Conformal modulus: the graph paper invariant
or
The conformal shape of an algorithm

J. W. Cannon, W. J. Floyd, and W. R. Parry

§0. Introduction.

This paper is an expository paper about our joint work, which the first author presented in a series of lectures at the University of Auckland (New Zealand), the University of Melbourne (Australia), and the Australian National University in Canberra (Australia). We express appreciation for the kindness and interest of all the many wonderful mathematicians and their families whom that author and his wife enjoyed during their visit. This final version of the paper includes a few of the questions and comments which arose during the discussions of those lectures. We thank the referees for numerous insightful comments.

The first section, which is our own nonproof of the Riemann Mapping Theorem, can be used as a good intuitive introduction to the long and fussy proof of our own combinatorial Riemann mapping theorem [CRMT]. In particular, it demonstrates the geometry underlying the classical conformal modulus of a quadrilateral or annulus.

The second section shows how the classical conformal modulus is applied to combinatorics, with the intent of preparing for the exposition of sections 3 and 4.

The third section shows that, under subdivision, a topological quadrilateral can develop wildly oscillating conformal modulus, a behavior which was perhaps not expected.

The fourth section reviews how combinatorial moduli apply to the study of negatively curved or Gromov word hyperbolic groups and shows by example how our work might be used to recognize a Kleinian group combinatorially.

The final section, section 5, concludes the paper with remarks and questions.

§1. Conformal moduli.

What is the geometry underlying the *modulus formula*,

$$M_\rho = (H_\rho)^2/A_\rho,$$

which comes from the theory of conformal mapping and gives the modulus M_ρ as a ratio which compares the square $(H_\rho)^2$ of a certain length H_ρ to an area A_ρ? Just as Dr. Strangelove came to love the bomb, so we have come to love this unintuitive

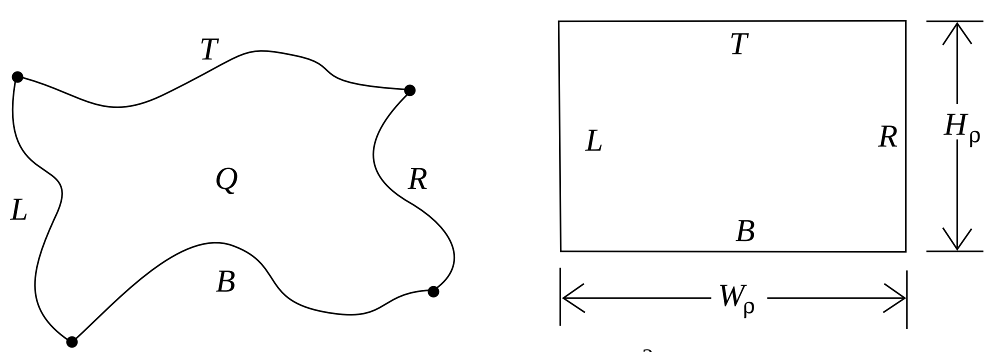
This work was supported in part by NSF research grants.

expression. Our aim is to explain to topologists and geometric group theorists its beautiful underlying geometry and applications.

To which length and area do the symbols H_ρ and A_ρ refer? In the sequel we shall consider the formula in four different settings: the classical modulus of a quadrilateral or annulus with a fixed Riemannian metric ρ; the conformal modulus of a quadrilateral or annulus obtained by optimizing the classical moduli over a family of Riemannian metrics; the combinatorial modulus of a tiled quadrilateral or annulus with a fixed weight function ρ; and the combinatorial conformal modulus of a tiled quadrilateral or annulus obtained by optimizing over a family of weight functions. The first two settings are considered in this section, and the next two settings are considered in section 2. The remainder of the paper is then devoted to applications of the combinatorial modulus to geometry and group theory.

Setting I. The classical, continuous setting. Let Q denote either a (compact) topological quadrilateral (disk with four distinguished boundary points) or a (compact) topological annulus, with Q having a Riemannian metric ρ. Call two opposite edges of the boundary of Q the *top* and *bottom* of Q. (In the quadrilateral case, the four distinguished points of the boundary of Q divide this boundary into four edges, two forming top and bottom, the other two forming the sides. In the annulus case, the two boundary curves of Q are considered opposite edges, the top and bottom of the annulus.) The top and bottom are also called the *ends* of Q. Then H_ρ denotes the Riemannian distance between the top and bottom of Q and A_ρ denotes the Riemannian area of Q.

It is easy to understand the geometric meaning of M_ρ in the case where Q is, as a topological quadrilateral, a true Euclidean rectangle or Q, as an annulus, has the shape of a right circular cylinder. Then top and bottom have obvious geometric meaning, and the distance H_ρ between top and bottom is the geometric height of Q. The rectangle or right circular cylinder Q has area A_ρ which is the product of H_ρ with the width or circumference W_ρ of Q. Thus M_ρ is the ratio $(H_\rho)^2/A_\rho = H_\rho/W_\rho$ which obviously measures the geometric proportions or shape of the rectangle or cylinder. See Figure 1.



$$\frac{H_\rho^2}{A_\rho} = \frac{H_\rho H_\rho}{W_\rho H_\rho} = \frac{H_\rho}{W_\rho}$$

Topological quadrilateral Geometric rectangle

Figure 1. The modulus of a geometric rectangle

Thus we see that the modulus M_ρ is a generalized measure of the shape of the quadrilateral or annulus Q. This measure of shape is obviously invariant under scaling of the metric ρ since the height H_ρ scales by the given scale factor and the area by the square of that same factor. It is precisely this invariance under scaling that dictates the powers of H_ρ and A_ρ used in the formula.

Setting II. The conformal setting. A *conformal change of metric* multiplies a given metric on Q not by a global scale factor but by an infinitesimal scale factor. The Riemannian metric of Q, which we have been calling ρ up to this point, we now assume fixed and put it into the background without any explicit name. We now reinterpret the symbol ρ as denoting not a Riemannian metric on Q but rather a positive function on Q which serves as the local scale factor of a conformal change of metric. The product of ρ with our fixed but unnamed Riemannian metric on Q gives a new Riemannian metric on Q, conformally equivalent to the old one, hence a new height denoted H_ρ, a new area denoted A_ρ, and a new modulus $M_\rho = (H_\rho)^2/A_\rho$.

It is an easy matter to create a conformal invariant from the modulus M_ρ. We simply define the *conformal modulus* of Q to be

$$M = \sup_\rho M_\rho,$$

where ρ varies over all possible local scaling functions ρ.

What is the geometric meaning of this conformal modulus? We shall see that within it hides the wonderful Riemann mapping theorem. The discussion which we shall give will be correct as regards intuitions and conclusions and is even capable of completion, but as it stands, it relies heavily on what Mike Shapiro would call snake oil, that standard merchandise of the travelling salesman intended to attract and perhaps fool or mislead the gullible. Our intent is to demonstrate that the Riemann mapping theorem is natural and beautifully geometric. We consider only the case where Q is a topological quadrilateral.

The graph-paper analysis of the conformal modulus.

Step 1. Optimal weight functions and metrics. There is an amazing fact: the supremum which defines the conformal modulus is finite, and, in fact, this supremum is actually realized as a maximum. That is, there is a positive function ρ on Q such that $M = M_\rho$. We shall call such a function ρ an *optimal weight function* on Q. There is a most marvelous proof of the fact that there exists an optimal weight function, but the margin of this page is too narrow to hold it. Classically its existence was proved by means of the Dirichlet principle, a principle which was relied upon by Gauss, Dirichlet, and Riemann, the use of which was eventually justified under appropriate conditions by Hilbert. The principle has numerous modern incarnations, but is essentially a compactness principle applied to a potentially noncompact, infinite dimensional space of functions. For the remainder of the analysis of conformal modulus, ρ will be a fixed optimal weight function on Q. We shall multiply the old Riemannian metric on Q by this local scaling function ρ and obtain a new Riemannian metric which we call *optimal*. We shall analyze the resulting Riemann surface, which we continue to call Q.

Step 2. The *flows* or *vertical* lines of our graph paper. Let p denote an arbitrary point of Q. We claim that p is on a path joining top and bottom of Q

which has length equal to the distance from top to bottom. If not, then we could reduce the local scale function ρ near that point without changing the height of Q. Such a local scale reduction would reduce the area A_ρ, leave the height H_ρ unchanged, and increase the modulus M_ρ, a contradiction.

We call a path joining the ends of Q which has minimal length a *flow line*. We have proved that Q is filled with flow lines. We shall think of the flow lines as the *vertical* lines of graph-paper coordinates for Q.

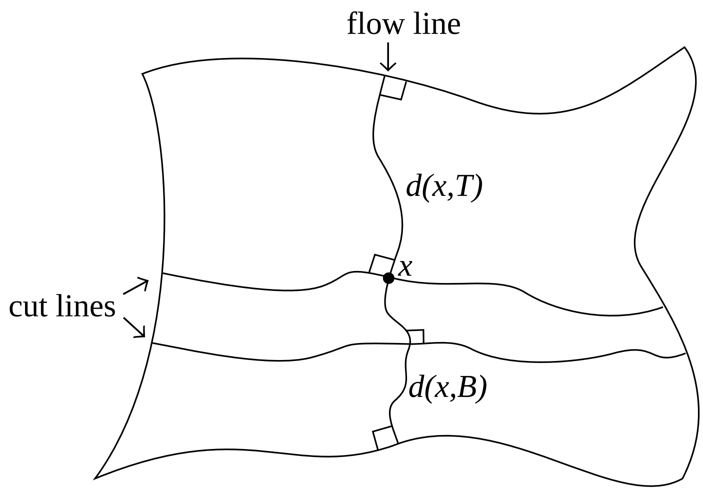


Figure 2. Flow lines and cut lines

Step 3. The *cuts* or *horizontal* lines of our graph paper. The Riemann surface Q with its new optimal metric is also filled by natural *level lines* or *cut lines* which we call *horizontal* and which we individually define as the set of points at constant distance from the bottom of Q. See Figure 2. We think of the cut lines as the *horizontal* lines of graph-paper coordinates for Q.

That the cut lines are actually topological segments joining the sides of Q we see as follows:

The top of Q is the level line corresponding to the constant distance H_ρ from the bottom of Q by step 1.

All other level sets must therefore correspond to levels between 0 and H_ρ. Each such level set is obviously compact, separates top from bottom, and, by our argument, is at each point arcwise accessible from the component of the complement containing the top and also from the component of the complement containing the bottom. A beautiful theorem from plane topology implies that this level set is therefore, in the case of the quadrilateral, an arc (and in the case of the annulus, a simple closed curve) separating top from bottom.

It is important to note that each flow line intersects each cut line orthogonally. Otherwise the flow line could be shortened and still join the ends of Q, a contradiction.

Step 4. The dissection of Q into planar strips of individually constant vertical height. We choose a finite subfamily of the cut lines in Q dividing Q into finitely many strips, each of individually small vertical height, each individual strip being so narrow that it is approximately planar. In keeping with our labelling of the top and bottom of Q as the ends of Q, we call the images of the cut lines (that we used to divide Q into strips) the ends of the strips.

