that changes  $\tau_1$  into  $\tau_2$ . We shall describe how to extract from this sequence an exposed triangulation of the ball extending  $U(\tau_1, \tau_2)$  containing  $d(\tau_1, \tau_2)$  tetrahedra.

Imagine that there is a planar base with triangulation  $\tau_1$  drawn on it. Suppose the first diagonal flip replaces diagonal  $(a, c)$  with diagonal  $(b, d)$ . Create a flat quadrilateral that is the same shape as quadrilateral  $(a, b, c, d)$ . On the back side of the quadrilateral draw diagonal  $(a, c)$ . On the front draw diagonal  $(c, d)$ . Now place the quadrilateral onto the base in the appropriate place with diagonal  $(a, c)$  down and  $(b, d)$  up. Looking at the base we see a picture of a triangulation which is the result of making the first diagonal flip. For each successive move we create another quadrilateral and place it onto the base. After placing  $d(\tau_1, \tau_2)$  such quadrilaterals we will see  $\tau_2$  when we view the base.

The triangulation of the ball that we construct has one tetrahedron for each quadrilateral. The tetrahedra are glued together according to the way the quadrilaterals are stacked. Two triangles are identified with each other if they face each other in the stack of quadrilaterals.

To finish the proof we need only verify that the resulting triangulation is an exposed triangulation of the ball extending  $U(\tau_1, \tau_2)$ . The fact that it is a triangulation of the ball is made clear by inflating each quadrilateral so that it turns into a tetrahedron. The resulting collection of tetrahedra is homeomorphic to a ball. (This is where we use the assumption that  $\tau_1$  and  $\tau_2$  have no diagonal in common.) The fact that the triangulation extends  $U(\tau_1, \tau_2)$  is obvious because the boundary triangles are exactly those of  $\tau_1$  and  $\tau_2$ .  $\square$

What the proof of Lemma 5 tells us is that for every sequence of diagonal flips from  $\tau_1$  to  $\tau_2$  there is an exposed triangulation of the ball extending  $U(\tau_1, \tau_2)$ . In fact, the same triangulation of the ball may result from many different sequences of moves from  $\tau_1$  to  $\tau_2$ . It is not the case that every exposed triangulation of the ball extending  $U(\tau_1, \tau_2)$  comes from a sequence of diagonal flips. In fact, it is possible to construct exposed triangulations of the ball with the property that no tetrahedron touches the boundary on more than one face, whereas in a triangulation obtained by the construction in the proof of Lemma 5 some tetrahedra touch the boundary on at least two faces.

Let  $t(n)$  be the maximum of the quantity  $t(\sigma)$  over all  $n$ -vertex four-connected triangulations  $\sigma$  of the sphere. (A triangulation is said to be



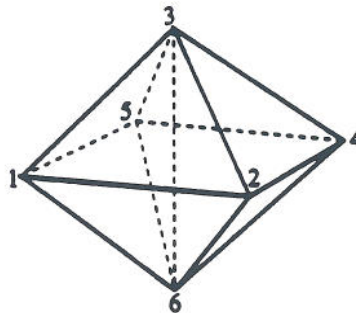
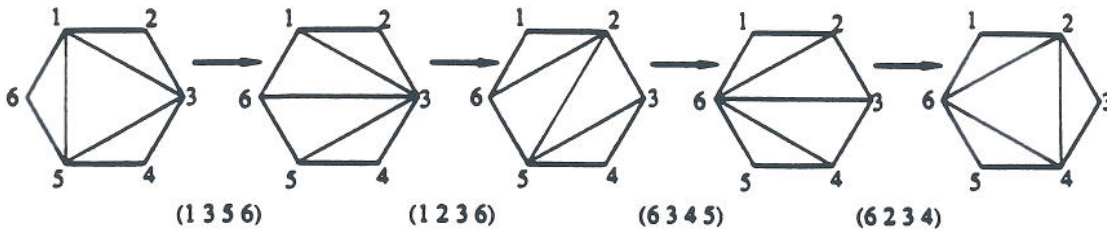
$k$ -connected if its 1-skeleton, regarded as a graph, is  $k$ -connected. An undirected graph is  $k$ -connected if deletion of any  $k - 1$  vertices leaves the graph connected.)

**Lemma 6.**  $t(n) \leq d(n)$ .

*Proof.* Let  $\sigma$  be an  $n$ -vertex four-connected triangulation of the sphere such that  $t(\sigma)$  is maximized. By a theorem of Hassler Whitney [15] or its generalization, Tutte's Theorem [13], any four-connected triangulated planar graph must have a Hamiltonian circuit. Draw the triangulation on a sphere. Cut the sphere along the edges of the Hamiltonian circuit. This separates the sphere into two disks, each of which is triangulated. Let these two triangulations be  $\tau_1$  and  $\tau_2$ . Now  $\sigma = U(\tau_1, \tau_2)$ . By the preceding discussion and Lemma 5,

$$t(n) = t(\sigma) = t(U(\tau_1, \tau_2)) \leq d(\tau_1, \tau_2) \leq d(n). \quad \square$$

To make these concepts more concrete, consider the two triangulations  $\tau_1$  and  $\tau_2$  of a hexagon whose diagonals form a triangle. (See Figure 7.) The triangulation obtained by gluing  $\tau_1$  and  $\tau_2$  together is the boundary of the octahedron. There are six paths of length four between  $\tau_1$  and  $\tau_2$  (see Figure 4). Each of these paths gives rise to a triangulation of the octahedron. Three different triangulations of the octahedron are obtained in this way. (Each is produced by two different paths from  $\tau_1$  to  $\tau_2$ .) These triangulations are the ones in which a single edge has been added between a pair of opposite vertices. The octahedron cannot be triangulated with fewer than four tetrahedra because no tetrahedron can contact more than two faces of the boundary.



The four tetrahedra: 3 6 1 5, 3 6 2 1, 3 6 4 2, 3 6 5 4

FIGURE 7. A sequence of diagonal flips and the corresponding triangulation.



By Lemma 6, an upper bound on  $d(n)$  is an upper bound on  $t(n)$ , and a lower bound on  $t(n)$  is a lower bound on  $d(n)$ . In the remainder of this paper we show that  $2n - 10 \leq t(n)$  for all sufficiently large  $n$ . Combining this with Lemma 2 gives  $t(n) = d(n) = 2n - 10$  for all sufficiently large  $n$ .

### 3. LOWER BOUNDS ON $t(n)$

Our approach to deriving accurate lower bounds on  $t(n)$  is geometric rather than combinatorial. We convert the combinatorial objects described in the previous section into geometric objects. We then infer properties of the combinatorial objects from the properties of the geometric objects. The following two paragraphs summarize our approach.

Let  $\sigma$  be an  $n$ -vertex triangulation of the sphere that is four-connected. Suppose  $\sigma$  is the boundary of a polyhedron  $P$  in  $R^3$  such that the vertices of  $P$  are on the unit sphere. Let  $T$  be an exposed triangulation of the ball extending  $\sigma$ . For each tetrahedron  $\Delta$  of  $T$  there is a geodesic tetrahedron  $\Delta'$  whose vertices are the appropriate vertices of  $P$ . The sums of the volumes of these geodesic tetrahedra must be at least the algebraic volume enclosed by  $P$ . Let  $V_\Delta$  be the volume of the largest tetrahedron that can be inscribed in a sphere. Let  $\text{vol}(P)$  be the volume of  $P$ . We conclude that at least  $\text{vol}(P)/V_\Delta$  tetrahedra are required to cover  $P$ . In other words:

$$\frac{\text{vol}(P)}{V_\Delta} \leq t(n).$$

In Euclidean space, this inequality does not give interesting bounds because the ratio of the volume of a sphere to the volume of the largest tetrahedron inscribed in the sphere is a small constant. In hyperbolic space, however, this method does lead to useful results. This is because in hyperbolic space the volume of a tetrahedron is bounded above by a constant  $V_0$ , while the volume of a polyhedron can grow linearly as a function of the number of vertices. Our problem is thus reduced to finding a polyhedron  $P$  with  $n$  vertices in hyperbolic space that has large volume.

First we present the necessary fundamentals of hyperbolic geometry. These ideas are described in more detail in Coxeter's book [1], Milnor's paper [6], and an expository article by Thurston and Weeks [11].

**3.1. Hyperbolic geometry.** In hyperbolic geometry there are many lines through a given point parallel to a given line, the sum of the angles of a triangle is less than 180 degrees, and the circumference of a circle is greater than  $\pi$  times the diameter. There are various ways of mapping hyperbolic space into Euclidean space. These mappings enable us to draw pictures on Euclidean paper of hyperbolic polygons, but these pictures are distorted; two congruent hyperbolic triangles may not look congruent when mapped into Euclidean space.

