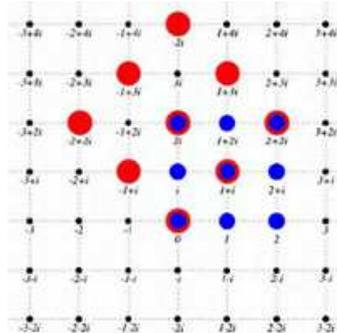


Indra's Pearls

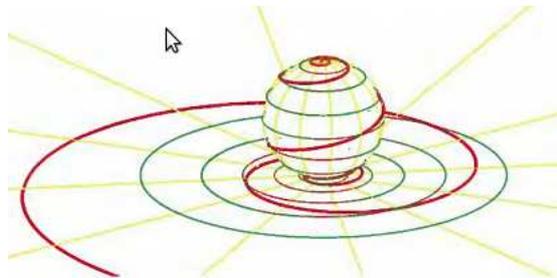
There is an ancient Buddhist metaphor attributed Tu-Shun (who lived around 600 B.C.) involving something referred to as Indra's Jeweled Net. In this metaphor one is supposed to imagine a vast net with the following properties: at each juncture there is a jewel, and each jewel reflects all other jewels. Thus each jewel is reflected infinitely in all of the others (the metaphor is supposed to show how everything in the universe is intimately connected and intertwined). In the book "Indra's Pearls", by David Mumford, this sort of phenomenon is studied using mathematical ideas. This book is the source of information for this paper (mostly chapters three through six). Before we can begin to look at examples similar to Indra's Jeweled Net a few preliminaries are needed. The basic building block of this subject is something called Möbius transformations- that is what will be addressed first.

Möbius transformations are the maps from the Riemann sphere to itself that send circles to circles, and preserve angles between circles that intersect. Clearly all scalings, rotations, and translations will do this, and thus are Möbius transformations. What may not be quite as obvious is that the function  $f(z) = \frac{1}{z}$  also satisfies these properties. Thus compositions of these maps are also Möbius transformations. Moreover, with compositions of these maps one can find all transformations that satisfy the needed conditions. In general a Möbius transformation can be written in the form  $f(z) = \frac{az+b}{cz+d}$ , where  $a, b, c, d \in \mathbb{C}$ , and is defined as long as  $ad - bc \neq 0$ . Notice that if two transforms are composed the result is a transform: let  $T(z) = \frac{az+b}{cz+d}$  and  $\tilde{T}(z) = \frac{a'z+b'}{c'z+d'}$  then  $T(\tilde{T}(z)) = \frac{a \frac{a'z+b'}{c'z+d'} + b}{c \frac{a'z+b'}{c'z+d'} + d} = \frac{(aa'+bc')z+(ab'+bd')}{(ca'+dc')z+(cb'+dd')}$ . If the maps are written as a matrix then this is just matrix multiplication:  $T((\tilde{T})(z)) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$  where  $z = x + yi$ . The requirement that  $ad - bc \neq 0$  implies that the determinant of a matrix is nonzero and thus every Möbius map has an inverse. It follows that these maps form a group. There is one last restriction that is placed on these matrices- that the determinant should equal one. Nothing is lost in this requirement because  $\frac{az+b}{cz+d} = \frac{k(az+b)}{k(cz+d)}$ , so multiplying the matrix by a scalar won't change the map. So for convenience all matrices are taken to be of determinant one.

So what do Möbius maps look like? A few examples are in order. One of the simplest of these is the map  $f(z) = kz$ , where  $k \in \mathbb{C}$ . For a concrete example consider the map  $f(z) = (1+i)z$ . This map rotates points counter-clockwise  $\frac{\pi}{2}$  radians around the origin and expands by a factor of  $\sqrt{2}$ . Here is a picture of this map acting on nine points in the complex plane.



