THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING IN MATHEMATICS

On the ring lemma

Jonatan Vasilis

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY AND GÖTEBORG UNIVERSITY
Göteborg, Sweden 2008

On the ring lemma

Jonatan Vasilis
Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden

Printed at the Department of Mathematical Sciences, *idem*,
Göteborg, Sweden 2008
Preprint 2007:43
ISSN 1652-9715

On the ring lemma
Jonatan Vasilis
Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University

Abstract

The sharp *ring lemma* states that if $n \geqslant 3$ cyclically tangent discs with pairwise disjoint interiors are externally tangent to and surround the unit disc, then no disc has a radius below $c_n = (F_{n-1}^2 + F_{n-2}^2 - 1)^{-1}$ – where F_k denotes the k^{th} Fibonacci number – and that the lower bound is attained in essentially unique *Apollonian* configurations.

Here we give a proof by transforming the problem to a class of *strip configurations*, after which we closely follow Aharonov's and Stephenson's method of proof [3].

Generalizations to three dimensions are discussed, a version of the ring lemma in three dimensions is proved, and a natural generalization of the extremal two-dimensional configuration – thought to be extremal in three dimensions – is given. The sharp three-dimensional ring lemma constant of order n is shown to be bounded from below by the two-dimensional constant of order $n - 1$.

Keywords: ring lemma, circle packing, sphere packing, Apollonian.

Contents

Introduction 1

Consider the unit disc in $\mathbf{R}^2$ and a disc externally tangent to it. Successively add discs that are externally tangent to both the unit disc and the previous disc, stopping when the first disc is intersected and the unit disc is enclosed by the discs (figure 1). If n discs with pairwise disjoint interiors surround the unit disc, how small can such a disc be? Let c_n denote the infimum of the radii. The sharp Rodin-Sullivan *ring lemma* gives that c_n is a reciprocal integer and also that the infimum is attained in essentially unique configurations.

The observation that $c_n > 0$ was made by Rodin and Sullivan [16] in their proof of Thurston's conjecture regarding the convergence of the discrete Riemann mapping to its classical counterpart. Shortly afterwards, Hansen [12] derived a configuration in which the infimum is attained and gave a nonlinear recursion formula for calculating c_n.

Aharonov [1] used the Descartes circle theorem – see below – to give a closed expression for c_n and also to prove that c_n is a reciprocal integer for all $n \geqslant 3$.

Later, Aharonov and Stephenson [3] (originally published in [2]) gave a more detailed proof of the extremal configuration – including uniqueness – and also showed that the extremal configuration is a special case of a more general class of so-called Apollonian circle packings. They also expressed c_n in terms of Fibonacci numbers, $c_n = (F_{n-1}^2 + F_{n-2}^2 - 1)^{-1}$. Furthermore, they showed that the extremal configuration is also extremal with respect to angles between the centers of two tangent discs, as seen

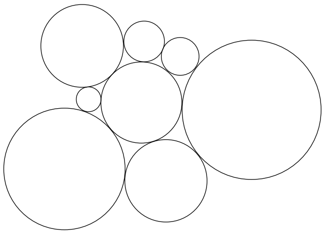


FIGURE 1 *Seven discs surround the unit disc. How large must the smallest disc be?*

from the origin. Stephenson has also given a different proof [21], closer to that of Hansen.

Acknowledgments

The author would like to thank his advisor, T. Lundh, for the encouragement to study generalizations of the ring lemma, as well as for countless suggestions during the course of the work. Also, several improvements of the presentation were made possible by helpful comments provided by P. Sjögren.

2 Preliminaries

In general, we let r, r', r_i and so on denote the radii of correspondingly marked discs D, D', D_i. Half-planes are considered to be discs of infinite radius in $\mathbf{R}^2 \cup \{\infty\}$, and two such discs have disjoint interiors if they only intersect at infinity, in which case they are also tangent. All discs are closed.

Definition. *Discs* $D_1, \ldots, D_n$, $n \geqslant 3$, *that have pairwise disjoint interiors and are externally tangent to the unit disc* D_0 *are said to* surround D_0 *if, after renumbering the discs counterclockwise by their tangency with* D_0*:*

(i) D_i *and* D_{i+1} *are tangent, and*

(ii) $\alpha_i \leqslant \pi$, *where* α_i *is the length of the counterclockwise arc of* D_0 *from* $D_i \cap D_0$ *to* $D_{i+1} \cap D_0$,

where $D_{n+1} = D_1$ *and* $i = 1, \dots, n$.

Tangencies imposed by (i) are said to be ordinary *tangencies, whereas other tangencies between* $D_1, \dots, D_n$ *are said to be* extra *or* non-trivial *tangencies.*

Remark. If at most one of the discs has an infinite radius, condition (ii) is equivalent to requiring that the discs separate D_0 from infinity. □

If a configuration of n discs surrounding the unit disc has extra tangencies, we get a *subconfiguration* of $k < n$ discs surrounding the unit disc by removing the redundant discs.

Rodin and Sullivan proved [16] the following result.

The ring lemma. *For each* $n \geqslant 3$*, let* c_n *denote the infimum of the radii among* n *discs surrounding the unit disc. Then* $c_n > 0$.

The sharp ring lemma states that $c_n = (F_{n-1}^2 + F_{n-2}^2 - 1)^{-1}$, where F_k is the k^{th} Fibonacci number, and that this lower limit is attained in essentially unique configurations, see section 3 for the precise statement.

We will use the Descartes circle theorem in the following special case, where the balls are required to have pairwise disjoint interiors. Just as half-planes are viewed as discs of infinite radius, half-spaces are considered to be balls of infinite radius in $\mathbf{R}^3 \cup \{\infty\}$, having disjoint interiors if they only intersect at infinity, in which case they are also tangent.

The Descartes circle theorem. *Suppose* $N + 2$ *pairwise tangent balls in* $\mathbf{R}^N$, $N \geqslant 2$*, have pairwise disjoint interiors and inverse radii* $b_1, \dots, b_{N+2}$. *Then* $N \sum_i b_i^2 = (\sum_i b_i)^2$.

See [6] or [15] for an elementary proof[1] and [9–11, 13, 14] for considerable generalizations. There is a converse statement [6, 14] – that is, given radii that satisfy the equation we can construct the corresponding balls – but we will only need this easy fact [15] in dimensions two and three: given $N + 1$ pairwise tangent balls in $\mathbf{R}^N$ with pairwise disjoint interiors, there are precisely two spheres that are tangent, possibly internally, to each of

1 The original theorem has been rediscovered many times, but Soddy and Gosset brought it to modern attention through the poems [19] and [8].

the given balls without intersecting the interior or coinciding with the boundary of any of the balls.

By introducing an extra tangency and using the Descartes circle theorem, one may show that the ring lemma constant satisfies $c_{n+1} < c_n$, but we will not need this in order to prove the sharp ring lemma.