One mapping of two-dimensional hyperbolic space into the Euclidean plane is called the *upper half-plane model*. In this model all of hyperbolic space is mapped into the upper half of the complex plane (the points with positive



imaginary parts). This mapping is conformal, which means that angles are preserved. The geodesics (straight lines) in hyperbolic space are mapped into the semicircles with centers on the real axis and the vertical half lines with ends on the real axis. Most of the area of hyperbolic space is mapped into the region near the real axis. See Figure 8.

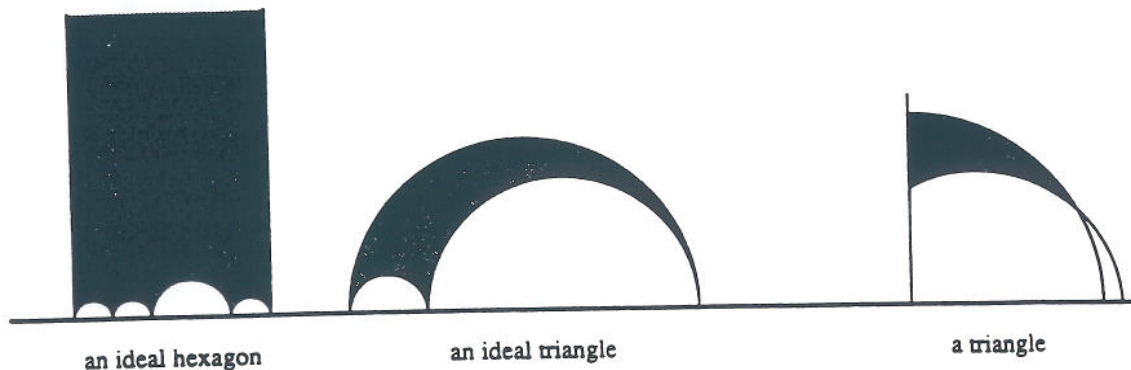


FIGURE 8. The upper half-plane model of two-dimensional hyperbolic space.

The area of a triangle in hyperbolic space is  $\pi - \Sigma$ , where  $\Sigma$  is the sum of the internal angles. An *ideal triangle* is one with three distinct vertices on the real axis or at infinity. All ideal triangles have area  $\pi$ . In fact all ideal triangles are congruent, that is, any ideal triangle can be transformed to any other by a rigid motion. (The rigid motions of the space form a group known as the group of orientation-preserving isometries.)

The upper half-space model of three-dimensional hyperbolic space consists of the complex plane plus all the points above the plane in Euclidean three-space plus a point at infinity. The complex plane plus the point at infinity is sometimes called the *sphere at infinity*. A geodesic in hyperbolic three-space is mapped to a semicircle perpendicular to the complex plane or a straight half line perpendicular to the complex plane going to infinity. The geodesic surfaces are mapped to hemispheres with centers in the complex plane and half planes orthogonal to the complex plane.

An *ideal hyperbolic tetrahedron* is a tetrahedron in which all the vertices are distinct and on the sphere at infinity. Any hyperbolic tetrahedron can be transformed by a rigid motion to one in which three of the vertices are at 0, 1, and  $\infty$  and the other vertex is at a point  $z$  in the complex plane. (This motion is possible because all four of the triangles of the tetrahedron are ideal and any ideal triangle can be moved to any other. Note that despite this fact, not all ideal hyperbolic tetrahedra are congruent.) The tetrahedron then looks like three vertical flat walls above the Euclidean triangle  $(0, 1, z)$ , bounded below by part of a hemispherical bubble.

The hyperbolic cross section of the vertical chimney, in the hyperbolic metric, scales in a way that decreases with increasing height. (Most of the volume of hyperbolic space is near the complex plane.) It can be seen by integrating



that the volume of an ideal tetrahedron is finite. Let this volume be denoted by  $v(z)$ . There are explicit formulas for  $v(z)$ , from which it can be seen that the maximum is attained at the point  $z = \omega$ , where  $\omega$  is defined as  $\omega = e^{2\pi i/6}$ . (See [6] for a discussion of how to compute hyperbolic volumes.) The tetrahedron of maximum volume is the most symmetrical one. Its base triangle  $(0, 1, \omega)$  is equilateral, its dihedral angles are all 60 degrees, and its volume is  $v(\omega) = V_0 = 1.014941606 \dots$ .

**3.2. The volume of hyperbolic polyhedra.** Let  $\sigma$  be a triangulation of the sphere. For concreteness think of  $\sigma$  as though it is embedded in the sphere in some particular way. For any mapping of the vertices of  $\sigma$  to distinct points in three-dimensional hyperbolic space (or Euclidean space for the purposes of this discussion), there is a continuous map  $f$  from the sphere into hyperbolic space that (1) maps the vertices of  $\sigma$  to the appropriate places, and (2) maps every triangle of  $\sigma$  one-to-one onto a geodesic triangle in hyperbolic space. (Just as in Euclidean space, a simplex is *geodesic* in hyperbolic space if it and all of its subsimplices are in geodesic surfaces of appropriate dimension.) In other words,  $f$  maps the surface of the sphere into the surface of some hyperbolic polyhedron.

**Lemma 7.** *Let  $\sigma$  be a triangulation of the sphere. Let  $z$  be the fundamental 2-cycle of  $\sigma$  (the sum of all its triangles with counterclockwise orientation). Let  $f$  be a map from  $\sigma$  into hyperbolic 3-space that maps each triangle of  $\sigma$  into a geodesic triangle. Let  $P$  be the hyperbolic polyhedron defined by  $f(\sigma)$ . Let  $\text{vol}(P)$  be the hyperbolic algebraic volume enclosed by  $f(z)$ . Then*

$$\frac{\text{vol}(P)}{V_0} \leq t(\sigma).$$

*Proof.* Let  $T$  be an exposed triangulation of the ball extending  $\sigma$  containing  $t(\sigma)$  tetrahedra. Let  $f$  be a map from the vertices of  $\sigma$  to the vertices of  $P$  as described above, with the additional property that  $f$  maps all of the triangles of all of the tetrahedra of  $T$  to geodesic triangles in hyperbolic space.

By the definition of  $T$ , there is a 3-chain  $c$  on  $T$  such that  $\partial c = z$ , and the coefficients of  $c$  are  $\pm 1$ . The algebraic volume of  $P$  is the integral of the wrapping number  $w(z, f) = \text{degree}(c, f)$ . Since the degree map is a linear operator on  $c$  and integration is a linear operator, we may separate this calculation into the sum of several terms, one for each tetrahedron of  $T$ . The contribution to this sum of each tetrahedron of  $T$  is its coefficient times the volume of the tetrahedron (degree is  $\pm 1$  inside, 0 outside). No term exceeds  $V_0$ , hence the lemma is true.  $\square$

We have now reduced the problem of finding lower bounds on  $t(n)$  to that of finding  $n$ -vertex hyperbolic polyhedra with large volumes. The remainder of this section is devoted to constructing such polyhedra.



**3.3. A preliminary bound:**  $2n - O(n^{1/2})$ . There is a tessellation of hyperbolic space consisting of copies of the simplex of maximal volume. This tessellation can be constructed by starting with some maximal simplex, reflecting it through its faces, reflecting these through their faces, and so on. For any finite union of these tetrahedra whose boundary is a sphere, we obtain a polyhedron. The triangulation we have is automatically a minimal extension of its boundary, since all the simplices are disjoint and have maximal volume. Indeed, it is the unique minimal extension: in any minimal triangulation, every tetrahedron must have maximal volume, hence it is determined by any of its faces. This determines, one-by-one, where the tetrahedra have to be.

Consider the special case in which all the simplices have a common vertex. (In general, we call a triangulation that is obtainable by coning to some vertex a *cone-type triangulation*.) We may assume that the common point is the point at infinity in the upper half space model. When drawn in this way, each of the tetrahedra lies above an equilateral triangle in the tessellation of the complex plane by equilateral triangles. Any set of triangles whose union is homeomorphic to a disk will define such a polyhedron. Consider the case when the bounding polygon is hexagonal, with  $k$  edges on a side. The hexagon contains  $6k^2$  triangles; hence the polyhedron contains  $6k^2$  tetrahedra. (See Figure 9.) The hexagon has  $3k^2 + 3k + 1$  vertices, so the polyhedron has  $3k^2 + 3k + 2$  vertices (including the one at infinity). In particular we obtain

$$t(n) \geq 6k^2 = 2(3k^2 + 3k + 2) - O(k) = 2n - O(n^{1/2}).$$

Note that by using other triangulations we can actually get explicit lower bounds for each  $n$ , not just those  $n$  of the form  $3k^2 + 3k + 2$ .

