December 21, 1994

What is a surface?

Frank Morgan
Williams College

A search for a good definition of surface leads to the rectifiable currents of geometric measure theory, with interesting advantages and disadvantages.

For details and references see Morgan's *Geometric Measure Theory: a Beginner's Guide*, Academic Press, 2nd edition, 1995.

1. What is an inclusive definition of a general surface in R^3? We want to include smooth embedded manifolds with boundary, as in Figure 1, and we want to be able to allow singularities, as in the cube and cone of Figure 2. We might allow any smooth embedded *stratified manifold,* i.e., a set which is a smooth embedded 2-dimensional manifold, except for a subset which consists of smooth embedded curves, except for a set of isolated points. We might go farther and allow any set which is a smooth embedded 2-dimensional manifold except for a set of 2-dimensional measure 0.

Unfortunately for such surfaces it is hard to prove the existence of area-minimizers or solutions to other geometric variational problems, because such classes of surfaces are not closed under the limit arguments used to obtain such solutions. What good is a sequence of surfaces with areas approaching an infimum, if there is no limit surface realizing the least area? Moreover, there are more general sets which deserve to be called surfaces. We begin with a lower-dimensional example of a compact, connected 1-dimensional "curve" in R^2 which is not a stratified manifold, and actually has a singular set of positive 1-dimensional measure. The construction is based on a Cantor set C in R^1 of positive measure, differing from the usual Cantor set in the rapidly diminishing size of the segments removed.

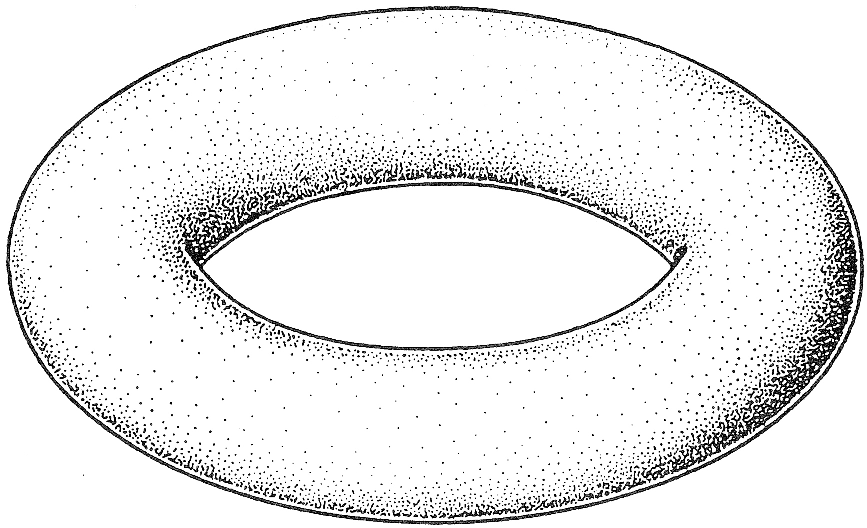

Figure 1
The nicest surfaces are smooth, embedded manifolds.

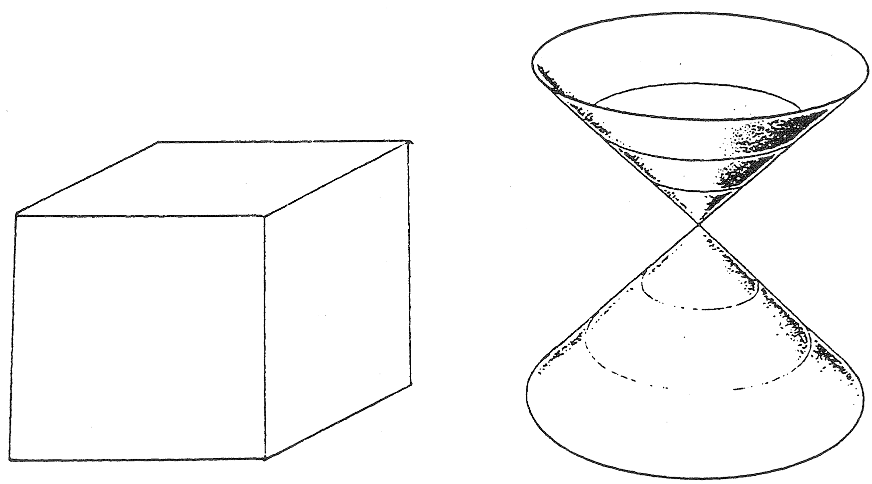

Figure 2
Surfaces such as the cube and cone have mild singularities.