2.1 Apollonian configurations

The sharp ring lemma, which we will give a proof of below, gives that the sharp value of the ring lemma constant is attained in essentially unique configurations. Those configurations, which we will call *Apollonian configurations*, are defined as follows.

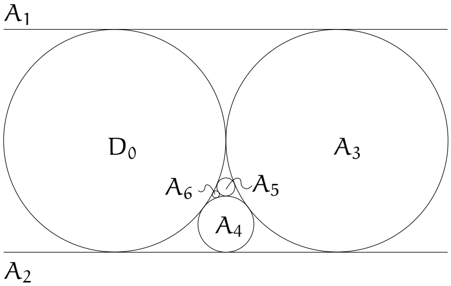
We recursively construct a configuration of discs $A_1, A_2, \dots$ such that for each $n \geqslant 3$, the discs $A_1, \dots, A_n$ surround the unit disc D_0 (figure 2). First, we let A_1, A_2, A_3 be discs with pairwise disjoint interiors that are externally tangent to the unit disc and where A_1, A_2 have infinite radii, hence A_3 has unit radius. Given discs $A_1, \dots, A_n$, $n \geqslant 3$, we let A_{n+1} be externally tangent to D_0, A_{n-1} and A_n.

Definition. *A configuration of* $n \geqslant 3$ *discs surrounding the unit disc is said to be an* Apollonian configuration *if it is equal to* $A_1, \dots, A_n$ *– as defined above – up to reflection and rotation.*

We see that the Apollonian configurations of three discs are unique up to rotation, and that Apollonian configurations of order four and higher are uniquely determined by the position of the third and fourth largest discs, corresponding to A_3 and A_4, respectively.

2.2 Method of proof

We will prove the sharp ring lemma using a compactness argument in *strip configurations*, defined below. A more 'dynamic' approach, which we will not use, is as follows. Noting that given $k \in \{3, \dots, n\}$, the radius of A_k in an Apollonian configuration of n discs cannot be increased if $A_1, \dots,$ A_{k-1} are fixed, it would be natural to try to successively modify the radii in an arbitrary configuration in order to reduce it to a configuration



An Apollonian configuration of six discs $A_1, \ldots, A_6$ surrounding the unit disc D_0. The radius of A_6 is the smallest possible for six discs surrounding the unit disc. FIGURE 2

having this property. That is, given discs $D_1, \ldots, D_n$ surrounding D_0 we want to find a permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$ such that in the k^{th} step the radius of D_{i_k} is increased – forcing $D_{i_{k+1}}, \ldots, D_{i_n}$ to change in order to maintain the surround property – until D_{i_k} touches the already increased $D_{i_1}, \ldots, D_{i_{k-1}}$. For this smaller set of configurations, the sharp ring lemma can be verified using a monotonicity property of the Descartes circle theorem.

The difficulty with this approach is determining in what order to modify the discs, while preventing the smallest radius from increasing after all the adjustments. Given the properties of the Apollonian configurations, a natural candidate is to index by size, so that $r_{i_k} \geqslant r_{i_{k+1}}$. However, one must be more careful as the following simple counterexample shows (figure 3). Let $D_1, \ldots, D_6$ be discs surrounding the unit disc D_0, numbered clockwise by their tangency with D_0, having radii $r_1 = r_2 = +\infty$, $r_3 = 1$, $r_5 = \frac{1}{5}$ and where D_3 and D_5 share an extra tangency; we see that this gives $\min(r_4, r_6) < \frac{1}{12}$. The largest disc that can be increased without decreasing the radius of a larger disc is D_5, and increasing the radius of this disc as much as D_1, D_2 and D_3 allow, we see that the new radii satisfy $\min(r_4, r_6) = \frac{1}{12}$.

This method of proof was pursued in more detail by Hansen [12] and Stephenson [21]. Here we will reduce the problem to a simpler class of

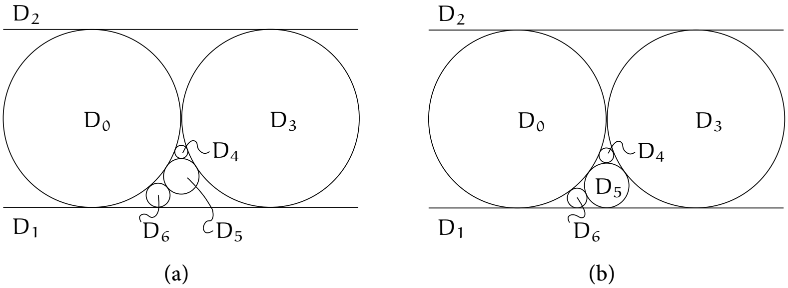


FIGURE 3 *The radii of the discs* D_1, D_2, D_3 *in (a) cannot be increased without decreasing a larger disc. Increasing the largest remaining disc* D_5 *until it is stopped by* D_1, D_2, D_3 *gives the configuration in (b), where* $\min(r_1, \dots, r_6)$ *has increased.*

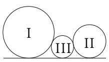
strip configurations and then closely follow Aharonov's and Stephenson's method of proof [3].

Definition. *Sequentially tangent discs* $D_1, \dots, D_n$, $n \geqslant 2$, *are said to lie in a* strip configuration *if all discs lie between two parallel straight lines* L_1 *and* L_2, *have pairwise disjoint interiors and are tangent to* L_1, *and moreover* D_1, D_n *are tangent to* L_2.

A tangency between D_i *and* D_{i+1}, $i = 1, \dots, n-1$, *is said to be* ordinary, *and other tangencies, except those with* L_1, *are said to be* extra *tangencies.*

3 The sharp ring lemma in two dimensions

Aharonov and Stephenson proved the sharp ring lemma using two key lemmas [3, lemmas 13 and 14], which in turn are proved by transforming to a strip configuration. We prefer to transfer the original problem to a single strip configuration – which Aharonov and Stephenson used to describe the general Apollonian packing – and then proceed using their method of proof.



