Cannonballs and Honeycombs

Thomas C. Hales

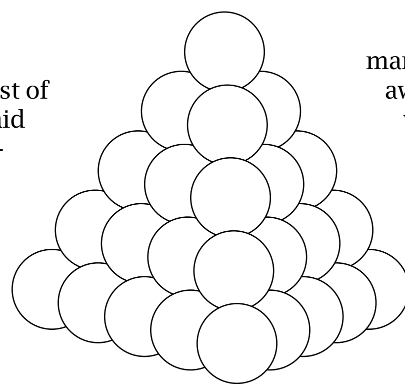

Figure 1. An optimal arrangement of equal balls is the face-centered cubic packing.

When Hilbert introduced his famous list of 23 problems, he said a test of the perfection of a mathematical problem is whether it can be explained to the first person in the street. Even after a full century, Hilbert's problems have never been thoroughly tested. Who has ever chatted with a telemarketer about the Riemann hypothesis or discussed general reciprocity laws with the family physician?

Last year a journalist from Plymouth, New Zealand, decided to put Hilbert's 18th problem to the test and took it to the street. Part of that problem can be phrased: Is there a better stacking of oranges than the pyramids found at the fruit stand? In pyramids the oranges fill just over 74% of space (Figure 1). Can a different packing do better?

The greengrocers in Plymouth were not impressed. "My dad showed me how to stack oranges when I was about four years old," said a grocer named Allen. Told that mathematicians have solved the problem after four hundred years, Allen was asked how hard it was for him to find the best packing. "You just put one on top of the other," he said. "It took about two seconds."

Not long after I announced a solution to the problem, calls came from the Ann Arbor farmers market. "We need you down here right away. We can stack the oranges, but we're having trouble with the artichokes."

To me as a discrete geometer there is a serious question behind the flippancy. Why is the gulf so large between intuition and proof? Geometry taunts and defies us. For example, what about stacking tin cans? Can anyone doubt that parallel rows of upright cans give the best arrangement? Could some disordered heap of cans waste less space? We say certainly not, but the proof escapes us. What is the shape of the cluster of three, four, or five soap bubbles of equal volume that minimizes total surface area? We blow bubbles and soon discover the answer but cannot prove it. Or what about the bee's honeycomb? The three-dimensional design of the honeycomb used by the bee is not the most efficient possible. What is the most efficient design?

This article will describe some recent theorems that might have been proved centuries ago if only our toolbag had matched our intuition in power. This article will describe the proof that the pyramid stacking of oranges is the best possible. But first I explain a few terms. A *sphere packing* always refers to a packing by solid balls. (The subject of sphere packings should more properly be called ball packings.) *Density* is the fraction of a region of space filled by the solid balls. If this region is bounded, this fraction is the ratio of the volume

Thomas C. Hales is professor of mathematics at the University of Michigan at Ann Arbor. His e-mail address is `hales@math.lsa.umich.edu`.

of the balls to the volume of the region. If any ball crosses the boundary of the region, only the part of the ball inside the region is used. If the region is unbounded, the density of the intersection of the region with a ball of radius R is calculated, and the density of the full region is defined as the lim sup over R.

Harriot and Kepler

The pyramid stacking of oranges is known to chemists as the *face-centered cubic packing*. It is also known as the cannonball packing, because it is commonly used for that purpose at war memorials. The oldest example I have seen is the pyramid of cannonballs from the sixteenth century that rests in front of the City Museum of Munich. Formulas for the number of cannonballs stacked in mounds have been known this long. In the sixteenth century Walter Raleigh gave his mathematical assistant, Thomas Harriot, the task of finding the formula. Harriot did this without difficulty.

As Harriot grew in reputation as a scientist, spheres became a favorite topic of his. To him, atoms were spheres. To understand how they stack together is to understand nature. Numbers were spheres. In the tradition of Pythagoras, triangular numbers are stacked like billiard balls set in a triangle. Harriot drew Pascal's triangle, but with a sphere packing with that many spheres replacing each number.

In 1606 Kepler complained that his recent book on optics was based entirely on theology. He turned for help to Harriot, who had been conducting experiments in optics for years. In fact, Harriot's knowledge of optics was so advanced that he had discovered Snell's law—twenty years before Snell and forty years before Descartes. Harriot supplied Kepler with valuable data in optics, but he also tried to persuade Kepler that the deeper mysteries of optics would be unfolded through atomism. Unlike Kepler, Harriot was an ardent atomist, believing that the secrets of the universe were to be revealed through the patterns and packings of small, spherical atoms. Kepler was skeptical. Nature abhors a vacuum, and between the atoms lies the void.

Harriot persisted. Kepler relented. In 1611 Kepler wrote a little booklet, *The Six-Cornered Snowflake*, that influenced the direction of crystallography for the next two centuries. This slender essay was the "first recorded step towards a mathematical theory of the genesis of inorganic or organic form."[1] In a discussion of sphere packings he constructed the face-centered cubic packing. Kepler asserted that it would be "the tightest possible, so that in no other arrangement could more pellets be stuffed into the same container." This assertion has come to be known as the Kepler Conjecture. It went without a proof for nearly four hundred years, until August 1998, when I gave a proof with the help of a graduate student, Samuel P. Ferguson.

[1] *L. L. Whyte, in* The Six-Cornered Snowflake, *Oxford Clarendon Press, Oxford, 1966.*

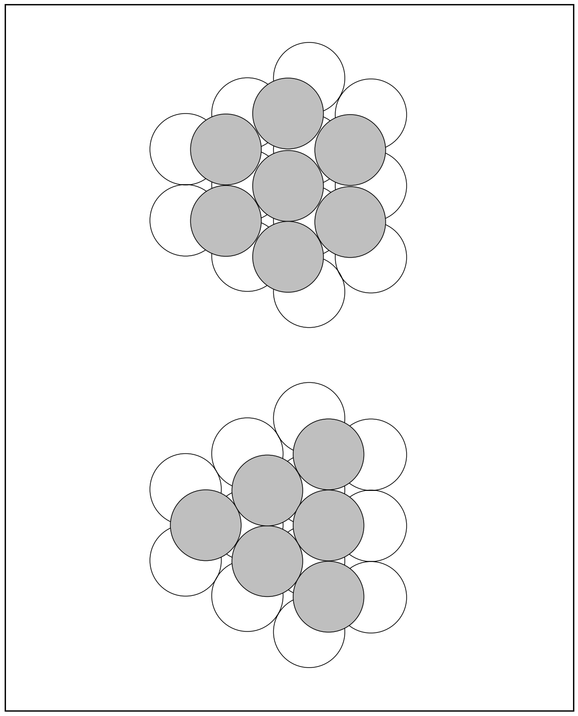

Figure 2. There are two optimal ways to place one layer of spheres upon another. Many different optimal packings can be constructed by varying the placement of each successive layer.