2. Cantor set C of positive measure. To construct a Cantor set C of positive measure, start with the unit interval as in Figure 3a. Remove the middle 1/4. From each of the remaining two pieces, remove an open segment in the middle of length 1/16. At the k^{th} step, remove from each of the remaining 2^k pieces an open segment in the middle of length $\frac{1}{4} \cdot 2^{-2k}$. In countably infinitely many steps you remove total length

$$\frac{1}{4} \sum_{k=0}^{\infty} 2^k 2^{-2k} = \frac{1}{2},$$

leaving a compact Cantor set C of length 1/2 as in Figure 3b.

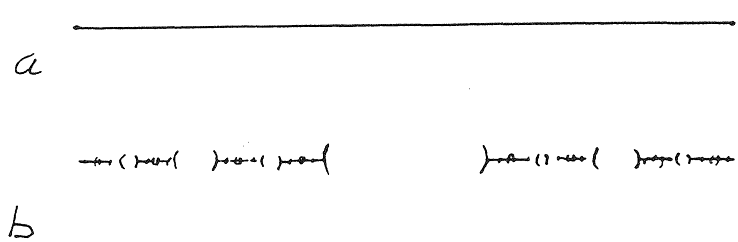

Figure 3
Start with the unit interval. Removing infinitely many open segments of rapidly decreasing length leaves a Cantor set C of positive measure.

3. A "curve" with a singular set of positive measure. We now construct a curve which is singular on our Cantor set C of positive measure. Let A consist of a nice smooth path from (0,0) to (1,0) above the x-axis, together with its reflection below the x-axis, as in Figure 4a. Replace every segment removed in the previous construction of the Cantor set C by a suitably scaled copy of A, as in Figure 4b. The resulting "curve" is an embedded manifold except for the singular set C, which has 1-dimensional measure 1/2.

Figure 4
Inserting bifurcating paths A into the gaps of a Cantor set C of positive measure yields a "curve" with a singular set of positive 1-dimensional measure

Similarly one could add infinitely many handles of finite total area to the sphere to produce a 2-dimensional surface S with a singular Cantor set of positive 2-dimensional measure, as in Figure 5. Note that S is the limit of smooth submanifolds without boundary, namely, spheres with a large finite number of handles.

Figure 5
Adding infinitely many handles of finite total area to the sphere can produce a surface S with a singular Cantor set of positive 2-dimensional measure.

4. Rectifiable sets. A good class of 2-dimensional subsets of R^3 which includes all of the surfaces we have considered or would ever want to consider and has nice closure properties under limit operations is the *rectifiable sets* of H. Federer. What do they include? To begin with they include the image of any reasonable function f from planar domains into R^3. The defining function f is not required to be differentiable, merely *Lipschitz:*

$$|f(x) - f(y)| \leq C|x-y| , \tag{1}$$

allowing for example the upper cone as the image of

$$\begin{aligned} f : R^2 &\rightarrow R^3 \\ x &\mapsto (x, |x|). \end{aligned}$$

This Lipschitz condition (1), by bounding the amount of stretching, is just right for producing surfaces of finite area. To obtain the full class of rectifiable sets, allow arbitrary (measurable) subsets of countable unions of such images of Lipschitz functions, as long as the total area remains finite. Such rectifiable sets include our example of a sphere with infinitely many handles.

It is a very fortunate theorem of real analysis that Lipschitz functions are differentiable almost everywhere. As a result, although rectifiable sets can be quite intricate, they turn out to have a kind of measure-theoretic "approximate" tangent plane at almost every point, which is good enough for doing lots of geometry. For example, one can define an orientation of a rectifiable set simply as a measurable orientation of (almost every) tangent plane. So far there is no coherence from point to point, and a smooth piece of surface has infinitely many "orientations." We will see that incoherent orientations can be detected by the extra boundary they introduce.

5. The structure theory. The fundamental role of rectifiable sets is exposed by a general structure theorem of Besicovitch and Federer. The theorem says that every subset E of R^3 of finite 2-dimensional measure can be decomposed as a rectifiable set and a *purely unrectifiable set* which is invisible from almost all directions (whose projections onto almost all planes have measure 0). So rectifiable sets form a fundamental and inclusive class of surfaces. The question is whether we can do geometry with them.