Consider three sequentially tangent discs D_I, D_{III}, D_{II} – where the numbering is chosen to conform with the lemma below – that have pairwise disjoint interiors and are externally tangent to a half-plane in the

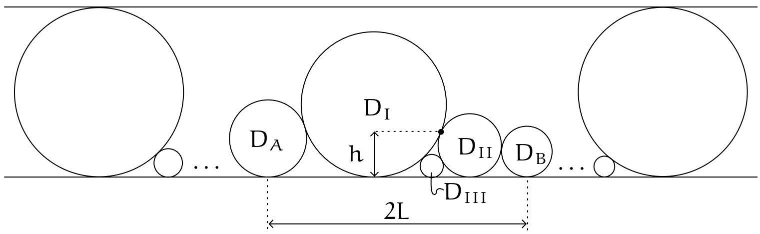


FIGURE 4

The configuration of the lemma, where the discs D_A, D_I, D_{II}, D_B *are sequentially tangent, and* D_{III} *is wedged between and tangent to* D_I *and* D_{II}. *If only* D_I, D_{II} *and* D_{III} *vary, then the height* h *or radii* r_I, r_{II}, r_{III} *are minimal only when* D_I *and* D_{II} *have an additional extra tangency.*

given order. Since the center-to-center distance, as measured parallel to the boundary of the half-plane, between two of the tangent discs with radii r, r' is $2\sqrt{r\,r'}$ and since the discs D_I, D_{II} have disjoint interiors, we see that $\sqrt{r_I r_{III}} + \sqrt{r_{II} r_{III}} \geqslant \sqrt{r_I r_{II}}$. Hence, the smallest possible radius of D_{III} – given r_I and r_{II} – is determined by $\sqrt{r_{III}} = \frac{\sqrt{r_I r_{II}}}{\sqrt{r_I}+\sqrt{r_{II}}}$, so that D_I and D_{II} are tangent, with D_{III} wedged in between.[2]

Lemma. *Suppose* D_I *and* D_{II} *are non-outermost discs that share an ordinary tangency in a strip configuration. Construct a new strip configuration by adding a disc* D_{III} *that shares an ordinary tangency with* D_I *and* D_{II}. *Let* h *be the distance from the common line in the strip configuration to the tangency between* D_I *and* D_{II}. *Suppose that* h *or one the of radii* r_I, r_{II}, r_{III} *is minimal among the strip configurations obtained by varying* D_I, D_{II} *and* D_{III}, *while maintaining the extra tangency between* D_I *and* D_{II}. *Then* D_I *or* D_{II} *has an additional extra tangency.*

Proof. Let D_A and D_B be discs such that $D_A, D_I, D_{III}, D_{II}, D_B$ are sequentially tangent discs with ordinary tangencies, and let 2L be the center-to-center distance between D_A and D_B, as measured parallel to the common line in the strip configuration (figure 4).

2 This resulting configuration is a Japanese *sangaku* (算額) problem [7, problem 1.1.1], [17], engraved in the year 1824 on a tablet in the Gunma prefecture.

We have that $L = \sqrt{r_A r_I} + \sqrt{r_I r_{II}} + \sqrt{r_{II} r_B}$, so that

$$\sqrt{r_I} = \frac{L - \sqrt{r_{II} r_B}}{\sqrt{r_A} + \sqrt{r_{II}}}.$$

It is now clear that all configurations can be parameterized by $r_{II} \in [a, b]$, $0 < a \leqslant b$ constants, where $r_{II} \in \{a, b\}$ gives D_I or D_{II} an additional extra tangency. Furthermore, by differentiating we see that r_I is strictly decreasing with r_{II}, so the case when r_I or r_{II} is minimal follows.

By above for r_{III} and by geometry for h, we get

$$\sqrt{r_{III}} = \Big(\frac{1}{\sqrt{r_I}} + \frac{1}{\sqrt{r_{II}}}\Big)^{-1}, \qquad h = 2\Big(\frac{1}{r_I} + \frac{1}{r_{II}}\Big)^{-1},$$

and the claim is that r_{III} and h have no minima for $r_{II} \in \,]a, b[$. To show this we let $t = \frac{1}{\sqrt{r_{II}}}$ and note that it is sufficient that

$$\begin{aligned}
&\frac{d^2}{dt^2}\Big(\frac{1}{\sqrt{r_I(t)}} + t\Big) = \frac{2L(\sqrt{r_A r_B} + L)}{(Lt - \sqrt{r_B})^3},\\
&\frac{1}{2}\frac{d^2}{dt^2}\Big(\frac{1}{r_I(t)} + t^2\Big) = \Big(\frac{d}{dt}\Big(\frac{1}{\sqrt{r_I(t)}}\Big)\Big)^2 + \frac{1}{\sqrt{r_I(t)}}\frac{d^2}{dt^2}\Big(\frac{1}{\sqrt{r_I(t)}}\Big) + 1,
\end{aligned}$$

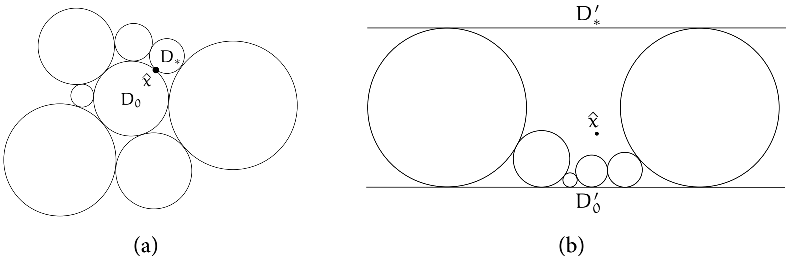
are strictly positive for $t \in \big]\frac{1}{\sqrt{b}}, \frac{1}{\sqrt{a}}\big[$. Since $Lt - \sqrt{r_B} > 0$, the claim follows. ∎

Proposition. *If a configuration of n discs surrounds the unit disc, then no disc has a radius below $c_n = (F_{n-1}^2 + F_{n-2}^2 - 1)^{-1}$ where F_k is the k^{th} Fibonacci number. Furthermore, equality is achieved if and only if the configuration is an Apollonian configuration.*

Proof. Consider $n \geqslant 3$ discs surrounding the unit disc D_0. Take $\hat{x} \neq 0$. The reflection[3] Φ in $S(\hat{x}, 1)$ – that is, the circle centered at $\hat{x}$ with unit radius – is a Möbius transformation[4] satisfying $\Phi^{-1} = \Phi$ and $\Phi\big(S\big((|\hat{x}| \pm$

3 In $\mathbf{R}^N \cup \{\infty\}$, the reflection Φ in the $(N-1)$-dimensional sphere $S(x_0, r)$ is defined as $\Phi(x) = x_0 + \frac{r^2}{|x - x_0|^2}(x - x_0)$ if $x \neq x_0, \infty$ and $\Phi(x_0) = \infty$, $\Phi(\infty) = x_0$.

4 A transformation in the full Möbius group in the sense of Ahlfors [4], it is not sense-preserving. In all dimensions, a generalized sphere – that is, a sphere or a plane – is mapped to a generalized sphere.



(a) *Seven discs surrounding the unit disc* D_0. *The intersection between* D_0 *and a disc* D_* *is marked* $\hat{x}$. (b) *The strip configuration given by reflecting the discs in (a) in the unit circle centered at* $\hat{x}$. FIGURE 5

$r)\frac{\hat{x}}{|\hat{x}|}, r)\big) = P\big(|\hat{x}| \pm \frac{1}{2r}\big)$, where $P(\alpha) = \big\{x \in \mathbf{R}^2;\ x \cdot \frac{\hat{x}}{|\hat{x}|} = \alpha\big\} \cup \{\infty\}$ is a straight line.