What does one do with cut lines? We take scissors and cut Q along them, thereby dissecting Q into finitely many essentially planar strips of individually constant height. We then approximate each individual strip almost isometrically by an exactly planar strip of the same constant height. See Figure 3.

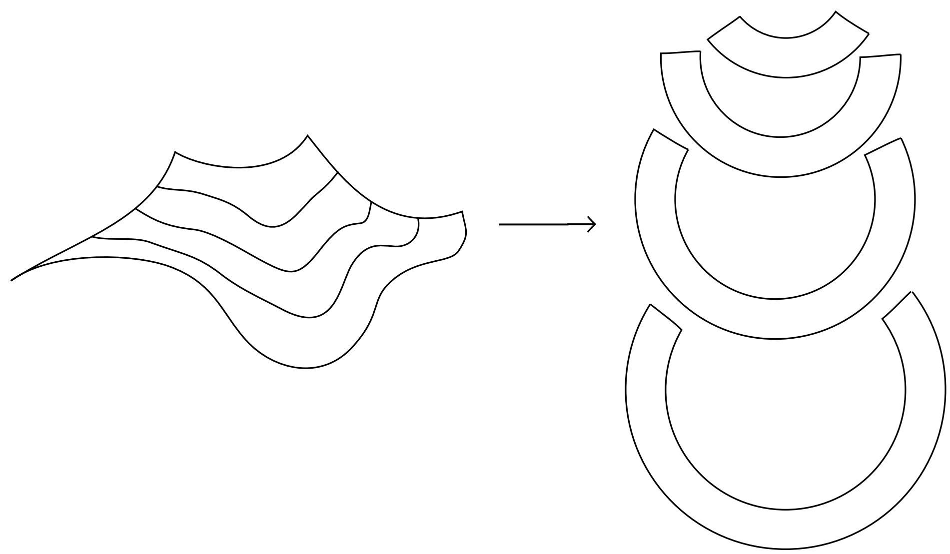


Figure 3. Sample planar strips of individual constant height

The reader can experiment and see how planar strips which twist in the plane but have constant individual heights can be pasted together to form essentially arbitrarily shaped surfaces. Why is this so? How would you dissect an arbitrary surface, say a curved tree leaf or a large portion of an apple or orange peel, into planar strips of constant individual height? A cantaloupe is particularly nice to work with.

Why does a planar strip of constant height twist and turn? It does so because the geodesics orthogonal to the ends (that is, the minimal paths or flow lines joining the ends) locally converge or diverge from one another. This twisting and turning is precisely what creates curvature in the global surface; said another way, the twisting and turning is a combined expression of the curvature of the global surface and the geodesic turning of the cut lines within that global surface.

Step 5. Do our planar strips actually twist? The reader will by now have figured out how to show that essentially arbitrary surfaces can be dissected into (almost) planar strips of individually constant height and will have noted that, in general, those strips, when put into the plane (almost) isometrically, twist and turn. But ours is not an arbitrary surface. Our surface has an *optimal* metric and we have dissected it along *cut lines* or *level lines* with respect to that optimal metric. Is it possible that these *optimal* strips of individually constant height twist and turn?

Let us suppose first that one of these strips, which we denote by S, turns, but

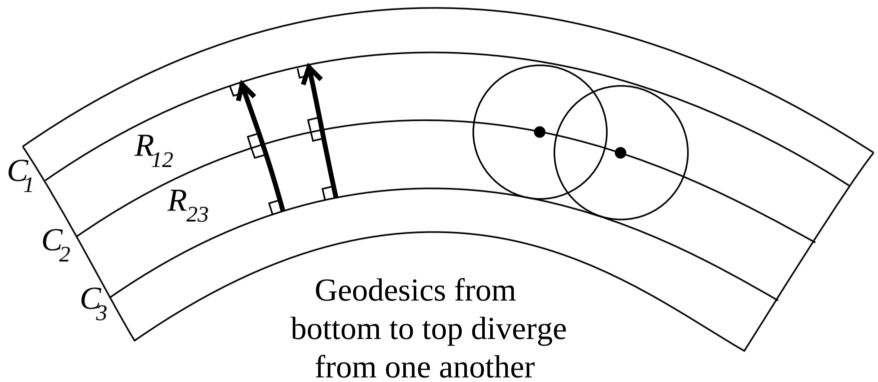


Figure 4. Substrips S_1 and S_2 of the same height in a planar strip S

only in one direction.

The following description is more complicated than it needs to be in order that we can later modify the description minimally in a more general context.

Since the strip S turns in only one direction, we may assume that geodesics orthogonal to the cuts in S diverge from one another as one goes from bottom to top. We construct three curves in S, each separating the top and bottom of S, as follows. The first is a cut C_1 in S. By the geodesic divergence property, the curve C_1 is concave down. Our second curve C_2 is formed by pushing the points of C_1 downward along flows a small distance but retaining the property of being concave down. Our third and final curve C_3 is formed as a portion of an envelope. Namely, we cover C_2 by circular disks centered on C_2 and tangent to C_1. The boundary of the union has two components which separate the ends of S, namely C_1 itself above C_2 and another curve C_3 below C_2. See Figure 4.

In this simple context, here is a simpler description of our three curves: C_1 is a cut in the interior of S, and C_2 and C_3 are cuts at distances ϵ and 2ϵ below C_1, respectively.

We summarize here the essential facts: (1) The three curves bound two regions, namely R_{12} between curves C_1 and C_2 and R_{23} between curves C_2 and C_3. It follows from the concavity property of C_2 that the corresponding areas satisfy the inequality $A_{12} > A_{23}$. (2) The region R_{23} is at least as thick as the region R_{12} in the sense that, if one starts at a point p of C_2, then $d(p, C_1) \leq d(p, C_3)$.

Theorem. There are positive multipliers λ and μ such that if our metric is multiplied by λ in R_{12} and by μ in R_{23}, then the area of S is decreased and the height of S is not decreased.

Remark. The local scaling indicated by the theorem can be applied to all of Q if Q has such a strip S, but that would increase the modulus M_ρ, a contradiction. We therefore conclude as a corollary of the theorem that Q has no such strip S which twists in only one direction. On the other hand, if S twists even locally in any direction, then the construction of C_1, C_2, and C_3 can be carried out locally in the following way. Choose C_1 in the interior of S as before, namely as a cut.

Assuming that S twists, there will be a subarc A of C_1 which is, say, concave downward as before. Now form C_2 from C_1 as before but only pushing points of A downward along flows and leaving the other points of C_1 unmoved. Make sure that C_2 remains concave downward along the image of A. Form the envelope about C_2 as before, noting that the envelope will pinch to height 0 except along A. Then one has precisely the essential facts listed above, from which follows the same theorem. Therefore, we have the following corollaries to the theorem.

Corollary. The planar strips into which we have dissected Q do not twist and turn; their tops and bottoms are parallel Euclidean lines. The flows are parallel to one another and perpendicular to the parallel tops and bottoms. Since the side points of Q lie on flows which join top to bottom, it follows that the sides are in fact flows.

Corollary. The Riemann mapping theorem. The quadrilateral Q with its optimal metric is a Euclidean rectangle. The conformal modulus of the quadrilateral Q with its original metric is the modulus of this rectangle. The flow lines in this rectangle are precisely the vertical lines in the rectangle and the cut lines are precisely the horizontal lines. (They form Euclidean graph paper.)

Proof of the Theorem. An easy but reasonably complex geometric argument will show that the height does not decrease if we choose our positive multipliers λ and μ subject to the conditions that $\lambda + \mu = 2$ and that $\lambda < 1$ ($\mu > 1$).

We begin however with the completely elementary argument that if, in addition, we choose λ and μ sufficiently close to 1, then the area of S will in fact decrease. These two verifications will complete the proof of the theorem and its two corollaries.

We work with the metric modified by the multipliers λ and μ in regions R_{12} and R_{23}, respectively. We call this the new metric. We have a resulting new height and area. The heights and areas before the change are called the old height and areas.

Area can be decreased by choosing λ and μ sufficiently close to 1. By construction, we have $A_{12} = k \cdot A_{23}$, with $k > 1$. Since $k > 1$, we may choose λ with $0 < \lambda < 1$ and with λ so near 1 that

$$(1 + k)(1 + \lambda) - 4 > 0.$$

We then define $\mu = 2 - \lambda$ as required above. It is now an easy exercise in arithmetic to prove that the change in area is given by

$$(\lambda^2 A_{12} + \mu^2 A_{23}) - (A_{12} + A_{23}) = [(1 + k)(1 + \lambda) - 4](\lambda - 1)A_{23},$$

which, by our choices, is negative.

Height does not decrease. Let P denote a path of minimal new length joining the top and bottom of S.

There is a subpath P_1 irreducible from the top to the curve C_1. Here we mean irreducible in the sense that no proper subpath of P_1 joins the top to the curve C_1. It has new length equal to old length and this length is necessarily the distance t from C_1 to the top.

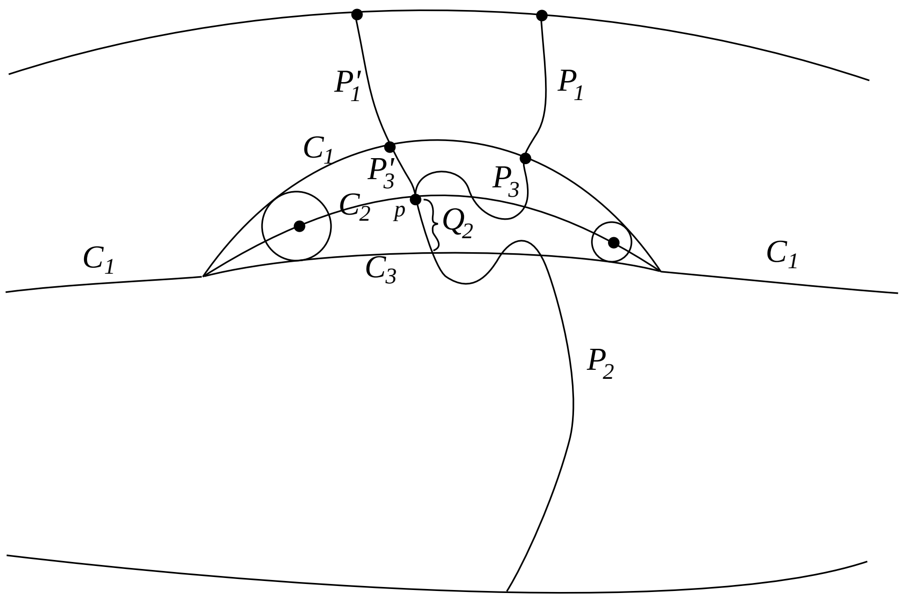


Figure 5. Calculation of new height

There is a subpath P_2 irreducible from C_2 to the bottom.

If P_2 does not begin in $C_2 \setminus C_1$, then it begins on C_1 and everywhere has length evaluated with respect to new metric at least as large as the old metric (perhaps at points lengths are multiplied by $\mu > 1$). In this case the length is at least the distance b from C_1 to the bottom. That is, if P_2 begins in C_1, the length of P is at least $t + b$, which is the old height of S and we are done.