6. Hausdorff metric unsuitable. How can you say when two rectifiable sets are close together? The standard Hausdorff metric distance between two compact subsets of $\mathbf{R}^3$ is defined as the greatest distance between any point of one and any point of the other. Rectifiable sets need not be compact, and the Hausdorff metric is practically useless, as shown by the following example of radically different rectifiable sets S_0 and S_1 which are close together in the Hausdorff metric. Let S_0 be the unit sphere. Let S_1 be a countable collection of tiny spheres of radius at most ε, centered in S_0, dense in S_0, with total area ε. Then the Hausdorff distance between S_0 and S_1 is at most ε, even though S_0 is the round sphere of area 4π and S_1 is a fragmented (though in some sense boundaryless) set of area at most ε.

7. Rectifiable currents. So how can we apply geometric concepts to rectifiable sets? The way to define boundary and topology for oriented rectifiable sets is to view them as *currents,* linear functionals on smooth differential forms φ. Since an oriented rectifiable set S has an approximate oriented tangent plane $\overrightarrow{S}$ at almost every point, one can integrate a differentail form φ over S:

$$S(\varphi) = \int_S \varphi(\overrightarrow{S}),$$

and thus view S as a current. The space of currents so arising from rectifiable sets is called the space $\mathcal{R}_2\mathbf{R}^3$ of 2-dimensional rectifiable currents. One allows integral multiplicities, but finite total area and compact support.

The general concept of currents was a generalization, due to G. de Rham, of distributions. H. Federer and W. Fleming introduced rectifiable currents in 1960 in their foundational paper on geometric measure theory, which won the AMS Steele prize for fundamental importance.

8. Boundary. The boundary of a current S, denoted ∂S, may be defined as an abstract current by the formula

$$\partial S(\varphi) = S(d\varphi).$$

By Stokes's theorem, this definition agrees with the usual one for smooth oriented manifolds with boundary. Of course there is no reason in general that the boundary ∂S of a rectifiable current S should be a rectifiable current. If it happens to be, then the original current S is called an *integral current.*

The boundary of the unit disc with the standard orientation is the unit circle with the standard orientation, and therefore the unit disc is a nice integral current. Suppose however the disc is decomposed into the infinitely many concentric annuli

$$A_n = \{1/(n+1) < r \leq 1/n\}$$

of Figure 6 and they are given alternating orientations. Then the boundary includes infinitely many circles (with multiplicity 2), has infinite length, hence is not a rectifiable current, and therefore the disc with this incoherent orientation is not an integral current. This example shows how incoherent orientations may be detected by the additional boundary they introduce.

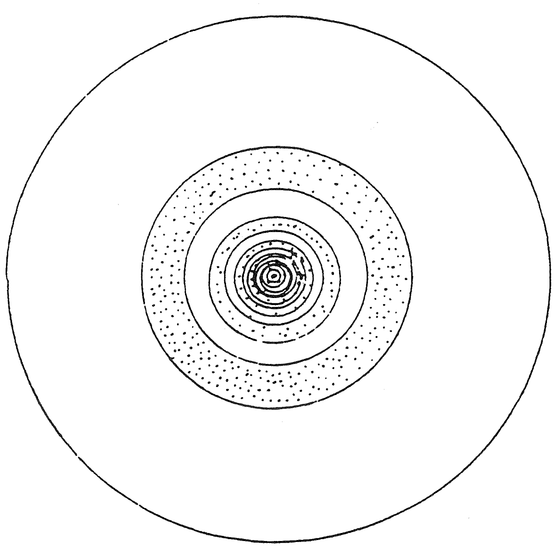

Figure 6
Giving alternating orientations to concentric annuli creates lots of extra boundary.

9. Topology and flat norm. The notion of boundary leads to a topology on the space of rectifiable currents given by H. Whitney's *flat norm* $\mathfrak{F}$ on currents, defined by

$$\mathfrak{F}(S) = \inf\{\text{area } T + \text{vol } R : S - T = \partial R\}.$$

For example, the two discs D_1, D_2 of Figure 7 are close together in this topology because their difference $S = D_2 - D_1$ together with a thin band T bounds a region R of small volume. (Whitney wanted to distinguish the flat norm from another larger sharp norm: as a music major, he borrowed the terms "flat" and "sharp," indicating lower or higher notes, from musical terminology.)

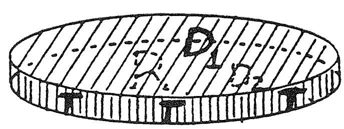

Figure 7
The two discs D_1, D_2 are close together in flat norm $\mathfrak{F}$ because their difference $S = D_2 - D_1$ together with a thin band T bounds a region R of small volume.