Now let D_* be one of the surrounding discs and $\hat{x}$ its intersection with D_0. Transforming the configuration using Φ gives a strip configuration (figure 5), and the distance between $\hat{x}$ and the half-planes $D'_* = \Phi(D_*)$, $D'_0 = \Phi(D_0)$ is $h_* = \frac{1}{2r_*}$, $h_0 = \frac{1}{2r_0}$, respectively. Hence, $\frac{r_*}{r_0} = \frac{h_0}{h_*}$, so that $\frac{r_*}{r_0}$ is minimal precisely when $\frac{h_0}{h_*}$ is.

Step 1: Extremal configurations are uniquely attained for minimal h_0. Without changing $\frac{r_*}{r_0}$, we may rescale so that $h_0 + h_* = 2$.

Since D'_0 and $\hat{x}$ are separated by discs, h_0 must be larger than or equal to the shortest distance between D'_0 and the intersection between two of the discs in the strip configuration, where equality is possible only if $\hat{x}$ is the intersection between two discs. However, placing $\hat{x}$ at the intersection between any two discs in the strip configuration gives two half-planes under Φ as well as preserves the surround property.

Hence, we may assume that $\hat{x}$ is the intersection between two discs in the strip configuration, and we should find the smallest value of h_0 in strip configurations with $n-1$ discs between the lines. Given the smallest value of h_0, we have that $c_n = \frac{h_0}{2-h_0}$.

Since, by the ring lemma, $c_n > 0$ we immediately get a strictly positive lower bound on h_0 as well as the radii in the strip configurations.

We can also see this directly by considering sequentially tangent discs $\widetilde{D}_1, \ldots, \widetilde{D}_{n-1}$ in a strip configuration, where the outermost discs $\widetilde{D}_1$ and $\widetilde{D}_{n-1}$ have unit radius, and observe that $\sqrt{\widetilde{r}_1 \widetilde{r}_i} \leqslant \sqrt{\widetilde{r}_1 \widetilde{r}_2} + \cdots + \sqrt{\widetilde{r}_{i-1} \widetilde{r}_i}$. From the fact that there exists a constant $C_1 > 0$ with $\widetilde{r}_1 \geqslant C_1$ – in fact, $\widetilde{r}_1 = 1$ – we recursively get that $\widetilde{r}_i \to 0^+$ as $\widetilde{r}_2 \to 0^+$. However, since $\widetilde{r}_{n-1} = 1$, we see that there must be a lower bound $C_2 > 0$ on the radius $\widetilde{r}_2$, and, by repeating the argument, we get $\widetilde{r}_i \geqslant C_i > 0$.

In order to simplify the notation, we drop the prime and allow ourselves below to ambiguously also let D_i denote $D_i' = \Phi(D_i)$, and so on.

Step 2: The minimal radius in strip configurations is uniquely attained in Apollonian configurations.

Before showing that Apollonian configurations uniquely determine the smallest height h_0, we show that they uniquely determine the smallest radius among all the strip configurations.

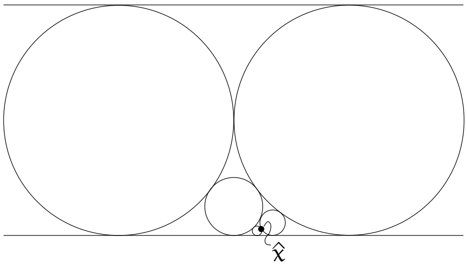
Let s_n be the infimum of all radii in strip configurations with $n - 1$ discs between the two lines and let a_n be the corresponding infimum over radii in Apollonian configurations of n discs under Φ, in both cases after rescaling so that the straight lines are at a distance 2 from each other. The latter is attained by letting $\hat{x}$ in the original configuration be the intersection between D_0 and the smallest disc, and we see that the configuration of $n - 1 \geqslant 3$ discs between the lines contains the configuration of $n - 2$ discs[5] (figure 6).

The cases $n = 3, 4$ are trivial. Suppose the claim is true for $n = 3, \ldots, k$, and consider the case $n = k + 1$, having $k \geqslant 4$ discs between two lines. By compactness, an extremal configuration is attained, giving the smallest possible radius. Let $\widetilde{D}$ be the smallest disc in the strip configuration and call its neighbors D_I, D_{II} with radii $s_{k+1} = \widetilde{r} \leqslant r_I \leqslant r_{II}$.

We show that $r_I \geqslant s_k = a_k$ and $r_{II} \geqslant s_{k-1} = a_{k-1}$, so that, by the considerations before the lemma, $\widetilde{r} \geqslant a_{k+1}$ where the minimality of $\widetilde{r}$ guarantees equality and, hence, also guarantees that $r_I = a_k$, $r_{II} = a_{k-1}$.

---

5 Furthermore, the configuration is a subset of the Ford circles. In particular, the coordinates of the tangencies with the common line can be arranged in a Farey sequence [5].



An Apollonian configuration with six discs surrounding the unit disc (figure 2) after reflection in a disc with unit radius centered at the intersection between the unit disc and the smallest disc. FIGURE 6

Clearly, only the larger disc D_{II} can be an outermost disc, and we first assume it is not. The minimality of $\tilde{r}$ gives that $\tilde{D}$ does not lie in a proper subconfiguration, hence it has no extra tangencies, and, using the lemma twice, we see that D_I and D_{II} have extra tangencies. In particular, D_I is in a k-configuration, so that $r_I \geqslant s_k = a_k$.

Consider first the case where D_I and D_{II} are not tangent. Since $\tilde{D}$ does not lie in a proper subconfiguration, the extra tangency of the larger disc D_{II} must encompass both D_I and $\tilde{D}$. Hence, D_{II} is in a p-configuration, where $p \leqslant k-1$, so in particular $r_{II} \geqslant s_{k-1} = a_{k-1}$, which is the desired inequality.

Secondly, consider the case when D_I and D_{II} are tangent, with $\tilde{D}$ wedged in between, so that D_I and D_{II} share an extra tangency. By the lemma, D_I or D_{II} has an additional extra tangency. If D_{II} has the additional extra tangency, it lies in a p-configuration, with $p \leqslant k-1$, so that $r_{II} \geqslant s_{k-1} = a_{k-1}$, and, since $r_{II} \geqslant r_I$, we get the same inequality if instead D_I has an additional extra tangency.

The easy case where D_{II} is an outermost disc remains. Obviously $r_{II} = 1 \geqslant s_{k-1} = a_{k-1}$, and using the lemma once we also see that D_I has an extra tangency, so that $r_I \geqslant s_k = a_k$.

Consequently, $s_{k+1} = \tilde{r} = a_{k+1}$ and also $r_I = a_k$, $r_{II} = a_{k-1}$, in particular D_I and D_{II} must be tangent. Hence, D_I is a disc in a k-configuration with $r_I = a_k = s_k$, and, by the uniqueness for k-configurations, it must be an Apollonian configuration, giving uniqueness in the case $k+1$ as well.