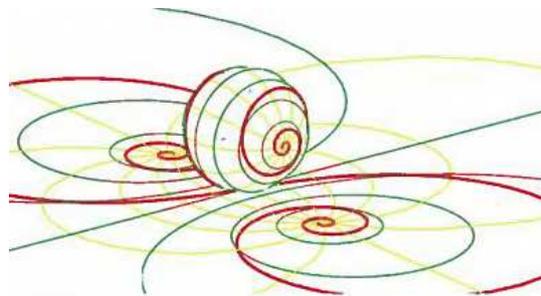
The red points in the figure are the images of the blue points under  $f(z)$ . What happens if the map is repeated? The points are then rotated and expanded again. Similarly inverses can be taken, which will contract and rotate the points in the opposite direction. If one looks at the orbit of a point under this map it looks like points spiraling out from around the origin. Connecting these points in a continuous line gives a picture like this one:



In this picture the red line spiraling out in the plane is the orbit of a 1-parameter subgroup under the map. It should be clear in the plane that the origin is a fixed point (as  $f(0) = 0$ ), but this picture shows there is another fixed point. In the picture the plane is inverse stereographically projected onto the Riemann sphere and it becomes obvious that the point at infinity is also a fixed point. This makes sense because  $f(\infty) = \infty$ . The point at infinity is called a sink because iterations of  $f$  moves other points closer to the point at infinity. Likewise the origin is called a source because iterations

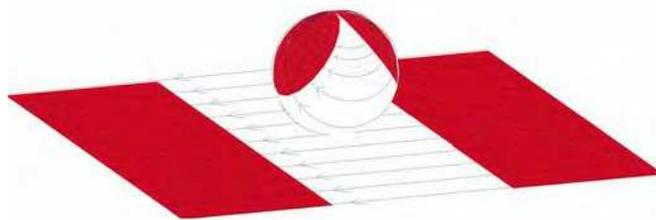
of  $f$  moves points further away from the origin.

Suppose we want a similar type of map but instead the fixed points being at  $0$  and  $\infty$  we want them to be at  $-1$  and  $1$ , and more specifically we want  $-1$  to be the source and  $1$  to be the sink. This can be accomplished by conjugation. Define  $H(z) = \frac{z-1}{z+1}$ . Then  $H(0) = -1$  and  $H(\infty) = 1$ . So  $H$  moves the source and sink of  $f$  to the desired coordinates. The map  $HfH^{-1}$  should give the desired result-  $HfH^{-1} = \frac{z\iota+2}{z(-2-\iota)-\iota}$ . Notice that  $-1$  and  $1$  are fixed points of this map. A picture of this (similar to the last, with the Riemann sphere above the plane) immediately shows what this map is doing.



The red lines show the orbit of a point spiraling away from  $-1$  toward the point  $1$ . When looking at this on the Riemann sphere an interesting observation is that the sphere is simply rotated  $\frac{\pi}{2}$  radians. Maps of this form (i.e. conjugate to  $f(z) = kz$  for some  $k \in \mathbb{C}$ ) are called **loxodromic** if  $|k| > 1$ , and **elliptic** if  $|k| = 1$  (i.e. are just rotational).

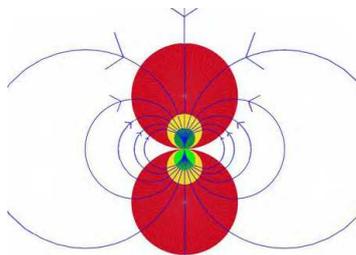
Now let's look at maps that translate the plane. These are maps of the form  $f(z) = k+z$ . On the plane this is pretty easy to visualize- every point is just moved over  $x$  units and up  $y$  units, where  $k = x + yi$ . But what happens on the sphere?



In the above figure the plane, with the sphere situated above it, is shown. In this picture a purely horizontal shift is being considered ( $y = 0$ ). The

white part of the plane matches the amount corresponding to how much the plane is being shifted. As the plane is shifted (to the left) the white strip gets shoved into the red strip on the left and the red strip on the right is moved over to cover where the middle white strip is. Viewed on the Riemann sphere the right and left side red strips correspond disks that are tangent at the point of infinity. As the plane is translated to the left the right red disk expands to also cover the area that is white. The white strip gets shoved over, along with the left disk, and shrink to occupy the space where the left red disk is. If this is iterated then a pattern of concentric “scallop schells” will be formed on the left side of the sphere. Likewise if the inverse transform (translating to the right) is iterated the same sort of pattern will occur on the right side of the sphere.