10. The compactness theorem. The almost miraculous payoff from the notions of boundary and topology from currents is Fleming's compactness theorem inside a ball B in $\mathbf{R}^3$:

$$\{\text{Rectifiable currents } S \text{ in } B : \text{area } S \leq c, \text{ length } \partial S \leq c\}$$

is compact under the flat norm $\mathcal{F}$. In other words, any infinite sequence of our rectifiable surfaces in the room you are sitting in, with bounds on the area and boundary length, has a convergent subsequence. Consequently, reasonable geometric variational problems have solutions, as we will now illustrate with area-minimizing surfaces. No such compactness holds for smooth submanifolds.

11. Existence of area-minimizing surfaces. *Let C be a closed bounded rectifiable curve of any number of components in* R^3. *Then C bounds a rectifiable current of least area.*

Proof. The curve C lies in some large ball B about 0. It is easy to show C bounds some rectifiable current, for example the cone over C as in Figure 8. Let S_i be a sequence of rectifiable currents bounded by C with areas converging to the infimum. By projecting the surfaces back into B if necessary, which does not increase area, we may assume all the S_i lie in B. By the compactness theorem, we may assume the S_i converge to some limit S. It is easy to show that area $S \leq \underline{\lim}$ area S_i, and hence S provides an area-minimizing surface.

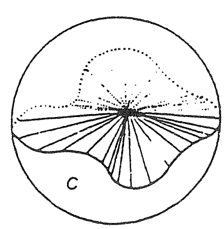

Figure 8
Any closed rectifiable curve C bounds some surface, for example the cone over C. J. Bredt.

12. The regularity theorem. For a given boundary curve, we now have an area minimizer S in the class of rectifiable currents. The big remaining question is whether S is a reasonable surface. The answer due to Fleming (1962) and R. Hardt and L. Simon (1979) sounds too good to be true:

An area-minimizing surface (rectifiable current) bounded by a smooth curve in R^3 is a smooth submanifold with boundary.

Thus it turns out that although we allowed all kinds of singularities, area-minimizing rectifiable currents do not have any. Such complete regularity fails for other classes of surfaces, such as classical mappings of the disc. For the boundary pictured in Figure 9, a circle with a tail, the area-minimizing disc passes through itself. The area-minimizing rectifiable current has higher genus, has less area, and is embedded. It flows from the top, flows down the tail, pans out in back onto the disc, flows around front, and flows down the tail to the bottom. There is a hole in the middle that you can stick your finger through. Incidentally, this surface exists as a soap film, whereas the least-area disc does not.

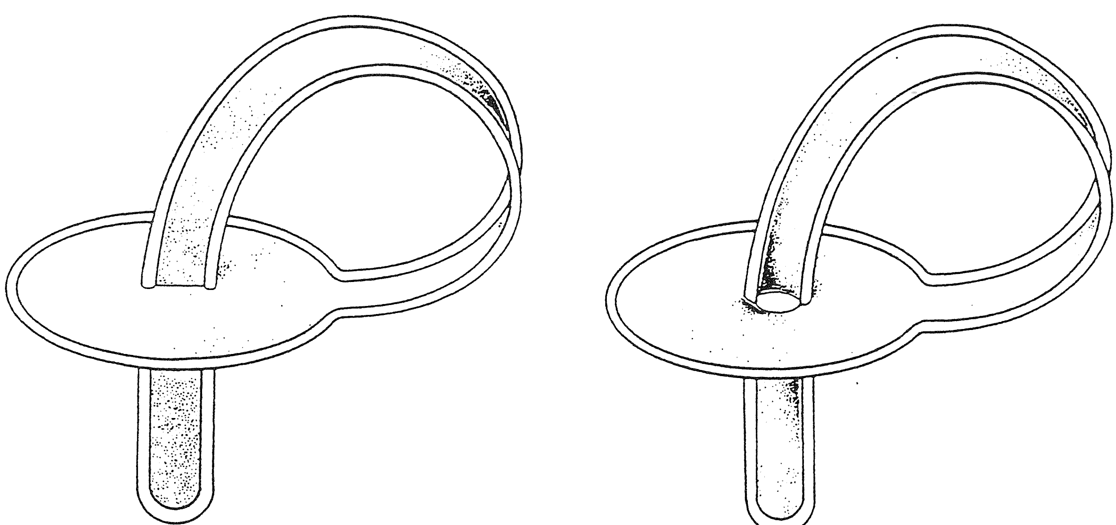

Figure 9
A classical area-minimizing disc need not be embedded, but the area-minimizing rectifiable current S always is embedded. Here S has higher genus and less area than the disc. J. Bredt.