The claim follows by induction.

Step 3: The minimal height h_0 is uniquely attained in Apollonian configurations.

By compactness, an extremal configuration is attained, giving the minimal height h_0 for $n-1$ discs between two lines. As noted above, $\hat{x}$ is the intersection between two discs D_I, D_{II}.

Recall that $h_0 = 2\left(\frac{1}{r_I} + \frac{1}{r_{II}}\right)^{-1}$ and by step 2, we have that $r_I, r_{II} \geqslant s_n = a_n$. Furthermore, since h_0 is minimal, the lemma requires D_I or D_{II} to have an extra tangency or to be an outermost disc; in either case we get $\max(r_I, r_{II}) \geqslant s_{n-1} = a_{n-1}$. By the minimality of h_0 we get $h_0 = 2\left(\frac{1}{a_n} + \frac{1}{a_{n-1}}\right)^{-1}$, so in particular $\min(r_I, r_{II}) = s_n = a_n$.

Hence, Apollonian configurations give the smallest height h_0, and since an extremal radius $s_n = a_n$ is attained in such configurations we get uniqueness by step 2.

Extremal radii

Letting $x_n = \frac{1}{\sqrt{a_{n+1}}}$, we showed above that $x_n = x_{n-1} + x_{n-2}$, $n \geqslant 3$, where $x_1 = x_2 = 1$; hence $x_n = F_n$, the n^{th} Fibonacci number, and $a_n = \frac{1}{F_{n-1}^2}$. The height for $n-1$ discs, corresponding to n discs surrounding the unit disc, is then given by $h_0 = 2\left(\frac{1}{a_n} + \frac{1}{a_{n-1}}\right)^{-1} = 2\left(F_{n-1}^2 + F_{n-2}^2\right)^{-1}$, and it immediately follows that the extremal radius, given by $\frac{h_0}{2-h_0}$, is $c_n = (F_{n-1}^2 + F_{n-2}^2 - 1)^{-1}$. ∎

The ring lemma in three dimensions 4

Introduction 4.1

A generalization of the ring lemma to three dimensions should determine the infimum of the radii for a set of n balls with pairwise disjoint interiors surrounding the unit ball. It is not clear, however, in what sense a finite set of balls should 'surround' the unit ball. For instance, given any finite number of balls tangent to the unit ball, one can always find a smooth curve starting on the unit sphere that escapes to infinity without passing through any of the balls.

Surrounding and hiding packings 4.2

All balls are closed, and we remind that we view half-spaces as balls of infinite radius in $\mathbf{R}^3 \cup \{\infty\}$, having disjoint interiors if they only intersect at infinity, in which case they are also tangent.

For each set of balls $B_1, \ldots, B_n$ we define a combinatorial complex $K(B_1, \ldots, B_n)$ of vertices, edges and faces as follows: start with n vertices $v_1, \ldots, v_n$, add edges between v_i, v_j, $i \neq j$ if B_i, B_j are tangent, $i, j = 1, \ldots, n$, and faces given by every set of three edges corresponding to three pairwise tangent balls.

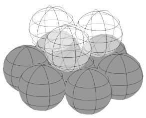
Definition. *Suppose* $B_1, \ldots, B_n$, $n \geqslant 4$, *are balls with pairwise disjoint interiors that are externally tangent to the unit ball* B_0. *We say that the balls* $B_1, \ldots, B_n$ surround B_0 *if* $K(B_1, \ldots, B_n)$, *as defined above, has a subcomplex, containing all the vertices, that triangulates the unit sphere, and if the triangulation can be embedded without overlap in the unit sphere in such a way that:*

- *(i)* v_i *is the point of tangency between* B_0 *and* B_i,
- *(ii) each edge is a shortest path on the unit sphere between its endpoints, and*
- *(iii) the faces are spherical triangles* T_{ijk} *satisfying* $\operatorname{area}(T_{ijk}) \leqslant 2\pi$.

If, additionally, every straight half-line starting at the origin intersects $B_1 \cup \cdots \cup B_n$, *we say that* $B_1, \ldots, B_n$ hide B_0.

Condition (iii) gives the convexity property that between any two points in T_{ijk} there is a shortest path on the unit sphere that is contained in T_{ijk}, and consequently that any shortest path strictly shorter than π between points in T_{ijk} is contained in T_{ijk}.

Remark. In two dimensions, unless two of the discs are half-planes, no curve starting on the unit disc can escape to infinity without passing through one of the surrounding discs. Hence, a hide property is less interesting in two dimensions. □



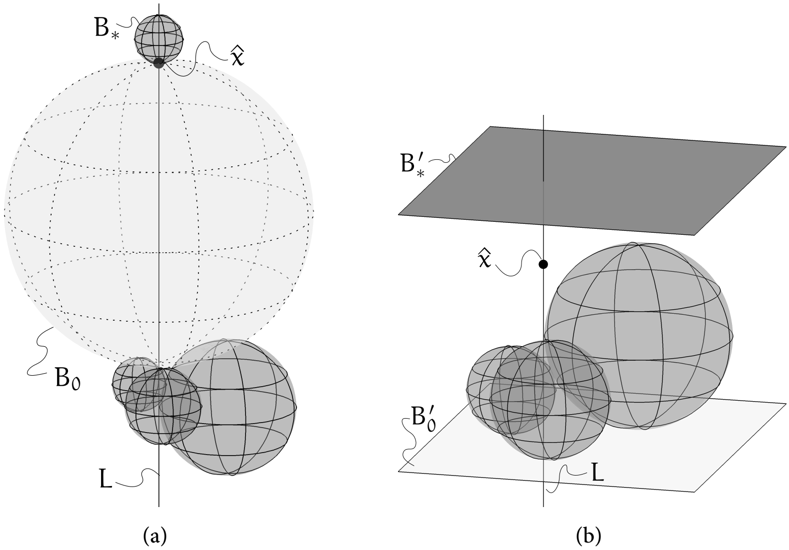
Remark. Portions of lattice structures such as hexagonal close packing or face-centered cubic – lattices which are of great practical importance – do not satisfy this definition. In fact, both packings can be realized as stacked layers of identical balls forming a hexagonal pattern, making each ball tangent to twelve others. If we select a ball, corresponding to the unit ball and consider the balls tangent to it, six of the balls form a closed chain around the original ball and the other six are divided between two layers, each containing three balls. The above definition only considers groups of three pairwise tangent balls, and we see that we get six 'holes' formed between groups of four cyclically tangent balls. We may, however, add a ball in each 'hole' in order to satisfy the definition. □

4.3 Lemma for the lower bound

Take $\hat{x} \neq 0$. Denote by Φ the reflection in $S(\hat{x}, 1)$ – that is, the sphere centered at $\hat{x}$ with unit radius – which is a Möbius transformation satisfying $\Phi^{-1} = \Phi$ and $\Phi\big(S\big((|\hat{x}| \pm r)\frac{\hat{x}}{|\hat{x}|}, r\big)\big) = P\big(|\hat{x}| \pm \frac{1}{2r}\big)$, where $P(\alpha) = \big\{x \in \mathbf{R}^3;\ x \cdot \frac{\hat{x}}{|\hat{x}|} = \alpha\big\} \cup \{\infty\}$ is a plane. Furthermore, we let r, r', r_i and so on denote the radii of correspondingly marked balls B, B', B_i.