Theorem *(Kepler Conjecture). No packing of balls of the same radius in three dimensions has density greater than the face-centered cubic packing.*

The face-centered cubic packing is constructed by setting one layer of balls upon another. Each layer is a regular pattern of balls in a grid of equilateral triangles. But there are other sphere packings that have exactly the same density ($\pi/\sqrt{18} \approx 0.74$) as the face-centered cubic packing. The best-known alternative is the hexagonal close-packing. Its individual layers are identical to those of the face-centered cubic, but the layers are staggered to produce a different global packing with the same density (Figure 2). There is no simple list of all packings of density $\pi/\sqrt{18}$, but there are many other possibilities. The density of the packing in space is defined as a limit, and removing one, two, or a hundred balls from the packing will not affect the limiting density. Nor will removing an entire infinite layer of balls.

This article gives the broad outline of the proof of the Kepler Conjecture in the most elementary terms possible. The proof is long (282 pages), and every aspect of it is based on even longer computer calculations. A jury of twelve referees has been deliberating on the proof since September 1998. Nobody has raised doubts about the overall correctness of the proof. And yet, to my knowledge, no

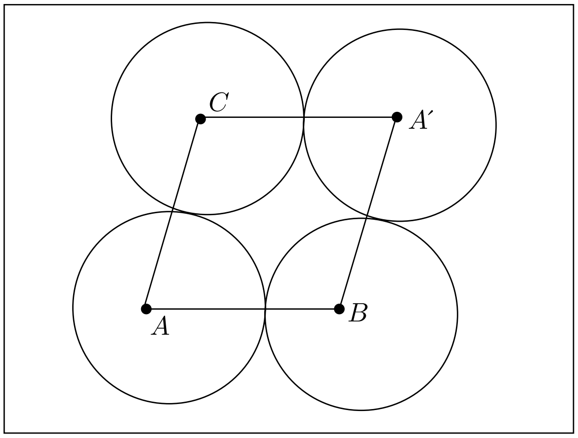

Figure 3. In the optimal lattice packing, a ball in the layer above the layer shown will touch the three balls A (or A'), B, and C.

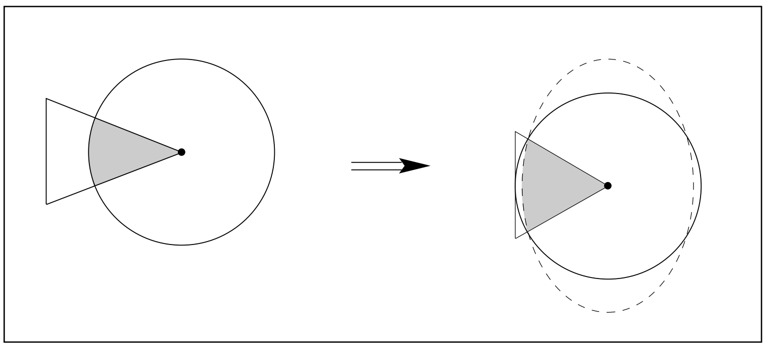

Figure 4. The linearly transformed region on the right occupies a smaller fraction of the equilateral triangle than the fraction inside the unit disk.

one has made a thorough independent check of the computer code.

Gauss

Gauss was the first to prove anything about the Kepler Conjecture. He showed that if all of the centers of the balls of a packing are aligned along the points of a lattice, then it can do no better than the face-centered cubic packing. Gauss's name confers an undeserved prestige to this elementary result. The proof takes only a few lines and requires no calculations. In the best case it will certainly be true that two balls will touch each other. Once two balls touch, the lattice constraint forces the balls to touch along long parallel strings of balls, like a thick row of marshmallows on a roasting stick. In the best case it will also certainly be true that two of the long parallel beaded strings will touch. The lattice constraint forces the balls to be laid out in identical parallel plates. The centers of four balls in the plate form a parallelogram, as shown in Figure 3. The parallel plates should be set one on the other so that the plates are as close as possible. A ball D of the next layer is set in the pocket between three balls A, B, C in the layer below so that it touches all three. The triangle ABD formed by the centers is equilateral.

We now change our point of view. We view all of the balls as arranged in planes parallel to ABD. In each of those layers the centers of the balls repeat the pattern of the equilateral triangle ABD. The balls of one layer should be nestled in the pockets of the layer before so that each ball rests on three below it. The lattice this describes is the face-centered cubic.

Thue

The two-dimensional version of Kepler's conjecture asks for the densest packing of unit disks in the plane. If we inscribe a disk in each hexagon in the regular hexagonal tiling of the plane, the density of the packing is $\pi/\sqrt{12} \approx 0.9069$. Thue's theorem, announced in 1890, affirms that this is the highest density possible. There is a common misconception that the proof of Thue's theorem is not elementary. The proof here is based on an idea of Rogers's and does not require calculus. The ideal presentation of this proof would be one that develops interactively on the computer screen without written words. But I never found the time to write the computer program, and I resort to words.

Take an arbitrary packing of the plane by nonoverlapping disks of radius 1. We will partition the plane into regions and will show that each region has density at most $\pi/\sqrt{12}$. Center a larger circle of radius $2/\sqrt{3}$ around each disk. Whenever two of these large circles intersect, draw the segment between the two points of intersection and draw two congruent isosceles triangles with this segment as base and vertices at the centers of the two circles. There cannot be a point interior to three large circles. Indeed, in the extreme case, three large circles meet at a point, the circumcenter, if the centers are the vertices of an equilateral triangle of edge 2.