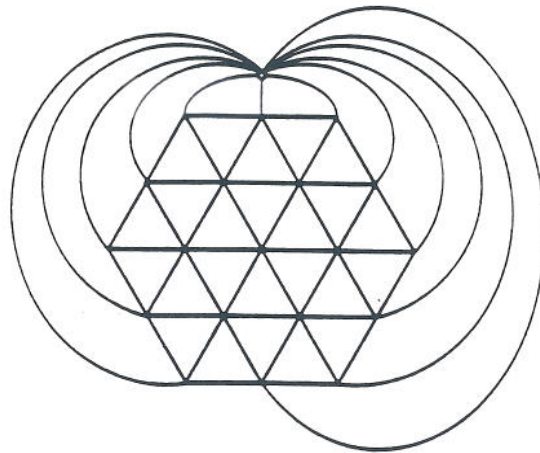


FIGURE 9. The boundary of the polyhedron used to show that  $t(n) \geq 2n - O(n^{1/2})$ .

**3.4. A better lower bound:**  $2n - O(\log(n))$ . To construct polyhedra that require more simplices for a given number of vertices we must eliminate the vertex of high degree. (Roughly speaking, polyhedra in which the vertices are spread over the sphere at infinity as uniformly as possible have the largest volumes.)



A natural sequence of triangulations for this purpose can be derived from a regular icosahedron. Divide each face of the icosahedron into  $k^2$  equilateral triangles, giving  $20k^2$  triangles in all and  $n = 10k^2 + 2$  vertices.

We need to map the vertices of the triangulation of the sphere defined above into hyperbolic space in such a way that the resulting polyhedron has a large volume. The Riemann mapping theorem gives a way to do this. Corresponding to the icosahedron, there is a subdivision of the sphere into triangles bounded by segments of great circles, obtained by projecting the edges of the icosahedron out to the sphere. The Riemann mapping theorem implies that there is a unique conformal map of the faces of the icosahedron to the spherical triangles, sending vertices of the icosahedron to the corresponding vertices of the spherical triangles. By symmetry the maps determined on individual triangles piece together to give a map  $h$  of the entire surface of the icosahedron to the sphere. This map is conformal everywhere except at the vertices of the icosahedron. Note that it is conformal even on the edges of the icosahedron because they can be flattened out (locally) in the plane. Define the ideal hyperbolic polyhedron  $P(k)$  to have its vertices at those places on the sphere at infinity in hyperbolic space to which  $h$  maps the vertices of the subdivided icosahedron.

**Lemma 8.** *The volume of  $P(k)$  is  $2nV_0 - O(\log(n))$ , where  $n = 10k^2 + 2$ .*

*Proof.* To make the estimate of volume, pick a vertex of  $P(k)$  of degree six that is as far as possible from vertices of degree five and arrange this vertex to be at infinity in the upper half-space model. (This is a rigid rotation of  $P(k)$  in hyperbolic space.) Triangulate  $P(k)$  as the union of cones from the vertex at infinity to the triangles with all vertices finite. (This is a cone-type triangulation.) Call a vertex of  $P(k)$  "bad" if it is a vertex of the icosahedron or if it is the vertex mapped to infinity. Now  $h$  can be thought of as a map from the icosahedron to  $\mathbb{C}$  (the complex plane) that is conformal everywhere except at the bad vertices. For large  $k$  the triangles far away from bad vertices get mapped by  $h$  to triangles that are nearly equilateral (because  $h$  is conformal). Figure 10 shows how the vertices of  $P(k)$  near a vertex of the icosahedron get mapped to the complex plane by  $h$ . If the vertices were vertices of true equilateral triangles then the tetrahedra formed by coning them to infinity would all be congruent to the tetrahedron of maximal volume. We must show that the deficit caused by the fact that the triangles are not quite equilateral is small.

The shape of a triangle  $\Delta$  with vertices  $p, q, r$  is conveniently described by a complex number  $s(\Delta) = (r - p)/(q - p)$ , its *shape parameter*. (The triangle  $(0, 1, s(\Delta))$  is congruent to  $\Delta$ .) The volume of the hyperbolic tetrahedron  $C(\Delta)$  formed by coning  $\Delta$  to the point at infinity is a function of the shape parameter of  $\Delta$ ,  $\text{vol}(C(\Delta)) = v(s(\Delta))$ . Note that  $z$ ,  $1/(1 - z)$ , and  $(z - 1)/z$  describe similar triangles, so  $v$  takes the same value at these three points. Since  $v$  attains its maximum at  $\omega$ , the first derivative of  $v$  at  $\omega$  is 0, and the second derivative is the same in every direction because of symmetry. Thus, by Taylor's theorem, the volume deficit of  $C(\Delta)$ , defined as  $V_0 - \text{vol}(C(\Delta))$  (where  $V_0$  is



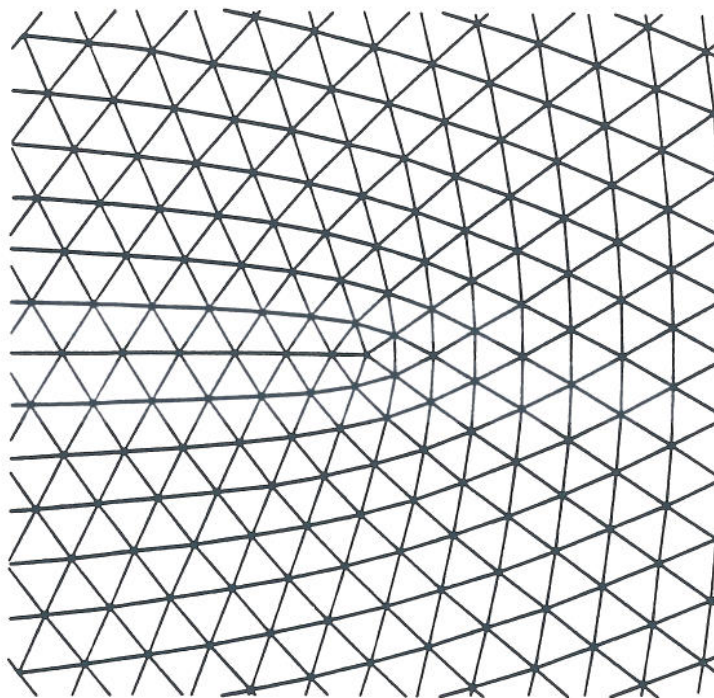


FIGURE 10. The way  $P(k)$  (near a vertex of the icosahedron) gets mapped to the plane by  $h$  in the proof of Lemma 8. This is also the way that the boundary of a blemish of type (c) gets embedded in the plane in the proof of Theorem 1.

the maximum volume of a tetrahedron), satisfies

$$V_0 - \text{vol}(C(\Delta)) = K_1(|s(\Delta) - \omega|^2) + O(|s(\Delta) - \omega|^3),$$

for a certain constant  $K_1$ . Our goal is to show that the cumulative volume deficit is  $O(\log(n))$ .

We now digress to estimate the deficit of a triangle obtained by applying an arbitrary complex analytic function  $f$  to the vertices of an equilateral triangle. Subsequently we will apply this result using the map  $h$ .

Consider the equilateral triangle  $\Delta_t$  with vertices  $0$ ,  $t$ , and  $\omega t$  in the complex plane and suppose that  $f$  is a holomorphic (complex analytic, or conformal) embedding of the disk of radius  $R > |t|$  about  $0$  into the complex plane. How can we estimate the volume deficit of the tetrahedron spanned by  $\infty$  together with the images of the vertices of  $\Delta_t$  under  $f$ ?

In general, we may change  $f$  by postcomposing with a translation so that  $f$  fixes the origin. Let us expand  $f$  as a power series

$$f(z) = az + bz^2 + cz^3 + \dots$$



and solve for the coefficients of the power series expansion of the shape parameter  $s(t) = f(\omega t)/f(t)$  of the image triangle. We can write

$$\begin{aligned} s(t) &= \omega + At + Bt^2 + \dots \\ &= \frac{a\omega t + b\omega^2 t^2 + c\omega^3 t^3 + \dots}{at + bt^2 + ct^3 + \dots} \end{aligned}$$

and solve for the coefficients. We obtain

$$\omega b + Aa = b\omega^2$$

and

$$\omega c + Ab + Ba = c\omega^3.$$

The term of real interest is

$$A = -b/a = -f''(0)/2f'(0)$$

since  $\omega^2 - \omega = -1$ . The next term is

$$B = \frac{1}{a} \left( c(\omega^3 - \omega) - \frac{-b}{a} b \right) = -(\omega + 1) \frac{c}{a} + \frac{b^2}{a^2}.$$

Thus, the shape of the image triangle is

$$s(t) = \omega - tf''(0)/2f'(0) + O(t^2).$$

How does the error term depend on  $f$ ? First, we claim that the error is uniformly bounded by  $O(t^2)$  independent of  $f$  defined on a fixed disk of radius  $R$ . In fact, the set of all holomorphic embeddings of the disk of radius  $R$  into  $\mathbb{C}$  is compact in the appropriate topology; that is, any sequence of embeddings has a subsequence that converges to an embedding. The errors could not get worse and worse, or else the limit function would not have an estimate of the form  $O(t^2)$ .