Lemma. *Suppose* $B_1, \ldots, B_n$ *surround* B_0. *Let* B_* *be one of the surrounding balls and* $\hat{x}$ *its intersection with* B_0. *Denote by* Φ *the reflection in* $S(\hat{x}, 1)$. *If the spherical triangle* T_{pqr} *contains the antipode of* $\hat{x}$, *then* $\frac{1}{r_i'} + \frac{1}{r_j'} \geqslant 4$ *for all* $i, j \in \{p, q, r\}$ *with* $r_i', r_j' < +\infty$ *and* $i \neq j$, *where* r_i' *is the radius of* $\Phi(B_i)$.

Proof. Transforming the configuration using Φ gives a kind of three-dimensional 'strip configuration' where the balls tangent to B_* form a



Portion of a packing when transformed by Φ, the reflection in $S(\hat{x}, 1)$. *In (b),* $\hat{x}$ *and* B'_* *have been shifted towards* B'_0 *for clarity. The line* L *– which in (a) passes through* $\hat{x}$ *and its antipode – is invariant as a set under the transformation.* FIGURE 7

closed chain of balls tangent to the half-spaces $B'_* = \Phi(B_*)$ and $B'_0 = \Phi(B_0)$.

For purposes of orientation, we consider B'_0 to lie below B'_* – a point x lies below another point x', and x' above x, if x is closer to the boundary of B'_0 than x' is – and in the original configuration we let $\hat{x}$ be the north pole of B_0 (figure 7).

To simplify the notation, we renumber so that $(p, q, r) = (1, 2, 3)$. If B_i, B_j, $i \neq j$, are tangent at $\widetilde{t}_{ij}$ we let $t_{ij} = \frac{\widetilde{t}_{ij}}{|\widetilde{t}_{ij}|}$, which is the point on B_0 closest to the tangency between B_i and B_j.

Suppose first that $B_* \neq B_i$, $i = 1, 2, 3$. Referring to the 'strip configuration' – figure 7 (b) – consider the three planes $P'_{12}, P'_{13}, P'_{23}$ that are parallel to the boundary of B'_0 and where P'_{ij} passes through $B'_i \cap B'_j$. Aiming to show that $\hat{x} = \Phi(\infty)$ lies on or above at least one of the planes, we assume

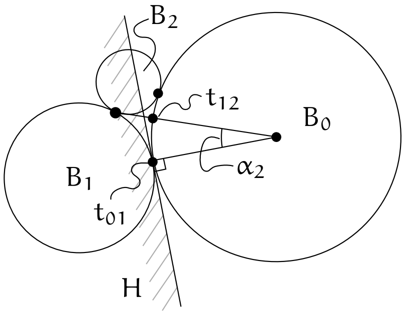


FIGURE 8 *The plane through the tangencies between* B_0, B_1 *and* B_2 *determines a great circle on* B_0. *Noticing that the half-space* H, *which encloses* B_1 *and is tangent to the same ball at* $B_0 \cap B_1$, *contains* $B_1 \cap B_2$, *we see that the angle* α_2 *is acute.*

this is not the case. Then the planes P'_{ij} are mapped by Φ to spheres that contain $B_i \cap B_j$, are tangent to B_* at $\hat{x}$ and enclose $B_* \backslash \{\hat{x}\}$, and it follows that t_{12}, t_{13} and t_{23} lie in the interior of the northern hemisphere. We obtain the required contradiction by showing that the spherical triangle T_{123} cannot enclose the south pole $-\hat{x}$, contradicting the construction.

Since the points of tangencies between the balls B_1, B_2, B_3 lie on the northern hemisphere of B_0, we see that at most one of them can be tangent to the southern hemisphere of B_0. If all balls B_1, B_2, B_3 are tangent to the northern hemisphere, we obtain the contradiction immediately, hence we may assume without loss of generality that B_1 is tangent to the interior of the southern hemisphere of B_0. Consider first the plane through the tangencies between B_0, B_1 and B_2 (figure 8). The plane passes through the center of B_0 and determines a great circle on B_0. We see that the length α_2 of the shortest path on the unit sphere between t_{01} and t_{12}, satisfies $\alpha_2 < \frac{\pi}{2}$, and the same holds for the correspondingly defined length α_3.

The points t_{12} and t_{13} lie on the northern hemisphere, and we claim that since the lengths α_2, α_3 satisfy $\max(\alpha_2, \alpha_3) < \frac{\pi}{2}$, the spherical triangle T_{123} cannot enclose $-\hat{x}$. First, note that it is sufficient to consider the spherical triangle contained in T_{123} that has vertices t_{12}, t_{13}, t_{01}, where we may assume that t_{12} and t_{13} lie on the equator. Suppose now that the south pole is contained in this smaller triangle. Since the triangle does not contain a great circle, a shortest path on the unit sphere from t_{01} to the south pole – a path which lies within the triangle – may be extended within the triangle along its great circle until it intersects the side between t_{12} and t_{13} at a point x. The intersection is at a right angle, and, without loss of generality, we may assume that the shortest path between x and t_{13} is shorter than $\frac{\pi}{2}$ and also exclude the trivial case $x = t_{13}$. Letting a, b, c be the lengths of the sides opposite to x, t_{13} and t_{01}, respectively, the spherical Pythagorean theorem yields $\cos a = \cos b \cos c$. We have that $a < \frac{\pi}{2}$, $b \leqslant \pi$, $c \leqslant \frac{\pi}{2}$, hence $\cos a > 0$ and $\cos c \geqslant 0$, so that $b < \frac{\pi}{2}$, which is impossible since the corresponding shortest path passes through both the equator and the south pole.

Now consider the case where $B_* = B_i$ for some i; without loss of generality we may assume that $B_* = B_3$. Using the previous construction of P'_{12}, we again assume that $\hat{x}$ lies strictly below P'_{12}, and, as above, it follows that t_{12} lies on the northern hemisphere of B_0. Furthermore, we see that the projection $x \mapsto \frac{x}{|x|}$ on B_0 of the entire ball $B_3 = B_*$ lies on the northern hemisphere of B_0, hence so does t_{13}, t_{23}. Again this means that the spherical triangle T_{123} determined by B_1, B_2, B_3 cannot enclose $-\hat{x}$.