Just as with the previous scaling map we can get a different viewpoint of the translating map by conjugating. Since there is only fixed point on this map ( $\infty$ ), let's try moving the point of infinity to 0 and vice-versa. Notice the map  $H(z) = \frac{1}{z}$  does this. The conjugated map  $H(f(H^{-1}(z))) = \frac{1}{kz+1}$  then has red disk tangent at the origin (corresponding to the disks tangent to the point at infinity on the sphere).



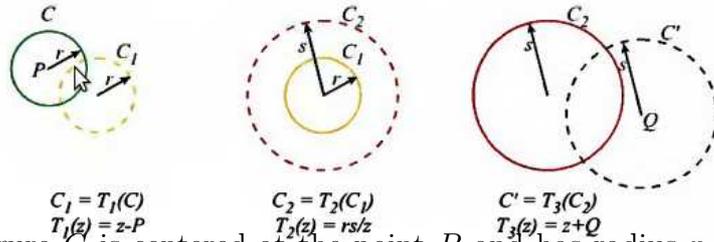
In this picture the blue lines correspond to the translation lines in the previous picture. Thus viewed on the plane  $HfH^{-1}$  pushes the top red disk and the white region outside of it onto the bottom red disk, and the bottom red disk down onto the bottom yellow disk. As this map is repeated the bottom disk is shoved further inside itself. As  $HfH^{-1}$  is iterated the disks shrink smaller, limiting a point at the origin. Also, if the inverse transform is repeated the bottom disk gets pushed up to the top one, and the top one gets pushed down into itself. Maps of this kind (that is conjugate to a translation) are called **parabolic**. Notice that parabolic maps have exactly one fixed point, whereas loxodromic and elliptic have two. As a side note, these maps can also be classified according to the trace of their matrix. Loxodromic maps will always have a trace not between  $-2$  and  $2$ , elliptic are

strictly between  $-2$  and  $2$ , and for parabolic maps the trace is  $\pm 2$ .

Now that we have a basic characterization of Möbius maps we can look at a variety of Möbius maps that will allow us to develop a certain type of map that will be the building blocks for Indra's pearls. The first Möbius transform we want to consider is known as the Cayley map, defined by  $K(z) = \frac{z-i}{z+i}$ . Notice that  $K(0) = -1$ ,  $K(1) = -i$ ,  $K(-1) = i$ , and  $K(\infty) = 1$ . Since Möbius maps send circles to circles, and the real line is a circle through infinity, it follows that the Cayley map sends the real line to the unit circle. In fact, it also follows that  $K$  maps points in the upper half plane to the inside of the unit circle (take for example  $K(i) = 0$ ).

With the above map in hand we can find the maps which take the unit disk to itself. Suppose we know the maps which send the upper half plane to itself, then conjugation by these maps of the Cayley map will take the disk to itself. So the question becomes what Möbius transforms send the upper half plane to itself? Such maps should send the real line to itself- this corresponds to ones of the form  $f(z) = \frac{az+b}{cz+d}$ , where  $a, b, c, d \in \mathbb{R}$ . This will work because if  $z$  is real then clearly  $f(z)$  will also be real. There is one catch though, consider the map  $f(z) = \frac{1}{z}$ .  $f(z)$  switches the half planes, sending the upper one to the lower one and vice-versa. But viewed as the matrix  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  but we are "saved" by the fact that the determinant of this matrix is  $-1$ . Requiring the determinant to be  $1$  ensures that  $f$  maps the upper half plane to itself. This group of maps is known as  $SL(2, \mathbb{R})$  and is a subgroup of  $SL(2, \mathbb{C})$ . Thus we can find that any map,  $T$ , that sends the unit disk to itself is of the form  $T = KGK^{-1}$ , where  $K$  is the Cayley map and  $G \in SL(2, \mathbb{R})$ . Carrying out the algebra this can be simplified down to  $T = \begin{bmatrix} u & v \\ \bar{v} & \bar{u} \end{bmatrix}$ . Where  $u$  and  $v$  satisfy  $|u|^2 - |v|^2 = 1$ . Notice that the trace of  $T$  has imaginary part  $0$ , so  $T$  is either elliptic or parabolic, but can't be loxodromic.