The dependence of the error term on  $R$  can now be easily deduced. A disk of radius  $R$  can be mapped to a disk of radius  $S$  by a complex affine map — that is, a complex linear map followed by a translation. The parameter  $t$  is multiplied by the ratio of the radii of the disks under such an affine map. Consequently, the error term above is  $O((t/R)^2)$ .

Now we return to the estimate of the volume deficit. For any point  $p$  on the icosahedron define  $inj(p)$ , the *injectivity radius* at  $p$ , to be the maximum radius of a disk in the Euclidean plane that can be isometrically embedded on the icosahedron with its center at  $p$ . Suppose we have a triangle in some subdivision of the icosahedron that has no vertices in common with the icosahedron and no vertex mapped to infinity. Let  $\varepsilon$  be the size of a side of the triangle (in the metric of the icosahedron). Then its volume deficit is not more than

$$(1) \quad K_2 \left( \varepsilon^2 \left| \frac{h''}{2h'} \right|^2 \right) + O \left( \left( \frac{\varepsilon}{inj(p)} \right)^3 \right).$$



Since  $\varepsilon^2$  is proportional to the area of the triangle, the total volume deficit can be approximated by an integral.

$$(2) \quad \int_{\text{good triangles}} K_3 \left| \frac{h''}{2h'} \right|^2 dA + O(1).$$

By good triangles we mean those that have no bad vertex. The contribution of the  $O((\varepsilon/\text{inj}(p))^3)$  term in (1) is at most a constant because on the good triangles  $\text{inj}(p)$  is at least  $\varepsilon$ . This has been included in the  $O(1)$  term in (2). The contribution of the bad triangles (those with a bad vertex) is also only a constant. The contribution of the part of the icosahedron that is farther than a fixed distance  $\varepsilon_0$  away from a bad vertex is also bounded by a constant. This is because the integrand is continuous and bounded except near the bad vertices. The only contribution left to evaluate is that of the annular regions of inner radius  $\varepsilon$  and outer radius  $\varepsilon_0$  centered on the bad vertices (where  $\varepsilon$  is the triangle mesh size).

Near the bad vertices  $h$  behaves like  $z^\beta$  where  $\beta = 6/5$  if the bad vertex is an icosahedron vertex and  $\beta = -1$  if the bad vertex is the one mapped to infinity. (The local coordinates are chosen so that the bad point is at the origin.) The entire deficit is estimated to within an additive constant by the sum (over bad vertices) of the integrals of  $K_3(1/4)(\beta - 1)^2(1/|z|^2)$  over annular regions centered at the bad vertices with fixed outer radius  $\varepsilon_0$  and inner radius approximately equal to the mesh size  $\varepsilon$ . The value of each of these integrals is

$$K_3 \frac{1}{4} (\beta - 1)^2 2\pi (\log \varepsilon_0 - \log \varepsilon).$$

Since  $-\log \varepsilon = \log(1/\varepsilon) = O(\log(n))$ , we have bounded the deficit by  $O(\log(n))$ . This completes the proof that  $\text{vol}(P(k)) = nV_0 - O(\log(n))$ .  $\square$

**3.5. The structure of minimal triangulations.** We shall now apply the fact that  $P(k)$  has a volume deficit that is small compared to the number of tetrahedra to completely determine the minimal extensions of  $P(k)$  to the ball, provided that  $k$  is sufficiently large.

**Theorem 1.** *For sufficiently large  $k$ , any exposed triangulation of the ball extending the boundary of  $P(k)$  having the minimum number of tetrahedra is a cone-type triangulation.*

*Proof.* Suppose that  $T$  is any minimal exposed triangulation of the ball extending the boundary of  $P(k)$ . For any fixed  $\varepsilon$ , at most  $O(\log(n))$  of the simplices of  $T$  can have volume less than  $V_0 - \varepsilon$ . In particular, only  $O(\log(n))$  of its simplices can touch the boundary on two faces, since such simplices have volume roughly  $(2/3)V_0$ .

Since  $T$  is minimal it can have no more tetrahedra than any cone-type triangulation. By coning to a vertex of degree six we obtain a triangulation with  $F - 6$  tetrahedra, where  $F$  is the number of faces. If there are  $m$  tetrahedra in



$T$  that have no face on the boundary then there are at least  $m+6$  tetrahedra of  $T$  that have two faces on the boundary. Since this quantity is  $O(\log(n))$ , we conclude that  $m = O(\log(n))$ . Thus, most of the tetrahedra of  $T$  have exactly one face on the boundary.

The three-dimensional triangulation near a vertex  $v$  has several possibilities. To analyze the possibilities, let  $B_v$  be the union of tetrahedra that have  $v$  as a vertex.  $B_v$  is a ball and its boundary is composed of two parts:  $N_v$ , the union of the triangles of  $P(k)$  containing  $v$  as a vertex, and  $Q_v$ , the union of the remaining triangles.  $N_v$  and  $Q_v$  are joined along a polygon  $p$ , which is a hexagon unless  $v$  is one of the twelve vertices of order 5, in which case it is a pentagon. It is convenient to color the faces of  $Q_v$  red or white, according to whether they are faces of the boundary  $P(k)$  of  $T$ . (This happens naturally if  $T$  is an apple with a red skin and a white interior.) All the faces of  $N_v$  are red.

The vertex  $v$  is an *ordinary* vertex if the triangulation of  $Q_v$  is isomorphic to that of  $N_v$ , with exactly one interior vertex and all triangles having one corner in the interior and an edge along  $p$ . It is a *cone* vertex if  $B_v = T$ , so that the entire triangulation is a cone-type triangulation to  $v$ . If  $v$  is neither an ordinary vertex nor a cone-vertex, it is an *extraordinary* vertex.

We shall show that if  $v$  is extraordinary, then either (1) at least one pair of triangles of  $N_v$  belong to a common tetrahedron, or (2) there is at least one white triangle of  $Q_v$  that does not have an edge on the boundary of  $Q_v$ . Suppose that  $v$  is extraordinary and that (2) is false; that is, all white triangles of  $Q_v$  have an edge on the boundary of  $Q_v$ . There is at least one white triangle of  $Q_v$  (otherwise  $v$  would be a cone vertex). If one of these white triangles has two edges on the boundary of  $Q_v$ , then two of the other sides of its tetrahedron are triangles of  $N_v$ . In this case statement (1) holds.

It remains to consider the case in which all of the white triangles of  $Q_v$  have exactly one of their edges on the boundary of  $Q_v$ . Two white triangles are said to be adjacent if they share an edge. Consider a maximal set of adjacent white triangles of  $Q_v$ . This set is either a cycle of white triangles or it is a sequence of white triangles bounded at each end by a red triangle. If the former case occurs it shows that  $v$  must have been an ordinary vertex. In the latter case, let  $e$  be the edge on one end of the sequence of white triangles separating it from a red triangle and let  $f$  be the edge on the other end of the sequence separating it from a red triangle. A cycle of length four on the boundary of polyhedron  $P(k)$  is formed by  $e$  and  $f$  along with two edges incident on  $v$ . In  $P(k)$  the only cycles of length four are the boundary of the union of two adjacent triangles. Therefore, the only possibility is that the length of the sequence of white triangles is either 1 or  $order(v) - 1$ . In the first case, the single white triangle would actually have been red (all of its edges are on the boundary of  $P(k)$ ), contrary to assumption. In the latter case,  $Q_v$  would have had one red triangle and  $order(v) - 1$  white triangles, making it an ordinary vertex after all.



It follows that there are at most  $O(\log(n))$  extraordinary vertices, since at each extraordinary vertex there is at least one corner of a simplex either touching  $\partial T$  on two faces or on no faces, and there are only  $O(\log(n))$  such simplices.

Let  $m$  be the map from the faces of the polyhedron to the vertices of the polyhedron that associates to a face  $f$  the fourth vertex of the tetrahedron of  $T$  having  $f$  as a face. If  $f$  and  $g$  are any two faces in  $N_v$  for any ordinary vertex  $v$ , then  $m(f) = m(g)$ .

Consider the partition of the faces of the polyhedron according to  $m(f)$ . Let us assume that there is no cone vertex. Any two distinct partition elements can be separated from each other by a cycle of edges passing only through extraordinary vertices. A simple curve of length  $m$  on the icosahedron separates the surface into two regions, at least one of which has area  $A$  less than the area of a circle of circumference  $m$  in the plane; that is,  $A \leq m^2/4\pi$ . It follows that one partition element must have  $n - O((\log(n))^2)$  faces, and the other  $O(\log(n))$  partition elements are of size  $O((\log(n))^2)$ .