Hence $\hat{x}$ – which is a distance $\frac{1}{2}$ from B'_0 – must lie on or above at least one plane P'_{ij}. Since the distance between P'_{ij} and B'_0 is $2\left(\frac{1}{r'_i} + \frac{1}{r'_j}\right)^{-1}$ just like in the two-dimensional case, we get the desired inequality. ■

Remark. We see that the proof gives a useful geometrical interpretation of the lemma: In the 'strip configuration' – figure 7 (b) – the point $\hat{x}$ is never closer to B'_0 than is the point of tangency between two balls, balls which in the original setting – figure 7 (a) – determine two vertices of a spherical triangle containing the antipode of $\hat{x}$. □

The Apollonian packings in three dimensions 4.4

The Apollonian packings – the extremal packings in the two-dimensional case – have a natural generalization to three dimensions. We start with four pairwise tangent balls $A_1, \ldots, A_4$ with pairwise disjoint interiors that are externally tangent to the unit ball A_0 and have radii $a_1 = a_2 = +\infty$ and $a_3 = a_4 = 1$, respectively. Then, in the smallest pockets, we add balls having pairwise disjoint interiors with the previous, starting with A_5 tangent to A_0, A_2, A_3 and A_4, or – equivalently in terms of radii – A_1 instead of A_2. The ball A_6 is tangent to A_0, A_3, A_4 and A_5; A_7 is tangent to A_0, A_4, A_5 and A_6, or equivalently A_3 instead of A_4. Having determined $A_1, \dots, A_n$, $n \geqslant 7$, the next ball A_{n+1} is tangent to A_0, A_{n-2}, A_{n-1} and A_n. The radius of the ball A_n is denoted a_n.

We see that an Apollonian packing $A_1, \dots, A_n$ of $n \geqslant 4$ balls defined in this way surrounds, but does not hide, A_0 in the sense above. However, the only unhidden portion of A_0 is along the great circle parallel to the two half-spaces, and, without breaking the surround property, we can add four balls with unit radius to hide the portion not already hidden by A_1 and A_2.

4.5 The ring lemma in three dimensions

Proposition. *For each* $n \geqslant 4$*, let* $c_n(\mathbf{R}^3)$ *denote the infimum over the radii among all balls* $B_1, \dots, B_n$ *surrounding the unit ball. The constant* $c_n(\mathbf{R}^3)$ *is bounded from below by the two-dimensional ring lemma constant for* $n-1$ *discs and from above by the smallest radius in an Apollonian packing of* n *balls, that is* $c_{n-1}(\mathbf{R}^2) \leqslant c_n(\mathbf{R}^3) \leqslant a_n$.

Proof. The upper bound follows immediately since an Apollonian packing of $n \geqslant 4$ balls surrounds B_0. Now take $n \geqslant 4$ and suppose that $B_1, \dots, B_n$ surround B_0.

Let $\hat{x}$ be the point of tangency between B_0 and one of the balls having the minimal radius among $B_1, \dots, B_n$. The distance between $\hat{x}$ and the half-spaces $B'_* = \Phi(B_*)$, $B'_0 = \Phi(B_0)$ is $h_* = \frac{1}{2r_*}$, $h_0 = \frac{1}{2r_0}$, respectively. Hence, $\frac{r_*}{r_0} = \frac{h_0}{h_*}$, so that $\frac{r_*}{r_0}$ is minimal precisely when $\frac{h_0}{h_*}$ is, and, without changing $\frac{r_*}{r_0}$, we may rescale so that $h_0 + h_* = 2$. Thus, for the lower bound, we should show that the smallest value of h_0 satisfies $\frac{h_0}{2-h_0} \geqslant c_{n-1}(\mathbf{R}^2)$.

By the geometrical interpretation of the lemma, we have that h_0 is greater than the distance between B'_0 and $B'_i \cap B'_j$, for some tangent balls B'_i, B'_j. We may take a non-self-intersecting edge path $v_{i_1}, \dots, v_i, v_j, \dots,$

ν_{i_k}, where $r'_{i_1} = r'_{i_k} = 1$. Hence we get a chain of k tangent balls B'_{i_1}, $\dots, B'_{i_k}$, and we may construct a straight chain where all tangencies with B'_0 lie on a straight line as follows. Start with B'_{i_1}, B'_{i_2} and move B'_{i_3} while keeping the tangency with B'_{i_2} so that its tangency with B'_0 lies on the line determined by $B'_{i_1} \cap B'_0$ and $B'_{i_2} \cap B'_0$. Note that the interiors of B'_{i_1} and B'_{i_3} are still disjoint, since the center-to-center distance increases with the center-to-center distance measured parallel to the boundary of B'_0. Recursively, this yields a straight chain.

Considering the plane through all the points of tangencies between $B'_0, B'_{i_1}, \dots, B'_{i_k}$, we get a two-dimensional strip configuration, so, by the two-dimensional ring lemma, the distance d between B'_0 and $B'_i \cap B'_j$ – which is less than h_0 – must satisfy $\frac{d}{2-d} \geqslant c_{k+1}(\mathbf{R}^2)$, and we see that $\frac{h_0}{2-h_0} \geqslant c_{k+1}(\mathbf{R}^2)$. Since there are $n-1$ balls between the half-spaces, of which we may discard at least one of the three or more balls that have unit radius, we see that $k \leqslant n-2$, so that $\frac{h_0}{2-h_0} \geqslant c_{n-1}(\mathbf{R}^2)$, giving the desired lower bound on $c_n(\mathbf{R}^3)$. ∎

Again, we may use the Descartes circle theorem to see that $c_{n+1}(\mathbf{R}^3) < c_n(\mathbf{R}^3)$. Letting $h_n = h_n(\mathbf{R}^3)$ be the infimum over the radii among all balls $B_1, \dots, B_n$ hiding the unit ball, we have that $c_n(\mathbf{R}^3) \leqslant h_n \leqslant a_{n-4}$, where the last inequality holds for $n \geqslant 8$.