Now we are finally ready for the main tool used to construct Indra's necklace- Möbius maps that pair circles. Suppose there are two disjoint circles  $C$  and  $C'$  in the plane. To pair the circles mean we find a Möbius map, say  $H$ , that takes everything outside of  $C$  and puts it inside of  $C'$ , and likewise everything inside of  $C$  is mapped to the outside of  $C'$ . One way to go about doing this is broken down into three steps and shown below:



$$C_1 = T_1(C)$$

$$T_1(z) = z - P$$

$$C_2 = T_2(C_1)$$

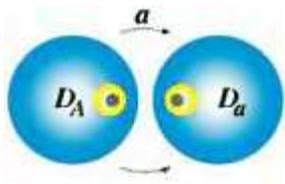
$$T_2(z) = rs/z$$

$$C' = T_3(C_2)$$

$$T_3(z) = z + Q$$

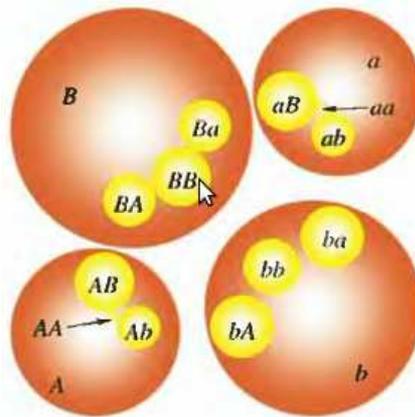
In the figure  $C$  is centered at the point  $P$  and has radius  $r$ , and  $C'$  is centered at  $Q$  with radius  $s$ . In the first step,  $T_1$ ,  $C$  is moved to the origin by subtracting off the point  $P$ - this circle is labelled  $C_1$ . In the next step  $T_2$ ,  $C_1$  is first made into a disk of radius 1, then inverted (mapping everything inside the unit disk to the outside, and everything outside to the inside), and then finally expanded to a circle of radius  $s$ - this circle is labelled  $C_2$ . The final operation is to move  $C_2$  over so that it is centered at  $Q$ . This is accomplished by the map  $T_3(z) = z + Q$ . When these maps are composed ( $H(z) = T_3(T_2(T_1(z)))$ , or  $H = T_3T_2T_1$ ) we get that  $H$  maps everything outside of  $C$  to the inside of  $C'$ , and everything inside of  $C$  to the inside of  $C'$ . But is this the only map that does this? No. During the middle step when the circle  $C_1$  was mapped onto the unit circle we could have then inserted one of the previous maps we discovered that send the unit circle to itself (i.e ones of the form  $T(z) = \frac{uz+v}{vz+u}$ ). Doing the algebra we find that the following map will pair the circles  $C$  and  $C'$ :  $M(z) = s \frac{\bar{u}(z-P)+r\bar{v}}{u(z-P)+rv} + Q$ . Not only does this map pair  $C$  and  $C'$ , but it also leaves a degree of freedom with  $u$  and  $v$  (remember that  $|u|^2 - |v|^2$  must equal 1). This extra degree of freedom allows one to pick a point inside  $C$  to be mapped to the point  $\infty$  outside of  $C'$ - this will be important later.

With this tool of being able to pair circles in hand we can start to look at Indra's pearls. Specifically we will be looking at what happens when maps that pair circles interact. To do this we will adopt some notation. Denote by  $a$  the map that pairs disks  $D_A$  and  $D_a$ , that sends the outside of  $D_a$  to the inside of  $D_A$ . Also denote  $a^{-1}$  by  $A$ . Consider two disjoint disks in the plane,  $D_A$  and  $D_a$ , that are paired by the map  $a$ .



In applying the map  $a$  once everything outside  $D_A$ , namely  $D_a$  is put inside  $D_a$ . So the yellow disk inside  $D_a$  corresponds to the map  $a$  being applied to the plane once. What happens if  $a$  then acts again on this? We get the little green disk inside of the yellow one. Iterations of  $a$  will produce concentric disks inside of  $D_a$  that seems to be limiting down to a point. This point that the disks are shrinking down to is called a fixed point of the map  $a$ . This is because this point will be left invariant under  $a$ . Similarly if the inverse map  $A$  is iterated there will be a sequence of concentric disks in  $D_A$  that shrink down to a fixed point. Thus  $a$  has a fixed point in  $D_a$  and  $A$  has a fixed point in  $D_A$ .