This gives our partition of space: regions outside all of the larger circles, the isosceles triangles, and the part inside the larger circles but outside all triangles. The regions outside all of the larger circles have density 0, which is certainly less than $\pi/\sqrt{12}$. The density of the interior of the larger circles is the square $3/4$ of the ratio $(1 : 2/\sqrt{3})$ of the smaller to larger radius, again less than $\pi/\sqrt{12}$. This inequality can be seen geometrically by drawing a hexagon that the small circle inscribes and the large one circumscribes. The hexagon has density $\pi/\sqrt{12}$, and this will be greater than the density inside the full larger circle.

To calculate the density of an isosceles triangle, we apply a linear transformation to the triangle (preserving ratios of areas and hence densities) to transform it into an equilateral triangle with edge $2/\sqrt{3}$. The transformation scales along the orthogonal axes through the base and altitude of the triangle and fixes the vertex v opposite the base of the isosceles triangle. The unit circle is transformed into an ellipse. The linear transformation preserves the lengths of the two equal edges of the isosceles triangle. Hence the ellipse cuts those edges at distance 1 from its center v. This identifies all four points of intersection of two conic sections, the ellipse and the unit circle centered at v (see Figure 4.) Knowing these points of intersections, we deduce that the intersection of the ellipse with the interior of the equilateral triangle is contained in a disk of radius 1 centered at v. In particular, the density of the equilateral triangle is increased by replacing the ellipse with a circle of radius 1. The equilateral triangles fit together to form hexagons with inscribed disks of radius 1. The density of these pieces is therefore $\pi/\sqrt{12}$. This completes the proof of Thue's theorem.

Three Dimensions

We return to three dimensions to discuss the proof of the Kepler Conjecture. To avoid the unpleasant boundary effects caused by finite packings, we study sphere packings that extend to all of Euclidean space. Sphere packings are determined by a countably infinite set of parameters which give the coordinates of the center of each sphere. It was realized in the 1950s that it should be possible to prove the Kepler Conjecture by looking at a finite number of balls at a time. With this in mind, we discuss finite clusters of balls.

Voronoi

Each ball in our packings should be painted one solid color from a finite color set. These colors are needed in the details of certain constructions to resolve degeneracies, to make piecewise smooth functions smooth, and to keep domains compact. These colors will help me avoid oversimplifications in my exposition. But by going into details about colors, I would obscure the main lines of the proof of the Kepler Conjecture. So now that we have established that the balls are colored, the reader is free to paint all the balls black.

Let $t > 1$ be a real number. We define a cluster of balls to be a set of nonoverlapping colored balls around a fixed ball at the origin, with the property that the ball centers have distance at most $2t$ from the origin. A cluster of n balls is determined by the $3n$ coordinates of the centers. These coordinates give a topology on the set $C = C(t)$ of all clusters, making it a compact set. Two clusters with a different number of balls or different colorings lie in different connected components of C.

The ball at the center of the cluster is contained in a *truncated Voronoi cell*. By definition the Voronoi cell is the set of all points that lie closer to the origin than to any other ball center in the cluster. The truncated Voronoi cell $V_t(p)$ is the intersection of the Voronoi cell with a ball of radius t at the origin. We have seen Voronoi cells already in the proof of Thue's theorem without calling them that. The regular hexagons that appear in the proof of Thue's theorem are the Voronoi cells of the optimal packing. And the large disks, sliced at times to form isosceles triangles, are truncated Voronoi cells ($t = 2/\sqrt{3}$). Truncation is purely a matter of convenience, making the volumes of Voronoi cells easier to estimate.

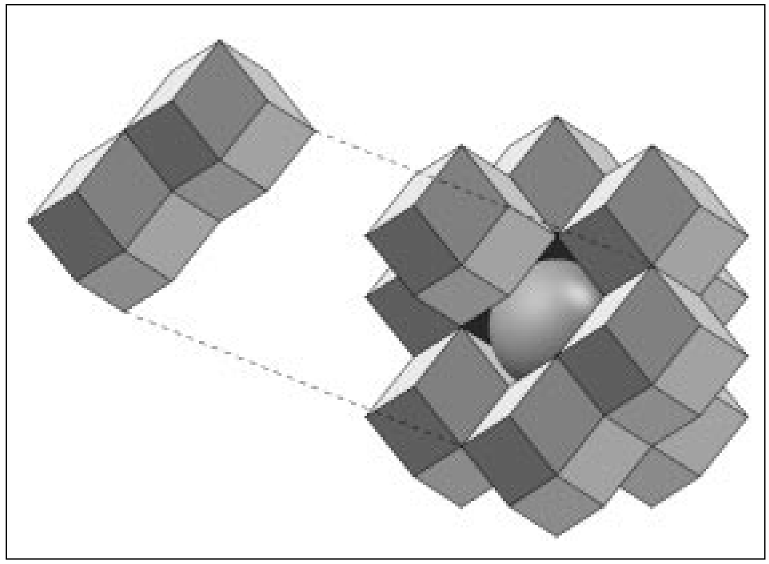

Figure 5. The face-centered cubic packing is produced by placing a ball inside each rhombic dodecahedron in the tiling.

Truncated Voronoi cells give a bound on the density of sphere packings. We place every ball of a packing inside its truncated Voronoi cell. Voronoi cells do not overlap; a point of intersection of two closed Voronoi cells, being equidistant from the center of two balls, lies on the boundary of both. The parts of space outside all truncated Voronoi cells do not meet any balls and have density 0. The density of the packing is no greater than its densest truncated Voronoi cell. That is, the greatest possible volume ratio of ball to truncated Voronoi cell is an upper bound on the density of a packing.