If P_2 begins at a point p of $C_2 \setminus C_1$, we argue as follows. The subarc P_3 of P which begins at $P_1 \cap C_1$ and ends at $p \in C_2 \setminus C_1$ has old length at least equal to the old distance from p to C_1 and its new length is calculated with local multiplier at least as large as λ. Hence we do not increase the length of P_3 if we replace it by a minimal path P_3' through R_{12} from the point p back to $q \in C_1$ (such a path exists in the circular disk about p used in forming the envelope about C_2). We can then replace P_1 by a minimal path P_1' from q back to the top of S.

We call the resulting minimal path P'. Note that it is the union of the three paths P_1', P_3', and P_2. We can now calculate the relative shortenings and lengthenings of these paths as we change from old metric to new metric. The path P_2 which is irreducible from C_2 to the bottom of S contains a subpath Q_2 irreducible from p to C_3 which necessarily passes through R_{23} and has old length at least equal to the old length of P_3'. The only shortening occurs in the path P_3' where the shortening has magnitude $(1-\lambda)$ times the old length of P_3'. We have at least the lengthening which occurs in the path Q_2 where the lengthening has magnitude $(\mu - 1)$ times the old length of Q_2. Since the old length of Q_2 is at least as large as the old length of P_3', and since $\lambda + \mu = 2$, easy arithmetic shows that the lengthening is at least as large as the shortening. $\square$

Remarks. The conformal modulus, as we have seen, precisely captures the variational geometry required to change an arbitrary quadrilateral conformally into a rectangle. The hardest step which we have omitted is the proof that there is in fact an optimal weight function; it is this step that Riemann himself failed to complete. The optimal weight function has, of course, an interpretation as the absolute value of the derivative of the Riemann mapping. That is, the modulus formula captures the notion of complex derivative without any mention of derivative or coordinates.

§2. Combinatorial modulus.

Setting III. The tiled quadrilateral or ring. We assume given as before a topological quadrilateral or ring Q, but we suppress any Riemannian metric on Q and retain only combinatorial data. We assume that Q is tiled by disks. That is, we assume given a structure on Q as a polyhedral 2-cell complex whose 2-cells form a tiling $T = \{t_i\}$ of Q. We assume given a function $\rho : T \to [0, \infty)$, not identically 0, which we call a *weight function* on T. We denote the image or weight of the tile t_i by ρ_i.

We shall use the weight function ρ to define approximate distances and areas in Q as follows. Let X denote a collection of tiles. Then the *length* of X is defined to be the sum of the ρ_i where $t_i \in X$. The *area* of X is defined to the the sum of the squares ρ_i^2 where $t_i \in X$. It is as though we were considering the elements of X as squares of edge ρ_i. There are two types of subsets of T which we wish to call paths. The first type is called a *fat path*. A fat path is associated with a topological path p in Q and is equal to the set of all tiles hitting p. The second type is called a *skinny path*. A skinny path is likewise associated with a topological path p in Q, but it involves an additional choice as well. A skinny path is any set of tiles covering p; it may or may not contain all of the tiles hitting p. A path (fat or skinny) is said to *join the ends* of Q if its associated topological path joins the ends of Q. The path is said to separate the ends of Q if its associated path either joins the sides of Q (the quadrilateral case) or its associated path is a closed curve which circles the annulus Q (annular case). We will often identify a collection p of tiles with its characteristic function $\chi_p : T \to \mathbf{R}$. By doing this, we can take linear combinations of paths.

It is now easy to define height, width or circumference, area, and combinatorial modulus for Q and ρ. The *combinatorial height* H_ρ of Q is the minimum length of a combinatorial fat path joining the ends of Q. The *combinatorial width* (quadrilateral case) or *combinatorial circumference* (annular case) of Q, both of which we denote by W_ρ, is the minimum length of a combinatorial skinny path separating the ends of Q. The *combinatorial area* A_ρ of Q is defined to be the area of T. Now the *combinatorial modulus* M_ρ of Q is defined by the same formula as before:

$$M_\rho = (H_\rho)^2 / A_\rho.$$

Setting IV. The combinatorial conformal modulus. We think of the weight function ρ as a local scaling function, a conformal change of metric, on the tiling T. In order to get a conformal invariant, we simply take the supremum over all weight functions ρ:

$$M = \sup_\rho M_\rho.$$

As before, this supremum is realized as a maximum by a weight function which we call *optimal*. But now this fact is not amazing but easy and we give the proof.

Theorem. Existence and uniqueness of combinatorial optimal weight functions.

(i) There is an optimal weight function ρ.

(ii) The optimal weight function is unique up to scaling.

(iii) The optimal weight function respects all combinatorial symmetries of the tiling T.

Proof. (i) Since the modulus is scale invariant, we may consider only those weight functions of area equal to 1. If we consider the weights of the tiles as forming the components of a vector, one coordinate for each of the finitely many tiles, then the weight functions of area 1 are precisely the vectors of length 1. The legal weight functions form therefore the (compact) set of nonnegative unit vectors in a high dimensional Euclidean sphere. The height associated with a vector is a continuous function of the vector itself. Hence the maximum possible height is attained on our compact set of weight functions. The vector on which it is attained is an optimal weight function.

(ii) Suppose there were two optimal weight functions of unit area, vectors v_1 and v_2. Then the average of v_1 and v_2 is a vector whose height is at least as large as that of v_1 and v_2, but its area is smaller since the average lies inside the unit sphere, a contradiction. Thus the optimal weight function is unique up to scaling.

(iii) If one were to transform an optimal weight function by a combinatorial symmetry of the tiling T which preserved the top and bottom of T, the height and area would remain unchanged so that the tranformed function would also be optimal. But uniqueness implies that the function has not been changed. $\square$

Remark. In the continuous case, proving the existence of an optimal weight function was difficult, essentially an infinite dimensional problem. In the combinatorial case, the existence is a finite dimensional compactness argument. Riemann's attempt to prove the existence of an optimal function in the continuous case depended on the Dirichlet principle which attempted to apply finite dimensional compactness arguments to infinite dimensional function spaces.

A tiled quadrilateral or annulus Q with its optimal combinatorial weight function can also be analyzed in terms of minimal (fat) flows and minimal (skinny) cuts whose corresponding and underlying topological arcs intersect one another like the coordinate lines of graph paper. Here is the theorem, independently discovered by us (see [2]), Oded Schramm, and John Robertson.

Characterization Theorem. The graph paper theorem: the geometric structure of the combinatorial optimal weight functions. Let ρ denote a weight function for Q with tiling T. Let $f_1, \ldots, f_m$ denote the minimal fat flows and $c_1, \ldots, c_n$ denote the minimal skinny cuts with respect to the tiling T. Then the following three conditions are equivalent and imply the fourth.

(i) The weight function ρ is optimal.

(ii) There are nonnegative real numbers $\alpha_1, \ldots, \alpha_m$ such that the vector ρ can be written as the linear combination

$$\rho = \sum_i \alpha_i \cdot f_i.$$

(iii) There are nonnegative real numbers $\beta_1, \ldots, \beta_n$ such that the vector ρ can be written as the linear combination

$$\rho = \sum_j \beta_j \cdot c_j.$$

(iv) There is a positive number k such that the vector $k\rho$ is an integral vector. The numbers α_i and β_j may also be chosen to be integral. (Of course, the flows and cuts are already vectors, each entry of which is 0 or 1.)

Geometric Corollary. The finite Riemann mapping theorem. Scale an optimal weight function ρ so that it is integral. Choose the α_i and β_j so that they are integral. By using parallel copies of the paths underlying flows and by using essentially parallel copies of the paths underlying cuts (they may have to coincide where they pass through a vertex in passing from one tile to another), we may think of each of the weights α_i and β_j as actually being equal to 1 so that ρ is actually a sum of fat flows and a sum of skinny cuts. Then the topological paths underlying the minimal flows may be realized in Q as nonintersecting paths joining top to bottom. The paths associated with minimal cuts may be realized in Q as noncrossing paths joining left to right which intersect one another only when they pass through a vertex from one tile to another. Each cut intersects each flow exactly once where they cross each other. Now pull the cuts and flows taut so that they form the grid of graph paper. As a consequence Q is pulled into exactly a Euclidean rectangle (or the graph paper on a right circular cylinder in the case of an annulus). In each individual tile there are exactly as many crossing flows as there are crossing cuts. Hence each tile is pulled into exactly a Euclidean square. That is, the optimal weight function shows how to square a rectangle. That is, there is a uniquely shaped rectangle determined by the tiling filled with square tiles corresponding to the tiles of the original tiling such that the tiling preserves, almost, the original combinatorics of the tiling. We say, almost, because two types of things can happen. It may happen that no minimal cut or flow goes through a given tile or tile-edge, in which case that tile or edge will collapse to size 0 in the squared rectangle. It may also happen that a skinny cut must go through a vertex of a tile, in which case that vertex will expand into a nondegenerate vertical interval.

We apply this corollary to an example. We could apply it to the tiling of Australia by its states. However that supplies us with a puzzle that has too few pieces. It is more interesting to consider the continental United States of America as a topological quadrilateral with one vertex in Maine, one in Florida, one in California, and one in the state of Washington. In Figure 6 we give both the standard map and the associated tiled rectangle. The relative size of the state images is perhaps a bit surprising. Texas, usually thought of as the largest and therefore the most important state finds its claims conformally a bit exaggerated.

Figure 6. A tiling and its squared rectangle

And New England, with all its historical claims to fame disappears entirely because it is cut off by New York. The state which is conformally most important? We would hope, of course, that it would be Utah. But in fact it is Utah's neighbor Idaho.

Algorithms for finding the combinatorial optimal weight functions.

There are nice algorithms for calculating optimal weight functions either approximately or exactly. For more details, see [2] and [7].

Algorithm 1. The convergent algorithm. A given weight function will by the characterization theorem fail to be optimal precisely because it is not a positive combination of minimal flows. This suggests the procedure of calculating the minimal flows of a weight function, and adding these minimal paths to the given weight function so that the function is more nearly a sum of minimal paths. An iteration actually converges *projectively* to the optimal weight function. The intuition is that the addition of the minimal paths to the original weight function makes the quadrilateral more and more nearly of constant height as should be the case with the optimum, and more and more nearly the sum of minimal paths, again as it should be with the optimum.

Algorithm 2. The cyclic algorithm. In the convergent algorithm, one has to scale to determine an actual function to which the algorithm converges projectively. But there is an exact algorithm which does not scale. It proceeds exactly as in the convergent algorithm, but it checks to see whether the sequence of minimal vectors added stage by stage becomes cyclic. A theorem ensures the existence of a stage after which the process does become cyclic. If the process has become cyclic, then the exact optimal weight function is given by the sum of added vectors over the cycle. If one guesses that the process has become cyclic, then one must check to see that the cycle sum attained is optimal, a check which is actually not difficult once one has the candidate weight function.

Algorithm 3. The hybrid algorithm. The convergent algorithm is really quite quick and can be processed on moderately sized examples thousands of times in a few seconds. The cyclic algorithm gives exact integral weight functions, but its running time increases exponentially in the complexity of the tiling. There is a hybrid algorithm which uses the convergent algorithm thousands of times to *guess* what the appropriate minimal paths should be for the optimal weight function. Once these paths have been guessed, one can set up a candidate system of linear equations whose solution will be the optimal weight function *provided one has guessed the correct paths*. The hybrid algorithm can often work very quickly and supply the exact optimal weight function.