The boundary of  $P(k)$  minus the large component is a collection of simply connected regions. Call each of these components a *blemish*.

For each triangle of the big component there is a tetrahedron of  $T$  from that triangle to a common vertex  $v^*$ . Thus  $T$  is almost a cone-type triangulation, with just a few possible bad spots. The places where  $T$  may disagree with the cone-type triangulation  $C(v^*)$  are in the volumes bounded by the triangulation of the big component and the blemishes. We know the triangulations of the boundaries of these volumes, but the interior triangulations are unknown. To complete the analysis, we shall show that the triangulations of the volumes of these blemishes in fact must also agree with  $C(v^*)$ , given that they are minimal triangulations.

The blemishes can be sorted into four types, depending on the type of boundary they have. Here are the four types.

- (a) The blemish contains no vertex of degree five, and does not contain  $v^*$ . In this case the boundary is a portion of the tessellation of the plane by equilateral triangles, with the boundary of these triangles coned to  $v^*$ .
- (b) The blemish contains no vertex of degree five, and  $v^*$  is among the vertices of the blemish. The boundary of the blemish is not a sphere; it may be chosen to be homeomorphic to a sphere with the north pole and south pole identified.
- (c) The blemish contains a vertex of degree five, and  $v^*$  is not among the vertices of the blemish. The boundary triangulation is obtained from a regular pentagon by first subdividing into five triangles, then subdividing these into congruent subtriangles, repeating this process, taking a subset of this, and then coning its boundary to one extra vertex  $v^*$ .
- (d) The blemish contains a vertex of degree five, and  $v^*$  is among the vertices of the blemish. This case is like (b) and (c) combined.



We shall show that none of these four blemishes can actually exist. That is, the unique minimum way to triangulate each of the blemishes is to cone to vertex  $v^*$ . We shall use four proofs, one for each of the blemish types. The impossibility of (b) and (d) follows from that of (a) and (c). The proofs that (a) and (c) are impossible are hyperbolic volume arguments similar to those used to prove the bounds of §§3.3 and 3.4. Each proof describes how the blemish is to be embedded in hyperbolic three-space, evaluates the volume of the embedding, and finally shows that the volume is so large that any other triangulation besides  $C(v^*)$  uses more simplices.

Case (a) is the easiest to resolve. We can embed the boundary of the blemish so that all its vertices except  $v^*$  agree with the vertices of the equilateral triangulation of the plane and  $v^*$  is mapped to the point at infinity. (The boundary of the blemish is the same as the boundary of one of the partial tessellations of space with maximal simplices described in §3.3.) The triangulation of the blemish obtained by coning to  $v^*$  contains only simplices of maximal volume, and these are nonoverlapping. Therefore, any triangulation of this blemish that uses a simplex of less than maximal volume would use at least one more simplex. The only way to avoid using a simplex of less than maximal volume is by coning to  $v^*$ . It is easy to see this by considering a simplex  $s$  with one of its vertices at  $v^*$  and having a face on the boundary of the blemish. If  $s$  is of maximal volume, its fourth vertex must be at the lattice point on the complex plane nearest its other two vertices. Removing  $s$  and iterating this process shows that the only triangulation exclusively using maximal simplices is a cone triangulation to  $v^*$ .

Case (b) can be handled in a similar fashion. Again embed the boundary of the blemish so that all its vertices except  $v^*$  agree with the vertices of the equilateral triangulation of the plane and  $v^*$  is mapped to the point at infinity. The resulting object resembles an annular cylinder in which the hole is hexagonal. Now each simplex of the cone triangulation  $C(v^*)$  has maximal volume, and these simplices are nonoverlapping. Again, this is the only possible minimal triangulation extending the boundary of the blemish.

There is another approach to eliminating type (b) blemishes. Let  $B$  be a type (b) blemish. By gluing six simplices to the boundary of  $B$  around  $v^*$ , we obtain a blemish  $B'$  of type (a). Suppose there were an alternative method of triangulating  $B$  that used no more simplices than coning to  $v^*$ . Then this alternative triangulation could be made into a noncone-type triangulation of  $B'$  using no more simplices than the cone-type triangulation of  $B'$ . We have already shown that such alternative triangulations do not exist.

It remains for us to deal with the twelve possible blemishes of types (c) and (d). The same argument used above implies that if we can show that a blemish of type (c) must be triangulated with a cone-type triangulation then one of type (d) must also. Thus it only remains to deal with the type (c) blemishes.

We can embed the boundary of any such blemish in hyperbolic space as follows. First we construct a tiling of the plane with nearly equilateral triangles



and one vertex of degree five. This is done by modifying the tessellation of the plane by equilateral triangles. First remove a wedge (with its apex at the origin) containing one sixth of the triangles. Now raise every point to the power  $6/5$ , which closes the gap left by removing the wedge. (Figure 10 shows a portion of this tiling.) Now select a connected subset of this tiling of the plane that contains the vertex of degree five, such that this subset is isomorphic to the portion of the boundary of the blemish coming from the boundary of  $P(k)$ . Place the vertex  $v^*$  at infinity to complete the embedding of the boundary of the blemish.

In this embedding, the simplices of the cone triangulation  $C(v^*)$  are not regular, but they are nearly regular. To evaluate the deficit of this structure one can do a numerical calculation. To simplify matters we embed the blemish in a larger structure, one in which the portion of the tiling of the complex plane is shaped like a regular pentagon. Let  $j$  be the number of edges in the subdivision of each edge of the pentagon, so that the pentagon has  $5j^2$  triangles. We shall show that the only way to minimally triangulate this extended blemish is to cone to  $v^*$ .

When we apply the volume estimate of §3.4 to the above mapping and evaluate the constants, we obtain the following formula for the deficit.

$$Def(j) = \frac{\pi}{60} \ln(j) + O(1).$$

Numerical calculation can be used to determine the behavior of the  $O(1)$  term. The results are shown in Table 1. The deficit increases by about .0363 each time  $j$  is doubled, and the  $O(1)$  term is approaching .08538... (Our formula for estimating the deficit did not take into account the deficit from the five tetrahedra around the core of the blemish; note that the error term is always less than the deficit accounted for by these central tetrahedra.)

TABLE 1

$j$	Triangles	Deficit	Estimated	Error
1	5	.087935	.000000	.087935
2	20	.122353	.036293	.086060
4	80	.158141	.072586	.085554
8	320	.194304	.108879	.085424
16	1280	.230564	.145172	.085391
32	5120	.266849	.181466	.085383
64	20480	.303140	.217759	.085381

If there were any extraordinary vertices in a minimal triangulation of the blemish, then there would have to be at least one simplex with two faces on the base of the blemish, since the total number of simplices is no more than the number of faces of the base. The deficit of a simplex touching two faces along an edge tends to  $V_0/3 = 0.3383$ . We calculated values of the deficit for simplices having two faces sharing an edge near the vertex of order 5; the smallest deficit of any such tetrahedron is 0.3333.



Consider a triangulation  $T$  of the blemish that has exactly one simplex with two faces on the base of the blemish, exactly one simplex with no faces on the base, and all other simplices coned from the base of the blemish to  $v^*$ . In order for this triangulation to use no more tetrahedra than the canonical one, the blemish must be large enough so that the deficit of the entire structure is at least 0.3333. Table 1 shows that the radius must therefore be at least 64. Because  $T$  has only two deviant tetrahedra (tetrahedra that are not coned to  $v^*$ ), it has at most eight extraordinary vertices. It therefore must be possible to embed this blemish in a pentagon of radius at most eight. Since eight is less than 64, such a triangulation  $T$  is impossible.

Before there is enough volume deficit for a second simplex touching two faces, the radius of the blemish would have to double at least eight more times, to more than 4000. Such a triangulation could be embedded in a pentagon of radius at most 16. The reasoning used above shows that this too is impossible. In general, no matter how many of these deviant simplices there are, the radius required to achieve the necessary deficit is much bigger than required to absorb that many simplices. This completes the proof of Theorem 1.  $\square$

If the sizes of the deficits were not quite so small, this argument would not work for small  $j$ , and we could only deduce that the minimal number of simplices was within an additive constant of the number for  $C(v^*)$ . For example, consider the sequence of subdivisions of the tetrahedron or of the octahedron instead of the icosahedron. For a tetrahedron, the numbers do not work out: in fact, the cone-type triangulations of subdivisions of the tetrahedron can be improved by first cutting off the corners. This fact manifests itself in Table 2 in which the deficit is more than 1 even for small values of  $j$ .