The Voronoi cells of the face-centered cubic packing are identical rhombic dodecahedra, as shown in Figure 5. Let v_{fcc} be the volume of the rhombic dodecahedron. The density of the face-centered cubic packing is the ratio of the volume of the unit ball to v_{fcc}:

$$\frac{4\pi/3}{v_{\text{fcc}}} = \frac{\pi}{\sqrt{18}} \approx 0.74.$$

The most distant vertices of the rhombic dodecahedron are $\sqrt{2}$ from the center. Thus, if $t \geq \sqrt{2}$, then truncation has no effect. If $t < \sqrt{2}$, then the truncation cuts into the rhombic dodecahedron and destroys the relation between its volume and our target $\pi/\sqrt{18}$. We fix the truncation at $t = \sqrt{2}$, its smallest useful value. The truncation now fixed, we drop t from the notation and write $V(p) = V_t(p)$.

The minimum volume of a Voronoi cell (either untruncated or truncated at $t = \sqrt{2}$) was recently determined by Sean McLaughlin. For this result he was awarded the AMS-MAA-SIAM Morgan Prize in January 2000. It confirms a conjecture made by L. Fejes Tóth nearly sixty years ago.

Theorem *(McLaughlin). The volume of the Voronoi cell of a sphere packing of a cluster p is uniquely minimized by a regular dodecahedron of inradius 1.*

The cluster of balls that gives the regular dodecahedron is a cluster with one ball at the center and twelve additional balls tangent to the one at the center, placed at the centers of the faces of the regular dodecahedron.

The ratio of the volume of the unit ball to the volume of the regular dodecahedron is an upper bound on the density of a sphere packing. This upper bound is about 0.75. In two dimensions the Voronoi cell of minimal volume is the regular hexagon, and it tiles the plane to form the optimal packing. In three dimensions the Voronoi cell of minimal volume no longer tiles. The locally optimal figure, the dodecahedron, no longer corresponds to the globally optimal figure, the tiling by rhombic dodecahedra. This is the source of complications in the proof of the Kepler Conjecture.

We add a correction term f to the minimization of the volume of Voronoi cells; namely, we define a continuous function f on C and consider the minimization problem

$$\min_p \left(\mathrm{vol}(V(p)) + f(p)\right).$$

We say that f is *fcc-compatible* if the minimum of $\mathrm{vol}(V(p)) + f(p)$ is v_{fcc}, the volume of the rhombic dodecahedron.

Let Λ be the set of centers of the balls in a general packing. For $\lambda \in \Lambda$ consider the cluster of balls centered at distance at most $2t = 2\sqrt{2}$ from λ. Translating the cluster to the origin, we obtain a cluster p_λ in C. Let $\Lambda_R = \Lambda \cap B_R$ be the set of all centers within distance R of the origin. We say that f is *transient* if

$$\sum_{\lambda \in \Lambda_R} f(p_\lambda) = o(R^3).$$

Assume that f is fcc-compatible and transient. By summing

$$v_{\mathrm{fcc}} \le \mathrm{vol}(V(p)) + f(p)$$

over Λ_R, we obtain

$$|\Lambda_R| v_{\mathrm{fcc}} \le \mathrm{vol}(B_R) + o(R^3).$$

Divide by $R^3 v_{\mathrm{fcc}}$ to get the density of a packing inside a ball of radius R.

$$\frac{|\Lambda_R|}{R^3} \le \frac{4\pi/3}{v_{\mathrm{fcc}}} + o(1) = \frac{\pi}{\sqrt{18}} + o(1).$$

Taking the limit as R tends to ∞, we obtain the bound $\pi/\sqrt{18}$ on the density of the packing.

That shows that the whole proof of the Kepler Conjecture follows if a transient fcc-compatible function f can be found. To establish fcc-compatibility, an extremely difficult nonlinear optimization problem on C must be solved. We select the function f with transience in mind, so that transience is automatically satisfied.

Fejes Tóth

Do correction terms with the required properties exist? The evidence suggests that they exist in abundance. The first correction term was proposed by L. Fejes Tóth in 1953, but his clusters were much larger than those used here. His clusters contain so many balls that fcc-compatibility has never been established. Nevertheless, his proposal represents a significant advance, because it gave the first evidence that the Kepler Conjecture could be solved through an optimization problem in a finite number of variables. He proposed in 1964 that computers might be used to determine the minimum. Thus, the general strategy of a proof was set.

The correction terms f are based on a careful study of the local geometry of sphere packings. Fejes Tóth's correction term $f(p)$ has the form

$$\sum_q a(p,q) v(q),$$

with q running over all centers of balls in the cluster p within a fixed distance of the center of the cluster. The term $v(q)$ is the volume of a truncated[2] Voronoi cell centered at q. The constants $a(p,q)$ sum to zero, $\sum_p a(p,q) = 0$, for all clusters p in a packing. This zero-sum condition leads to a cancellation of the terms in $\sum f(p)$, and hence to the transience of f. This correction term illustrates the general correction term strategy: f is constructed as sums of volumes that are added at p and subtracted again at q. The sum $\sum f(p)$ behaves like the telescoping series $1 - 1 + 1 - 1 + \cdots$, and each term of the sum is swallowed by the next in its path. Transience results from this cancellation of terms.

Of course, there is no need for the volumes that are shuffled between p and q to be Voronoi cells. One of many other possibilities is known as the Delaunay decomposition. In that decomposition an edge is drawn between two centers of balls if their Voronoi cells share a face. These edges form simplices, known as Delaunay simplices, and the simplices partition space.

When asked to name the most difficult part of the proof of the Kepler Conjecture, I answer

[2] *A different truncation constant is used in this approach.*

without hesitation that it was the design of the decomposition of space implicit in f. I had worked with Voronoi cells without success and had also tried Delaunay simplices. Both approaches became complicated beyond my ability to understand them. My progress stopped. Finally, one day in November 1994 I realized how to combine the two approaches into a hybrid decomposition that retained the best features of each. From that day on, I never waivered in my confidence that the Kepler Conjecture would eventually be solved by the hybrid approach.