Approximating the combinatorial conformal modulus. An exact optimal weight function for a tiling can be algorithmically calculated as indicated above in the cyclic and hybrid algorithms. However, the weight function can be very complicated even for simple tilings. It is often better, and almost always sufficient, to get a good approximation to the conformal modulus of the tiling. The convergent algorithm seems to work wonderfully with a wide range of examples. However, for theoretical work it is important to find more conceptual ways of approximating the modulus of a tiled quadrilateral or annulus. We give here four simple results in that direction.

The layer theorem ([3, Theorem 4.1.1]). Suppose a tiled quadrilateral Q is divided into a family Q_j of quadrilaterals by a finite collection of disjoint edge paths in the tiling T which join the two sides of Q. Then the combinatorial conformal modulus of Q is at least as large as the sum of the moduli Q_j.

Proof. Find an optimal weight function ρ_j for each of the quadrilaterals Q_j and scale them so that each of them has the same combinatorial width equal to 1. Note that for fixed index j, after the normalization which makes the width equal to 1, the combinatorial area and the combinatorial height and the combinatorial modulus of Q_j are all equal to one another. A minimal path joining the top and bottom of Q must cross each quadrilateral Q_j and hence will have length at least equal to the sum of the heights of the Q_j. The area will be exactly equal to the sum of the areas of the Q_j, which is equal to the sum of the heights of the Q_j. Hence the modulus will be at least the sum of the heights of the Q_j which is the sum of the moduli of the Q_j. $\square$

Remark. The cut theorem. The moral of the layer theorem is this: to find a lower bound on modulus, try to *constuct many cuts* (to force height to be high) *in such a way that the associated area is small.*

The rotation theorem. Suppose a tiled quadrilateral admits a symmetry which maps the top and bottom into opposite sides. Then the combinatorial modulus of the quadrilateral is at least 1.

Proof. Let ρ be an optimal weight function for Q. Apply the symmetry which takes the top and bottom of Q into opposite sides. Let τ be the image of ρ under the symmetry. Then τ is not necessarily an optimal weight function because the symmetry has not preserved the top and bottom of the quadrilateral. However, each fat flow has image under the symmetry a cut, and the τ-length of the flow is the ρ-length of the corresponding cut. Since $A_\rho = A_\tau$,

$$M = M_\rho = (H_\rho)^2/A_\rho \geq H_\tau^2/A_\tau = H_\tau^2/A_\rho \geq W_\rho^2/A_\rho.$$

Hence $H_\rho \geq W_\rho$ and $M \geq 1$. $\square$

The bounded overlap theorem. Suppose that Q has two tilings T and T', that no element of T intersects more than K elements of T', and that no element of T' intersects more than K elements of T. Then $M(Q,T) \leq K^3 \cdot M(Q,T')$.

Proof. Let ρ be an optimal weight function on the tiling T, and let ρ' be a weight function on T' which we shall construct by means of ρ. Then we attain heights H_ρ and $H_{\rho'}$, areas A_ρ and $A_{\rho'}$, and moduli M_ρ and $M_{\rho'}$. It will suffice to prove that $M_\rho \leq K^3 M_{\rho'}$.

Define the weight function ρ' on T' as follows. If $t' \in T'$, then choose $f(t') \in T$ such that $t' \cap f(t') \neq \emptyset$ and

$$\rho(f(t')) = \max\{\rho(t) \mid t \in T,\ t \cap t' \neq \emptyset\}.$$

Define $\rho'(t') = \rho(f(t'))$.

We can bound ρ-area below as follows.

$$\begin{array}{lcl} A_\rho & = & (1/K) \cdot K \sum_{t \in T} \rho(t)^2 \\ & \geq & (1/K) \cdot \sum_{t' \in T'} \rho f(t')^2 \\ & = & (1/K) A_{\rho'}. \end{array}$$

For the inequality in this calculation we have used the fact that, for each tile t of T, there are at most K tiles t' of T' with $f(t') = t$.

We can bound ρ-height above as follows. Let α denote a path which joins the ends of Q and which realizes the ρ'-height of Q. Then, with $L_\rho'(\alpha)$ denoting the

ρ'-length of α, we have

$$\begin{array}{rcl} H_\rho & \leq & L_\rho(\alpha) \\ & = & \sum\{\rho(t) \mid t \in T \text{ and } t \cap \alpha \neq \emptyset\} \\ & \leq & K \sum\{\rho(f(t')) \mid t' \in T' \text{ and } t' \cap \alpha \neq \emptyset\} \\ & = & K \sum\{\rho'(t') \mid t' \in T' \text{ and } t' \cap \alpha \neq \emptyset\} \\ & = & K \cdot L_{\rho'}(\alpha) \\ & = & K \cdot H_{\rho'}. \end{array}$$

The first inequality in this calculation follows from the deinition of H_ρ. For the second inequality, suppose that $t \in T$ with $t \cap \alpha \neq \emptyset$. Then there is a tile t' in T' with $t' \cap t \cap \alpha \neq \emptyset$. By the definition of f, $\rho(t) \leq \rho(f(t'))$. The second inequality now follows from the fact that each tile t' of T' intersects at most K tiles t of T.

From the inequalities of the last two paragraphs it is easy to see that

$$M_\rho = (H_\rho)^2/A_\rho \leq K^3 (H_{\rho'})^2/A_{\rho'} = K^3 M_{\rho'}. \quad \square$$

Corollary. Let Q_{opp} denote the quadrilateral Q with tiling T but with the two sides playing the role of top and bottom. Suppose that no tile of T intersects more than K tiles of T. Then

$$(1/K^3) \leq M(Q) \cdot M(Q_{\text{opp}}) \leq K^3.$$

Proof. Take an optimal weight function ρ for Q with integer weights and realize ρ as a sum of noncrossing skinny cuts. These skinny cuts may pass through vertices of the tiling, thereby "splitting" those vertices. Replace each such vertex by an arc and distribute the edges originally ending at such a vertex along the arc so that the skinny cuts can now be realized as fat cuts. This splitting process forms a new tiling T' for Q. We can do this locally in small neighborhoods of the vertices so that two tiles of T' do not intersect unless the corresponding tiles of T intersect. If t_1 and t_2 are tiles of T with $t_1 \cap t_2 \neq \emptyset$, then the corresponding tiles t_1' and t_2' of T' will have nontrivial intersection if $t_1 \cap t_2$ contains an edge but they may be disjoint if $t_1 \cap t_2$ is a vertex (or a union of vertices). That is, the two tilings T and T' satisfy the bounded overlap theorem with overlap constant K. We conclude that

$$(1/K^3) M(Q_{\text{opp}}, T) \leq M(Q_{\text{opp}}, T') \leq K^3 M(Q_{\text{opp}}, T).$$

Now we note that Q has precisely the same optimal weight function with respect to both T and T' since ρ squares them both. Here we are using the facts that the weight function ρ on T can be written as a sum of disjoint fat flows which are disjoint from the vertices, and these fat flows can be realized on T'. Since they are minimal for the weight function (which we still call ρ) that is their sum, ρ is an optimal weight function for T'. Hence $M(Q,T) = M(Q,T')$.

On the other hand, Q_{opp} with tiling T' is also squared by ρ, and therefore $M(Q,T') \cdot M(Q_{\text{opp}}, T') = 1$.

Our desired result follows immediately from the concluding equalities or inequalities at the end of the preceding three paragraphs. $\square$

§3. The combinatorial shape of an algorithm: the oscillating quadrilateral.

In our applications of combinatorial modulus to group theory, which we shall discuss in the next section, we are concerned with the behavior of the modulus of a quadrilateral when the quadrilateral is subdivided successively by some subdivision rule. The goal of this section is to show that almost anything can happen provided the subdivision rule is context sensitive. We shall construct a quadrilateral and a context-sensitive subdivision rule in such a way that the modulus oscillates wildly, first approaching infinity, then zero, then back again much closer to infinity, then back again, much closer to zero, and so on. The rule will be of sufficiently restricted type that, from one subdivision to the next the modulus will be able to change multiplicatively only by a uniformly bounded amount. Hence, in order to create our example, we shall have to be able to have the shapes or moduli change gradually for arbitrarily long times before they turn around and gradually change in the opposite direction for an even longer time.

In a sense, what we are discussing is the *asymptotic combinatorial shape of an algorithm.* There are many ways of associating geometric objects with algorithms. They may be said to reflect the "shape" of the algorithm.

We mention only two examples. First there is the state graph. The states of a machine might be vertices of a graph. Input to the machine carries a machine, *via an algorithm*, to another state. Join the two states by a directed edge labelled by the input. The resulting graph has many natural geometries which reflect the "shape" of the algorithm.

A second example appropriate to our setting is the subdivision geometry. Let T be a tiling of some surface. Let S be a subdivsion rule which applies to T. Think of S as a finite algorithm. Then S carries $T = T^1$ to T^2 to T^3, etc. Join tiles of T^n to their neighbors in T^n and to their offspring in T^{n+1}. The resulting graph has natural geometries which reflect the "shape" of the algorithm S. There are therefore negatively curved or Gromov hyperbolic subdivision rules to which geometry may be applied. They have a space at infinity usually equal to the original surface.

In reviewing Mandelbrot's book on the fractal geometry of nature, we suggested that one might appropriately broaden the definition of fractal geometry to include all methods of studying the asymptotic geometric shape of an algorithm, with Hausdorff dimension describing only a single invariant or aspect of that asymptotic shape. Mandelbrot was quite put off by our suggestion. We see in this section that one might also discuss not only the Hausdorff dimension but also the combinatorial and asymptotic conformal shape of certain algorithms. Other invariants might be studied as well.

Types of subdivision rules. We shall be intuitive rather than precise in our definitions.

Finite subdivision rules.

In a finite subdivision rule, we are given a finite number of types of cells-with-labels. Each type of cell subdivides according to a rule *which depends only on its label* into a finite number of subcells-with-labels from the same collection of types. The subdivision rule is required to be compatible from a cell to each of its faces. This compatibility ensures that subdivision is well-defined on a labelled cell complex. The action of subdivision is not only completely *local* but strictly *individual* in the sense that a cell does not sense the state of its neighbors.

Remark. In a true finite subdivision rule, it is impossible to send signals about. Each cell is a law unto itself. It takes no account of the behavior of its neighborhood. If we wish to send signals in a subdivision rule we need a more sensitive tool. One such is given by the next type of subdivision rule.

Context sensitive subdivision rules.

As with a finite subdivision rule, we are given a finite number of types of cells-with-labels. Each type of cell subdivides into a finite number of subcells-with-labels from the same collection of types. However, the subdivision rule is now allowed to depend not only on the label of the cell but also upon the combinatorics and labelling of a neighborhood of fixed finite size. One requires that this rule be compatible from a cell to each of its faces. This compatibility ensures that subdivision is well-defined on a labelled cell complex.

Remark. In a context sensitive subdivision rule, it is possible to send signals. Gossip can spread like wildfire. One can coordinate actions over long distances. Fashions can become dominant: "everyone is doing it."