TABLE 2

$j$	Triangles	Deficit	Estimated	Error
1	3	1.014942	0.000000	1.014942
2	12	1.520982	0.544397	0.976585
4	48	2.054948	1.088793	0.966155
8	192	2.596627	1.633190	0.963437
16	768	3.140336	2.177586	0.962750
32	3072	3.684560	2.721983	0.962578
64	12288	4.228914	3.266379	0.962535

For an octahedron, the numbers work easily to prove that the cone-type triangulation is minimal: a blemish with deficit  $V_0$  must have radius more than 11, at which size it is easy to see that one can replace it with smaller blemishes. See Table 3. However, in this case, the cone-type triangulation of the blemish is not unique. Already the blemish of radius 1 has deficit 0.395904. This blemish is the octahedron, which admits three different minimal triangulations (related by the symmetry of the octahedron). This octahedron is contained within the cone-type triangulations of all the larger blemishes, so none of them are unique.



TABLE 3

$j$	Triangles	Deficit	Estimated	Error
1	4	0.395904	0.000000	0.395904
2	16	0.566991	0.181466	0.385525
4	64	0.745656	0.362931	0.382725
8	256	0.926398	0.544397	0.382002
11	484	1.009655	0.627768	0.381887
16	1024	1.107681	0.725862	0.381819
32	4096	1.289101	0.907328	0.381774
64	16384	1.470555	1.088793	0.381762

Theorem 1 applies to a far wider variety of triangulations than  $P(k)$ . Our proofs of both Lemma 8 and Theorem 1 apply with almost no changes to any triangulation satisfying the following conditions: (1) All the vertices are of order five or six. (2) The triangulation can be drawn with equilateral triangles on the surface of a convex polytope whose shape is distorted from a sphere by only a "constant amount." (3) Within a radius of  $O(\log n)$  of any vertex of order five, the triangulation is isomorphic to Figure 10, and elsewhere it is isomorphic to the standard tessellation of the plane by equilateral triangles.

To prove a result analogous to Lemma 8 for this class of triangulations we first need to map the triangulation to a sphere, with a map that is conformal everywhere except at the vertices of order five. In the proof of Lemma 8 we used the Riemann mapping theorem. Here we must use a more general theorem called the uniformization theorem. It tells us that there is a map from the surface of any polytope to a sphere that is conformal everywhere except at the vertices of the polytope. Furthermore the fact that the polytope is distorted from a sphere by only a constant amount allows this map to have its derivative bounded below and its second derivative bounded above away from the vertices. These are the bounds required for the proof.

Given condition (3) on the triangulation and the  $O(\log n)$  bound on the deficit of the embedding, the same reasoning employed in the proof of Theorem 1 shows that the cone-type triangulation is the minimal for these triangulations. To prove that  $d(n) = t(n) = 2n - 10$  for all sufficiently large values of  $n$ , it only remains to show how to construct an appropriate triangulation for every sufficiently large  $n$ . We shall now give this construction.

First consider the triangulation of the plane by equilateral triangles, arranged so that one vertex is at the origin. Pick any other two vertices  $v_1$  and  $v_2$  so that the three are not collinear and mark off the sublattice they generate. There is a group of symmetries of the triangulation generated by  $180^\circ$  rotations about 0,  $v_1$ , and  $v_2$ . Each of these rotations preserves the lattice we have marked off. This group preserves not only the tessellation of the plane by equilateral triangles, but also a tessellation by parallelograms congruent to the one spanned by  $v_1$  and  $v_2$ . If the parallelograms are colored black and white in a checkerboard pattern, the coloring is also preserved by the group. If we fold up the



plane by the group action, we obtain as quotient space a polyhedron that has four vertices of order three (coming from  $0$ ,  $v_1$ ,  $v_2$ , and  $v_1 + v_2$ ), while all the other vertices are of order six. It can be realized in space as a tetrahedron, with a finer triangulation by equilateral triangles drawn on its surface. Its area is the sum of the area of the black and the white parallelograms. The number of vertices is the area of the tetrahedron, divided by the area of the smallest lattice parallelogram, plus two. Using this information, it is not hard to construct an example of this type with any even number of vertices bigger than two and where the vertices of order three are fairly far apart.

A slight modification of this construction gives rise to triangulated polyhedra with  $n$  vertices for any  $n \geq 8$ . In the new construction superimpose on the original equilateral triangulation a triangulation by equilateral triangles of half the edge length, arranged so that the origin is in the middle of an edge of a larger triangle. Call the vertices of the larger triangles *coarse vertices*, and all the remaining vertices *fine vertices*. Let the equilateral triangles formed by three neighboring coarse vertices be called *coarse triangles*, and define *fine triangles* analogously. Choose  $v_1$  and  $v_2$  so that  $v_1$ ,  $v_2$ , and  $v_1 + v_2$  are all fine vertices. Mark out the lattice generated by  $v_1$  and  $v_2$ . It is again the case that a rotation by  $180^\circ$  about any lattice point preserves both the coarse and the fine vertices. Now cut out from the plane all coarse triangles that touch a lattice point. The picture is now the plane minus a collection of parallelograms (each formed by a pair of missing coarse triangles). Form the quotient space by the group generated by  $180^\circ$  rotations about lattice points, as before. The four edges of each of the missing parallelograms fold up to form a bi-gon. Glue together the two edges of the bi-gon to form a single edge. The desired triangulation is that obtained by considering the coarse vertices and the coarse triangles. Each vertex is of order six except those near the missing parallelograms. Each of the four missing parallelograms creates a vertex of order five and a vertex of order four. (The process described here breaks down if  $n$  is less than eight, because for these small numbers the deleted parallelograms are not disjoint.) Figure 11(a) shows the first stage of the construction of a polyhedron of nine vertices. The fine triangles are not shown. Figure 11(b) shows the triangulation that results from this choice of lattice vertices.

The number of vertices in a triangulation formed by this process is just twice the area of a lattice parallelogram as measured in units of coarse parallelograms (each of which is two coarse triangles). To prove that we can obtain triangulations with any number of vertices, we need to show that there is a basis parallelogram for the lattice that has any desired half-integral area. The following paragraph describes how to do this in such a way that the lengths of the basis vectors are  $\Theta(n^{1/2})$ .

Choose a fine vertex as the origin. Then choose nonorthogonal  $x$  and  $y$  axes such that (1) they go through the origin, (2) they avoid all coarse vertices, (3) they are parallel to one of the sides of each coarse triangle, and (4) the  $x$