Hybrid correction terms are extremely flexible and easy to construct, and soon Samuel Ferguson and I realized that every time we encountered difficulties in solving the minimization problem, we could adjust f to skirt the difficulty. The function f became more complicated, but with each change we cut months—or even years—from our work. This incessant fiddling was unpopular with my colleagues. Every time I presented my work in progress at a conference, I was minimizing a different function. Even worse, the correction function in my early papers differs from the one in the final papers, and this required me to go back and patch the old papers. The correction function did not become fixed until it came time for Ferguson to defend his thesis, and we finally felt obligated to stop tampering with it. However, if I were to revise the proof to produce a simpler one, the first thing I would do would be to change the correction function once again. It is the key to a simple proof.

Combinatorial Structures

With f a fixed transient function, the only remaining problem is the minimization problem: show that the corrected volume $F(p) = \mathrm{vol}(V(p)) + f(p)$, for $p \in C$, is at least the volume v_{fcc} of the rhombic dodecahedron. The space of clusters C is so complicated that we cannot minimize F directly. We associate with each cluster $p \in C$ a planar graph that identifies the most prominent geometrical features of the cluster. Next, we give a lower bound on the value of $F(p)$ that depends only on the combinatorial structure of this planar graph. In most cases the combinatorial lower bound on $F(p)$ is greater than v_{fcc}, as desired. The combinatorial approximation to $F(p)$ is rather crude, and sometimes it fails to give the lower bound we want. By means of a computer search we can generate all planar graphs for which the combinatorial lower bound on $F(p)$ is less than v_{fcc}. If $F(p) < v_{\mathrm{fcc}}$, the planar graph of p must appear on this computer-generated list of possibilities. This list contains about 5,000 planar graphs.

The planar graph of a cluster p is easy to construct. Each edge of the graph corresponds to pairs of balls that are close to the ball at the origin and that are close to each other. Set $\tau = 4\sqrt{3/7} \approx 2.618$. Closeness is determined by a parameter[3] $T \in (2, \tau)$. If two ball centers have distance at most T from the origin and distance at most T from each other, we draw a circular arc on the unit ball at the origin connecting the two unit direction vectors pointing to the two ball centers (Figure 6).

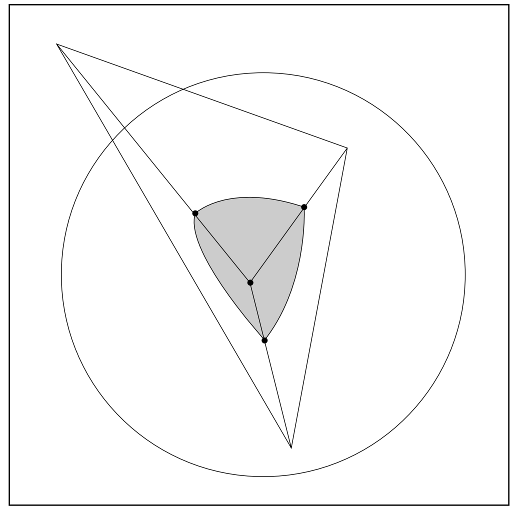

Figure 6. A tetrahedron cuts out a spherical triangle on the unit sphere.

These circular arcs do not meet, except at their endpoints. In fact, this is why we require that $T < \tau$. If $T = \tau$, we have the configuration of ball centers shown in Figure 7. The edges all have length 2 or τ, and the edges of length τ are marked with heavy lines. The edges A and B give crossing arcs on the sphere at the origin 0. For any smaller T, the arcs cannot cross.

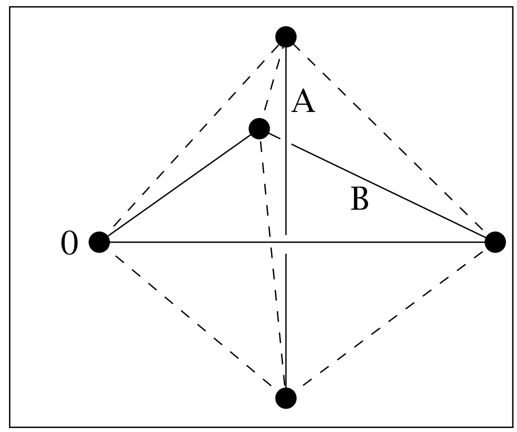

Figure 7. This configuration must be excluded because it produces overlapping edges on the unit sphere.

The planar graph—or more accurately, the spherical graph—associated with a cluster p is the one formed by this system of arcs on the unit sphere. For example, the planar graph associated with the face-centered cubic packing is an alternating pattern of equilateral triangles and squares, with two triangles and two squares meeting at every vertex. The planar graph associated with the regular dodecahedral cluster is the icosahedral graph with twenty triangles arranged five to a vertex.

In the complement of this collection of arcs on the unit sphere, we call the closure of each connected component a *standard region*. In simple cases, these standard regions are just spherical polygons on the unit sphere. But in general they may be much more complicated. The most difficult estimates in the proof of the Kepler Conjecture are those that establish that the complicated standard regions are never optimal. This is done by making a combinatorial approximation to the minimization problem. We calculate a lower bound

[3] *We use* $T = 2.51$*, but this is rather arbitrary.*

on $F(p)$ that depends only on the combinatorial structure of the planar graph associated with p.

Most of my time between 1995 and 1998 was spent working out a proof of these combinatorial lower bounds. Already in 1994 I knew what the bounds should be, but the proofs eluded me. The primary difficulty was with standard regions with a large number of sides. As the number of sides increased, the dimension became too large for me to handle.

The eventual proof was shaped by the capabilities of computers. If computers had been more powerful, the proof might be drastically shorter. If computers had been less powerful, I would still be working toward a solution. As a result of the development of computers, the proof of the Kepler Conjecture fifty years from now will likely be entirely different from what it is today.

A simple heuristic told me what I could get from a computer. My computer was generally able to prove statements about a single tetrahedron, but failed to prove anything about more complicated geometrical objects. In other words, my computer could tell me about the six-dimensional space parametrizing the edge lengths of a tetrahedron but was too slow to handle seven dimensions. Given that the Kepler Conjecture is an optimization problem in about seventy variables, I found this limitation to be a frustrating one. The challenge of the problem was to come to a thoroughly six-dimensional understanding of a seventy-dimensional space.