The oscillating quadrilateral. We are prepared to discuss a context-sensitive subdivision rule which, when applied to an appropriately tiled quadrilateral Q, leads to a wildly oscillating combinatorial conformal modulus.

The problems that we must face in the construction are these:

(1) How do we ensure that the modulus is getting big or small?

(2) How do we coordinate actions globally?

(3) How do we delay direction-changes in modulus growth for ever longer periods?

We begin by looking at the initial stages of the construction. The quadrilateral Q is shown in the top left of Figure 7, and the first eight subdivisions of Q are also shown in Figure 7. For the moment, ignore the "signals" S and C. Note that only the cells that border the circular boundary arc of Q subdivide properly. In the second subdivision, the cell that borders the top of Q subdivides into three cells and the cell that borders the right side of Q subdivides into two cells. In the next six subdivisions (and for 46 more after that) the cells that border the right side of Q subdivide into three cells and the cells that border the top of Q subdivide into two cells. Figure 8 shows an enlargement of the eighth subdivision of Q, together with seven disjoint cuts on it. The lengths of these cuts are $3, 4, \ldots, 9$. These cuts give seven disjoint layers with moduli $1/3$, $1/4$, $\ldots$, $1/9$. By the layer theorem, the modulus of the eighth subdivision of Q is $\sum_{n=3}^{9} 1/n \approx 1.329$. After we have done 53 subdivisions of Q, there are 52 disjoint cuts with lengths $3, 4, \ldots, 54$, and hence 52 disjoint layers with moduli $1/3$, $1/4$, $\ldots$, $1/54$. Again by the layer theorem, the modulus of this subdivision of Q is at least $\sum_{n=3}^{54} 1/n \approx 3.075$.

The idea behind this subdivision scheme is to switch, with successively longer periods between switching, between subdividing so as to produce a lot of disjoint cuts (in order to make the modulus large) and subdividing so as to produce a lot of disjoint flows (in order to make the modulus small).

The plan for estimating the modulus. We shall construct our rule in such a way that there are positive integers $n_0 = 2 < m_1 < n_1 < m_2 < n_2 < \dots$ having the following properties: After the subdivision has been performed m_i times, then there will be $d_i = m_i - n_{i-1}$ ("d" for "difference") disjoint layers joining the sides of Q that are a single tile high at each point and whose successive widths are j_i+1, j_i+2, $\dots$, j_i+d_i. The modulus of such a layer is clearly the reciprocal of the number tiles in the layer. By the layer theorem, the modulus of the quadrilateral will be at stage m_i at least the sum of the reciprocals of the widths listed, which sum is approximately the logarithm of $1 + d_i/j_i$. Hence these moduli will go to infinity provided the ratio d_i/j_i goes to infinity.

Similarly, after the subdivision has been performed n_i times, with $i > 0$, then there will be $e_i = n_i - m_{i-1}$ disjoint layers joining the top and bottom of Q that are a single tile high at each point and whose successive widths are k_i+1, k_i+2, $\dots$, k_i+e_i. The modulus of Q_{opp} will, by the same argument, be at least as large as approximately the logarithm of the ratio $1+e_i/k_i$. Hence these moduli will go to infinity provided the ratio e_i/k_i goes to infinity. Since each cell in our quadrilateral will at each stage intersect at most 10 cells, we can apply the bounded overlap theorem and its corollary to conclude that, if $M(Q_{\mathrm{opp}})$ is very large at subdivision n_i, then $M(Q)$ will be very near to 0.

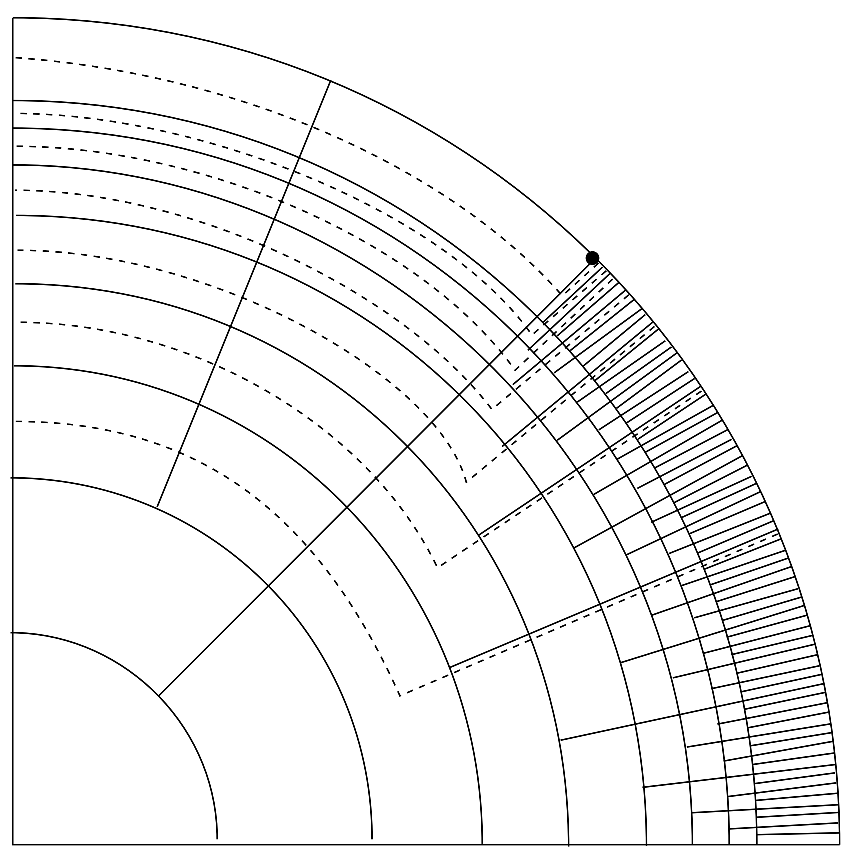


Figure 8. Layers in oscillating quadrilateral

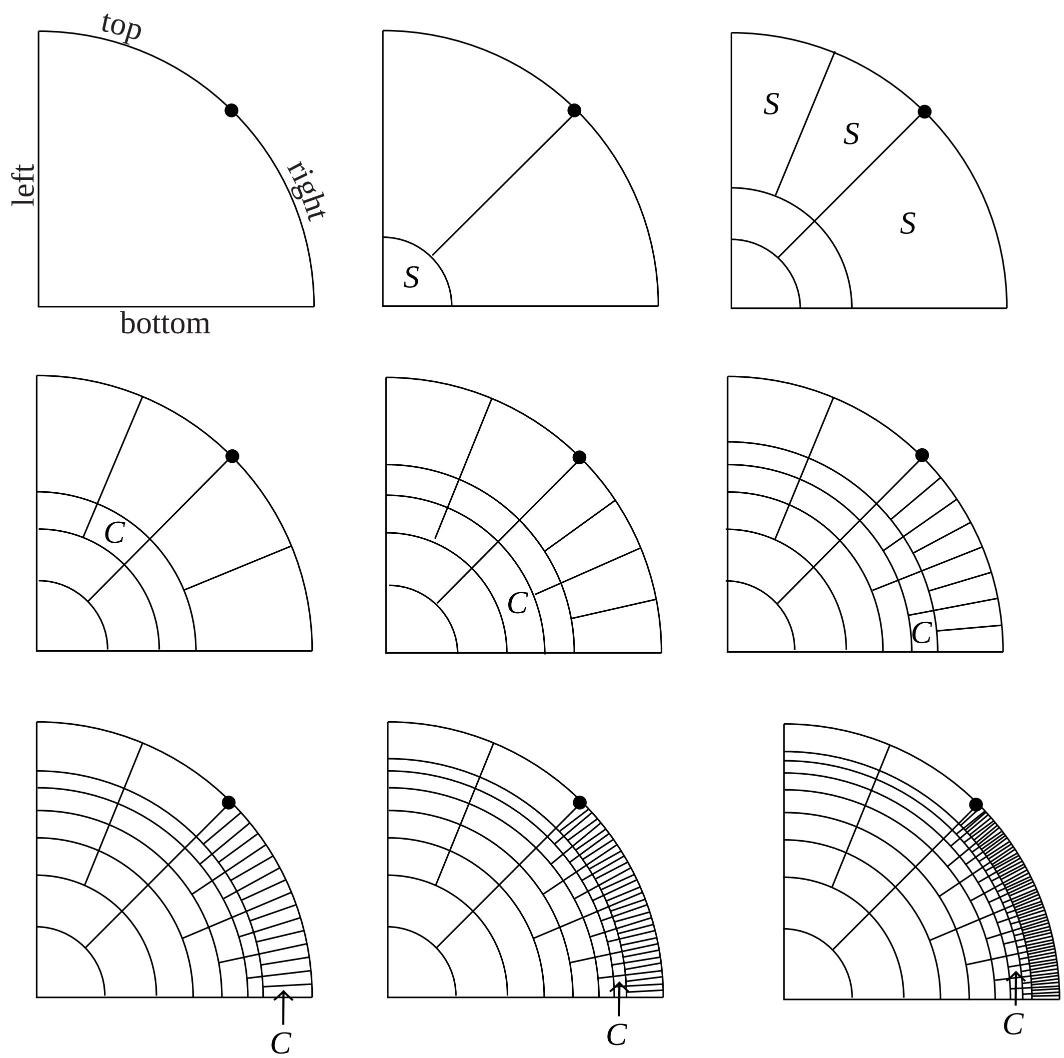


Figure 7. The oscillating quadrilateral

The geometry of the subdivisions. We have already seen the initial stages of the subdivision in Figure 7. Without giving details of how we arrive at subdivision n_{i-1}, we describe its geometry. The quadrilateral will be pictured as a radial sector in a cirle. The two radial sides of the sector will be considered, respectively, the left side and the bottom of the quadrilateral. The circular segment of the boundary will be divided in half at the middle to form, respectively, the top and right side of the quadrilateral. Each subdivision divides all of the cells adjacent to the circular segment which forms the union of the top and right side. At all stages between subdivision n_{i-1} to subdivision m_i, the cells adjacent to the right side are divided first by a circular segment into two cells that are radially adjacent and then the outer of the two cells is divided into two cells by a radial cut. That is, each of the cells adjacent to the right side is divided into three cells. On the other hand, the cells adjacent to the top are only divided by the circular segment into two radially adjacent cells. Thus the outer layer of the circular sector is dividing exponentially on the right side and is maintaining its size at a constant along the top.

When one finally reaches subdivision m_i, the exponential subdivision switches from right side to top. Cells adjacent to the top subdivide via a circular cut, then the outer of the two resulting cells divides via a radial cut. Cells adjacent to the right side subdivide only via the circular cut.

The last layer at stage n_{i-1} will have a large number of cells adjacent to the top, namely, a large power of 2, where the power is the sum of the integers e_h for $h \leq i-1$. Similarly this last layer will have a large number cells adjacent to the right side, namely, a large power of 2, where the power is the sum of the integers d_h for $h \leq i-1$.

Where do we find the disjoint layers joining the two sides after we have reached subdivision m_i? One simply traverses the layer of cells adjacent to the top and adds the first cell on the right side. One removes these cells from consideration, traverses the new top layer until the right half of the sector is reached, then follows an obvious path outward. And so on. See Figure 8.