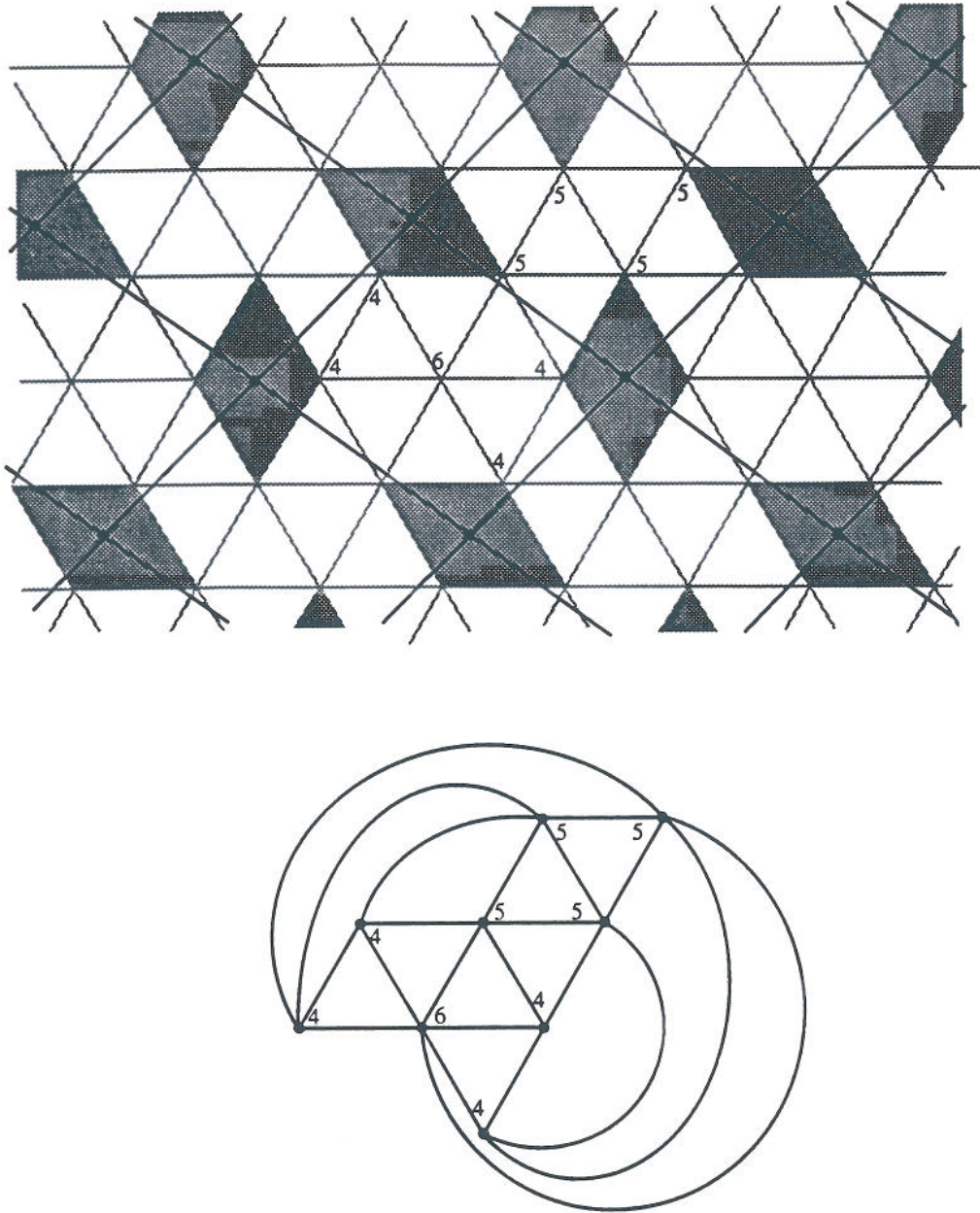


FIGURE 11. Constructing a triangulation of nine vertices, one of order six, four of order five, and four of order four. (a) shows the coarse triangles, the lattice points, and the lines between the lattice points. The shaded parallelograms are the ones that are removed. (b) shows the triangulation that results.

axis is  $60^\circ$  clockwise of the  $y$  axis. In the  $x$  direction let the unit distance be the length of a side of a coarse triangle. In the  $y$  direction let the unit distance be the length of a side of a fine triangle. Each point with integral coordinates is a fine vertex (although not all fine vertices have integral coordinates in this framework). We are seeking two vectors  $v_1 = (a, b)$  and  $v_2 = (c, d)$  such that



the area of the resulting parallelogram (in the metric of this coordinate system) is the desired value  $n$  and the parallelogram is nearly square. Let

$$a = 1, \quad b = \lfloor n^{1/2} \rfloor, \quad c = \lceil \frac{n}{b} - \frac{1}{2} \rceil, \quad d = cb - n.$$

The area of the parallelogram is  $bc - ad = bc - (cb - n) = n$ . It is also easy to see that all the sides are  $\Theta(n^{1/2})$  in length.

A further modification gives rise to triangulations with twelve order five vertices and all others of order six, as follows. First, consider a triangle  $T$  with coarse vertices and with sides  $a$ ,  $b$ , and  $c$  in counterclockwise order. Let  $p$  be the vertex between sides  $a$  and  $b$ . Around the vertex  $p$ , one can arrange six triangles, alternating between equilateral triangles and congruent copies of  $T$  arranged so that one of the other corners is at  $p$ . The union of the six triangles covers a hexagon  $H$  with sides of type  $a$ ,  $b$ ,  $c$ ,  $a$ ,  $b$ , and  $c$  in counterclockwise order, with opposite sides parallel. The hexagon admits a  $180^\circ$  symmetry. The point of symmetry is always on a coarse or fine vertex; which type depends on the original triangle  $T$ . For present purposes, it is desirable that  $T$  be nearly equilateral and that the point of symmetry be a fine vertex. We can construct such a  $T$  by starting with a large equilateral triangle on the coarse vertices and moving one of the vertices of this triangle to one of the six nearby coarse vertices. Figure 12 shows how the hexagon is formed from three copies of a triangle  $a$ ,  $b$ ,  $c$ , and three equilateral triangles.

Use  $H$  and  $T$  to construct the desired triangulation as follows. As in the previous construction, let the origin be a fine vertex, and choose vectors  $v_1$  and  $v_2$  on fine vertices so that  $v_1 + v_2$  is also a fine vertex. Form the lattice from these two vectors. Now, instead of removing the two coarse triangles near each lattice point, remove a copy of  $H$  centered there. Form the quotient by the same group of symmetries, generated by order two rotations about lattice points. These symmetries carry the copies of  $H$  to themselves so that the quotient is a surface with boundary, where each boundary component has three edges of types  $a$ ,  $b$ , and  $c$ . Glue a copy of  $T$  to each one. (In the previous construction, each missing parallelogram became a bi-gon, which we then closed up. Here the missing hexagon becomes a missing triangle, and we patch the hole with a copy of  $T$ .) The polyhedron we obtain has twelve vertices of degree five (one at each corner of each copy of  $T$  that is inserted), and all the other vertices have degree six.

It is easy to obtain such a polyhedron for any sufficiently large  $n$ . First construct a  $T$  that contains about  $\alpha n$  coarse vertices, where  $\alpha$  is some small fraction, say  $1/20$ . Let  $|T|$  be the number of coarse vertices in  $T$ , not counting boundary vertices. Construct  $H$ , and define  $|H|$  similarly. Let  $n' = n - 2|H| + 4|T|$ . Now choose  $v_1$  and  $v_2$  such that they would give a construction with  $n'$  vertices without replacing the hexagons by triangles. When this replacement is done, the number of vertices left is  $n$ .



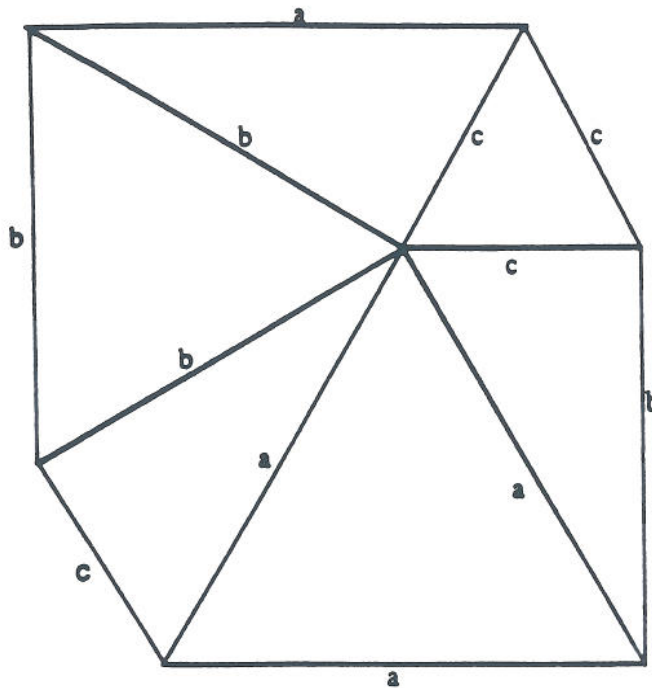


FIGURE 12. The way three copies of triangle  $(a, b, c)$  are glued together, along with three equilateral triangles, to form a hexagon. Note that the hexagon has  $180^\circ$  symmetry.

It is easy to see that the resulting triangulation has all the properties required for our proof of Theorem 1 to apply, and it has any sufficiently large number of vertices. Thus we have proved the following theorem.

**Theorem 2.**  $t(n) = d(n) = 2n - 10$  for all sufficiently large values of  $n$ .

#### 4. REMARKS AND QUESTIONS

Our results say nothing about small values of  $n$ . We conjecture that  $t(n) = 2n - 10$  for all  $n > 12$ , but we have been unable to extend our proof to show this. Probably some more concrete calculations of volumes and triangulations, for polyhedra with low values of  $n$ , would show this.

Empirically, it seems that for very many triangulations of the sphere, cone-type triangulations give the best possible extensions to a ball. It would be nice to have a good criterion of when this is the case. Even when cone-type triangulations are not the best possible, they provide a seemingly efficient method of finding a minimum triangulation: begin with a cone-type triangulation to a vertex  $v^*$  of maximal order. Now look for subpolyhedra where the cone vertex is not the vertex of minimal order for the boundary triangulation; replace the triangulation in the subpolyhedron by a different cone-triangulation. For instance, if one begins with a triangulation of the sphere having vertices  $v_1$  and  $v_2$ , of orders say fifteen and twelve, separated by a simple cycle of edges  $p$  of length



eleven or less, then one can begin by constructing the cone triangulation to  $v_1$ . The union of tetrahedra that touch  $v_2$ 's side of  $p$  form a subpolyhedron  $Q$ , which has a cone-type triangulation to  $v_1$  — but on the boundary of  $Q$ ,  $v_1$  has order the length of  $p$ , which is not more than eleven. Replacing the triangulation of  $Q$  by the cone to  $v_2$  reduces the number of tetrahedra. This idea can be iterated. There are sometimes situations where several moves of this type that do not reduce the number of tetrahedra, but maintain it at a constant, are necessary before a move of this type can reduce the number. It is possible to get from any triangulation to any other if we allow moves of this type that may increase the number of tetrahedra. We know of no instance, however, where in order to decrease the number of tetrahedra, it is first necessary to increase the number.