5,000 Cases

In some cases the crude combinatorial bounds were not good enough. One case turned out to be far more intricate than the others, and it became the subject of Ferguson's thesis. The remaining 4,999 or so planar graphs were analyzed individually. For each there is a large-scale nonlinear optimization problem to be solved. Minimize $F(p)$ subject to the constraint that the cluster be associated with the given planar graph. Nonlinear optimization problems of this size can be hopelessly difficult to solve rigorously. We might easily have come this close to a solution only to be thwarted in our attempts by nonlinearities. But a new observation carries us forward: the large-scale structure of the problem is linear and can be solved by linear programming methods.

The large-scale linearities of the problem can best be understood by turning back to the problem, solved by McLaughlin, of minimizing the volume of a truncated Voronoi cell. Here there is no correction term; that is, $f = 0$. To further simplify the exposition, let us assume that there is no truncation, so that the full Voronoi cell lies inside a ball of radius $\sqrt{2}$ at the origin. We divide the Voronoi cell into simplices according to some convenient scheme. To fix our attention, here is one possibility. Drop a perpendicular from each face (at a point on the face we will call v) to the center of the Voronoi cell. Drop a perpendicular from each edge (at a point w) of the face to v. The vertices of a simplex are the center of the Voronoi cell, the point v on the face, the point w on the edge of the face, and finally either endpoint of the edge. These simplices partition the Voronoi cell.

Instead of minimizing the volume of the Voronoi cell directly, we can introduce variables x_i representing the volumes of the individual simplices. We minimize the sum of the x_i (which is certainly linear in x_i), subject to the constraint that the pieces fit together. The assembly constraints are all linear. There are constraints of the form $z = z'$ (which is linear in z and z'), where z is the length of an edge of a simplex and z' is the length of a matching edge of a second simplex that shares an edge with the first. There are constraints of the form $\alpha_1 + \alpha_2 + \ldots + \alpha_k = 2\pi$ (linear in α_i) that stipulate that the dihedral angles of the simplices around an internal edge should sum to 2π, and there are similar linear constraints for edges that lie on a face of the Voronoi cell. The problem becomes a massive linear programming problem.

There is a flaw in this argument: there are unavoidable nonlinear constraints. The volumes x_i and the dihedral angles α_i are *nonlinear* functions of the edge lengths of a simplex. Nevertheless, the large-scale structure of the problem is linear. The nonlinear relations are the relations that hold for a simplex in isolation. These nonlinearities involve a small number of variables and can be treated by computer according to the heuristic principle enunciated above that the computer can tell us whatever we want to know about a single simplex. In particular, the computer verifies inequalities between volumes, dihedral angles, and edge lengths that can be used as linear substitutes for the nonlinear relations.

If we now go back to the Kepler Conjecture, so that the correction term f is nonzero, the large-scale structure of the problem is still linear. In fact, the function f is defined geometrically as a linear combination of volumes. Knowing that linear programs would be an essential part of the proof, I was careful to choose a function f adapted to that end. By breaking larger objects into smaller objects, we can express the minimization problem in terms of simple quantities such as areas of spherical triangles and volumes of simplices, subject to linear assembly constraints. All nonlinearities of the problem are confined to a small number of variables.

Linear Programs

The linear part of the problem was solved with a linear programming package. A typical linear optimization problem involves about 200 variables and perhaps 2,000 constraints. I estimate that nearly 10^5 linear programming problems of this size were solved as part of the solution. This is a small calculation in comparison with industrial applications of linear programming.

Some variables represent distances between balls in various finite clusters of balls. Other variables represent dihedral angles, volumes, solid angles, and corrected volumes of Voronoi cells. Some constraints express geometric relations between the variables. Other constraints restrict the lengths and angles so that physically realistic packings of balls are obtained. The linear programming problems minimize the corrected volume subject to these constraints. By checking that in every case the corrected volume is greater than the volume of the rhombic dodecahedron, the Kepler Conjecture is proved.

Honeycombs

If we turn to the next page after the Kepler Conjecture in Kepler's *Six-Cornered Snowflake*, we find a discussion of the structure of the bee's honeycomb. The rhombic dodecahedron was discovered by Kepler through close observation of the honeycomb. The honeycomb is a six-sided prism sealed at one end by three rhombi. By sealing the other end with three additional rhombi, the honeycomb cell is transformed into the rhombic dodecahedron.

During the eighteenth century, mathematicians made extensive studies of the isoperimetric properties of the honeycomb and believed the honeycomb to be the most efficient design possible. For example, in 1743 C. MacLaurin in his investigation of the rhombic bases of the honeycomb concluded, "The cells, by being hexagonal, are the most capacious, in proportion to their surface, of any regular figures that leave no interstices between them, and at the same time admit of the most perfect bases." However, the obvious answer provided by the honeybee turned out to be incorrect. Upsetting the prevailing opinion, L. Fejes Tóth discovered that the three-dimensional honeycomb cell is not the most economical (Figure 8). The most economical form has never been determined.

The cannonball packing of balls leads to honeycomb cells. It is also related to more general foam problems. If we tile space with hollow rhombic dodecahedra and imagine that each has walls made of a flexible soap film, we have an example of a foam.

Kelvin

The problem of foams, first raised by Lord Kelvin, is easy to state and hard to solve. How can space be divided into cavities of equal volume so as to minimize the surface area of the boundary? The rhombic dodecahedral example is far from optimal. Kelvin proposed the following solution. Truncated octahedra fill space (see Figure 9). In fact, their cross-sections are regular octagons, and the octagons tile the plane except for square holes. Each truncated octahedron contains square plugs, so that the next layer of octahedra plugs the square holes of the previous layer.

Kelvin found that by warping the faces of the truncated octahedra ever so slightly, he could obtain a foam with smaller surface area than the cells of the truncated octahedra. This was Kelvin's proposed solution. It satisfies the conditions Plateau discovered for minimal soap bubbles. Everyone seemed satisfied with Kelvin's solution; only a proof of optimality was missing.