The rules for obtaining the subdivision. We return to the initial subdivisions shown in Figure 7. This time we expand our discussion to include the signals. In order to be able to make use of the symmetry in the construction, in the next few paragraphs we will refer to the top as the left outer arc and to the right side as the right outer arc. The first subdivision divides the sector first by a circular arc into two radially adjacent cells, then divides the outer cell by a radial segment which ends at the point which divides the top from the right side. The single cell which resides at the center of the circle now becomes the controlling cell. It signals the outer two cells to start subdividing, with the cells along the left outer arc dividing as described earlier into three cells, and the cells along the right outer arc dividing into two cells. This kind of subdivision will continue at the outmost circular edge until a different signal arrives to indicate a switch, from subdivider into three to subdivider into two, or vice versa.

The counting signal. Whenever the switch signal arrives at the outer circular edge which tells the left outer arc to begin subdividing into twos and the right outer arc into threes, we are at some stage n_{i-1} of the subdivision. At that point, the outer left cell, which can recognize itself as outer left cell from its environment alone, begins a counting signal which moves across its circular layer from left to right at a rate of one cell per subdivision. We think of this signal as counting the size of its half of the circular outer layer (before further subdivision). The left "half" of the layer being counted will have size exactly j_{i-1}. When the signal gets to the rightmost edge of the circular sector, the direction of signal motion will change to radially outward, and its speed will double so that it is moving fast enough to catch the leading edge of the subdivision. When it reaches the leading edge, it changes speed and direction again. It now moves circularly to the left along the circular layer that was outmost when it reached the edge but at only one cell per subdivision. It is now counting to a very large power of 2, with the power being larger than the previously counted size j_{i-1}. When it finishes counting its layer, the signal again reaches the leftmost radial edge of the subdivision, changes direction and speed once more, moving now radially inward toward the control cell at the center of the circle. It eventually gets there.

The switching signal. When the control cell at the center receives the counting signal, it sends switching signals outward along all radial paths at a rate of two layers per subdivision so that the signal can move fast enough to catch the leading edge of the subdivision. All of the leading edge cells will receive the switching signal at the same time. Those on the left outer arc will switch to subdividers into three, those on the right outer arc become switchers into two. We have now reached stage m_i of the subdivision. The cell that is outermost and rightmost which has just been changed to a subdivider into two recognizes all of these properties from its environment and begins a counting signal circularly left along its level. See Figure 7. The counting and subdividing cycle begins anew with processes switched about the center radial segment, left and right essentially interchanged.

Verification that subdivision continues without switching for sufficiently long time intervals. Suppose that the right outer arc is subdividing into threes, the left into two. Then before switching, the number of subdivisions, denoted d_i, will be more than 2 raised to the power j_{i-1}. Thus the ratio d_i/j_{i-1} goes to infinity very rapidly.

Simplifying the counting signals? We clearly have subdivided many more times before switching than we needed to do. The most obvious simplification would be to count only across the first half of the circular sector, then send the signal back to the central controlling cell. But the geometry suggests that the moduli stay bounded if that is done. Is that so? How can the counting be modified? How randomly can the moduli be switched?

Remark. We observe three interesting things in this example. First, signals can only be sent at a linear rate; hence one must retain enough linearly growing pathways connecting the example to allow coordinating signals. Second, the existence of exponential pathways orthogonal to the linear pathways allowed us to create long signal delays. Third, edges and boundaries can play important roles in the behavior of otherwise indistinguishable cells.

Question. How would one change the model so that signals travel only at speed 1? How would one change the model so that there is only one kind of cell?

Gromov has suggested that this wildly oscillating growing crystal might serve as a naive early model for biological differentiation and growth. He suggests in particular the problem of building a finite local replacement rule that will grow into a *homunculus*. He suggests that the finite rules might be discovered *evolutionarily* by doing computer experiments.

Asymptotic shape questions about groups.

Remark. Gromov also suggests that one might try constructing finitely presented groups, perhaps groups of curvature ≤ 0, with rather arbitrary growth functions in the sense that the successive shapes of the metric balls in the group should be quite arbitrary. In particular, the flats supply polynomial growth spots in which signals might in some sense be sent, the negative curvature spots might allow one to act for a long time before causing some strange collapse, etc. (We have no idea how such an idea might be carried to completion.)

§4. Recognizing groups of constant curvature combinatorially: an example.

Our theory of combinatorial moduli was developed to the end of recognizing Kleinian groups combinatorially. We have not yet brought the project to completion, but this section gives an outline of the progress to date and shows by example how the results obtained can be used to recognize combinatorially that certain groups are Kleinian.

We first state the principal conjecture, then review the definitions in terms of which that conjecture is stated.

Main Conjecture. A group G is Kleinian if and only if it is negatively curved (in the large) (Gromov word hyperbolic) and has as its space ∂G at infinity the 2-sphere S^2.

Definition of Kleinian groups. We define a *Kleinian group* to be a group that acts isometrically, cocompactly, and properly discontinuously on 3-dimensional hyperbolic space.

All Kleinian groups are of course negatively curved in the large because their Cayley graphs are quasi-isometric with hyperbolic space:

Groups that are negatively curved (in the large) (that is, Gromov hyperbolic). A finitely generated group G is *negatively curved* (in the large) or *Gromov word hyperbolic* if its Cayley graph with respect to some finite generating set (hence with respect to all finite generating sets (theorem)) has uniformly thin triangles in the following sense: there is a nonnegative number δ such that if abc denotes a geodesic triangle in the Cayley graph Γ of G, and if x is any point of any side, say ab of abc, then there is a point y of the union $ac \cup bc$ of the two other sides of the triangle which lies within δ of x.

Our main theorem is that the conjecture is true if one imposes one extra modulus condition on the group. Again, we state the result before explaining all of the terms involved.

Main Theorem ([4, Theorem 8.2] and [5, Theorem 2.3.1]). A group G is Kleinian if and only if it is negatively curved, has the 2-sphere S^2 as its space at infinity, and satisfies the following additional condition:

> *Modulus Condition:* If Q is any annulus in $S^2 = \partial G$, then the sequence $D(n)$ of finite coverings of $S^2 = \partial G$ by combinatorial disks, defined below, assigns a sequence of combinatorial moduli to Q which is bounded away from 0.

Remark. The quasiconformal shape (modulus) of a closed annulus can be measured combinatorially and analytically. The analytic result is never 0. The theorem states that the group is Kleinian if and only if the combinatorial result is also asymptotically nonzero.

The modulus condition does not depend on the finite generating set used in defining Γ. It does not depend on the base point chosen below which is used in defining the covers by combinatorial disks. The covers by combinatorial disks can be replaced by commensurable coverings without changing the condition. (In this setting two sequences $\{S_n\}$ and $\{T_n\}$ of tilings are commensurable if there is a constant K such that S_n has bounded overlap (K) with T_n and T_n has bounded overlap (K) with S_n for each n.) The modulus condition can also be stated in terms of Gromov's coarse quasiconformal structure on ∂G. Whatever the statement, enough facts are known to show that the condition is a group invariant.

Combinatorial Disks. Given any negatively curved group G and any finite generating set for G, the space ∂G at infinity has a natural sequence of open covers given by its combinatorial disks. The coverings $D(n)$ by combinatorial disks are not *tilings* of Q. (They are shinglings: that is, finite covers by compact connected sets. They may overlap.) They are, however, enough like tilings that the theory developed in Section 2 applies almost without change. They are called combinatorial disks not because they are known to be topological disks but for two reasons: (1) The analogous construction in hyperbolic space itself yields round topological disks; and (2) the construction in some Cayley graphs yields topological disks. We shall now describe this sequence of coverings. See [5] for more details.

A *point* at infinity is represented by a geodesic ray in the Cayley graph. Two geodesic rays represent the same point at infinity provided that asymptotically they remain a bounded distance apart (thinness of triangles implies that they then actually remain distance $\leq 2\delta$ apart asymptotically).

Given a ray $R : [0, \infty) \to \Gamma$ representing a point at infinity and a positive number n, we may define the *combinatorial half space* $H(R, n)$ to be the set of points $x \in \Gamma$ such that the distance $d(x, R[n, \infty))$ is at least as small as the distance $d(x, R[0, n])$. If one used the same metric definition in classical hyperbolic space, one would obtain a true hyperbolic half space.

Finally we can define the *combinatorial disk* $D(R, n)$ to be the set of points at infinity (set of equivalence classes of geodesic rays) such that a representative ray $S : [0, \infty) \to \Gamma$ asymptotically lies arbitrarily far (infinitely far) within $H(R, n)$.

The collection of combinatorial disks forms a basis for a topology on ∂G which is finite-dimensional, metrizable, and compact.

Fix a base point, say the identity vertex of Γ, and consider only those rays which begin at the base point. Then the set $D(n) = \{D(R, n)\}$ of combinatorial disks, where R varies over the rays R which begin at the base point and where n is a fixed positive integer, forms a finite open cover of ∂G.

Combinatorial Moduli. If $\partial G = S^2$, and if Q is a topological quadrilateral or ring in ∂G, then we may define the *combinatorial modulus* $M_n(Q)$ in terms of the cover $D(n)$ of Q. We consider only the set $T_n(Q)$ of combinatorial disks which intersect Q. We then formally define the modulus by precisely the formulas used in section 2.

With these definitions completed, we can state the modulus condition more precisely:

Modulus Condition Restated: If Q is a topological quadrilateral or annulus in $\partial G = S^2$, then the sequence $M_n(Q)$ is bounded away from 0.

It is actually enough to prove a weaker condition.

Weaker Modulus Condition ([4, Axiom 0]): For each point $p \in \partial G = S^2$ and each neighborhood N of p, there is an annulus Q which surrounds p in N (separates p from the complement of N) which satisfies the modulus condition.

Remark. That the weaker condition is sufficient implies that one need only check the modulus condition for a finite number of appropriately chosen annuli, or even for only a finite number of quadrilaterals (conjecturally, one appropriately chosen quadrilateral). This reduction to checking only finitely many annuli or quadrilaterals depends on the great homogeneity at infinity entailed by the action of the group.

Example. We shall now describe the original example which motivated much of our work. This example was the simplest group we knew that was cocompact, discrete, and hyperbolic in dimension 3. Our precise definitions were all designed to mirror (albeit imperfectly) features of this example. There is no exact correspondence between the features of our example and our definitions, though one can with effort distort either the definitions or the example into the orbit of the other. We hope rather that the reader will accept the example as Plato's reality and the definitions as the imperfect shadows in the cave.

From this example we will naturally see a *real* 2-sphere at infinity corresponding to Gromov's constructed boundary, we will see round circles at infinity corresponding approximately to our combinatorial disks at infinity, and, after much fussing we could realize that the circles appear in recursive patterns. The discovery of the recursion demanded several days of tedious picture drawing, but the discovery led to automatic group theory. The attempts to pass from combinatorial "circle" data to round real circle data led to the combinatorial Riemann mapping theorem. In our conjectural view, the example has only two features that differ from those of the generic Kleinian group: the example has too much symmetry, and the tiles, geometrically realized, are too smooth. We expect, in general, much more twisting in the generic global pattern which realizes itself locally, under subdivision, in fractal tiles — quasidisks.