If the boundary curve is badly knotted, it bounds no embedded disc, and the area-minimizing surface will necessarily have high genus. The presumed area-minimizing (orientable) surface bounded by a trefoil knot is pictured in Figure 10.

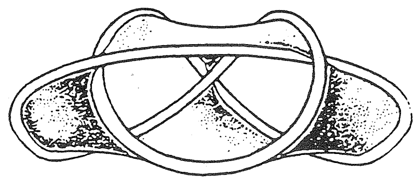

Figure 10
The area-minimizing (orientable) surface bounded by a trefoil knot is presumably this embedded surface of genus 1. J. Bredt.

13. Higher dimensions. The theory of rectifiable currents generalizes to m-dimensional surfaces in R^n. Area-minimizing hypersurfaces remain smooth submanifolds through R^7; for $n \geq 8$, area-minimizing hypersurfaces in R^n can have (n-8)-dimensional singular sets. In higher codimension, singularities occur, even for 2-dimensional surfaces in R^4.

14. Failings of rectifiable currents. For all their virtues, rectifiable currents have their failings too. Rectifiable currents must be oriented, while many physical surfaces need not be oriented. W. Ziemer's *rectifiable currents modulo two* yield a similar theory of unoriented surfaces.

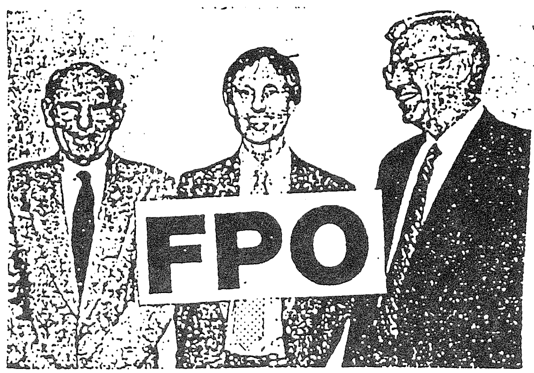

Figure 11
Bill Ziemer (right), who introduced flat chains modulo 2, with his thesis advisor, Wendell Fleming (left), and the author (center), at a celebration in Ziemer's honor at Indiana in 1994. Photo courtesy of Ziemer.

Physical surfaces such as soap films often consist of pieces of surface meeting along whole singular curves. These curves, although not part of the given boundary, unfortunately count as boundary for rectifiable currents. Explaining the structure of soap films required a new theory of *(M,ε,δ)-minimal sets* developed by F. Almgren and J. Taylor.

Crystal surfaces often exhibit an infinitesimal corrugation well modeled by the *varifolds* of Almgren and W. Allard.

15. Open questions. There are many fundamental open questions. For example, for the existence of area-minimizing surfaces in R^3, is there a simple direct proof that stays inside the class of smooth submanifolds? Are area-minimizing surfaces in general dimensions stratified manifolds, or can they have fractal singular sets?

Frank.Morgan@williams.edu

Kelvin's 100-year-old Conjecture Disproved by Weaire and Phelan

Excerpted from the new edition of Geometric Measure Theory: a Beginner's Guide by Frank Morgan, to appear in May, 1995.

This year brings striking news of the disproof of Lord Kelvin's 100-year-old conjecture by Denis Weaire and Robert Phelan of Trinity College, Dublin. Kelvin sought the least-area way to partition all of space into regions of unit volumes. (Since the total area is infinite, least area is interpreted to mean that there is no area-reducing alteration of compact support preserving the unit volumes.) His basic building block was a truncated octahedron, with its six square faces of truncation and eight remaining hexagonal faces, which packs perfectly to fill space as suggested by Figure 13.13.1. (The regular dodecahedron, with its twelve pentagonal faces, has less area, but it does not pack.) The whole structure relaxes slightly into a curvy equilibrium, which is Kelvin's candidate. All regions are congruent.