Kelvin and his supporters were wrong, as the physicists D. Phelan and R. Weaire showed in 1994. They produced a foam with cavities of equal volume with a considerably smaller surface area than the Kelvin foam. The Phelan-Weaire foam contains

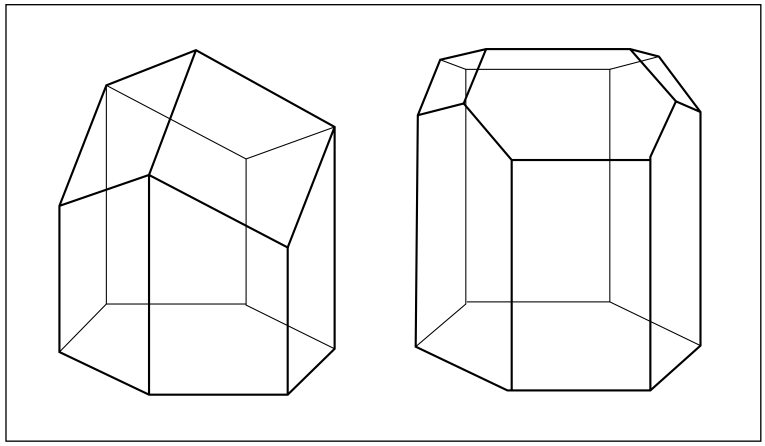

Figure 8. In 1964 a cell (right) was discovered that is more efficient than the bee's honeycomb (left).

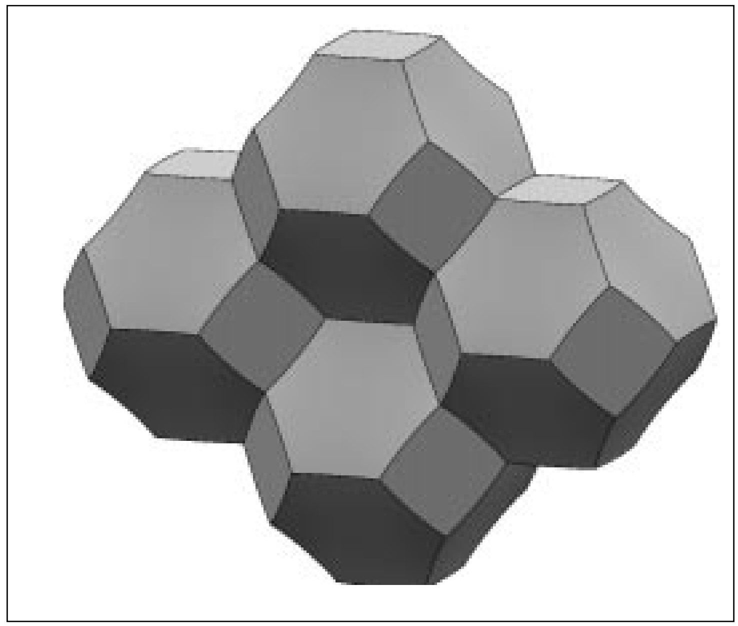

Figure courtesy of John Sullivan.

Figure 9. Kelvin conjectured that the optimal partition of space into equal volumes is obtained by warping the tiling of space by truncated octahedra.

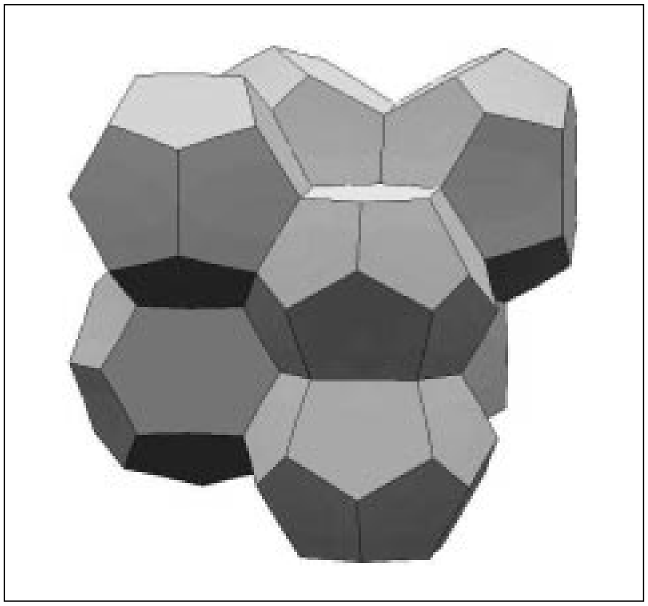

Figure courtesy of John Sullivan.

Figure 10. In 1994 R. Phelan and D. Weaire found this counterexample to the conjecture of Kelvin stated in the caption of Figure 9.

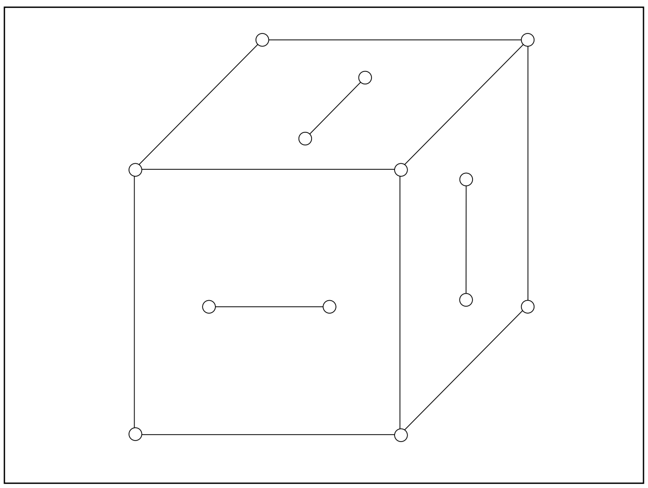

Figure 11. The centers of the Phelan-Weaire cells are located at the corners, at the marks on the faces, and at the center of the cube.

two different types of cavities, one with 14 sides, the other with 12. (See Figure 10.) Imagine a cube with a small ball at each vertex and one in the center. Add two balls to each face as shown in Figure 11. Form the Voronoi cells around each ball, adjusted so that they all have the same volume. If one warps the faces of this configuration ever so slightly, the Phelan-Weaire counterexample to Kelvin's conjecture is obtained.