Numerical values 4.6

To calculate the radii a_i in the Apollonian packing, it is easiest to use the Descartes circle theorem, which gives that $3(b_0^2 + b_{n-2}^2 + b_{n-1}^2 + b_n^2 + b_{n+1}^2) = (b_0 + b_{n-2} + b_{n-1} + b_n + b_{n+1})^2$ where $b_i = a_i^{-1}$, and, as in [1], we subtract the equations for n and $n+1$ to get – using the fact that $b_{n+2} \neq b_{n-2}$ – the linear recurrence equation $b_{n+2} = b_0 - b_{n-2} + b_{n-1} + b_n + b_{n+1}$, which is related to [18, A052527]. We have that $b_0 = 1$, $b_1 = b_2 = 0$, $b_3 = b_4 = 1$, hence $b_n \in \mathbf{N}$. Solving the recurrence equation gives that

$$b_n = C_1(\alpha+\beta)^n + \frac{C_2}{(\alpha+\beta)^n} + (-1)^n(C_3 \cos n\gamma + C_4 \sin n\gamma) - 1,$$

TABLE 1 *The sharp constant* $c_n(\mathbf{R}^3)$ *in the three-dimensional ring lemma satisfies* $c_{n-1}(\mathbf{R}^2) \leqslant c_n(\mathbf{R}^3) \leqslant a_n$.

n	$1/a_n$	$1/c_{n-1}(\mathbf{R}^2)$
4	1	1
5	3	4
6	6	12
7	10	33
8	19	88
9	33	232
10	57	609
11	100	1596

where

$$\alpha = \frac{\sqrt{13}+1}{4}, \qquad \beta = \sqrt{\frac{\sqrt{13}-1}{8}}, \qquad \gamma = \arccos 2\beta,^{-1}$$

$$C_{1,2} = \frac{1}{2} + \frac{1}{\sqrt{13}} \mp \frac{1}{2}\sqrt{\frac{2\sqrt{13}+7}{13}},$$

$$C_3 = 1 - \frac{2}{\sqrt{13}}, \qquad C_4 = -\sqrt{\frac{2\sqrt{13}-7}{13}},$$

and we see that $b_n = \left[C_1(\alpha+\beta)^n\right] - 1$, $n \geqslant 4$, where $[x]$ is the integer closest to x.
The first few upper and lower bounds of $c_n(\mathbf{R}^3)$ are summarized in table 1.

4.7 Discussion

There are several potential ways to improve the lower bound. The present method may be improved by using a better upper estimate in the following problem.

Problem. Given a triangulation with n vertices of a two-dimensional topological disc, how long can the shortest edge path be that is non-self-intersecting, that starts and ends on the boundary, and that passes through a given edge? $\square$

In the proposition, we only noted that one never has to pass through more than $n-1$ vertices.

Still, one should probably not expect significant improvements – in particular not for large n – using only two-dimensional estimates. Instead one should exploit the fact that each interior ball B' in the 'strip configurations' is tangent to three or more sequentially tangent balls, $B'_1, \ldots, B'_k$. Assuming all radii r'_i are bounded from below, the lower bound on r' given by the requirement that B' is tangent to two of the balls, is less than what the stronger[6] requirement of tangency with all of the balls gives.

In two dimensions, each disc has exactly two required tangencies, and varying those two discs is in essence what gives the sharp ring lemma. In three dimensions, however, each ball can have arbitrarily many tangencies that are part of the triangulation, and this is – intuitively speaking – the reason a sharp result cannot be immediately obtained in the same manner as in two dimensions. It seems likely, however, that using more than three tangencies for the smallest ball is inefficient, but a proof is somewhat complicated due to the sensitivity of ball packings: there is no immediate guarantee that a change of radius will maintain the surround property or even the tangency pattern.

Conjecture. $c_n(\mathbf{R}^3) = a_n$, $h_{n+4} = a_n$, attained in essentially unique configurations. $\square$

We see that our estimate yields $c_4(\mathbf{R}^3) = 1$, and considering the 'strip configuration', it is easy to see that $c_5(\mathbf{R}^3) = \frac{1}{4}$.

Finally we note that the lemma in section 4.3 does not directly show that an extremal three-dimensional packing has two half-spaces. Although one may move $\hat{x}$ to the intersection between two balls in the 'strip configuration' without increasing the ratio $\frac{r_*}{r_0}$, it is not obvious that doing so preserves the surround property.

---

6 In fact, we have the following result due to Soddy [20]: Given three pairwise tangent balls B_1, B_2, B_3 with pairwise disjoint interiors, a fourth ball B_4 can be tangent to all three balls only if its radius is sufficiently large. In fact, the minimal radius is attained precisely when B_4 lies in the plane through the points of tangencies between B_1, B_2, B_3, and the minimal radius can be calculated using the *two*-dimensional Descartes circle theorem.

References

[1] D. Aharonov, *The sharp constant in the ring lemma*, Complex Var. Theory Appl. **33** (1997), 27–31.

[2] D. Aharonov and K. Stephenson, *Geometric sequences of discs in the Apollonian packing*, Algebra i Analiz **9** (1997), 104–140.

[3] ———, *Geometric sequences of discs in the Apollonian packing*, St. Petersburg Math. J. **9** (1998), 509–542.

[4] L. V. Ahlfors, *Möbius transformations in several dimensions*, Ordway Professorship Lectures in Mathematics, School of Mathematics, University of Minnesota, 1981, revised third printing.

[5] T. M. Apostol, *Modular functions and Dirichlet series in number theory*, second ed., Graduate Texts in Mathematics, vol. 41, Springer-Verlag, New York, 1990.

[6] W. S. Brown, *The kiss precise*, Amer. Math. Monthly **76** (1969), no. 6, 661–663.

[7] H. Fukagawa and D. Pedoe, *Japanese temple geometry problems – San Gaku*, Charles Babbage Research Centre, 1989.

[8] T. Gosset, *The kiss precise*, Nature **139** (1937), 62.

[9] R. L. Graham, J. C. Lagarias, C. L. Mallows, A. R. Wilks, and C. H. Yan, *Apollonian circle packings: geometry and group theory – I. The Apollonian group*, Discrete Comput. Geom. **34** (2005), 547–585.

[10] ———, *Apollonian circle packings: geometry and group theory – II. Super-Apollonian group and integral packings*, Discrete Comput. Geom. **35** (2006), 1–36.

[11] ———, *Apollonian circle packings: geometry and group theory – III. Higher dimensions*, Discrete Comput. Geom. **35** (2006), 37–72.

[12] L. J. Hansen, *On the Rodin and Sullivan ring lemma*, Complex Var. Theory Appl. **10** (1988), 23–30.

[13] J. C. Lagarias, C. L. Mallows, and A. R. Wilks, *Beyond the Descartes circle theorem*, Amer. Math. Monthly **109** (2002), 338–361.

[14] J. G. Mauldon, *Sets of equally inclined spheres*, Canad. J. Math. **14** (1962), 509–516.

[15] D. Pedoe, *On a theorem in geometry*, Amer. Math. Monthly **74** (1967), 627–640.

[16] B. Rodin and D. Sullivan, *The convergence of circle packings to the Riemann mapping*, J. Differential Geom. **26** (1987), 349–360.

[17] T. Rothman, *Japanese temple geometry*, Sci. Amer. **278** (1998), 85–91.

[18] N. J. A. Sloane, *The on-line encyclopedia of integer sequences*, http://www.research.att.com/~njas/sequences/, 2007.

[19] F. Soddy, *The kiss precise*, Nature **137** (1936), 1021.

[20] ———, *The bowl of integers and the hexlet*, Nature **139** (1937), 77–79.

[21] K. Stephenson, *Introduction to circle packing*, Cambridge University Press, Cambridge, 2005.