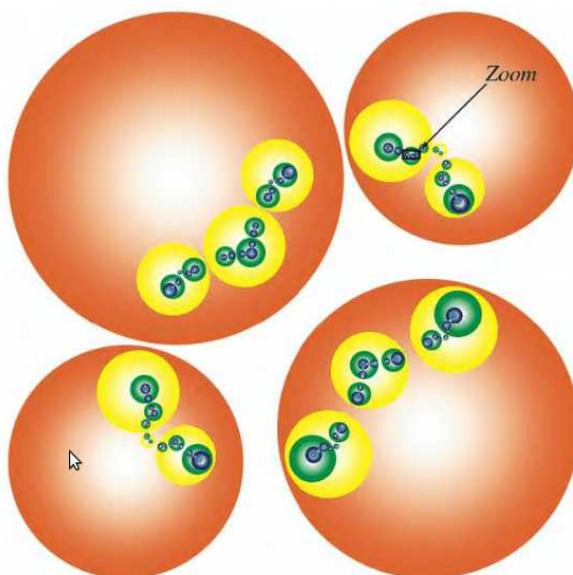
Now suppose there are four disjoint disks in the plane,  $D_a, D_A, D_b,$  and  $D_B$ , that are paired by the respective maps  $a$  and  $b$ . Let's look at what happens when the pairing maps are applied to these disks.



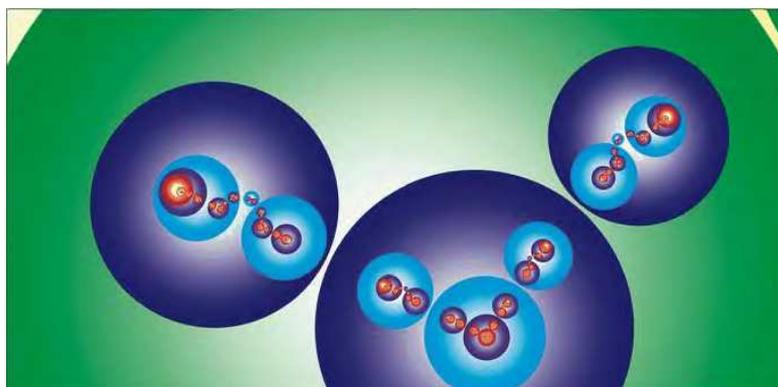
This picture shows the four disjoint disks and the image of the four Möbius maps ( $a, A, b, B$ ). The disks are simply labeled  $A, a, B,$  and  $b$  and  $B$  (which does leave ambiguity in distinguishing between maps and disks). Let's look at the image of  $b$  in this picture.  $b$  maps everything outside of disk  $B$  onto disk  $b$ . Since the disks  $A, a,$  and  $b$  are outside of disk  $B$  they get mapped onto disk  $b$ . Then the disk labeled  $bA$  is the image of disk  $A$  under the map  $b$ , the disk labeled  $bb$  is the image of disk  $b$  under the map  $b$ , and the rest are labeled similarly. The image of the white region outside of the disks under  $b$  is the inside of the disk  $b$ , excluding the three image disks inside ( $bA, bb,$  and  $ba$ ). Also, the image of the disk  $B$  under  $b$  is just the region outside of disk  $B$ .