A finite version of the problem might be more tractable. What shape minimizes surface area if the foam contains only finitely many bubbles of equal volume? If there is a single bubble, the problem is the classical isoperimetric problem. The sphere uniquely minimizes the area of a surface enclosing a given volume. The problem of two bubbles, known as the double bubble conjecture, was solved only recently by J. Hass, M. Hutchings, and R. Schlafy.[4] The problem for more than two bubbles is still unsolved.

The Kepler Conjecture and the Kelvin problem are both special cases of a more general foam problem. Phelan and Weaire ask us to imagine that the soapy film walls have a measurable thickness. We interpolate between the Kepler and Kelvin problems with a parameter w (measuring the wetness of the film) that gives the fraction of space filled by the thick film walls, and $1 - w$ is the fraction filled by the cavities. If the foam is perfectly dry, then $w = 0$ and the film walls are surfaces. The Kelvin problem asks for the most efficient design. When the foam becomes sufficiently wet, w is close to 1 and the cavities of the foam can be independently molded. The isoperimetric inequality dictates that they minimize surface area by forming into perfect spheres. The Kepler problem asks for the smallest value of w for which every cavity is a perfect sphere.

R. Weaire was writing a book on sphere packings when I finished the proof of the Kepler Conjecture, and we began to correspond. Under his influence, I turned to the planar version of the foam problem. This problem goes back over two thousand years. What is the most efficient partition of the plane into equal areas? The Honeycomb Conjecture asserts that the answer is the regular hexagonal honeycomb.

Pappus

Around 36 B.C. the Roman scholar Marcus Terentius Varro wrote a book on agriculture in which he discusses the hexagonal form of the bee's honeycomb. There were two competing theories of the hexagonal structure. One theory held that the hexagons better accommodated the bee's six feet. The other theory, supported by the mathematicians of the day, was that the structure was explained by the isoperimetric property of the hexagonal honeycomb. Varro writes, "Does not the chamber in the comb have six angles…The geometricians prove that this hexagon inscribed in a circular figure encloses the greatest amount of space."

This ancient proof has been lost, unless it was the proof presented a few centuries later by Pappus of Alexandria in the preface to his fifth book. The argument in Pappus is incomplete. In fact, it involves nothing more than a comparison of three suggestive cases. It was known to the Pythagoreans that only three regular polygons tile the plane: the triangle, the square, and the hexagon. Pappus states that if the same quantity of material is used for the constructions of these figures, it is the hexagon that will be able to hold more honey.

[4] *The double bubble conjecture,* Electron. Res. Announc. Amer. Math. Soc. ***1** (1995), no. 3, 98–102.*

Pappus's reason for restricting his attention to the three regular polygons that tile are not mathematical (bees avoid dissimilar figures). He also excludes gaps between the cells of the honeycomb without mathematical argument. If the cells are not contiguous, "foreign matter could enter the interstices between them and so defile the purity of their produce."[5]

In 1943 L. Fejes Tóth gave a proof of the Honeycomb Conjecture under the hypothesis that all the cells are convex polygons. He stated that the general conjecture had "resisted all attempts at proving it." In 1999 I found the first general proof.

***Theorem** (Honeycomb Conjecture). Any partition of the plane into regions of equal area has perimeter at least that of the regular hexagonal honeycomb tiling.*

The main ingredient of the proof is a new isoperimetric inequality which has the regular hexagon as its unique minimum.

My expectations of mathematics have been shaped by the Kepler Conjecture. I have come to expect every theorem to be a monumental effort. I was psychologically unprepared for the light 20-page proof of the Honeycomb Conjecture. It makes no significant use of computers and took less than six months to complete. In contrast with the years of forced labor that gave the proof of the Kepler Conjecture, I felt as if I had won a lottery.

The honeycomb problem is preparation for the real challenge, the Kelvin problem. What is the most efficient partition of space into equal volumes? Is it the Phelan-Weaire foam?

In this year of renewed interest in Hilbert's problems, many mathematicians are proposing new lists of problems. My submission to Hilbert's millennial list is the Kelvin problem. It has a rich history. Its solution will require new ideas from geometric measure theory. Frank Morgan predicts that the Kelvin problem might take a century to be solved. For starters: Does an optimal solution exist?

References

[1] S. P. Ferguson and T. C. Hales, A formulation of the Kepler conjecture, LANL e-print Archive `math.MG/9811072; http://xxx.lanl.gov/`.

[2] T. C. Hales, An overview of the Kepler conjecture, LANL e-print Archive `math.MG/9811071; http://xxx.lanl.gov/`.

[3] T. C. Hales, The Honeycomb conjecture, LANL e-print Archive `math.MG/9906042; http://xxx.lanl.gov/`.

[4] T. C. Hales and S. McLaughlin, A proof of the dodecahedral conjecture, LANL e-print Archive `math.MG/9811079; http://xxx.lanl.gov/`.

[5] *T. Heath,* A History of Greek Mathematics*, Vol. II, Oxford, 1921, p. 390.*

About the Cover

The mathematical structure of the hexagonal honeycomb has intrigued humankind for over two thousand years. Pappus attributes great wisdom to bees for choosing the hexagonal form for the cells in the honeycomb. The theme of the sagacity of the bees is repeated throughout the centuries, and the bee in *Arabian Nights* asserts that "Euclid himself could learn from studying the geometry of my cells."

Mathematicians, as early as the first century B.C., have remarked on the efficiency of the design of the honeycomb. As the Roman scholar Marcus Terentius Varro expressed it, "The geometricians prove that this hexagon inscribed in a circular figure encloses the greatest amount of space." This isoperimetric property of the honeycomb has continued to intrigue mathematicians today. The article "Cannonballs and Honeycombs" describes a modern version of this ancient observation in two dimensions.

In three dimensions, the bee's honeycomb is not the most efficient design possible, as was shown by L. Fejes Tóth (see page 447).

—T. C. H.

"Bees on a Honeycomb". Photograph by Peter Poulides for Tony Stone Images.