Our goals are modest. We first give the geometric description of the group and observe both the real sphere at infinity and the real circular disk at infinity. We simply *assert* the nature of the recursions of circles without proof. We then prove that the resulting sequence of tilings of S^2 formed by intersecting the given circular disks satisfies the Weak Modulus Condition. If all of these *real* objects were the constructed objects of the main Theorem, it would follow that the group is Kleinian, a fact already known at the start. But the point is that this Kleinian nature of the group would have been deduced from the *combinatorial* properties of the group.

As example, we take a group G known to be Kleinian for geometric reasons and indicate how one might verify the conditions of the Main Theorem without knowing the original geometric realization. The argument has some similarities with the proof that was given in Section 7 of [3]. The group has twelve generators of order 2 associated with the twelve faces of a dodecahedron. For each edge of the dodecahedron there are two adjacent faces with their two generators. The product of these two generators has order two. In other words, our group is a particularly nice Coxeter group (and is called the *right-angled dodecahedron group* since it can be realized as the group generated by reflections in the twelve faces of a regular right-angled dodecahedron in hyperbolic 3-space). This Coxeter group can also be realized as the fundamental group of an orbifold whose quotient space is the dodecahedron. One can recursively build either the Cayley graph or the universal covering orbifold, even though it is infinite, by hand. This is possible because negatively curved groups have a recursive structure at infinity (an automatic structure, say) which in practice can often be discovered manually. What is that structure at infinity?

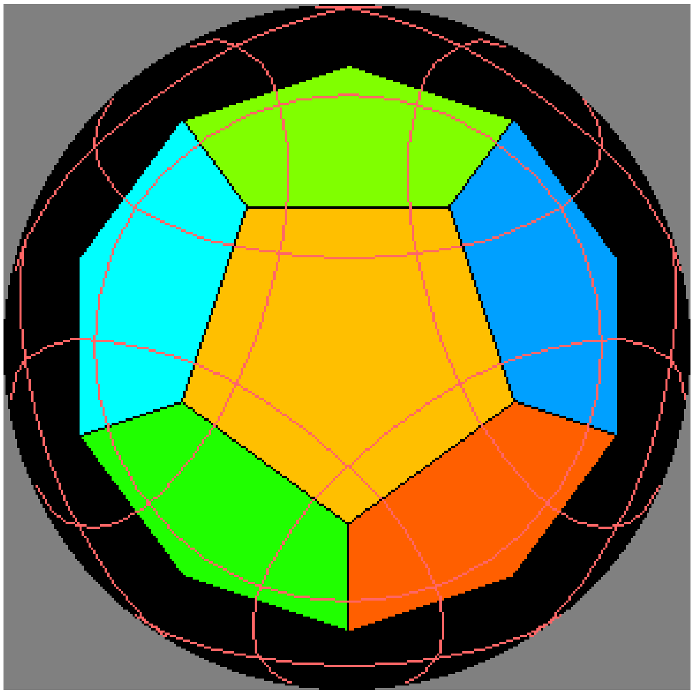


Figure 9. The initial circle pattern

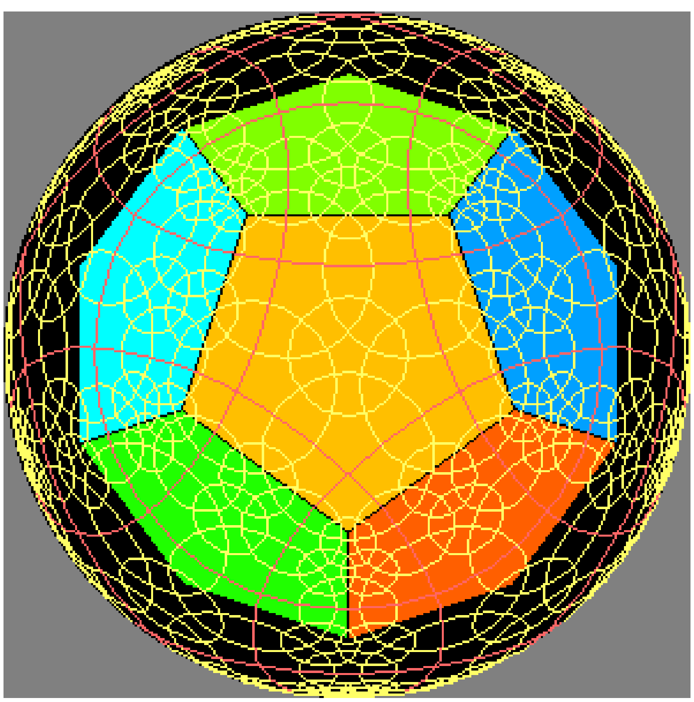


Figure 10. The circle pattern at level 1

The right-angle feature of the group allows one to associate with the group an infinite family of planes tiled by right-angled pentagons, four meeting at a vertex. These planes exist in the universal covering orbifold. These planes persist to infinity and define a family of simple closed curves at infinity. At least intuitively, and certainly commensurably, these circles correspond to the boundaries of combinatorial disks at infinity. Just as combinatorial disks can be partitioned into families by the parameter n, so also can these circles be partitioned by the parameter n which measures the combinatorial distance from a base point to the plane which defines the circle. According to this partition, the circle patterns become recursive. This recursion is indicated in Figures 9 and 10, which were produced by SnapPea from a modification written by Jeff Weeks.

Note that the circle patterns divide the plane into combinatorial triangles, quadrilaterals, and pentagons, each with its own subdivision rule. Thus we obtain from the intersection pattern an actual sequence of *tilings* of the 2-sphere at infinity. By commensurability considerations (the bounded overlap theorem of section 2) it is enough to check the weak modulus condition for this sequence of tilings.

It is an easy matter to check that the pentagonal tiles arise in three ways, namely as isolated singletons, as pairs sharing an edge, and as four pentagons situated about a vertex. See Figure 10.

It is possible to expand each pentagonal tile canonically to a pentagon (in the sense of a topological disk with five distinguised vertices) that is not covered exactly by tiles but in such a way that, with respect to our sequence of tilings, the expanded pentagons are combinatorially identical with one another at all levels of subdivision. See Figure 11, where the boundary of the expanded pentagon is drawn in bold. Figure 11 only indicates the boundary of the expanded pentagon. One has to subdivide once more for it to become apparent how the boundary is defined. This is shown in Figure 12.

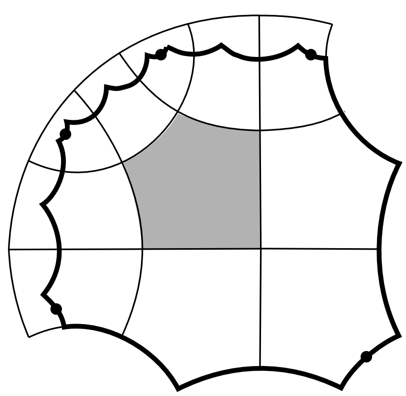


Figure 11. A shaded pentagon and its expanded pentagon

Each of these expanded pentagons is rotationally symmetric about a center point (see Figure 12) and reflectionally symmetric about an axis through a vertex and the midpoint of the opposite side.

Verification of the Weak Modulus Condition. As noted above, the tile coverings of the expanded pentagons respect, at all levels of subdivision, a dihedral

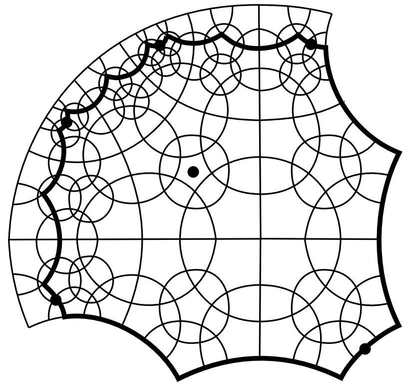


Figure 12. An expanded pentagon in the subdivision

symmetry of order 10. If you consider an expanded pentagon as a quadrilateral by thinking of a pair of adjacent edges of the pentagon as one edge of the quadrilateral and thinking of each of the other edges of the pentagon as one edge of the quadrilateral, then the quadrilateral will have reflectional symmetry. By the symmetry property of optimal weight functions, the same is true of those functions. In addition, the rotation theorem of section 2 implies that if we consider any two nonadjacent edges of such an expanded pentagon as the left and right sides of the quadrilateral and consider any fixed level of subdivision, it is possible to canonically construct a family of skinny cuts creating relatively large height but relatively small area. For this application of the rotation theorem, we need to view the expanded pentagon as a quadrilateral in two different ways.

These cuts (or naturally truncated portions of the same) will naturally match up with the corresponding skinny cuts (likewise truncated if necessary) of any adjacent pentagon, again by symmetry. Figure 13 shows the expanded pentagons of three shaded pentagons, and Figure 14 shows the corresponding cuts.

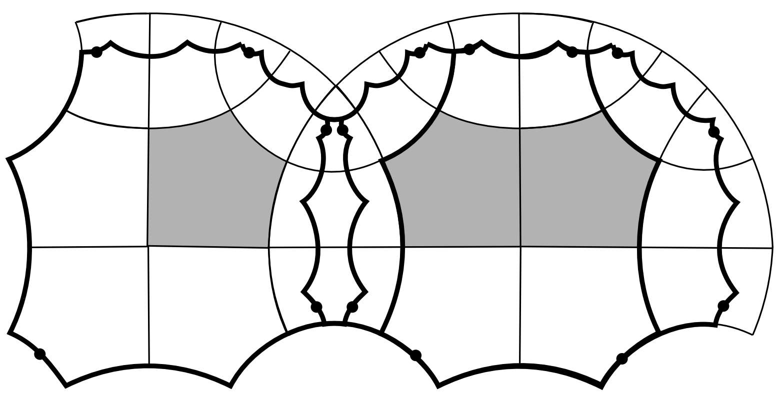


Figure 13. Expanded pentagons for three shaded pentagons

Thus, any cyclic sequence of pentagons is circled by a cycle of skinny cuts, of nontrivial height and moderately small area. A simple example is shown in Figure 15. In general, one needs crossing cuts in some pentagons in order to “turn the corner”.