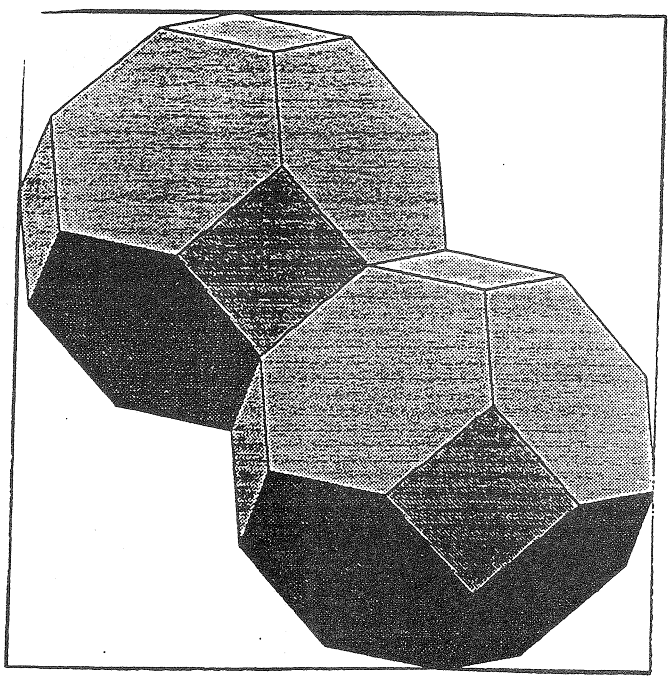

Figure 13.13.1. Lord Kelvin conjectured that the least-area way to partition space into unit volumes uses relaxed truncated octahedra. Graphics by Ken Brakke on the Brakke Evolver from Brakke's early Geometry Center report.

Weaire and Phelan recruited two different building blocks from certain chemical "clathrate" compounds: the regular dodecahedron and a tetrakaidecahedron with 12 pentagonal faces and 2 hexagonal faces. The tetrakaidecahedra are arranged in three orthogonal stacks, stacked along the hexagonal faces, as in Figure 13.13.2. The remaining holes are filled by dodecahedra. Again, the structure is allowed to relax into a stable equilibrium. Computation on the Brakke Evolver shows an improvement over Kelvin's conjecture of about 0.3%. Weaire and Phelan thus provide a new conjectured minimizer. Weaire's popular account in *New Scientist* gives further pictures and details.

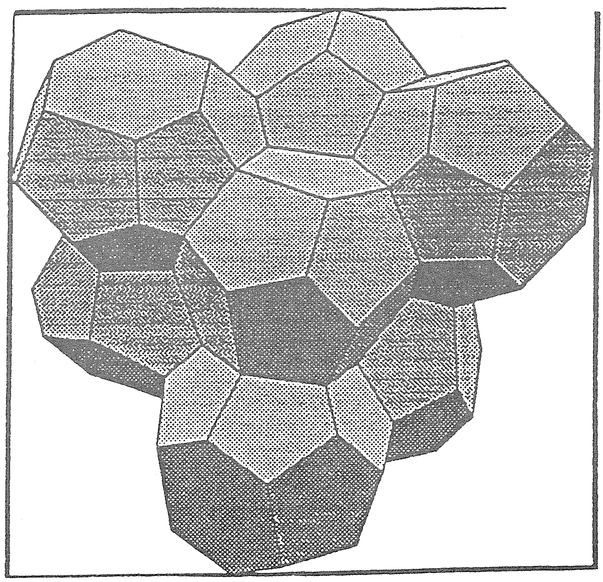

Figure 13.13.2. The relaxed stacked tetrakaidecahedra and occasional dodecahedra of Weaire and Phelan beat Kelvin's conjecture by about 0.3%. Graphics by Ken Brakke on the Brakke Evolver from Brakke's early Geometry Center report.

In greater detail, the centers of the polyhedra are at the points of a lattice with the following coordinates modulo 2:

0	0	0
1	1	1
.5	0	1
1.5	0	1
0	1	.5
0	1	1.5
1	.5	0
1	1.5	0

Given a center, the corresponding polyhedral region is just the "Voronoi cell" of all points closer to the given center than to any other center. The relaxation process also needs to slightly adjust the volumes to make them all 1.

Incidentally, Kelvin's particular truncated octahedron is actually a scaled "permutohedron," the convex hull of the 24 permutations of (1,2,3,4) in $\mathbf{R}^3 = \{ x \in \mathbf{R}^4 : \Sigma x_i = 10 \}$.

Proving the new Weaire-Phelan conjecture could take a while. After all, the single bubble was proved minimizing by Schwartz in 1884, and the double bubble remains conjectural in 1994. Will Weaire-Phelan's "infinite bubble" take another century?

Brakke, Kenneth A, Century-old soap bubble problem solved! *Imagine That!* 3 (Fall, 1993), The Geometry Center, University of Minn., 1-3.

Weaire, Denis, Froths, foams and heady geometry, New Scientist, May 21, 1994.