A more complete development of our method would allow other types of blemishes that are not nearly round and that touch the boundary triangulation in a surface that is not necessarily a disk. The estimates would be much stronger in this case. We conjecture that such a method would show that if there are no vertices of order less than four, and if the injectivity radius is always bigger than some fixed constant, that the cone-type triangulation is a minimal triangulation.

The role that geometry plays in this problem may seem mysterious. There should be an analysis that is entirely combinatorial. It may help clarify the relation to combinatorics if we point out the relation of this question to network flow problems, and to the “max flow min cut” principle.

To develop this idea, consider the  $(n - 1)$ -simplex  $\Delta^{n-1}$ , which has  $n$  vertices  $v_0, \dots, v_{n-1}$ . This circular order defines a Hamiltonian circuit in the 1-skeleton of  $\Delta^{n-1}$ , which we can think of as identified with the boundary of our original polygon. The sum of the oriented edges of this circuit is a 1-cycle  $\alpha$ . A triangulation of the disk defines a 2-chain  $\sigma$  such that  $\partial\sigma = \alpha$ , since each triple of vertices of  $\alpha$  spans a unique triangle in the 2-skeleton of  $\Delta^{n-1}$ .

Consider two different triangulations of the disk, giving two different 2-chains  $\sigma_1$  and  $\sigma_2$ . Then  $z = \sigma_1 - \sigma_2$  is a 2-cycle. Define the quantity  $g(z)$  to be the minimum  $L^1$  norm of a 3-chain  $T$  such that  $\partial T = z$ . (The  $L^1$  norm of a chain is the sum of the absolute values of its coefficients.) Both the minimum number of elementary moves to get from  $\sigma_1$  to  $\sigma_2$  and the minimum number of 3-simplices it takes to extend the triangulation they define to the ball give upper bounds for  $g(z)$ . Geometrically, if  $z$  comes from a triangulation of the boundary of a ball, then any 3-chain  $T$  such that  $\partial T = z$  can be mapped into the ball, giving a possibly multiple-valued triangulation of the ball, where tetrahedra with fractional weights are allowed, but the weights of overlapping tetrahedra must add up to 1 at each point.

Unlike the problem of finding minimal triangulations, the problem of finding a chain  $T$  such that  $\partial T = z$  with minimum  $L^1$  norm is a linear problem: it is a question of minimizing a convex function subject to a linear constraint. Therefore there is a dual problem. The dual space of the space  $C_k$  of  $k$ -chains



of a simplicial complex is denoted  $C^k$ , the space of  $k$ -cochains. A  $k$ -cochain is a linear function on the  $k$ -chains, and it is determined by its values on the oriented simplices, since they span the space (and a choice of an orientation on each simplex gives a basis). The coboundary map  $\delta$  maps  $C^k$  to  $C^{k+1}$ ; when  $\beta$  is an oriented  $k+1$ -simplex and  $c$  is a  $k$ -cochain,  $\delta(\beta)$  is  $c$  applied to the sum  $\partial\beta$  of the oriented faces of  $\beta$ . A cochain whose coboundary is 0 is called a cocycle.

The dual problem is this: given  $z$ , what is the minimum  $L^\infty$  norm  $h(z)$  of a 3-cocycle  $V$  such that  $\langle V, T \rangle = 1$ , where  $T$  is any chain such that  $\partial T = z$ ? Here,  $\langle \cdot, \cdot \rangle$  is the dual pairing of cochains with chains. This formulation depends on the fact that the simplex  $\Delta^{n-1}$  is *acyclic*; that is, in any dimension except 0, every cycle is a boundary. Therefore, if  $T'$  is any other chain such that  $\partial T' = z$ , the difference  $T - T'$  is a 3-cycle, so there exists a 4-chain  $a$  with  $\partial a = T - T'$ . Therefore

$$\langle V, T \rangle - \langle V, T' \rangle = \langle V, \partial a \rangle = \langle \delta V, a \rangle = \langle 0, a \rangle = 0.$$

The solutions to the dual problems satisfy  $g(z) \cdot h(z) = 1$ . A good example of a chain  $T$  gives an upper bound for  $g(z)$ ; a good example of a cocycle  $V$  gives a lower bound. The quantity  $g(z)$  is like a minimum cut; the quantity  $1/h(z)$  is like a maximum flow.

A mapping  $f$  of  $\Delta^{n-1}$  into hyperbolic space defines a 3-cocycle  $V_f$ , which assigns to any 3-chain the algebraic volume of its image. (We should divide  $V_f$  by the total volume to make it literally fit the formulation above.) If we allowed ourselves to push the triangles around so that they were curved, we could obtain every possible 3-cocycle in this way. In most cases, no single map that takes triangles to geodesic triangles gives the best cocycle for our problem. Presumably, some convex combination of cocycles  $V_f$  of the form we used would give a cocycle of norm even closer to the minimum. Our actual argument made use of the fact that we only allowed integral cycles  $T$ , with no fractional weighting.

The quantity  $g(z)$  was introduced and studied in a more general topological context by Gromov (see [12]). He proved that for a closed oriented hyperbolic manifold  $M^n$ , the minimum  $L^1$  norm of an  $n$ -cycle representing the fundamental class of  $M$  is exactly the volume of  $M$  divided by the maximal volume of a hyperbolic  $n$ -simplex. To approach the best bound in this case of a closed manifold, it is definitely necessary to use simplices of fractional weights. Perhaps this work will be useful in analyzing best possible actual triangulations of 3-manifolds.

#### REFERENCES

1. H. M. S. Coxeter, *Non-Euclidean geometry*, Univ. Toronto Press, Toronto, 1942.
2. K. Culik and D. Wood, *A note on some tree similarity measures*, Inform. Process. Lett. 15 (1982), 39-42.
3. A. K. Dewdney, *Wagner's theorem for torus graphs*, Discrete Math. 4 (1973), 139-149.



4. M. Gardner, *Time travel and other mathematical bewilderments*, Freeman, New York, NY, 1987.
5. D. E. Knuth, *The art of computer programming, Vol. 3: Sorting and searching*, Addison-Wesley, Reading, MA, 1973.
6. J. Milnor, *Hyperbolic geometry: the first 150 years*, Bull. Amer. Math. Soc. **6** (1982), 9–24.
7. J. Pallo, *On rotation distance in the lattice of binary trees*, Inform. Process. Lett. **25** (1987), 369–373.
8. D. D. Sleator and R. E. Tarjan, *Self-adjusting binary search trees*, J. Assoc. Comput. Mach. **32** (1985), 652–686.
9. D. D. Sleator, R. E. Tarjan, and W. P. Thurston, *Rotation distance, triangulations, and hyperbolic geometry*, Proc. 18th Annual ACM Symposium on Theory of Computing, May 28–30, 1986, Berkeley, CA, Association for Computing Machinery, NY, pp. 122–135.
10. R. E. Tarjan, *Data structures and network algorithms*, Soc. Indust. Appl. Math., Philadelphia, PA, 1983.
11. W. P. Thurston and J. R. Weeks, *The mathematics of three-dimensional manifolds*, Scientific American **251** (1984), 108–120.
12. W. P. Thurston, *The geometry and topology of three-dimensional manifolds*, Princeton Univ., Dept. of Math., 1979.
13. W. T. Tutte, *A theorem on planar graphs*, Trans. Amer. Math. Soc. **82** (1956), 99–116.
14. K. Wagner, *Bemerkungen zum Vierfarbenproblem*, J. Deutschen Math.-Verein. **46** (1936), 26–32.
15. H. Whitney, *A theorem on graphs*, Ann. of Math. **32** (1931), 378–390.

**ABSTRACT.** A rotation in a binary tree is a local restructuring that changes the tree into another tree. Rotations are useful in the design of tree-based data structures. The rotation distance between a pair of trees is the minimum number of rotations needed to convert one tree into the other. In this paper we establish a tight bound of  $2n - 6$  on the maximum rotation distance between two  $n$ -node trees for all large  $n$ . The hard and novel part of the proof is the lower bound, which makes use of volumetric arguments in hyperbolic 3-space. Our proof also gives a tight bound on the minimum number of tetrahedra needed to dissect a polyhedron in the worst case and reveals connections among binary trees, triangulations, polyhedra, and hyperbolic geometry.

COMPUTER SCIENCE DEPARTMENT, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PENNSYLVANIA 15213

COMPUTER SCIENCE DEPARTMENT, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08544  
AT&T BELL LABORATORIES, MURRAY HILL, NEW JERSEY 07974

MATHEMATICS DEPARTMENT, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08544