Furthermore, the height and area depend only on the length of the cycle and not on the depth of the subdivision. Thus the weak modulus condition is satisfied

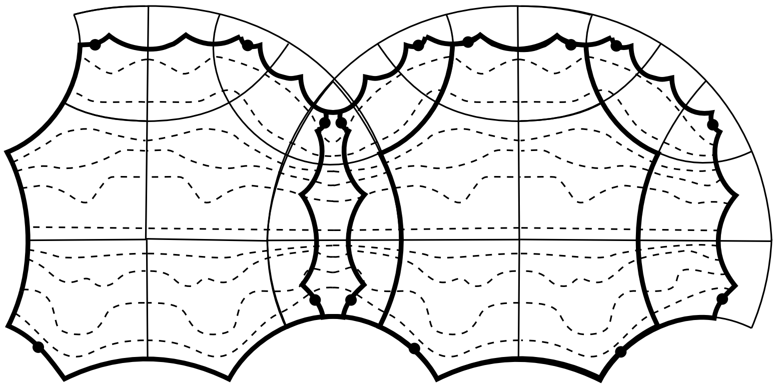


Figure 14. Cuts matching up by symmetry

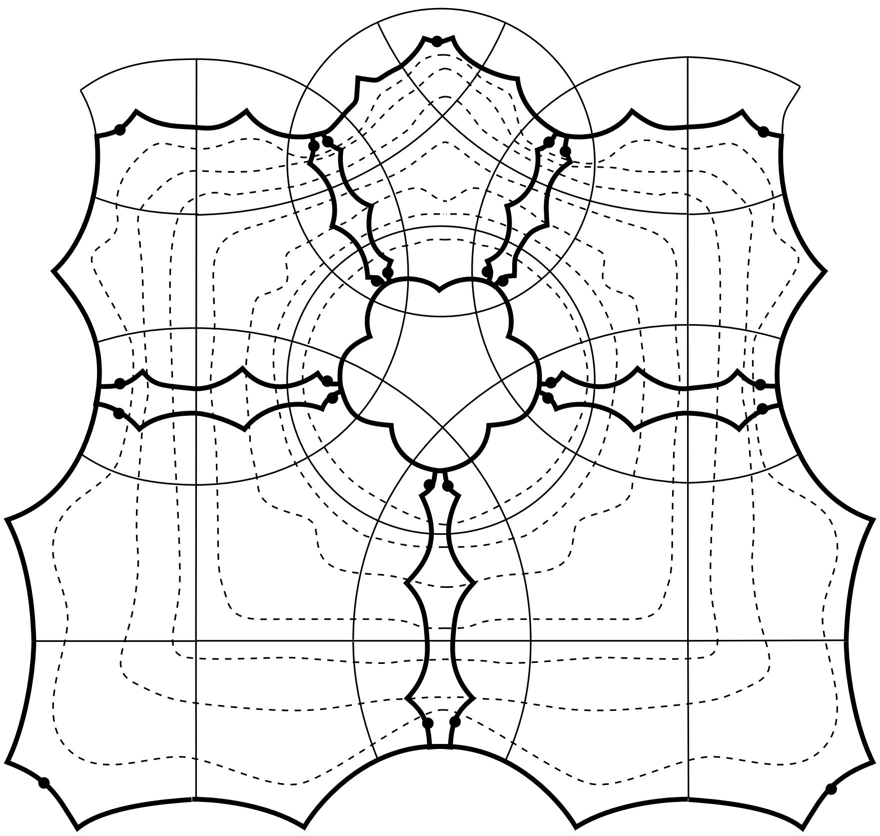


Figure 15. A cycle of cuts

and the group is Kleinian by the Main Theorem.

§5. Remarks and Questions.

How generic are the conditions we have used? With a negatively curved group there is always a natural sequence of coverings, but not necessarily by tiles. The tiles overlap and therefore should be thought of as *shingles*. There is always a natural *subdivision rule* obtained simply by moving outward along rays away from the base point in the Cayley graph. This recursion is always defined by finite rules. But the resulting subdivision rules do not seem to be exactly like the kind of subdivision rules we described in the previous section. In particular, there is some fuzziness (in our minds) about what happens at the edges of shingles.

Furthermore, in a negatively curved group, the shingles given by the combinatorial disks subdivide exponentially in every direction so that it is impossible to send coordinating signals about in the subdivision: the signals can only move at a linear rate while the shingles are dividing geometrically so that a signal always stays in the *light cone* of the place it started. That is, a signal can have infinite influence through time, but only local influence through space. Therefore we have the following question:

Question. Suppose that one has a subdivision rule which, when applied to a tiled quadrilateral, subdivides exponentially in every direction. Must the successive moduli of the quadrilateral necessarily either converge to 0 or infinity or remain in a bounded interval?

Question. In a negatively curved group, must there always be shingles that are almost rotationally symmetric? Under what conditions must there be reflectional symmetry? dihedral symmetry?

Rubinstein and Mosher have both suggested that cubulated 3-manifolds and their behavior at infinity might be a particularly rich source of examples of subdivision rules associated with negatively curved groups.

Question. Can our program be carried out for cubulated 3-manifolds?

Local replacement rules. In a finite subdivision rule or even in a context sensitive subdivision rule a vertex can only subdivide into a vertex though its label may change. An edge can only subdivide into a finite collection of edges and vertices. And so forth. Boundaries between cells are preserved over time. Some of the rules that are discovered by the algorithmic building of universal covers do not preserve such boundaries. These other natural rules might be described as local replacement rules: a vertex may be replaced by an entire cell of higher dimension or by an entire complex; likewise, cells of higher dimension might be replaced by complexes of even different dimensions, even by the empty set.

With a *local replacement rule*, we are given a finite number of types of cells-with-labels. Each type of cell is supplied with a labelled *replacement complex* by a rule which may or may not be context sensitive. Compatibility now becomes a real issue. We must know not only how to replace a labelled cell but also how to sew the entire union of replacement complexes together to form the *total replacement complex*. We choose to call this identification rule the *folding rule* in imitation of John Stallings's terminology used in constructing Cayley graphs or trees (essentially by Todd-Coxeter coset enumeration techniques). The situation may require that the image of a replacement complex actually undergo collapse during the folding process, again as suggested by the Todd-Coxeter procedure.

Note. We may think of every finite subdivision rule or context-sensitive subdivision rule as a local replacement rule. In this setting, we should probably accept the convention that every vertex arises not just from itself but also from the subdivision of every cell which contained it at the previous stage, and similarly for cells of higher dimension. In this manner, compatibility is subsumed under the folding rules. This convention allows us to distinguish finite valence subdivision rules as those which have bounded folding as described below.

Remark. With a local replacement rule, the topological type of the total complex can change with time. The total complex can grow forever or collapse catastrophically after innumerable generations of growth. Or the shape of the total complex can grow in an algorithmically unpredictable fashion. All of these possibilities are demonstrated by the construction of the Cayley graph of a group with unsolvable word problem by the method of Todd-Coxeter coset enumeration.

Negatively curved subdivision and replacement rules.

With all of the rules that we have described so far, given a beginning complex or *seed*, it is possible to associate a *history complex*. This complex consists of countably many levels together with connecting edges. Level 0 is the seed complex. Level n is the total replacement complex (after folding, if required) under subdivision or replacement for level $n-1$. The barycenter of each cell from the total replacement complex is joined by an edge to the barycenter of the cell (cells) in the previous level which spawned it.

A rule is said to have *bounded folding* if the history complex is locally finite of bounded degree. Such a complex can be assigned a metric which is essentially a *word metric*: that is, the distance between points is essentially the number of cells one must cross in going from point to point.

The subdivision rule most known to topologists is that of *barycentric subdivision*. Note that a vertex as it is subdivided is in the replacement complex of more and more 1-cells, 2-cells, etc. Hence this rule does not have bounded folding. A subdivision rule with bounded folding must have what has previously been called *finite valence*.

A rule is said to be *negatively curved* (in the large) or *Gromov hyperbolic* if it has bounded folding and its history complex with word metric has uniformly thin triangles in the sense of Gromov.

Remark. Just as negatively curved groups have many characterizations, so also do negatively curved subdivision rules. The most natural characterization requires first bounded folding and then that there be exponential subdivision in every direction. This characterization corresponds roughly to what has been called exponential divergence of geodesics in negatively curved spaces. Although in a negatively curved replacement rule it is theoretically possible to send signals, those signals can only travel at a linear rate while the subdivision is proceeding at exponential rate in every direction. As a consequence, a signal can affect ever increasing numbers of cells in future generations, but those cells affected will always be in the *light cone* of the individual sending the signal.

Conjecture. If a planar subdivision rule is negatively curved, then the modulus of a quadrilateral under subdivision either goes to 0 in the limit or goes to ∞ in the limit or oscillates in a finite interval.

Generalized cellular automata. A *cellular automaton* has an underlying cellular space in which the cells have labels. Then the labels change in a context sensitive way from generation to generation. Usually one of the labels is considered to indicate inactivity of the cell and only finitely many cells are to have the active label at any generation. One begins with a *seed* of active cells and traces the labelling pattern through time.

Cellular automata, as a model of organisms responding to their environment, were introduced by John von Neumann. He showed that their structure was rich enough to model a universal Turing machine. The most widely known cellular automaton is, of course, John Conway's game of life.

One can interpret a cellular automaton as a context sensitive subdivision rule in which the only effect of subdivision is to change labels.

All of our subdivision and replacement rules may be considered either special cases of or generalizations of the notion of cellular automaton. We particularly like to think of the local replacement rules as being *generalized cellular automata* which create their own underlying cell structure on the fly.

Almost convex groups. A finitely generated group is said to be *almost convex* if there is an integer k such that any two elements of the Cayley graph which lie in the metric ball of (arbitrary) radius n in the graph and which lie within distance 2 of each other in the graph also lie within distance k of one another in the ball of radius n.

Exercise. Show that the Cayley graph of an almost convex group can be built efficiently by a local replacement rule with bounded folding in the following sense: the total complex at generation n should be the ball of radius n in the Cayley graph together with an additional labelling; all nontrivial replacements performed in moving from the n ball to the $n+1$ ball should take place only at the surface of the n ball; all folding that takes place should be local and take place only at the surface of the n ball.

Exercise. (Essentially a tautology) A group is almost convex if and only if it can be constructed efficiently in the sense described in the previous paragraph by a local replacement rule with bounded folding.

Question. Is there a good family of constructions leading to almost convex groups? How does one construct an almost convex group? How does one recognize an almost convex group?

Remark. Although the property of almost convexity is not a group invariant, it can be made group invariant by the usual cheat: there exists a finite generating set with respect to which the group is almost convex.

References

[1] Cannon, J. W., *The combinatorial Riemann mapping theorem*, Acta Math. **173** (1994), 155–234.

[2] Cannon, J. W., Floyd, W. J., Parry, W. R., *Squaring rectangles: the finite Riemann mapping theorem*, The Mathematical Heritage of Wilhelm Magnus — Groups, Geometry & Special Functions, Contemporary Mathematics, vol. 169, Amer. Math. Soc., Providence, 1994, pp. 133-212.

[3] Cannon, J. W., Floyd, W. J., Parry, W. R., *On the conformal invariance of tiling-systems*, preprint.

[4] Cannon, J. W., Floyd, W. J., Parry, W. R., *Sufficiently rich families of planar rings*, preprint.

[5] Cannon, J. W., Swenson, E. L., *Recognizing constant curvature discrete groups in dimension*

3, preprint.

[6] Gromov, M., *Hyperbolic groups*, Essays in Group Theory (S. M. Gersten, ed.), MSRI Publications, vol. 8, Springer-Verlag, Berlin Heidelberg New York, 1987, pp. 75–264.

[7] Schramm, O., *Square tilings with prescribed combinatorics*, Israel J. Math. **84** (1993), 97–118.

Department of Mathematics, BYU, Provo, UT 84602, U.S.A.

Department of Mathematics, VPI&SU, Blacksburg, VA 24061, U.S.A.

Department of Mathematics, EMU, Ypsilanti, MI 48197, U.S.A.