The four original disks are referred to as level one disks, and the twelve

disks inside the first four are referred to as level two disks. A very natural thing to do is iterate the maps to draw further level disks. Doing this we get a picture that looks like the following:



Inside each of the level one disks there are three other disks (level two) which are the images of the Möbius transformations applied once. Also inside each of the level two disks there are three other disks which are the images of compositions of two Möbius transformations (e.g.  $ab$ ,  $BA$ , or  $bb$ ) applied to one of the original level one disks. This pattern repeats as the maps are iterated and composed in different combinations. Indeed, here is the zoomed in section of the box marked zoom in the above picture:



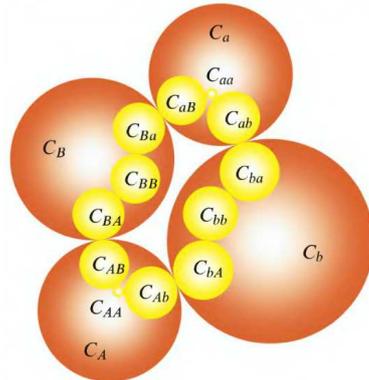
It looks like there are points that the successive level disks are shrinking down to, but it's not quite as obvious as when there was only one pair of disks. Indeed there are limit points, but they are sort of "scattered" in a way reminiscent of the Cantor set. This is true because the size of the disks (radii) at a given level, say  $l$ , are all less than  $Ck^{-l}$  where  $C$  is some positive constant, and  $k > 1$ . Or in other words the size of the disks shrinks at an exponential rate relative to the level of the disk. Thus the limit set must be points.

Compositions of these pairing maps still pair disks (and Möbius transformations as we saw earlier) so this forms a group. A collection of maps that pair circles (defined on the Riemann sphere) is called a **classical Schottky group** (as a side note the general Schottky group is a collection of transformations that pair disjoint shapes homeomorphic to a circle). Elements of this group are compositions of these transformations. This group acts on the Riemann sphere, and the fixed points of the groups are the points to which the disks shrink down to when all compositions of these functions are iterated. A disk at any level can be described by a string consisting of the letters  $A, a, B$ , and  $b$ . For example the string  $AbbaBaB$  corresponds to the image of the disk  $B$  under the map  $AbbaBa$  (using the same ambiguous looking notation that was used earlier). Thus points that are fixed under both maps  $a$  and  $b$  simultaneously can be thought of, or represented by, infinite strings of the letters  $a, A, b$ , and  $B$ - or put another way as the boundary of the Cayley graph for the free group on two generators.

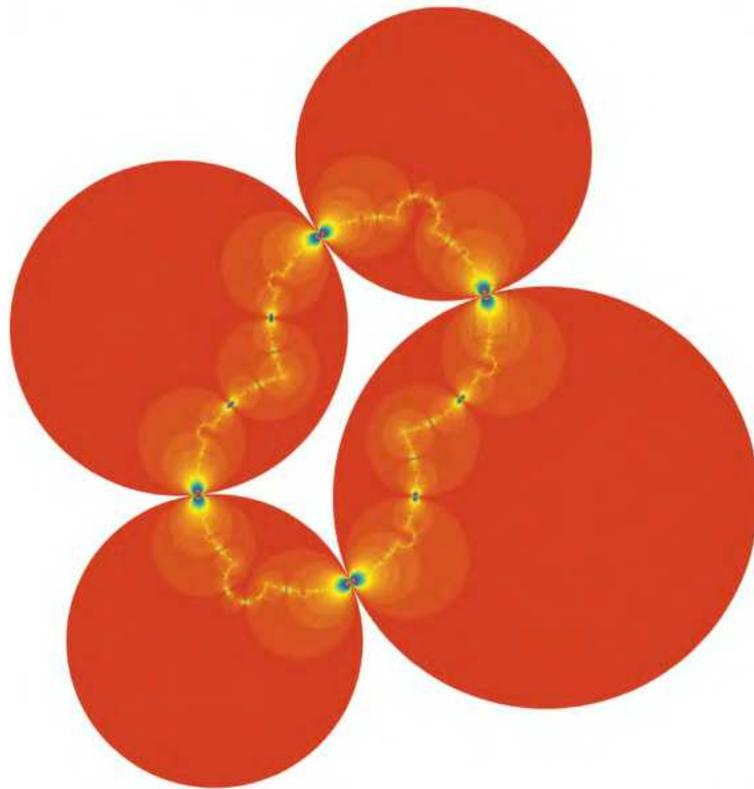
At this point it should seem fairly obvious that a Schottky group is isomorphic to the free group on  $n$  generators, where  $n$  is the number of transformations being used that pair circles. To see this suppose that this isn't true. Then there is some reduced word, say  $a_1^{i_1}a_2^{i_2}\dots a_k^{i_k}$ , where each  $a_i$  is one of the Möbius transformations that pairs circles, and  $i_i \in \mathbb{Z}$ , such that  $a_1^{i_1}a_2^{i_2}\dots a_k^{i_k} = 1$ . Pick some point,  $p$ , that is not in any of the disjoint disks. Then consider  $a_1^{i_1}a_2^{i_2}\dots a_k^{i_k}(p)$ . Since  $p$  is not in any of the disks it is first mapped by  $a_k$  into one of the disks. If  $i_k > 1$  then  $p$  is still mapped into the same disk. Since  $a_1^{i_1}a_2^{i_2}\dots a_k^{i_k}$  is a reduced word  $a_k$  and  $a_{k-1}$  are not inverses, so  $a_{k-1}$  must map  $p$  into another disk. Continuing this and following  $p$  through the chain of compositions we can see that  $p$  will always be inside a disk after the initial map  $a_k$  and will never get outside of one (as a result of inverse maps never being next to each other because the word is reduced). Thus  $a_1^{i_1}a_2^{i_2}\dots a_k^{i_k}(p)$  is inside some disk. But  $p$  started outside of all disks, so it we get a contradiction and it must be that  $a_1^{i_1}a_2^{i_2}\dots a_k^{i_k} \neq 1$ . Thus the

group is free.

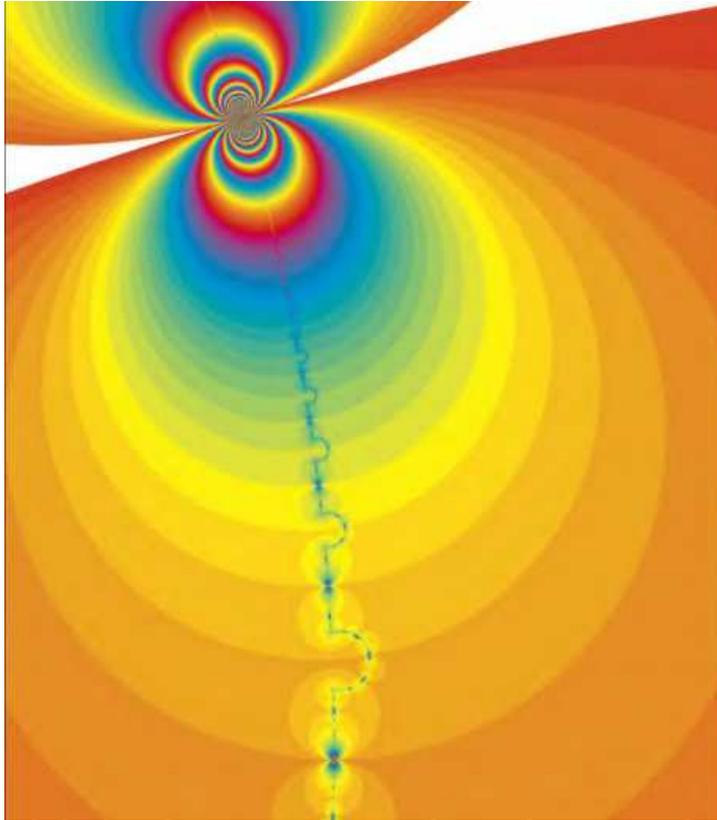
The last thing to be addressed is Indra's necklace. The term necklace should invoke the idea of something homeomorphic to a circle. Indeed this will be accomplished by looking at a pair of tangent circles. The desired setup looks like this:



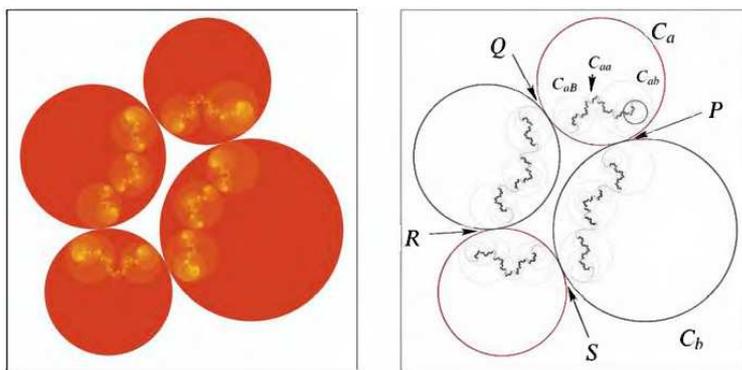
From this picture, and considering what was previously done with this sort of picture, it should be obvious what it is we want to do next- iterate the pairing Möbius maps. Doing this we get a beautiful picture:



and to get an even better view here is a zoom of the small rectangle on the first picture:



To accomplish this there are only two things that we need to worry about. The first is that the disks need to line up properly. To see what is meant by this look at the following picture where this condition isn't met.

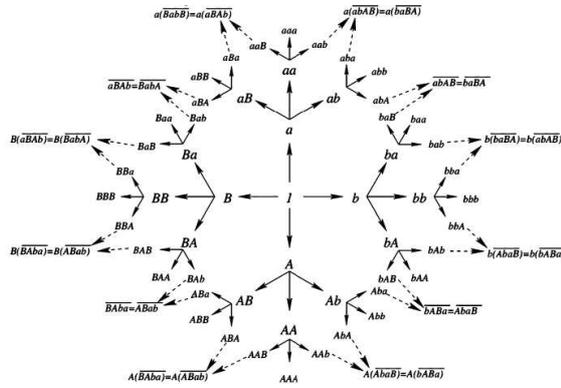


The four paired disks are arranged as the ones above, but something is wrong with the transformation that causes the limit sets inside of the disks to not

line up with the others where they touch. Notice that the image of the points  $R$  and  $S$  by the transformation  $a$  are not where we want them to be-  $Q$  and  $P$ , respectively. In order that the points line up under the maps the following is required:  $a(R) = Q$ ,  $a(S) = P$ ,  $b(R) = S$ , and  $b(Q) = P$ . Writing some of these using their inverses we get the following requirements:  $A(P) = S$ ,  $B(S) = R$ ,  $a(R) = Q$ , and  $b(Q) = P$ . Notice that this is the same as saying that  $a(b(A(B(P)))) = P$ . In other words  $P$  is a fixed point for the commutator  $abAB$ . Similarly  $R$  must be a fixed point for the commutator  $baBA$ , and with the other two fixed points the story is the same. Clear back at the first of the paper there was a degree of freedom mentioned in maps that pair circles. This is where that degree of freedom is needed to guarantee that we can line up the concentric disks in the appropriate manner.

The second requirement to get Indra's necklace is that the commutator maps must all be parabolic. Remember that parabolic maps have only one fixed point. This is needed because where the disks touch (and the maps need to interact in just the right way to line everything up) the maps need to shrink the disks down to the same point in both the disks, where they are tangent. If the commutator maps were loxodromic then there would be two fixed points and the disks wouldn't shrink down on both sides to the single tangent point.

A **quasifuchsian group** is a collection of Möbius transformations such that the fixed point (or limit) set of the group is homeomorphic to a circle. Thus maps that pair tangent circles in the above manner are a quasifuchsian group. There is one interesting thing mentioned about this group in "Indra's Pearl", although I'm not sure if it is of any value. In the case of the Schottky group where all off the paired circles were disjoint one could think of fixed points in the group as being represented by an infinite word of letters in the group. Furthermore every fixed point had a unique infinite word associated with it. Here that is not quite the case, as there is a shared point between circles that touch, and there are disks shrinking down to that point from both sides. In the case of four paired circles this corresponds to the image (point) of  $D_a$  under the map  $BabABabABabA\dots$ , denoted by  $a(\overline{BabA})$ , to be equal to the image (point) of  $D_a$  under the map  $aBAbabABabABab\dots$ , denoted by  $a(\overline{aBAb})$ . Similarly  $B(\overline{BAb}) = B(\overline{ABab})$ . The others are shown in the Caley graph below.



Since words are considered to be finite I don't know what this means (if anything at all), but I do think it is something that is noteworthy.

To conclude I would just like to once again state the reference for this paper: "Indra's Pearls", by David Mumford. It should also be noted that all images in this paper came from the same book.